



UNIVERSIDAD DE DEUSTO

SALOMÓN

Un nuevo enfoque para la mejora de procesos de negocio
mediante la producción inteligente basada en modelos
predictivos de control híbridos y autoadaptativos

Tesis doctoral presentada por JAVIER NIEVES
dentro del Programa de Doctorado en
SISTEMAS DE INFORMACIÓN

Dirigida por el Dr. PABLO GARCÍA BRINGAS

Bilbao, Julio de 2012



UNIVERSIDAD DE DEUSTO

SALOMÓN

Un nuevo enfoque para la mejora de procesos de negocio
mediante la producción inteligente basada en modelos
predictivos de control híbridos y autoadaptativos

Tesis doctoral presentada por JAVIER NIEVES
dentro del Programa de Doctorado en
SISTEMAS DE INFORMACIÓN

Dirigida por el Dr. PABLO GARCÍA BRINGAS

El doctorando

El director

Bilbao, Julio de 2012

Para los que saben que
hay 10 tipos de personas,
las que saben binario
y las que no. #b-team

Resumen

Se define un proceso productivo como toda aquella actividad que recibiendo en su entrada una serie de materias primas es capaz de modificarlas con el fin de conseguir un nuevo producto. De no ser por este tipo de actividades, la sociedad actual no hubiera podido crecer en la de la manera en lo que lo está haciendo.

La mayor problemática es que el halo de misticismo que envolvía los procesos productivos en la antigüedad sigue haciéndolo hoy en día. Además, otro tipo de exigencias como pueden ser las medidas de calidad o el cumplimiento de unas pautas medioambientales complican los procesos ya existentes. Sin dejar de lado que cada vez el mercado se encuentra más globalizado, las pequeñas mejoras que se aporten pueden convertirse en una mejora competitiva que impulse la producción de aquellas empresas que lo incorporen.

A la luz de estos hechos, la comunidad científica lleva muchos años desarrollando nuevos métodos de supervisión y control de plantas, - un ejemplo es el sistema de supervisión, control y adquisición de datos (SCADA) -. Otro tipo de sistemas son capaces de situarse en un horizonte temporal posterior al actual, intentando prever qué es lo que va a suceder en la planta. Este tipo de sistemas son los Modelos Predictivos de Control. A pesar de existir soluciones comerciales y académicas, siguen identificándose una serie de limitaciones, como la imposibilidad de trabajar con un sistema multivariable, la dificultad de adaptación de la solución a los cambios del proceso y la no adecuación de la no linealidad del proceso de negocio a los modelos desarrollados.

Con el fin de cubrir las necesidades que ya han identificado las empresas en cuanto a calidad, reducción de costes o eco-fabricación, así como, la capacidad de obtener una mejora competitiva, se pretende diseñar, desarrollar, evaluar e implantar un sistema que sea capaz de readaptar la producción para cumplir con las restricciones definidas en el proceso.

En este contexto, formulamos la siguiente hipótesis *«Es posible modelar el flujo de negocio de un proceso productivo como una nube de conocimiento a través de la creación de un modelo predictivo de control híbrido (lineal y no lineal) basado en técnicas de aprendizaje automático actuales, manteniéndolo actualizado y llevando el control de la producción para la optimización de alguno de sus parámetros críticos.»*

Para su validación se desarrollan las siguientes tareas que permitirán crear el sistema completo anteriormente definido y que se centran en las siguientes problemáticas: (i) el desarrollo de los análisis sobre predictores estadísticos basados en técnicas de aprendizaje automático y la forma de unirlos con el objetivo de mejorar el sistema de predicción, (ii) definir y diseñar el método que permitirá mantener actualizados en todo momento los modelos predictivos, (iii) determinar la forma de decidir las acciones que hay que llevar a cabo y, una vez hecho, comunicarlo a los operarios o a otros sistemas de control ya incorporados en la planta y (iv) evaluar y contrastar la solución alcanzada.

En la actualidad, los procesos productivos están totalmente integrados en la sociedad. De este modo, los avances en el estado de la técnica, como el aquí detallado, no sólo afectarán a aquellas personas directamente relacionadas con las plantas de producción, sino que, de forma indirecta, esas mejoras llegarán a toda la sociedad. El resultado de un proceso productivo, un producto manufacturado, puede formar parte de un sistema mucho más grande, un sistema que cualquiera de nosotros puede acabar utilizando.

Abstract

The manufacturing process is defined as any activity that getting as its input a set of raw materials is able to modify them in order to achieve a new product. This kind of activities, as simple as they may look, is the main reasons of the growth and evolution of our society, as we know it.

In its beginnings, the ancient manufacturing process was a magic surrounded activity. Nowadays, it still happens the same. Besides, other requirements, such as quality measures or environmental guidelines, complicate the whole process. In addition, we cannot forget that we are living in a globalised market and every little improvement to any of the multiple processes deployed at any plant, could become a competitive boost.

In light of this background, the scientific community has spent many years developing new methods for supervising and controlling a manufacturing plant - one example is the Supervisory, Control And Data Acquisition (SCADA) -. Other systems are capable of being two steps ahead, trying to foresee what will happen in the plant. Those are commonly referred as Model Predictive Control systems. Although there are commercial and academic solutions, the experts have identified a number of limitations such as the inability to work with a multivariable system, the difficulty to adapt the solution to changes in the manufacturing process and the predictive models that do not adapt to the nature of the business process.

In order to meet the needs that companies have already identified in terms of quality, cost reduction and eco-manufacturing, as well as the ability to obtain a competitive boost, we have designed, developed, evaluated and implemented a system that is able to readjust the manufacturing process, working with the constraints previously defined.

Accordingly, we formulated the following hypothesis *‘It is possible to model the business flow of a manufacturing process as a knowledge cloud through the creation of an hybrid (linear and nonlinear) model predictive control, based on current machine learning techniques, keeping it updated and carrying the manufacturing control to optimise some of its critical parameters.’*

For its validation, we have developed the following tasks with the aim of creating the complete system defined above. Hence, we have focused our efforts towards completing the following tasks: (i) developing statistical classifiers based on machine learning techniques and thinking the way of joining them to improve the prediction system, (ii) defining and designing the method that will keep up to date the models, (iii) determining which actions have to be carried out and, once decided, informing operators or other control systems already incorporated in the plant and, finally, (iv) evaluating and comparing the developed solution.

Nowadays, manufacturing processes are fully integrated into society. Therefore, improvements in the state of the art, as detailed here, do not only affect those people directly involved in production plants. In other words, the result of a manufacturing process, a manufactured product, can be part of a much larger system than any of us may end up using it.

Agradecimientos

Una vez llegados a este punto y escribiendo estas líneas como la guinda final del pastel que es esta tesis doctoral, no me queda otra cosa que dar las gracias a todos aquellos que la han hecho posible.

En primer lugar, no puedo olvidarme del Dr. Pablo García Bringas, mi director, jefe y amigo. Su formación, ayuda y dedicación a lo largo de esta carrera de fondo han sido la vitamina extra que ha fortalecido los músculos que han permitido conseguir este hito en mi carrera profesional. Además, no puedo dejar de lado la oportunidad que me brindó para entrar en el mundo de la investigación a través del laboratorio S³Lab - *Laboratory for Smartness, Semantics and Security*, consiguiendo despertar en mí una pasión que hasta aquel momento era desconocida para aquel becario de desarrollo Oracle. Gracias, Pablo; esto se merece una cerveza *Sapporo*.

En segundo lugar, no podría olvidarme del equipo humano con el que cuenta el laboratorio S³Lab, el núcleo, el #b-team, esas cuatro personas entre las que yo mismo me encuentro y que, según cuenta la leyenda, miden más de dos metros, portan mentes brillantes y son capaces de hacer una propuesta de proyecto en únicamente un par de horas; por supuesto, las autoridades pertinentes no tuvieron otra opción que aceptarla. El primero de los integrantes es el incommensurable cuasi doctor Borja Sanz Urquijo, mi compañero de depósito, mi compañero de risas, esa *putilla tecnológica* que todo buen laboratorio necesita y, por qué no decirlo, nuestro particular *Boogle Calendar*. Borja, qué pena que esta relación no empezara en cuarto de carrera. Gracias por todo tu esfuerzo, tu dedicación, tus viajes en el *boomovil*, y otras muchas cosas. Gracias, compi. A su lado, inseparable, está el Dr. Carlos Laorden Gómez. Su destreza con el teléfono no tiene parangón, su impresionante *networking* no deja de sorprendernos. Pero además de esto, es un gran investigador, capaz de lo inimaginable. Ninguna pega tiene este chico; bueno, una sí: es de Vitoria. Y para él guardo una última pregunta

relacionada con su tesis. Concretamente, ¿por qué sigo recibiendo *spam* en mi bandeja de entrada? Gracias, Carlos, por ser como eres y por actuar de estrella del norte guiando al marinero perdido hacia su estancia europea, entre otras cosas. En último lugar, tenemos al Dr. Igor Santos Grueiro, la que ostenta el actual récord de velocidad en la obtención de un doctorado, y la persona que más me ha ayudado en esta tesis doctoral. Si no hubiera sido por él, seguiríamos sin saber que es un *paper*. Además, también ha sido ha sido mi confidente *nerd*. Esas conversaciones de barra de bar sobre programación no se olvidarán nunca. Y las discusiones sobre algoritmos genéticos y compresión bayesiana ya han quedado registradas para los siglos de los siglos como las más interesantes del Facebook. Gracias, Igor, eres un crack.

El laboratorio S³Lab está compuesto por mucha más gente, pero me gustaría destacar y agradecer su trabajo a la sangre nueva que ha revitalizado el laboratorio. Vosotros también sois parte de este #b-team: Félix Brezo Fernández, Patxi Galán García, Xabier Ugarte Pedrero, Iker Pastor López y José Gaviria de la Puerta. Gracias chicos, haré todo lo que esté en mi mano para que consigáis llegar a escribir estas líneas cuanto antes.

En tercer lugar, debo decir que esta investigación no hubiera sido posible sin la ayuda de Azterlan-IK4, Centro Tecnológico de Investigación Metalúrgica, que me hizo descubrir que la Inteligencia Artificial era una rama apasionante, mucho más de lo que pudimos apreciar en la carrera. Asimismo, ellos han sido la fuente que me ha aportado todos los datos utilizados en esta investigación. Por lo tanto, gracias, Azterlan-IK4, y más concretamente, gracias, Argoitz Zabala.

En cuarto lugar, me gustaría agradecer tanto a la Universidad de Deusto, como a la Università degli studi di Bergamo, la oportunidad que me brindaron de descubrir cómo se hace investigación en otros países de la Unión Europea. No puedo olvidarme del Dr. Giuseppe Psaila y sus hijas Laura y Elena, la Dra. Pasqualina Potenza, Paolo Fosci y Simone Mutti. Me hicisteis pasar tres meses maravillosos descubriendo Italia y a los italianos. Os considero de mi familia y espero vuestra visita. *Grazie mille, amici.*

En último lugar, pero no por ello menos importante, quisiera agradecer a mi familia, y especialmente a mis padres, todo el esfuerzo

que han hecho para que fuera un hombre de provecho, por comprarme mi primer ordenador con aquella guía de *Informática para Torpes*, y por esforzarse al máximo para que su hijo pudiera conseguir su sueño, saber cómo funcionaba un ordenador. Igualmente, tengo que brindar este trabajo a mis amigos que están preparando ya la fiesta con el hombre vestido de mono y esperando que cante “Dr. Nieves, Dr. Nieves”. Especialmente, mi más sentido agradecimiento a Nere, a Juan, a Iñigo, a Lore e Igor, que han estado ahí tanto en los buenos momentos, como en los malos. Gracias.

No quería olvidarme de ninguno, así que también se lo agradezco a (pon tu nombre aquí) _____.

Gracias de corazón.

Índice general

Índice de figuras	ix
Índice de tablas	xiii
1 Introducción	1
1.1 Perspectiva de los procesos industriales	3
1.1.1 Visión actual del proceso productivo	3
1.1.2 Modelos Predictivos de Control, una solución para el control de los procesos productivos	5
1.1.2.1 Concepto general	5
1.1.2.2 Limitaciones	7
1.1.2.3 La siguiente generación	7
1.1.3 Otras aproximaciones	8
1.2 Retos a superar	10
1.3 Relevancia del campo de investigación	12
1.4 Mejora / solución propuesta	13
1.4.1 Hipótesis de trabajo	13
1.4.2 Objetivos	14
1.4.3 Arquitectura global de la solución	16
1.5 Desarrollo de la tesis	19
1.5.1 Metodología de investigación	19
1.5.2 Métricas para la evaluación	21
1.6 Estructura del documento	22

ÍNDICE GENERAL

2	Modelo Predictivo de Control y otras aproximaciones	27
2.1	Concepto general	28
2.1.1	Descripción de la estrategia MPC	28
2.1.2	Formulación de un MPC	32
2.2	Los Modelos Predictivos de Control a fondo	34
2.2.1	Diseño y afinación del control	34
2.2.2	Modelos del proceso	35
2.2.2.1	Modelos no lineales basados en principios fundamentales	36
2.2.2.2	Modelos lineales empíricos	36
2.2.2.3	Modelos no lineales empíricos	38
2.2.2.4	Modelos lineales basados en principios fundamentales	41
2.2.3	Modelado del MPC e identificación de la tecnología	41
2.2.4	Tecnología de control MPC	44
2.2.4.1	Estimación del estado	46
2.2.4.2	Determinación del estado al que llegar	47
2.2.4.3	Determinación del camino de optimización	49
2.2.4.4	Formulación de restricciones	50
2.2.4.5	Trayectorias de entrada y salida	51
2.2.4.6	Horizonte de salida y parametrización de entradas	52
2.3	Una breve historia de los Modelos Predictivos de Control	53
2.3.1	Modelos Predictivos de Control lineales en la industria, los orígenes	54
2.3.1.1	LQG	54
2.3.1.2	IDCOM	55
2.3.1.3	DMC	56
2.3.1.4	QDMC	56
2.3.1.5	IDCOM-M, HIECON, SMCA y SMOC	57
2.3.1.6	RMPCT y DMC-plus	59
2.3.2	Desarrollos teóricos de la tecnología MPC no lineal	60

ÍNDICE GENERAL

2.3.2.1	Estabilidad	60
2.3.2.2	NMPC de horizonte infinito	61
2.3.2.3	NMPC con la restricción de igualdad de estado terminal a cero	62
2.3.2.4	NMPC contractivo	62
2.3.2.5	NMPC de modo dual	63
2.3.2.6	NMPC de horizonte cuasi-infinito	63
2.3.2.7	NMPC robusto	64
2.3.3	Implementaciones industriales de la tecnología MPC no lineal	65
2.3.3.1	Modelos estado-espacio	65
2.3.3.2	Realimentación de la salida	66
2.3.3.3	Modelos de entradas / salidas	66
2.3.3.4	Modelo de principios fundamentales	67
2.3.3.5	Optimización de estados estacionarios y dinámi- cos	67
2.4	Aplicaciones de los Modelos Predictivos de Control	68
2.5	Limitaciones de la tecnología existente	70
2.6	Modelo Predictivo de Control, la siguiente generación	73
2.7	Aproximaciones mediante métodos de aprendizaje automático	75
2.8	Sumario	76
3	La fundición, un antiguo proceso rodeado de magia	79
3.1	El proceso metalúrgico de la fundición	80
3.1.1	Materiales utilizados en los procesos de fundición	82
3.1.2	El proceso de fundición y las variables que lo afectan	83
3.2	Defectos y características a predecir	86
3.2.1	Microrrechupes	87
3.2.2	Propiedades mecánicas	88
3.3	Sumario	91
4	Predicción del estado estacionario $t + 1$	93
4.1	Métodos de aprendizaje supervisado	95

ÍNDICE GENERAL

4.1.1	Redes bayesianas	96
4.1.1.1	Método de aprendizaje <i>escalar colinas</i>	98
4.1.1.2	Algoritmo de aprendizaje <i>K2</i>	98
4.1.1.3	Clasificador bayesiano ingenuo (<i>Naïve Bayes</i>)	100
4.1.1.4	El método <i>Tree Augmented Naïve</i> (TAN)	101
4.1.2	K vecinos más próximos	102
4.1.3	Redes neuronales artificiales	104
4.1.4	Máquinas de soporte vectorial	106
4.1.5	Árboles de decisión	109
4.1.5.1	El algoritmo C4.5	111
4.1.5.2	El método de bosques aleatorios	112
4.2	Combinación de métodos de aprendizaje automático	113
4.2.1	Combinación por voto	115
4.2.1.1	La mayoría decide	115
4.2.1.2	Producto de probabilidades	116
4.2.1.3	Media de probabilidades	117
4.2.1.4	Probabilidad máxima	118
4.2.1.5	Probabilidad mínima	119
4.2.2	Grading	119
4.2.3	Stacking	121
4.2.4	Multiesquema	122
4.3	Validación empírica	123
4.3.1	Metodología general	123
4.3.2	Predicción mediante clasificador único	128
4.3.2.1	Microrrechupes	128
4.3.2.2	Propiedades mecánicas	132
4.3.3	Predicción mediante meta-clasificador	137
4.3.3.1	Microrrechupes	137
4.3.3.2	Propiedades mecánicas	143
4.4	Discusión de los resultados	148
4.5	Sumario	150

5	Adaptando los modelos al proceso cambiante	153
5.1	Capacidades estadísticas para la adaptación de modelos	155
5.1.1	Redes bayesianas para comprimir datos	156
5.1.2	Generación de información a través del método de Montecarlo	157
5.2	Adaptación de los modelos que forman un MPC	161
5.2.1	Compresión bayesiana basada en métodos de Montecarlo	161
5.2.1.1	Aprendizaje de la red bayesiana	163
5.2.1.2	Ordenación de los nodos para la generación de información	164
5.2.1.3	Selección de la distribución de probabilidades y su representación	165
5.2.1.4	Generación del nuevo estado	166
5.2.2	Consiguiendo un sistema MPC autoadaptativo	167
5.3	Validación empírica	168
5.3.1	Metodología general	169
5.3.2	Predicción sobre conjuntos de datos reducidos	172
5.3.2.1	Microrrechupes	172
5.3.2.2	Propiedades mecánicas	188
5.3.3	Predicción sobre conjuntos de datos acumulados	205
5.3.3.1	Microrrechupes	206
5.3.3.2	Propiedades mecánicas	225
5.4	Discusión de los resultados	245
5.5	Sumario	246
 6	 Generación de nuevos parámetros de producción	 249
6.1	Programación evolutiva	252
6.1.1	Inicialización	255
6.1.2	Creación de nuevas generaciones	256
6.1.2.1	Cruce de individuos	257
6.1.2.2	Mutación de individuos	257
6.1.2.3	Selección de individuos	258

ÍNDICE GENERAL

6.1.3	Valoración de individuos	259
6.2	Validación empírica	261
6.2.1	Metodología general	261
6.2.2	Realimentación de la planta a través de programación evolutiva paralela	264
6.2.2.1	Microrrechupes	265
6.2.2.2	Propiedades mecánicas	269
6.3	Discusión de los resultados	273
6.4	Sumario	275
7	Conclusiones	277
7.1	Síntesis de la validación del sistema	278
7.2	Resumen de los resultados alcanzados	281
7.3	Aplicaciones de la investigación	285
7.4	Limitaciones del modelo	287
7.5	Trabajo futuro	288
7.5.1	Mejora del proceso de captura de datos	289
7.5.2	Optimización temporal del proceso de realimentación de la planta	290
7.5.3	Evolución a sistemas distribuidos, Modelos Predictivos de Control basados en agentes	291
7.5.4	Perfeccionar las infraestructuras de trabajo	291
7.6	Consideraciones finales	293
8	Conclusions	295
8.1	Synthesis of the system's validation	295
8.2	Summary of the achieved results	299
8.3	Research applications	302
8.4	Limitations of the model	304
8.5	Avenues for further work	305
8.5.1	Improving the data gathering process	305
8.5.2	Time optimisation in the feedback process	306

ÍNDICE GENERAL

8.5.3	Evolving into distributed systems, agent-based Model Predictive Control systems	307
8.5.4	Improving the working infrastructure	308
8.6	Final considerations	309
	Bibliografía	311
	Índice alfabético	335

Índice de figuras

1.1	Arquitectura de la solución global para conseguir el desarrollo de un MPC híbrido y autoadaptativo.	17
2.1	Flujo de trabajo general que siguen los controladores MPC para cada una de las ejecuciones que llevan a cabo.	45
2.2	Genealogía aproximada de los MPC lineales de la industria. . .	53
3.1	Representación del proceso productivo de una fundición. . . .	85
3.2	Proceso de análisis de una pieza mediante rayos X para la detección de microrrechupes.	87
3.3	Proceso de análisis de la resistencia máxima a la tracción de una pieza.	90
4.1	Red bayesiana de ejemplo para el control de los aspersores. . .	97
4.2	Ejemplo de un clasificador KNN en un espacio bi-dimensional. .	103
4.3	Representación gráfica de una red neuronal de tipo MLP. . . .	105
4.4	Representación gráfica de la división de un espacio bi-dimensional mediante SVM.	107
4.5	Representación gráfica de un árbol de decisión.	110
4.6	Resultados en la predicción del estado estacionario $t + 1$ en la detección de microrrechupes.	129
4.7	Tasas de error relativas a la predicción del estado estacionario $t + 1$ en la detección de microrrechupes.	131
4.8	Resultados en la predicción del estado estacionario $t + 1$ en la detección de problemas en la carga máxima.	133

ÍNDICE DE FIGURAS

4.9	Tasas de error relativas a la predicción del estado estacionario $t + 1$ en la detección de problemas en la ca carga máxima. . .	135
4.10	Resultados en la predicción del estado estacionario $t + 1$ en la detección de microrrechupes haciendo uso de la combinación de clasificadores.	138
4.11	Tasas de error relativas a la predicción del estado estacionario $t + 1$ en la detección de microrrechupes haciendo uso de la combinación de clasificadores.	140
4.12	Resultados en la predicción del estado estacionario $t + 1$ en la detección de posibles problemas en la carga máxima utilizando los métodos de combinación de clasificadores.	145
4.13	Tasas de error relativas a la predicción del estado estacionario $t + 1$ en la detección de problemas en la carga máxima mediante la combinación de clasificadores.	147
5.1	Procedimiento general para la generación de un conjunto de datos reducido a través de la compresión bayesiana basada en métodos de Montecarlo.	163
5.2	Precisión alcanzada por los modelos que hacen uso de un conjunto de datos reducido con el fin de predecir problemas de microrrechupes	174
5.3	Tasas de error alcanzadas por los modelos que hacen uso de un conjunto de datos reducido con el fin de predecir problemas de microrrechupes	177
5.4	Resultados apilados de los experimentos que prueban los modelos que hacen uso de un conjunto de datos reducido para detectar problemas de microrrechupe	179
5.5	Precisión obtenida por los modelos aprendidos con un conjunto de datos reducido y enfrentados al conjunto de datos original para la detección de problemas de microrrechupes	181
5.6	Tasas de error obtenidas por los modelos aprendidos con un conjunto de datos reducido y enfrentados al conjunto de datos original para la detección de problemas de microrrechupes . .	183

ÍNDICE DE FIGURAS

5.7	Resultados aplicados de los experimentos que enfrentan los modelos generados con conjuntos de datos reducidos y el conjunto de datos original para la detección de problemas de microrrechupes	186
5.8	Precisión alcanzada por los modelos que hacen uso de un conjunto de datos reducido con el fin de predecir problemas en la carga máxima	190
5.9	Tasas de error alcanzadas por los modelos que hacen uso de un conjunto de datos reducido con el fin de predecir problemas en la carga máxima	193
5.10	Resultados apilados de los experimentos que prueban los modelos que hacen uso de un conjunto de datos reducido para detectar problemas en la carga máxima	196
5.11	Precisión obtenida por los modelos aprendidos con un conjunto de datos reducido y enfrentados al conjunto de datos original para la detección de problemas en la carga máxima	198
5.12	Tasas de error obtenidas por los modelos aprendidos con un conjunto de datos reducido y enfrentados al conjunto de datos original para la detección de problemas en la carga máxima	201
5.13	Resultados aplicados de los experimentos que enfrentan los modelos generados con conjuntos de datos reducidos y el conjunto de datos original para la detección de problemas en la carga máxima	204
5.14	Precisión alcanzada por los modelos que hacen uso de un conjunto de datos acumulado con el fin de predecir problemas de microrrechupes	208
5.15	Tasas de error alcanzadas por los modelos que hacen uso de un conjunto de datos acumulado con el fin de predecir problemas de microrrechupes	211
5.16	Resultados apilados de los experimentos que prueban los modelos que hacen uso de un conjunto de datos reducido para detectar problemas de microrrechupe	214
5.17	Precisión obtenida por los modelos aprendidos con un conjunto de datos reducido y enfrentados al conjunto de datos original para la detección de problemas de microrrechupes	218

ÍNDICE DE FIGURAS

5.18 Tasas de error obtenidas por los modelos aprendidos con un conjunto de datos reducido y enfrentados al conjunto de datos original para la detección de problemas de microrrechupes . . .	221
5.19 Resultados aplicados de los experimentos que enfrentan los modelos generados con conjuntos de datos reducidos y el conjunto de datos original para la detección de problemas de microrrechupes	224
5.20 Precisión alcanzada por los modelos que hacen uso de un conjunto de datos reducido con el fin de predecir problemas en la carga máxima	228
5.21 Tasas de error alcanzadas por los modelos que hacen uso de un conjunto de datos reducido con el fin de predecir problemas en la carga máxima	232
5.22 Resultados apilados de los experimentos que prueban los modelos que hacen uso de un conjunto de datos reducido para detectar problemas en la carga máxima	234
5.23 Precisión obtenida por los modelos aprendidos con un conjunto de datos reducido y enfrentados al conjunto de datos original para la detección de problemas en la carga máxima	238
5.24 Tasas de error obtenidas por los modelos aprendidos con un conjunto de datos reducido y enfrentados al conjunto de datos original para la detección de problemas en la carga máxima . . .	241
5.25 Resultados aplicados de los experimentos que enfrentan los modelos generados con conjuntos de datos reducidos y el conjunto de datos original para la detección de problemas en la carga máxima	244

Índice de tablas

2.1	Resumen de las aplicaciones lineales MPC por áreas y proveedores industriales.	69
2.2	Resumen de las aplicaciones MPC no lineales por áreas y proveedores industriales.	69
3.1	Resumen de los procesos productivos desarrollados sobre metales.	81
6.1	Resumen de los resultados obtenidos en lo referente a las instancias tratadas y corregidas para los defectos de microrrechupes.	265
6.2	Resumen de los resultados obtenidos en lo referente a las variables del proceso modificadas con el fin de solucionar y evitar los problemas de microrrechupes.	266
6.3	Resumen de los resultados obtenidos en lo referente a las generaciones necesarias para obtener la solución óptima que elimina los problemas de microrrechupes.	267
6.4	Resumen de los resultados obtenidos en lo referente a los tiempos de ejecución para obtener 2000 generaciones de individuos que han tratado de eliminar los defectos de microrrechupe. . .	268
6.5	Resumen de los resultados obtenidos en lo referente a las instancias tratadas y corregidas para los defectos en las propiedades mecánicas.	269
6.6	Resumen de los resultados obtenidos en lo referente a las variables del proceso modificadas con el fin de solucionar y evitar los problemas en las propiedades mecánicas.	270

ÍNDICE DE TABLAS

6.7	Resumen de los resultados obtenidos en lo referente a las generaciones necesarias para obtener la solución óptima que elimina los problemas en las propiedades mecánicas.	271
6.8	Resumen de los resultados obtenidos en lo referente a los tiempos de ejecución para obtener 2000 generaciones de individuos que han tratado de eliminar los defectos en las propiedades mecánicas.	272

«Debemos intentar comprender el comienzo del Universo a partir de bases científicas. Puede que sea una tarea más allá de nuestras capacidades, pero al menos deberíamos intentarlo.»

Stephen Hawking
(1942–presente)

1

Introducción

LOS procesos productivos datan de épocas ancestrales. Procesos que han quedado representados tanto en versículos bíblicos y jeroglíficos egipcios, como en los grabados de antiguas vasijas griegas. A lo largo de todo este tiempo, los procesos productivos han permitido transformar las materias primas en productos elaborados. En la antigüedad, el proceso era mucho más primitivo, por supuesto, en la actualidad, y debido al avance y evolución de la sociedad, el proceso ha sido mejorado. Los cambios más relevantes que se han producido a lo largo de las últimas décadas son: (i) el aumento drástico de la competencia y (ii) las limitaciones impuestas por la legislación medioambiental, que han dotado de una complejidad extra al proceso de producción.

Los cambios mencionados implican que se debe llevar a cabo un control estricto de los procesos productivos. Con este objetivo, y gracias a la ayuda de múltiples disciplinas del conocimiento, es posible anticiparse y solucionar situaciones adversas que se dan durante el proceso. En particular, los aspectos fundamentales sobre los que trabajar son los siguientes: (i) reducir los costes, (ii) maximizar la tasa de producción, (iii) minimizar el stock, (iv) controlar la fluctuación de los precios y (v) evitar fallos en el servicio del producto debidos a defectos ocultos.

El objetivo de un proceso de manufacturado es servir de aprovisionamiento a otras industrias, como por ejemplo, la industria metalúrgica, que provee pie-

1. Introducción

zas a otras industrias, como la industria automovilística o la naval. Además, y debido a la naturaleza de esas industrias, el proceso cuenta con grandes controles de seguridad, que certifiquen el buen funcionamiento de los productos manufacturados, ya que el menor de los defectos podría ser fatal. Habitualmente, las medidas adoptadas para determinar que el proceso ha sido finalizado libre de fallos son la realización de controles exhaustivos de la producción y diferentes técnicas de simulación. Desafortunadamente, muchas de estas técnicas únicamente pueden ser aplicadas cuando el producto ya ha sido finalizado. Es por ello que la detección de un producto defectuoso en las etapas finales del proceso se traduce en un incremento de costes del mismo, tanto por el deterioro de la imagen corporativa como por los desechos generados.

Con el fin de solucionar la problemática presentada, se busca una forma de determinar el resultado final del proceso antes de que los productos finales hayan sido manufacturados completamente. Éste es el punto en el que los Modelos Predictivos de Control permiten ayudarnos a la consecución del objetivo de mejora continuo. Bajo esta denominación se agrupan una serie de algoritmos capaces de llevar el control de un modelo de proceso concreto. Este tipo de algoritmos trata de mantener la producción de la planta sobre una trayectoria ideal, modificando los parámetros de la producción para satisfacer al máximo las condiciones para las que fue diseñado originariamente el sistema.

No siempre los resultados son ideales. A lo largo del proceso pueden darse alteraciones que hagan que el resultado final no sea el adecuado. Para solucionarlo, estos algoritmos son capaces de adelantar en el instante de tiempo t cómo se encontrará la planta, de seguir así, en el instante de tiempo $t + 1$, es decir: en el futuro. En caso de detectar algún problema, el sistema se encargará de realizar la corrección de los parámetros manteniendo la producción a lo largo de una trayectoria óptima. Originariamente, este tipo de sistemas fueron desarrollados para satisfacer las necesidades específicas del control de centrales eléctricas y refinerías de petróleo. Sin embargo, en la actualidad, la tecnología de los Modelos Predictivos de Control se utiliza en una amplia variedad de áreas como en la producción química, el procesamiento de alimentos, la automoción y las aplicaciones aeroespaciales.

Utilizando estos sistemas, pretendemos ser capaces de elaborar productos que no desfallezcan por problemas de calidad que no hubieran sido detectados durante su fabricación, así como reducir el coste del propio proceso de producción. Es por ello que la solución adoptada, y que aquí presentamos, se trata de un Modelo Predictivo de Control que deja de lado las formulas matemáticas específicas y que desarrolla las mismas operaciones a través de modelos de

1.1 Perspectiva de los procesos industriales

clasificación basados en las técnicas de aprendizaje automático generalistas y de vanguardia.

El resto del capítulo está organizado como sigue. La sección 1.1 describe, de un modo conciso y breve, la situación actual de los procesos productivos y apunta hacia cómo pueden ser mejorados a través de la tecnología MPC. La sección 1.2 describe cuáles son los retos para la generación de un MPC que supere las limitaciones inherentes de la tecnología y se enmarque dentro de lo que espera de ella. Posteriormente, la sección 1.3 detalla cómo de relevante es esta investigación para la sociedad actual. La sección 1.4 establece la hipótesis fundamental de esta tesis doctoral y enumera los objetivos a cumplir en esta investigación. También, describe brevemente la arquitectura global del sistema planteado. Finalmente, la sección 1.5 describe la metodología de investigación que se ha utilizado y las métricas para la validación empleadas y la sección 1.6 detalla cómo se encuentra organizada esta tesis doctoral.

1.1 Perspectiva de los procesos industriales

Esta sección revisa la situación actual que viven los procesos industriales, comenzando por las modificaciones que han sufrido desde sus orígenes, hasta determinar cómo los Modelos Predictivos de Control pueden ayudar a conseguir la producción *cero defectos*.

1.1.1 Visión actual del proceso productivo

Durante las últimas décadas, los procesos industriales se han enfrentado a un cambio importante en el mercado [SS05]. En todo el mundo, la competencia se ha incrementado drásticamente y la legislación ambiental es mucho más dura, sobre todo cuando se habla del consumo de recursos naturales. Las limitaciones ambientales impuestas por la legislación también han dado lugar a un aumento significativo de la complejidad del proceso productivo, así como de los costes de los equipos de producción. Por lo tanto, los sistemas más avanzados de apoyo al proceso tienen la obligación de explotar las libertades disponibles a lo largo del proceso productivo, en un objetivo general de mejora continua.

Una de las razones principales de los cambios es la globalización del mercado, que se ha conseguido gracias a hitos clave acaecidos en el ámbito de las telecomunicaciones, el transporte y la automatización avanzada. Todas las

1. Introducción

mejoras han surgido debido a los rápidos avances en electrónica, informática y tecnología de la información. Como consecuencia, los procesos industriales de hoy en día deben enfrentarse a un entorno altamente competitivo. El mercado se ha desarrollado a partir de la oferta de los proveedores hasta llegar a estar impulsado íntegramente por la demanda del mercado. Estos cambios tienen consecuencias para los productores; el mercado les obliga a responder con rapidez y fiabilidad a la demanda de los productos.

Además, los productos tienen que ser entregados en muy poco tiempo, en ventanas de tiempo estrictamente ajustadas a la calidad adecuada esperada y teniendo en cuenta el volumen de la producción marcado por la cantidad de producto solicitado. De esta forma, se origina una competencia asombrosa entre proveedores de productos hasta tal punto que el mercado puede llegar a acabar saturado, en ocasiones. Si los productos cumplen las especificaciones impuestas, el precio se convertirá en el principal factor de discriminación para que un consumidor se decante por comprar el producto a un proveedor determinado o a otro. Tras la optimización de la producción, los costes se nivelan, aportando una mejora de negocio significativa a aquellos que hayan optimizado sus procesos.

En la actualidad, las empresas funcionan de formas diferentes [SS05]. Muchas de las industrias siguen operando basadas en las peticiones que reciben (i.e., trabajan bajo pedido). Esto implica la inexistencia de una conexión directa entre la demanda real del mercado y la producción. Por otra parte, otras industrias a menudo fabrican los productos en un ciclo de secuencias fijas. De esta forma, cualquier cambio puede repercutir impredeciblemente, dada la naturaleza tan dispar que caracteriza a las empresas.

Con el fin de estar preparado para estos drásticos cambios, se necesita un control estricto de los procesos de producción. De este modo, al procesar todas y cada una de las operaciones que se realizan a lo largo del proceso se puede prever cuál va a ser el funcionamiento, debido a que la naturaleza del proceso hace que sea totalmente predecible y reproducible. Teniendo en cuenta esta previsión del proceso y trabajando sobre diferentes factores de la producción, los resultados económicos de toda empresa pueden ser mejorados.

En el caso que aquí presentamos, nos vamos a centrar en minimizar los costes de la producción a través de la reducción de los problemas de fabricación. La esencia del razonamiento consiste en que, si no se producen elementos defectuosos no se pierde dinero por rehacerlos, o desecharlos, o por daños en la imagen de la compañía. Del mismo modo, se pueden reducir los costes asociados a las pruebas de calidad de los productos ya finalizados. No quiere decir

1.1 Perspectiva de los procesos industriales

que esas pruebas desaparezcan, pero sí que ahora la maquinaria especializada para esas pruebas no será tan numerosa y la cantidad de análisis desarrollados se verá reducida, significando una disminución de los costes asociados a estas tareas.

1.1.2 Modelos Predictivos de Control, una solución para el control de los procesos productivos

1.1.2.1 Concepto general

Un Modelo Predictivo de Control (MPC a partir de ahora) se define como un algoritmo o conjunto de algoritmos que permite controlar cada una de las etapas del proceso productivo de la planta. Para conseguir su objetivo, un MPC basa su funcionamiento en la predicción de cómo se encontrará la planta en un instante de tiempo determinado localizado en un futuro cercano. Los MPC trabajan de forma iterativa. En otras palabras, calculan en cada uno de sus ciclos la situación de la planta, determinando si existe alguna variación sobre la producción marcada como idónea. En caso de encontrarla, se modificarán los valores del proceso y se comprobará de nuevo el estado futuro de la planta. Este proceso se repetirá hasta encontrar los valores adecuados para mantener el proceso controlado.

El éxito de esta tecnología como un paradigma de control de procesos se puede atribuir a tres factores importantes. El primero y principal es la incorporación de un modelo de proceso explícito en el cálculo del control. Esto permite que el controlador, en principio, pueda tratar directamente con todas las características importantes de la dinámica del proceso. En segundo lugar, el algoritmo considera el comportamiento de la planta en un horizonte de futuro temporal. Esto significa que los efectos de las perturbaciones del proceso se pueden prever y eliminar, permitiendo al controlador de la unidad de la planta centrarse en dirigirla a lo largo de una trayectoria deseada. Por último, el controlador considera las entradas del proceso, el estado y las limitaciones de la producción directamente en el cálculo del control de la planta. Esto significa que es casi imposible violar las restricciones del proceso, ya que el control es más estricto. La inclusión de las limitaciones, o restricciones, es lo que distingue los MPC de otros paradigmas de control de procesos.

Por otro lado, aunque los procesos de fabricación son intrínsecamente no lineales, la gran mayoría de los modelos desarrollados y que pueden verse detallados en los siguientes artículos e informes [RRTP78, CR80, GFH88, QB97]

1. Introducción

se basan en modelos dinámicos lineales. Existen varias razones posibles para ello [QB00]. Primero, los modelos lineales empíricos se pueden identificar porque funcionan de una manera directa, a partir de los datos obtenidos mediante una serie de pruebas del proceso. Segundo, la mayoría de las aplicaciones hasta la fecha han sido utilizadas en procesos de refinera, donde el objetivo es, en gran medida, mantener el proceso en un estado de equilibrio deseado en lugar de tener que desplazarlo rápidamente de un punto a otro. En otras palabras, el proceso productivo llevado a cabo en las refineras se mantiene constante, mientras que en otros entornos, el proceso sufre múltiples cambios debido a la gran cantidad de productos diferentes que pueden llegar a ser manufacturados. Finalmente, usando un modelo lineal, la consecución del objetivo *cero defectos* puede alcanzarse mediante técnicas como la programación cuadrática, para la que es sencillo encontrar algoritmos fiables [Wri97].

Por el contrario, hay casos en que los efectos no lineales son lo suficientemente importantes como para justificar el uso de la tecnología no lineal sobre los MPC (NMPC, de la voz inglesa “*Non-linear Model Predictive Control*”). Éstos últimos son una variante de los MPC originales que, en lugar de utilizar funciones matemáticas lineales para la predicción del estado de la planta, hacen uso de funciones no lineales para desempeñar la misma tarea. Una pequeña problemática, definida por Camacho [CB04], es que los sistemas para Modelos Predictivos de Control deben encontrar su solución en un número finito de iteraciones, algo no tan sencillo en los modelos no lineales. Camacho también indica que los NMPC son habitualmente utilizados para entornos con poca tasa de recogida de información durante el proceso productivo. Sin embargo, cada vez son más usados en entornos con altas tasas de captura.

La literatura sobre las investigaciones en MPC es muy extensa. Los retos teóricos y prácticos relacionados con esta tecnología se resumen en varios artículos. Los tres documentos sobre MPC presentados en la conferencia del PCCh son un excelente punto de partida [LC97, May97, QB00]. Además, Qin y Badgwell [QB00] presentaron una breve historia de esta tecnología, así como una encuesta sobre los modelos industriales, algoritmos y aplicaciones; algo que los profesionales de la industria pueden encontrar especialmente útil. Camacho y Bordons resumen las propiedades teóricas de los algoritmos para los MPC en [CB04]. Por otra parte, Froisy debate sobre el pasado, presente y futuro de la tecnología en su revisión [Fro94]. Kwon proporciona una lista muy extensa de referencias en [Kwo94]. Una visión más reciente de la teoría del desarrollo de este tipo de modelos se puede encontrar en [NMS00]. Asimismo, mucha más información puede encontrarse en varios libros que han ido

1.1 Perspectiva de los procesos industriales

apareciendo sobre la tecnología de los MPC como [BWG91, Soe92, Mos95].

1.1.2.2 Limitaciones

Muchos de los MPC, algoritmos que incluso están implantados en la industria, sufren de limitaciones heredadas de sus tecnologías de base. Entre estos problemas, identificados en [ML91, MR93], se incluyen: (i) limitadas opciones entre las que elegir el modelo, (ii) realimentación ineficiente, (iii) pérdida de la estabilidad ante la variabilidad del proceso y (iv) solución ineficiente de la optimización dinámica de los modelos.

Además, las señales de la planta (datos que serán utilizados como entrada para un MPC) requieren la atención especial de ingenieros experimentados durante las fases del diseño y prueba del modelo. Por otro lado, otro de los problemas existentes es que no se dispone de ninguna herramienta para determinar si los datos recogidos son los adecuados para representar el comportamiento del proceso antes de comenzar con el diseño del MPC.

Asimismo, por lo general no se aborda la eficiencia estadística y la coherencia de los algoritmos. Igualmente, también existe una falta de métodos de validación del modelo que permitan comprobar si éste es el adecuado para el control, o si se ha producido un deterioro en su eficacia después de su puesta en marcha. Hasta donde llega nuestro conocimiento, tampoco encontramos ningún enfoque sistemático para la construcción de modelos dinámicos no lineales. Por consiguiente, se necesita una serie de directrices para realizar las pruebas de la planta con el objetivo de construir un modelo confiable.

Estas limitaciones pueden solucionarse, en parte, gracias a los modelos de aprendizaje automático.

1.1.2.3 La siguiente generación

Morari et ál. [ML97] recogen en sus investigaciones la visión que los comerciales de tecnologías industriales tenían sobre la tecnología MPC. Del mismo modo, nos hacen vislumbrar los diferentes caminos por los que dirigir la investigación dentro de esta área. A continuación se pasará a realizar una breve descripción de cada uno de ellos.

- **Formulación básica del controlador:** debido a la dificultad de expresar todos los objetivos de control pertinentes en función de un único

1. Introducción

objetivo, la nueva generación de MPC debería utilizar múltiples funciones objetivo.

- **MPC adaptativos:** existen fuertes incentivos del mercado para que la comunidad científica desarrolle un MPC capaz de autoajustarse. Esta característica es de vital importancia ya que, debido a la naturaleza cambiante de ciertos procesos productivos, los modelos desarrollados deben mantener los resultados alcanzados en las etapas iniciales mediante una adaptación de su comportamiento en las tareas de predicción y realimentación de las plantas.
- **MPC robustos:** los MPC industriales, es decir, aquellas soluciones realmente implementadas en entornos reales de producción, suelen confiar únicamente en la fuerza bruta de la simulación para evaluar los defectos de la planta. La robustez¹ podría reducir el tiempo necesario para adecuar y validar los algoritmos. Incluso, estos valores podrían ser combinados con las estimaciones de incertidumbre.
- **MPC no lineal:** viendo las necesidades actuales en el mundo industrial, la tendencia de desarrollo de este tipo de sistemas debe dirigirse hacia modelos no lineales (NMPC), ya que éstos se ajustan mucho mejor a los actuales procesos productivos.

1.1.3 Otras aproximaciones

Como se ha podido advertir a lo largo de la breve descripción de la tecnología MPC dada en la sección 1.1.2, no se contempla en el estado de la técnica el desarrollo de sus fórmulas o de sus modelos mediante clasificadores estadísticos obtenidos a través de un proceso de aprendizaje automático. No obstante, en la literatura sí que encontramos algunas aproximaciones que hacen uso de ellos, pero únicamente para las acciones de realimentación [QB00, QB00]. A pesar de no utilizarlos dentro de los propios modelos MPC, ciertas características de predicción y de reducción de ruido sí que han sido solucionadas. Así, el aprendizaje automático está siendo cada vez más utilizado para la optimización de los procesos de producción.

Concretamente, los clasificadores más utilizados en la literatura son las redes neuronales artificiales (ANN, del vocablo inglés *“Artificial Neural Net-*

¹Una rama de la teoría de control que, explícitamente, incorpora las mediciones de incertidumbre para su enfoque observación.

1.1 Perspectiva de los procesos industriales

works”), por ejemplo, para clasificar piezas de fundición [LSOdM98], para optimizar los parámetros de calidad [ZXD08], en la detección de las causas de defectos de fundición [PK03], y en otros problemas inherentes al proceso [Bha99].

Estos clasificadores no han sido los únicos utilizados; la comunidad científica también ha hecho uso de otros, como las redes bayesianas, esta vez como método previo a la utilización de redes neuronales artificiales [VMB02]. En este caso, el objetivo que buscaban los autores era la predicción de la cantidad de ferrita en las soldaduras de acero inoxidable. Sorprendentemente, existen experimentos que tratan de predecir las propiedades de las piezas moldeadas utilizando estos métodos de aprendizaje automático [Vor01], algo que nos animó a continuar con la línea de investigación que se presenta en esta tesis doctoral.

Del mismo modo, también existen experiencias exitosas con la utilización del algoritmo de los k vecinos más próximos, “*K-Nearest Neighbours*” en inglés. Uno de los ejemplos de su utilización se ha dado en la detección de fallos durante el proceso de fabricación de semiconductores [HW07].

Pero no sólo son utilizados clasificadores tan habituales como los ya mencionados; modelos que normalmente son utilizados para trabajar con textos, tales como las máquinas de soporte vectorial (en inglés “*Support Vector Machines*”) y los árboles de decisión, han sido la base para la solución de problemas de clasificación dentro del ámbito de los procesos de fundición. Más concretamente, ambos se han aplicado en (i) el diagnóstico de fallos [YZZ07, WY07], (ii) la predicción de series temporales [TVBDW⁺03], (iii) la identificación de errores en turbinas de gas [MLP08], (iv) la detección de intrusiones en la red [ZZ06] y (v) la predicción de las propiedades mecánicas de materiales de acero [WVG06].

En cuanto a la reducción de ruido en las entradas de datos, uno de los experimentos más recientes es el desarrollado por Pham et ál. [PMM⁺07]. Estos autores utilizaban el algoritmo de las abejas (una técnica muy similar a los algoritmos genéticos) con el fin de realizar la selección de características más relevantes de una manera mucho más comprensiva.

Como se puede observar, no se han desarrollado demasiadas investigaciones orientadas a los procesos de producción con la utilización de diferentes métodos de aprendizaje automático. A pesar de ello, estas aproximaciones muestran la tendencia que está surgiendo dentro de la comunidad científica. Así que, esta es la línea de trabajo que nos lleva a fusionar los dos métodos (i.e., Modelos

1. Introducción

Predictivos de Control y aprendizaje automático) que tan buenos resultados están dando.

1.2 Retos a superar

Dada la problemática actual en el campo de los MPC, y las líneas futuras que se esperan seguir en los próximos años (ver sección 1.1), surgen una serie de retos sobre los que se conduce nuestra investigación. Especialmente, nos centraremos en los siguientes:

- **Adquisición de los datos.** Cualquier sistema basado en el conocimiento adquirido mediante aprendizaje supervisado o semi-supervisado necesita de unos datos que constituyan su base de conocimiento. Por ello, uno de los primeros retos que surge es el desarrollo de un método para la adquisición de esta información. En muchos de los casos va a depender de la planta de producción con la que se trabaja y los sistemas que ya tenga implantados. El mejor de los escenarios es el de aquella planta que disponga de un sistema de control supervisor y adquisición de datos (SCADA, del vocablo inglés “*Supervisory Control And Data Acquisition*”) [DS99], ya que contaremos con información sobre la situación actual de la planta en tiempo real.

Por otra parte, en caso de no disponer de este tipo de sistemas, los datos pueden ser adquiridos de forma manual, mediante anotaciones realizadas en papel, o mediante software informático. Este tipo de sistemas, como el autor de esta tesis doctoral ha podido comprobar, es uno de los más habituales en el tejido empresarial vasco de la fundición. A pesar de ello, e independientemente del sistema utilizado para la extracción de la información, es necesaria la comprobación de los datos para descartar cualquier tipo de problema o ruido introducido en el proceso de adquisición. Además, no se trata de un proceso trivial; así lo ilustra la cantidad de trabajos existentes en la literatura [Cle81, OW84, KWL98, GNDC05].

- **Análisis y predicción de situaciones futuras.** La base de un MPC es averiguar qué es lo que va a suceder a lo largo del proceso de producción de antemano; en otras palabras, cómo se encontrará la planta en un instante de tiempo $t + 1$. Por ello, el segundo de los retos a superar es aquel que, basándose en diferentes técnicas de aprendizaje automático, nos permite prever el comportamiento de la planta. Para

ello, proponemos el desarrollo y utilización de un meta-clasificador multiobjetivo. Con el fin de superar este reto es necesaria la realización de múltiples de pruebas hasta dar con el método de predicción adecuado. En este dominio encontramos diversos autores que han trabajado en ello [VOF10, CD09, KD08].

- **Realimentación de datos (Adaptación continua del sistema).** El sistema tiene que estar siempre actualizado. Las bases de conocimiento de los sistemas de predicción tienen que estar siempre adaptadas y listas para obtener buenos resultados a pesar de la naturaleza cambiante de los procesos que se estén desarrollando en la planta. De esta forma, por una parte, las evidencias más antiguas deben mantenerse en el sistema pero con un menor nivel de importancia, mientras que por otro lado, asignamos un mayor nivel de significancia a la nueva información que estamos extrayendo del proceso de producción. Algunos de los métodos propuestos por la comunidad científica para la solución de este reto son los factores de desvanecimiento [FW95] y la compresión bayesiana [DM99].
- **Toma de decisiones para realimentar el proceso.** Un sistema como un MPC aporta una ventaja significativa a las empresas sólo con ser capaz de indicar qué es lo que va a suceder. Además, siempre será mejor que pueda aconsejar a los operarios de la planta o incluso poder comunicarse con aplicaciones desarrolladas por terceros, y que tienen la posibilidad de interactuar con la planta, los parámetros de configuración que deben ser introducidos en todas y cada una de sus máquinas con el fin de llevar a cabo una producción sin defectos. Al fin y al cabo, se trata de realizar un proceso de predicción a la inversa, es decir, en lugar de partir de los datos y saber qué es lo que sucede con el producto, se intenta partir de la situación ideal de cómo debe obtenerse el producto, para deducir cuáles son los requisitos que deben cumplirse. Existen aproximaciones parecidas dentro del mundo de las redes bayesianas y basadas en la utilización del análisis de sensibilidades [SE00].
- **Optimización de los resultados en términos de precisión y de falsos negativos.** El objetivo final de todos los retos es conseguir una optimización del proceso productivo, obteniendo los resultados más idóneos para la solución de los problemas del proceso. Esto significa que los porcentajes de precisión deben ser lo más altos posibles. Además, siempre debemos intentar minimizar los errores, desde una perspectiva general.

1. Introducción

Sin embargo, en el caso que nos ocupa, se va a dar más importancia a minimizar los falsos negativos¹. La razón fundamental es que cuando se da un falso negativo es porque se había detectado que el proceso iba a funcionar correctamente y no ha sido así, con lo que se produce un gasto inesperado que hace que el proceso no sea óptimo. Por otra parte, los falsos positivos quieren decir que se ha predicho que el producto no va a cumplir con lo marcado y se modifica el proceso, pero si ese producto se hubiera realizado no se hubiera producido ningún error. Estos últimos casos no son tan importantes para el hecho de la optimización, ya que lo que se está consiguiendo es un sistema con un alto grado de prudencia, aspecto interesante, debido a que se trata de una idea ampliamente extendida en los procesos de producción actuales.

1.3 Relevancia del campo de investigación

Los procesos productivos son de una importancia crucial para la sociedad actual, pero estos procesos no han surgido ahora; vienen ya de lejos. Sin embargo, la forma en la que se desarrollan los productos sigue bajo un halo de magia y misticismo que hasta los propios trabajadores del proceso reconocen. A pesar de que los procesos eran y siguen siendo algo primitivos, la evolución de la sociedad (produciendo mucha más competencia y apareciendo limitaciones de calidad y medioambientales impuestas por terceros) han hecho que esos procesos mejoren. A la luz de estos acontecimientos, la investigación que aquí se presenta contribuirá al avance de los procesos de manufacturación actuales. De esta forma, los procesos productivos pueden ser mejorados como se detalla a continuación:

1. Disponer de un proceso productivo mucho más controlado, evitando ese grado de misticismo y de imprecisión con el que se cuenta actualmente.
2. Mejorar los resultados de los procesos productivos frente a los agentes externos, es decir, frente a los requerimientos como la calidad o los límites medioambientales para la producción.
3. Mejorar los costes de los procesos de producción con el fin de hacer más prospero el negocio que incorpore esta nueva tecnología.

¹El resultado de una prueba ha sido indicado como exitosa, cuando, realmente, fue fallida.

4. Minimizar el stock de almacenamiento de productos para evitar realizar un trabajo innecesario y el coste adicional asociado por su almacenamiento.

En resumen, esta investigación podría convertirse en una mejora competitiva para una empresa. Pero no sólo esto; puede acabar ayudando tanto a las empresas, como, de una forma transitiva, a los consumidores (otras empresas o consumidores finales), que son los destinatarios de los productos manufacturados bajo este nuevo proceso de producción totalmente óptimo.

1.4 Mejora / solución propuesta

Dados los retos que acabamos de enumerar (para más información ver sección 1.2), trabajaremos en la eliminación de algunas de las limitaciones existentes en los sistemas MPC. Asimismo, estos retos nos inducen a seguir las tendencias recogidas por Morari et ál. [ML97]. En esta sección se pasará a definir la hipótesis inicial de la investigación. También se identifican y describen los objetivos principales, específicos y operacionales necesarios para superar esos retos, y la arquitectura de la solución global a presentar.

1.4.1 Hipótesis de trabajo

Debido a la falta de soluciones óptimas dentro del dominio de los procesos de producción, como se describió en la sección 1.1, se pretende realizar un avance del estado de la técnica con el afán de conseguir un proceso de producción totalmente optimizado para los parámetros que fueron identificados por los expertos de la producción. Como punto de partida se define la siguiente hipótesis fundamental:

Es posible modelar el flujo de negocio de un proceso productivo como una nube de conocimiento a través de la creación de un modelo predictivo de control híbrido (lineal y no lineal) basado en técnicas de aprendizaje automático actuales, manteniéndolo actualizado y llevando el control de la producción para la optimización de alguno de sus parámetros críticos.

1. Introducción

Realizamos la formulación de esta hipótesis con el objetivo de demostrar que existe la posibilidad de optimizar los procesos de producción a través de las técnicas que la informática ha ido compartiendo con la comunidad científica a lo largo de los años. Para ello, este desarrollo tiene que ser capaz de llevar a cabo una predicción multivariable de características y errores que se dan dentro del proceso productivo, y en caso de detectarlos, realizar el aviso a una tercera entidad (e.g. los operadores del sistema o un sistema SCADA [DS99]) que tenga la oportunidad de modificar los parámetros aplicados al proceso productivo manteniendo la producción dentro de una línea adecuada.

1.4.2 Objetivos

Teniendo en cuenta la hipótesis de trabajo definida hemos identificado y definido el siguiente objetivo principal:

Objetivo principal 1 *Con el fin de cubrir las necesidades que ya han identificado las empresas en cuanto a calidad, reducción de costes o eco-fabricación, así como la capacidad de obtener una mejora competitiva, se pretende diseñar, desarrollar, evaluar e implantar un sistema que sea capaz de readaptar la producción para cumplir con las restricciones definidas en el proceso.*

Este objetivo principal puede ser dividido en una serie de objetivos específicos que pasamos a enumerar.

Objetivo específico 1 *Desarrollar y evaluar un modelo general de predicción de defectos y características utilizando técnicas de aprendizaje automático.*

Objetivo específico 2 *Desarrollar y evaluar un modelo basado en técnicas de toma de decisiones y análisis de sensibilidades que permita ofrecer la información necesaria para que el proceso de producción sea modificado para mantener la producción dentro de las restricciones del proceso.*

Objetivo específico 3 *Desarrollar y evaluar una metodología que permita mantener el sistema totalmente adaptado en todo momento ante los posibles cambios que puedan darse en el proceso de producción.*

Nótese que los objetivos presentados acaban haciendo referencia a cada uno de los elementos que constituyen un MPC. El primero de ellos trata de identificar la manera en la que prever qué sucederá en el proceso productivo en un futuro cercano. El segundo es el objetivo que se centra en determinar las modificaciones que deben ser realizadas en los parámetros de producción con el fin de mantener el devenir de la planta sin la aparición de productos que no cumplan las restricciones del proceso. Y por último, el tercero de ellos es el objetivo específico que focaliza sus esfuerzos en determinar cómo mantener el sistema en concordancia con el proceso actual, es decir, evitar mantener unas bases de conocimiento obsoletas.

A fin de lograr la consecución de estos objetivos, hemos determinado una serie de objetivos operacionales, con un menor grado de abstracción, que quedan definidos de la siguiente forma:

Objetivo operacional 1 *Diseñar e implementar un marco de trabajo para la detección de defectos y características en las etapas iniciales del proceso productivo.*

Objetivo operacional 2 *Diseñar e implementar un sistema basado en un meta-clasificador que permita predecir la etapa de la planta en un periodo de tiempo $t + 1$ basado en la decisión tomada a través de múltiples predictores, maximizando la precisión y minimizando el error medido en tasas de falsos positivos y falsos negativos.*

1. Introducción

Objetivo operacional 3 *Diseñar e implementar un método de representación del conocimiento que sirva como base al sistema cuyo objetivo sea indicar los parámetros que deben ser introducidos en el proceso de producción para evitar anomalías.*

Objetivo operacional 4 *Diseñar e implementar un sistema que permita dar indicaciones de cómo configurar el proceso que se está llevando a cabo en la planta de producción.*

Objetivo operacional 5 *Diseñar e implementar un método que permita mantener el sistema actualizado frente a los cambios que se producen en la producción.*

Estos objetivos operacionales, desde un punto de vista totalmente relacionado con la investigación que aquí se presenta, muestran las tareas e indagaciones que hemos desarrollado para demostrar la hipótesis inicial.

1.4.3 Arquitectura global de la solución

Una vez definidos los objetivos de la investigación, mostramos a continuación la solución general que hemos planteado y que agrupa las diferentes tareas realizadas para la consecución de estos objetivos.

La Figura 1.1 ilustra la arquitectura desarrollada en esta investigación. Más concretamente, esta arquitectura es una composición de los siguientes módulos:

- **Lector de evidencias.** Elemento inicial del sistema y que se encarga de obtener qué es lo que está sucediendo en cada momento en la planta. Podemos considerar al lector de evidencias como el intermediario entre el MPC y la planta.
- **Preprocesador.** Es el elemento que se encargará de determinar las operaciones que hay que desarrollar para adaptar la evidencia recogida

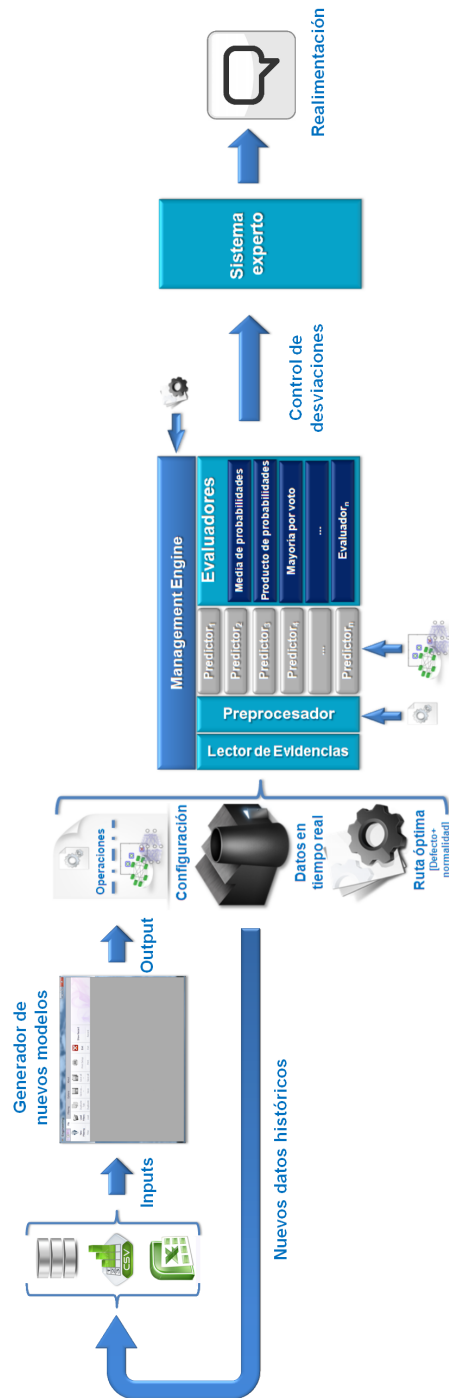


Figura 1.1: Arquitectura de la solución global para conseguir el desarrollo de un MPC híbrido y autoadaptativo.

1. Introducción

de la planta y mantener la coherencia con los modelos utilizados en la etapa de predicción.

- **Predictores.** Se generará un predictor por cada uno de los objetivos que hayan sido fijados en el arranque del sistema. Más concretamente, los predictores son los encargados de realizar el proceso de previsión de la planta. Con el ánimo de ser mucho más específicos, cada uno de los predictores utilizará diferentes modelos para la predicción de la misma característica.
- **Evaluador.** Éste es el sistema que se encargará de combinar los diferentes resultados obtenidos por cada uno de los modelos pertenecientes al predictor de una característica. Con el fin de poderse adaptar a diferentes procesos, se determina la necesidad de más de un método de unión, ya que para casos de uso diferentes la combinación óptima de clasificadores difiere entre ellos.
- **Sistema Experto.** Es el sistema encargado de determinar los cambios a realizar en el proceso productivo para que no se llegue a dar la situación anómala.
- **Controlador.** Se trata del elemento encargado de inicializar, configurar y orquestar todos y cada uno de los elementos que componen el sistema global. Es el núcleo del MPC. Además, este controlador será el encargado de determinar si la producción se está desviando de las restricciones impuestas por el proceso.
- **Generador de nuevos modelos.** Sistema concurrente que de forma periódica se encarga de regenerar los modelos que representan el conocimiento para evitar que el sistema quede obsoleto. Para poder llevar a cabo esta tarea, se han utilizado técnicas que priorizan las evidencias que han aparecido más recientemente.

El sistema funciona como se indica. En primer lugar, tras el arranque, las entradas recibidas para su etapa de inicialización son las siguientes: (i) los parámetros a monitorizar, (ii) la representación de la normalidad para cada uno de esos parámetros y (iii) las transformaciones aplicadas al conjunto de datos original.

Segundo. Una vez el controlador ha configurado y arrancado todos los sistemas, el lector de evidencias extrae la información actual de la planta modificándolos para adaptarlos a los modelos utilizados.

Tercero. Los resultados de los clasificadores son unidos mediante diversos métodos con el fin de crear una conciencia colectiva para la predicción de la situación de la planta.

Cuarto. Tomando las predicciones obtenidas se comprueba la existencia de una desviación de la normalidad. De ser así, es iniciada la etapa de adaptación de los parámetros de producción para que, realmente, no llegue a darse aquello que fue predicho.

Finalmente, existe una etapa concurrente a las anteriores que permite la generación de nuevos modelos adaptados al proceso que mantendrán el sistema actualizado a lo largo del tiempo (i.e., la requerida adaptación continua).

1.5 Desarrollo de la tesis

A continuación, se detalla la forma en la que ha sido desarrollado el trabajo de investigación presentado en este documento. En primer lugar enunciaremos la metodología de investigación empleada. En segundo lugar, identificaremos las métricas que serán utilizadas para poder realizar las etapas de evaluación y análisis.

1.5.1 Metodología de investigación

Con el fin de demostrar la hipótesis fundamental planteada (para más información ver sección 1.4.1) y mejorar el estado de la técnica actual en el campo de los MPC, se detalla a continuación la metodología que nos permitirá acercarnos a la consecución de los objetivos ya mencionados (ver sección 1.4). Así, más concretamente, la metodología utilizada es la siguiente:

1. **Identificación de la problemática y de los retos a superar:** el primero de los pasos de la metodología tiene el fin de escudriñar el trasfondo y el contexto del problema a solucionar. Es decir, se intenta tomar conciencia de la forma en la que afecta a la industria, qué es lo que se intenta solucionar y, si realmente lo que se cree como reto a solucionar es el verdadero reto o se encuentra encubierto y tiene repercusión directa en la sociedad.
2. **Adquisición de conocimiento:** el segundo de los pasos es la adquisición de conocimiento a alto nivel. Esta etapa es fundamental porque

1. Introducción

gracias a ella contaremos con la visión global necesaria en las etapas iniciales de la investigación. Posteriormente, cuando nos hayamos centrado y estemos trabajando en una temática mucho más concreta, deberemos realizar una segunda fase de adquisición de conocimiento. Ahora sí, ese proceso de adquisición será totalmente específico y estará dirigido a la ampliación del conocimiento sobre el proceso en concreto en el que se trabaja.

3. **División en retos:** basándose en la idea del *Divide y Vencerás*, en este paso de la metodología definiremos aquellos retos a afrontar. Gracias a esta etapa conseguiremos disminuir la dificultad de las tareas con las que nos enfrentaremos. Será una etapa clave, ya que si la división no se realiza correctamente puede poner en peligro el resultado final de la investigación. Para solucionar cada uno de los retos identificados, llevaremos a cabo los siguientes pasos:
 - *Adquisición de conocimiento específico:* una vez focalizado nuestro esfuerzo en un único aspecto, identificaremos la documentación más relevante sobre la temática a solucionar.
 - *Definición del experimento y las técnicas a utilizar:* llegados a este punto, definiremos qué se va a investigar, así como las técnicas a utilizar. Este paso marcará qué tipo de instrumentación va a ser necesaria para el desarrollo del experimento y la captura de las mediciones.
 - *Evaluación:* el siguiente de los pasos es aquél que nos permite llevar a cabo el experimento. Del mismo modo, durante su realización desarrollaremos el proceso de adquisición y ordenación de los datos resultantes. Los logros obtenidos en esta tarea serán utilizados como entrada en la siguiente de las etapas de la metodología.
 - *Análisis:* una vez finalizada la anterior etapa de experimentación, pasaremos a realizar el proceso de análisis sobre el problema específico. Más concretamente, se trata de realizar un primer procesamiento de la información recabada de la experimentación que, posteriormente, uniremos a todos los resultados obtenidos, sirviendo así de base para la interpretación global de los resultados.
4. **Análisis e interpretación de los resultados:** una vez encontrada la solución a todos los retos identificados, llevaremos a cabo la fusión de los resultados para poder discurrir las interpretaciones mediante el estudio

de relaciones. Más adelante, se pasará a formular de forma teórica los enunciados. Por último, se realizará la interpretación final de todos esos resultados.

5. **Difusión de los resultados:** en esta última tarea de la metodología se darán a conocer los resultados obtenidos a la comunidad científica, para así compartir el conocimiento, completando de esta forma el ciclo natural del proceso científico clásico.

1.5.2 Métricas para la evaluación

La evaluación desarrollada, y que presentamos en esta tesis doctoral, cuenta con un amplio recorrido en el área del aprendizaje automático. De este modo, las características y métricas en las que se centrarán los análisis de cada uno de los retos identificados a lo largo del proceso de la investigación (para más información sobre ellos ver la sección 1.2, así como para conocer más a fondo la metodología seleccionada véase la sección 1.5.1) serán aquellas que habitualmente son utilizadas en este campo y que se encuentran totalmente detalladas en [Bis06]. Más específicamente, a continuación se hará una breve descripción de las métricas que serán utilizadas a lo largo del proceso de evaluación y análisis de resultados.

- **Porcentaje de precisión:** con el fin de saber cómo de correcto y preciso es el modelo desarrollado, una de las primeras métricas a tener en cuenta es el porcentaje de precisión que es alcanzado. Esta métrica nos indicará la relación de instancias correctamente clasificadas frente a las incorrectas y representado mediante la utilización de una fracción de 100. En otras palabras, se indicará que n instancias de 100 han sido correctamente clasificadas.
- **Tasas de error:** para medir cómo de bueno es un clasificador no podemos utilizar únicamente su nivel de precisión, por ello, también tendremos en cuenta diferentes tasas de error que nos ayudarán a identificar el funcionamiento del comportamiento anómalo del modelo. Así, las tasas de error que utilizaremos son las siguientes:
 - *Error absoluto medio:* la tasa de error entre el conjunto de valores predichos X y el conjunto de valores reales Y (ambos con el tamaño del conjunto de datos de pruebas m). Esta tasa de error también

1. Introducción

es conocida como MAE (de la voz inglesa “*Mean Absolute Error*”) y se muestra en la ecuación 1.1.

$$MAE(X, Y) = \sum_{i=1}^m \frac{|X_i - Y_i|}{m} \quad (1.1)$$

- *Raíz del error cuadrado medio*: es una medida de uso frecuente para valorar las diferencias entre los valores pronosticados por el modelo y los valores observados a partir de los que se modela o estima. También es conocida como RMSE (del inglés “*Root Mean Square Error*”) y se muestra en la ecuación 1.2.

$$RMSE(X, Y) = \frac{1}{m} \cdot \sqrt{\sum_{i=1}^m (X_i - Y_i)^2} \quad (1.2)$$

- *Área bajo la curva de característica operativa del receptor*: la curva de característica operativa del receptor, o simplemente, curva ROC (del acrónimo inglés “*Receiver Operating Characteristic*”) es conocida como la representación gráfica de la sensibilidad frente a $(1 - \textit{especificidad})$ para un sistema clasificador binario según varía el umbral de discriminación. Sin embargo, y para nuestro proceso de experimentación, la interpretación que le vamos a dar es la de la representación que hay entre los falsos positivos y los falsos negativos. La medición la realizaremos a través del área que esa curva genera con el eje de abscisas, determinando que el valor óptimo que se podrá alcanzar es el de 1.

1.6 Estructura del documento

La presente tesis doctoral está organizada en los capítulos que se enumeran a continuación.

- **Capítulo 1. Introducción.** En este primer capítulo se introduce brevemente el contexto y la problemática existente en los procesos de manufacturación. Del mismo modo, se ilustra una de las soluciones que se están adoptando en el dominio empresarial –los Modelos Predictivos de Control– junto a la definición de cuáles son sus limitaciones y qué es lo

que se espera de este tipo de sistemas. Posteriormente, se pasa a enumerar los retos a los que se trata de dar solución a través del proceso de investigación detallado en esta tesis doctoral, además de determinar la hipótesis fundamental en la que se basa esta investigación. Junto a esto, el capítulo recoge los objetivos a cumplir. Además, se realiza la definición de la metodología que conducirá esta investigación, así como la definición de la arquitectura de la solución aportada en esta tesis doctoral.

- **Capítulo 2. Modelo Predictivo de Control (MPC) y otras aproximaciones.** En este capítulo se lleva a cabo una revisión de la literatura relacionada con el concepto de los Modelos Predictivos de Control. Más concretamente, se da a conocer qué es un MPC y la representación matemática del modelo general. Además, se desentrañan las singularidades del proceso de generación de un sistema de este tipo. Para ello, se realiza un análisis de las diferentes categorías en las que pueden englobarse estos sistemas y se ilustra cuál es el proceso de definición del modelo, así como las diferentes aproximaciones existentes para realizar las operaciones de control sobre una planta. También se realiza un breve estudio histórico sobre la evolución que han sufrido los sistemas MPC, tanto en entornos reales de producción como en el mundo científico-académico. Finalmente, se agrupan las limitaciones que tienen este tipo de sistemas y cuáles son las líneas de trabajo futuro que han sido identificadas, junto con otras aproximaciones para la mejora de los procesos productivos.
- **Capítulo 3. La fundición, un antiguo proceso rodeado de magia.** A pesar de que los sistemas MPC permiten llevar un control sobre cualquier tipo de planta de producción, para realizar la demostración empírica del funcionamiento del nuevo enfoque se ha seleccionado una planta de fundición de alta precisión. De esta forma, este capítulo muestra cómo es el proceso de producción llevado a cabo en una fundición, determinando las diferentes etapas en la que se divide una planta y cuáles son los materiales utilizados en el proceso. Asimismo, se identifican las variables que son relevantes para el proceso y se relacionan con las etapas de las que son extraídas. Finalmente, se analizan los defectos y características sobre los que se aplica el método descrito en esta investigación doctoral.
- **Capítulo 4. Predicción del estado estacionario $t + 1$.** En este capítulo se describen diferentes métodos para llevar a cabo la etapa de

1. Introducción

predicción de los estados estacionarios en un instante de tiempo futuro. En primer lugar hemos evaluado la utilización de métodos de aprendizaje supervisado con el fin de detectar dos defectos diferentes, los microrrechupes y los valores defectuosos en la carga máxima, demostrando que este nuevo enfoque es totalmente válido para ser incorporado como etapa inicial en los MPC. En segundo lugar, hemos evolucionado el sistema de predicción a través de la utilización de una serie de métodos para la combinación de múltiples clasificadores, permitiendo generar una inteligencia colectiva de clasificación de defectos con una naturaleza híbrida. Este nuevo enfoque también muestra interesantes resultados que indican que es la solución idónea para la eliminación de ciertas limitaciones inherentes a los MPC actuales.

- **Capítulo 5. Adaptando los modelos al proceso cambiante.** Los sistemas MPC tienen problemas para mantener los modelos actualizados. Este capítulo introduce un nuevo método que soluciona ese problema en el sistema MPC que proponemos en esta tesis doctoral. Concretamente, en primer lugar, hemos definido el método de adaptación que se basa en una técnica denominada compresión bayesiana basada en métodos de Montecarlo. Gracias a ella, somos capaces de envejecer las muestras históricas del proceso, dando una mayor importancia a las nuevas evidencias obtenidas del proceso ya modificado. En segundo lugar, hemos realizado la definición del proceso de adaptación automática para los modelos de aprendizaje automático utilizados en la predicción del estado estacionario $t + 1$. Finalmente, y con el fin de demostrar la utilidad de este método, hemos desarrollado los experimentos que nos permiten comprobar su funcionamiento. Específicamente, los experimentos se centran en comprobar los conjuntos reducidos obtenidos al aplicar esta técnica y testar si mantienen la misma distribución estadística que el conjunto de datos original en los dos defectos sobre los que venimos trabajando. Para completarlos, se prueba el método de adaptación basado en un aprendizaje acumulativo que utiliza la técnica desarrollada como envejecimiento de muestras. Los resultados alcanzados son altamente interesantes, mostrando que el método desarrollado puede ser la herramienta que solucione una de las mayores limitaciones con la que cuentan los sistemas MPC.
- **Capítulo 6. Generación de nuevos parámetros de producción.** A lo largo de este capítulo describiremos la solución adoptada para buscar los parámetros de producción más cercanos a lo que está sucediendo en la

planta pero con la seguridad de que evitamos los problemas identificados. Más concretamente, sentaremos las bases de la programación evolutiva y profundizaremos en el sistema que completa el MPC y que está fundamentado en la programación evolutiva paralela. Este nuevo sistema evoluciona hacia los sistemas de optimización basados en programación genética, al contar con una única solución para diferentes objetivos y procesos. Además, con el fin de solucionar los problemas de la planta de producción, el sistema de realimentación presentado hará uso de los sistemas predictivos híbridos presentados en el capítulo 4, consiguiendo así un sistema totalmente integrado y que incorpora las ventajas de los modelos híbridos. Finalmente, evaluaremos los resultados de aplicar esta aproximación al optimizar los valores del proceso productivo.

- **Capítulo 7. Conclusiones.** En el último de los capítulos realizamos una evaluación de los logros obtenidos en términos de comparación en el grado de consecución de la hipótesis y los objetivos. También, se discuten las aplicaciones de la investigación presentada y los campos de aplicación de la misma. Después, se lleva a cabo un debate sobre las limitaciones que presenta la solución adoptada, para más tarde, centrar las líneas de investigación futuras. Para terminar, se concluye el documento con una breve opinión sobre el devenir del futuro en el área de investigación seleccionado.

«La mejor manera de predecir el futuro es creándolo.»

Peter Ferdinand Drucker
(1909–2005)

2

Modelo Predictivo de Control (MPC) y otras aproximaciones

BAJO la denominación de Modelo Predictivo de Control (MPC, del inglés “*Model Predictive Control*”) se agrupan una serie de algoritmos capaces de llevar el control de un modelo de proceso en concreto. Este tipo de algoritmos trata de mantener la producción de una planta sobre una trayectoria ideal, alineando los parámetros de la producción con la intención de satisfacer las condiciones para las que fue diseñado originariamente el sistema. Sin embargo, a lo largo del proceso pueden darse ciertos cambios que hagan que el resultado final no sea el adecuado. Para solucionarlo, estos algoritmos son capaces de adelantar en el instante de tiempo t cómo se encontrará la planta en el instante de tiempo $t + 1$ de seguir así. En caso de detectar algún problema, el sistema se encargará de realizar la corrección de los parámetros manteniendo la producción a lo largo de esa trayectoria óptima. Originariamente este tipo de sistemas fueron desarrollados para satisfacer las necesidades específicas del control de centrales eléctricas y refinerías de petróleo. Sin embargo, en la actualidad la tecnología MPC se utiliza en una amplia variedad de áreas como en la producción química, el procesamiento de alimentos, la automoción, las aplicaciones

2. Modelo Predictivo de Control y otras aproximaciones

aeroespaciales, la metalurgia y la minería entre muchas otras.

Los avances en esta área han sido llevados a cabo tanto por las empresas, ya que tienen un gran interés, como por la comunidad científica. Es por ello que en este capítulo se muestra un estudio sobre los MPC centrándose en ambas partes, la académica y la industrial. Además se ahondará en los trabajos similares que únicamente utilizan técnicas de aprendizaje automático como base de sus optimizaciones.

El capítulo se organiza de la siguiente forma. La sección 2.1 introduce el concepto de los MPC a través de su definición y formulación. La sección 2.2 habla de las singularidades de los MPC. Asimismo, comienza a detallar la etapa de diseño y afinación de uno de estos sistemas, establece los diferentes tipos de MPC, el proceso de identificación de la tecnología a utilizar, y finalmente, muestra las diferentes tecnologías de control. Posteriormente, la sección 2.3 se centra en la historia de los MPC, dando detalles sobre los MPC lineales, las aproximaciones realizadas en el mundo académico y las implementaciones existentes en los entornos industriales actuales. La sección 2.4 proporciona una idea general de los procesos en los que se ha utilizado esta tecnología. Las secciones 2.5, 2.6 y 2.7 muestran la problemática actual de los sistemas MPC, las líneas futuras marcadas para su desarrollo, y otras aproximaciones existentes, respectivamente. Y por último, la sección 2.8 trata de resumir los conceptos recogidos a lo largo del capítulo.

2.1 Concepto general

2.1.1 Descripción de la estrategia MPC

Un Modelo Predictivo de Control (de la voz inglesa “*Model Predictive Control*”, MPC), es uno de los pocos métodos avanzados que tiene un impacto significativo en la ingeniería de control industrial. Este tipo de sistemas se aplican a la industria de los procesos [AT00, BCP00, AJGG⁺01] ya que permite: (i) el manejo de problemas de control multivariables de una forma natural, (ii) la gestión del fallo en los actuadores (ese determinado valor quedaría fijado como una restricción manejándose de forma automática y óptima en el algoritmo), (iii) determinar situaciones relacionadas con fases problemáticas del proceso, ya que se puede llegar a predecir el impacto de modificaciones del proceso actual viendo las salidas esperadas en el proceso y (iv) tener en cuenta la incertidumbre del modelo, lo que significa, que el controlador puede ser diseñado de manera que las especificaciones de rendimiento sean conocidas

a pesar de la variabilidad inherente en las operaciones de la planta (también conocido como el desajuste de plantas/modelo).

La literatura sobre los sistemas MPC es muy extensa, más concretamente, los retos teóricos y prácticos relacionados con la tecnología MPC que se resumen en varios artículos. No obstante, los tres documentos sobre MPC presentados en la conferencia del PCCCh en 1996 son un excelente punto de partida [LC97, May97, QB00]. Qin y Badgwell [QB00] presentaron una breve historia de la tecnología MPC y una encuesta sobre los MPC industriales, algoritmos y aplicaciones; algo que los profesionales en el control de la producción pueden encontrar especialmente útil. Camacho y Bordons resumen las propiedades teóricas de los algoritmos MPC en [CB04]. Morari y Lee debaten sobre el pasado, presente y futuro de la tecnología MPC [Fro94]. Kwon proporciona una lista muy extensa de las referencias en [Kwo94]. Además, podemos encontrar una visión mucho más reciente en [NMS00]. Por otra parte, varios libros han ido apareciendo sobre la tecnología MPC [BWG91, Soe92, Mos95] haciendo hincapié en las características de estos sistemas.

Los métodos para la predicción del control se desarrollan en torno a ciertas ideas comunes [Mac02, GSDD05] que son las siguientes:

- El uso explícito de un modelo para predecir las salidas del proceso en un horizonte temporal $t + 1$, es decir, situado en el futuro.
- Obtener una secuencia de control que minimice una función de coste, función que será el objetivo para el cual se encuentra optimizado el sistema.
- Aplicar señales de control calculadas en base a los resultados desplazados en el horizonte temporal hacia el futuro.

Los MPC, así como el amplio campo de los métodos de control, se desarrollan en torno a un conjunto de elementos básicos comunes que, a través de la modificación de los parámetros de configuración, permiten realizar el control con diferentes algoritmos. Principalmente, un MPC puede dividirse en los siguientes subsistemas. En primer lugar, el modelo de predicción, encargado de obtener la dinámica del proceso y de realizar las predicciones sobre el comportamiento futuro del sistema. En segundo lugar, la función objetivo (el coste) que será aquello que desplaza el modelo y el soporte para determinar las desviaciones del comportamiento del sistema en base a una trayectoria de referencia. Y finalmente, las propias señales de control, objetivo de la computación

2. Modelo Predictivo de Control y otras aproximaciones

y que mantendrán o modificarán el proceso de producción que se está llevando a cabo.

El éxito de la tecnología MPC como un paradigma de control de procesos se puede atribuir a tres factores importantes [ML97]. El primero y principal es la incorporación de un modelo de proceso explícito en el cálculo del control. Esto permite que el controlador, en principio, pueda tratar directamente con todas las características importantes de la dinámica del proceso. En segundo lugar, el algoritmo MPC considera el comportamiento de la planta en un horizonte de futuro temporal. Esto significa que los efectos de las perturbaciones se pueden prever y eliminar, permitiendo al controlador para la unidad de la planta se centre en dirigirla a lo largo de una trayectoria deseada. Por último, el controlador MPC considera las entradas del proceso, el estado y las limitaciones de la producción directamente en el cálculo del control de la planta. Esto significa que es casi improbable violar las restricciones del proceso, dando lugar a un control más estricto. La inclusión de las limitaciones es lo que distingue los MPCs de otros paradigmas de control de procesos.

Aunque los procesos de fabricación son intrínsecamente no lineales, la gran mayoría de las solicitudes MPC hasta la fecha se basan en modelos dinámicos lineales. Para estos casos, las medidas más comúnmente adoptadas son la respuesta a un impulso y los modelos derivados de la convolución. Hay varias razones que hacen que se hayan desarrollado este tipo de modelos. Primero, los modelos lineales empíricos pueden ser identificados de una manera directa a partir de los datos obtenidos de unas pruebas del proceso. Segundo, la mayoría de las aplicaciones hasta la fecha han sido utilizadas en procesos de refinería [QB00], donde el objetivo es, en gran medida, mantener el proceso en un estado de equilibrio deseado en lugar de moverse rápidamente de un punto de trabajo a otro. En otras palabras, el proceso productivo llevado a cabo en las refinерías se mantiene constante, mientras que en otros entornos, el proceso sufre múltiples cambios debido a la gran variedad de productos que son capaces de generar. Finalmente, usando un modelo lineal la consecución del objetivo “*cero defectos*” puede alcanzarse mediante la programación cuadrática, para la que es sencillo encontrar algoritmos fiables [Wri97]. Esto es importante porque el algoritmo de solución debe converger al mejor de los casos de una forma fiable y en un tiempo más que aceptable para poder ser realmente útiles en las aplicaciones de fabricación. Por estas razones, en muchos casos un modelo lineal proporciona la mayor parte de los beneficios posibles.

Sin embargo, hay casos en que los efectos no lineales son lo suficientemente importantes como para justificar el uso de la tecnología MPC no lineal [Hen98].

Estos incluyen al menos dos grandes categorías de aplicaciones:

- Regulador de problemas de control donde el proceso es altamente no lineal y está sujeto a grandes perturbaciones en numerosas ocasiones (por ejemplo para el control de pH).
- Los problemas de control en los puntos de trabajo cambian con frecuencia y se trabaja con una gama suficientemente amplia de la dinámica de procesos no lineales (fabricación de polímeros y síntesis del amoníaco entre otros).

Los temas como la viabilidad de la optimización de modelos, la estabilidad y el rendimiento han sido ampliamente estudiados, como así se atestigua en varios libros [BWG91, Soe92, Ber95, Cla96, CB97] y en otros tantos documentos¹. Además, se ha avanzado mucho en estas cuestiones para los sistemas no lineales [May97]. Sin embargo, para las aplicaciones prácticas aún quedan muchas preguntas que hacerse, incluyendo la confiabilidad y eficiencia del sistema. También, surge la aplicación de los MPC a sistemas híbridos integrando ecuaciones de la dinámica del proceso, los cambios en el proceso, las variables discretas, las condiciones de la lógica, las descripciones de heurística, y la priorización de restricciones; y así lo detallan Bemporad y Moran [BM99a].

Una cuestión fundamental de un MPC es su robustez frente a la incertidumbre del modelo y el ruido. Cuando decimos que un sistema de control es robusto, queremos decir que la estabilidad se mantiene y que las especificaciones de funcionamiento se cumplen para un determinado rango de variaciones del modelo y una clase de señales de ruido (rango de incertidumbre). Así, este tipo de problema ha sido abordado en el contexto de los MPC y recogido en [BM99b].

De esta forma, y debido a la naturaleza de esta estrategia de control, observamos que los MPC aportan unas importantes ventajas sobre otros métodos de control. Algunas de ellas son las siguientes [Bor00].

- Es de fácil uso por gente que no tiene amplios conocimientos en sistemas de control. Los conceptos son muy intuitivos y la forma en la que se realizan los ajustes también es relativamente simple. La complejidad se encuentra en otras etapas como, por ejemplo, el diseño.

¹Morari en 1994 indica que con una breve consulta existen alrededor de 128 referencias sobre el término *predictive control* entre los años 1991 y 1993. Posteriormente, una búsqueda similar retorna unas 2802 referencias entre los años 1991 y 1998. En la actualidad, esta búsqueda sigue aumentando.

2. Modelo Predictivo de Control y otras aproximaciones

- Se puede utilizar para controlar una gran variedad de procesos.
- Es muy útil para los sistemas multivariable.
- Se permite el uso de restricciones, que pueden ser incluidas durante el proceso de diseño del sistema.

Sin embargo, los MPC también tienen algunas desventajas, como por ejemplo, el alto coste computacional para obtener las leyes que controlan la planta. Pero el principal problema de esta estrategia radica en la dependencia de los modelos al sistema para el que fueron diseñados, es decir, se basa en el conocimiento previo del comportamiento del sistema, por lo que el rendimiento estará totalmente relacionado con la calidad de la representación de la planta.

2.1.2 Formulación de un MPC

En la mayoría de los casos presentados en la literatura, la formulación de los MPC es expresada en forma de espacio de estados [BM99b]. Más concretamente, siendo Σ el modelo de la planta a ser controlado a través de las ecuaciones de diferencias lineales de tiempo,

$$\Sigma : \begin{cases} x(t+1) = Ax(t) + Bu(t), & x(0) = x_0, \\ y(t) = Cx(t) \end{cases} \quad (2.1)$$

donde $x(t) \in \mathbb{R}^n, u(t) \in \mathbb{R}^m, y(t) \in \mathbb{R}^p$ denotan el estado, las entradas de control y la salida respectivamente. Y siendo $x(t+k, x(t), \Sigma)$ o, de forma más resumida, $x(t+k|t)$ aquello que define la predicción obtenida mediante la iteración del modelo mostrado en la ecuación 2.1 k veces desde el estado actual $x(t)$.

Sin embargo, y con el objetivo de presentar la estrategia de una forma genérica y simple, definimos

$$x_{k+1} = g(x_k, u_k) \quad (2.2)$$

siendo el mapeo de los estados $x_k \in \mathbb{X} \subseteq \mathbb{R}^n$ y las señales de control $u_k \in \mathbb{U} \subseteq \mathbb{R}^m$ para un sistema dado en el que $g : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ es la función del sistema de estados arbitrario y $k \in \mathbb{Z}_+$. Y donde

$$\mathbf{u}_k(x_k) \triangleq (u_{0|k}, u_{1|k}, \dots, u_{H_p-1|k}) \in \mathbb{R}^{H_p} \quad (2.3)$$

es la secuencia de entradas sobre una frontera de tiempo fijada H_p . Además, la secuencia de entradas admisibles con respecto al estado $x_k \in \mathbb{X}$ que es definido mediante

$$\mathcal{U}_{H_p}(x_k) \triangleq \{\mathbf{u}_k \in \mathbb{U}^{H_p} | \mathbf{x}_k \in \mathbb{X}^{H_p}\} \quad (2.4)$$

donde

$$\mathbf{x}_k(x_k, \mathbf{u}_k) \triangleq (x_{1|k}, x_{2|k}, \dots, x_{H_p|k}) \in \mathbb{X}^{H_p} \quad (2.5)$$

corresponde con la secuencia de estados generada al aplicar la secuencia de entrada definida en la ecuación 2.3 para el sistema de la ecuación 2.2 desde el estado inicial $x_{0|k} \triangleq x_k$, donde x_k es la medida de estimación del estado actual. Así, el enfoque del horizonte deslizante está basado en la solución del problema de optimización de bucle abierto (del inglés “*Open-loop Optimisation Problem*”, OOP) [BM99b]:

$$\min_{\{\mathbf{u}_k \in \mathcal{U}_{H_p}\}} J(\mathbf{u}_k, x_k, H_p) \quad (2.6a)$$

sujeto a

$$H_1 \mathbf{u}_k \leq b_1 \quad (2.6b)$$

$$G_2 \mathbf{x}_k + H_2 \mathbf{u}_k \leq b_1 \quad (2.6c)$$

donde $J(\cdot) : \mathbb{X}_f(H_p) \mapsto \mathbb{R}_+$ es la función de coste en el conjunto de estados factibles $\mathbb{X}_f(H_p) \subseteq \mathbb{X}$ [LHWB06]. H_p indica el horizonte de predicción o el horizonte de salida y G_2, H_i y b_i son las matrices de las dimensiones posibles. En el orden como han aparecido, la ecuación 2.5, $x_{k+i|k}$ indica la predicción del estado en el instante de tiempo $k+i$ hecho en k , partiendo desde $x_{0|k} = x_k$. Cuando $H_p = \infty$, la OOP lleva como nombre *problema de horizonte infinito*, mientras si $H_p \neq \infty$, la OOP se denomina *problema de horizonte finito*. Las restricciones que garantizan la estabilidad del sistema en el bucle cerrado deberían ser añadidas en las ecuaciones 2.6b y 2.6c.

Asumiendo que el OOP (ver ecuación 2.6) es factible para $x \in \mathbb{X}$, i.e., $\mathcal{U}_{H_p}(x) \neq \emptyset$, existe una solución óptima dada por la secuencia

$$\mathbf{u}_k^* \triangleq (u_{0|k}^*, u_{1|k}^*, \dots, u_{H_p-1|k}^*) \in \mathcal{U}_{H_p} \quad (2.7)$$

2. Modelo Predictivo de Control y otras aproximaciones

y posteriormente, los conjuntos de la filosofía de horizonte deslizante [Mac02, CB04]

$$u_{MPC}(x_k) \triangleq u_{0|k}^* \quad (2.8)$$

y haciendo caso omiso de las entradas calculadas a partir de $k = 1$ para $k = H_p - 1$, repetimos el proceso al completo en el siguiente paso temporal. De esta forma, la ecuación 2.8 es conocida dentro de la literatura de los MPC como la *ley de los MPC*. En resumen, el siguiente algoritmo describe brevemente el proceso de computación de la ley básica de un MPC.

Algoritmo 1: Computación de la ley básica de un MPC.

```
k=0
while funcionando do
   $x_{k+0|k} = x_k$ 
   $\mathbf{u}_k^*(x_k) \leftarrow$  resolver OOP 2.6
  Aplicar solo  $u_k = u_{K+0|k}^*$ 
   $k = k + 1$ 
end
```

2.2 Los Modelos Predictivos de Control a fondo

A continuación, en la presente sección se pretende indagar en el funcionamiento que tienen los MPC. Partiendo desde el punto de su definición y creación, pasando por etapas como la del modelado, identificación de la tecnología y cuáles son esas tecnologías para el control existentes. Nótese que todo aquello descrito en esta sección hace referencia a las técnicas habituales seguidas en entornos productivos reales. Sin embargo, y como se verá en la sección 2.5, estos métodos disponen de ciertas limitaciones que acaban promoviendo la presente investigación doctoral.

2.2.1 Diseño y afinación del control

El proceso de diseño y afinación de los sistemas MPC consta de los siguientes pasos [Inc93, Cor94, Inc95]:

- De los objetivos de control establecidos, definimos el tamaño del problema y determinamos las variables de control, las variables a manipular y

2.2 Los Modelos Predictivos de Control a fondo

las variables que no se pueden controlar.

- Probamos la planta de forma sistemática modificando las variables a manipular y las variables que no pueden ser controladas por el sistema pero sí configuradas. En ese momento se captura y se almacenan los datos en tiempo real para comprobar cómo se comportan las variables a controlar.
- Obtenemos el modelo dinámico, los principios fundamentales o los datos obtenidos de la prueba en la que empleamos un sistema de identificación.
- Configuramos el controlador MPC y establecemos los parámetros de configuración inicial.
- Probamos el controlador de forma *off-line* utilizando una simulación para comprobar el rendimiento de éste.
- Descargamos el controlador configurado en la máquina destino y ponemos a prueba las predicciones del modelo.
- Afinamos la puesta a punto según sea necesario.

Los paquetes MPC comerciales suelen ofrecer herramientas de software para ayudar en el diseño y control, el modelado y la simulación. Mayoritariamente, el esfuerzo debe dedicarse a la etapa de simulación para comprobar que el controlador tiene un rendimiento aceptable y la solidez que un sistema de control necesita. Por lo general se realizan pruebas para determinar que el comportamiento y las respuestas dadas se ajustan a las restricciones definidas. Sin embargo, habrá que terminar la afinación sobre la planta ya que la simulación no puede llegar a cubrir todas las eventualidades.

2.2.2 Modelos del proceso

El alcance técnico de los MPC está definido por el modelo del proceso y la forma en la que son utilizados. De esta forma, la primera dimensión en la que pueden ser clasificados son (i) los modelos *lineales* y (ii) *no lineales*. Pero por otra parte, los MPC también pueden ser clasificados según la forma en la que han sido generados. Más concretamente, encontramos otra división entre (i) los modelos *empíricos* derivados exclusivamente de los datos de prueba extraídos del proceso de manufacturación y (ii) los modelos *basados en principios fundamentales* generados a partir de una formulación matemática para la

2. Modelo Predictivo de Control y otras aproximaciones

representación del proceso de manufacturación. Mientras que los primeros de ellos tienen un desarrollo mucho más económico y no pueden ajustarse al proceso más allá de los datos que han sido utilizados para generarlos, los segundos tienen un desarrollo mucho más costoso y representan el proceso de una forma mucho más general. No obstante, este tipo de MPC son calibrados mediante la utilización de un conjunto de datos de prueba. Cabe destacar que la mayoría de los productos comerciales existentes se decantan por la generación de modelos empíricos.

Los MPC pueden ser clasificados en cualquiera de las cuatro combinaciones posibles, combinando los métodos de generación y la naturaleza del problema. A continuación se va a pasar a describir cada uno de los modelos en más detalle.

2.2.2.1 Modelos no lineales basados en principios fundamentales

No existen una gran cantidad de modelos MPC que pertenezcan a este tipo, únicamente dos tipos de aproximaciones [QB00]. La primera es la dedicada a aplicaciones en las que la monitorización principal son la masa y los balances de energía. La peculiaridad de este modelo es que ciertos parámetros del modelo que son desconocidos, como factores de transferencia térmica y las constantes de las reacciones cinéticas, son estimados de forma externa sobre un conjunto de datos de prueba (apoyándose en los métodos empíricos para su funcionamiento final) o de forma interna al modelo a través de la utilización de filtros de Kalman extendidos. Las aplicaciones típicas que utilizan este modelo se componen de entre 10 y 100 ecuaciones algebraicas diferenciales.

En segundo lugar, existe otra segunda aproximación mucho más genérica que puede ser utilizada en diferentes dominios. En este caso se trata de derivar el modelo de los cambios temporales que se dan. A pesar de ello, hay que tener especial cuidado en sistemas de producción que sean muy rígidos en sus configuraciones ya que pueden encontrarse problemas de ajuste el modelo.

2.2.2.2 Modelos lineales empíricos

Este tipo de modelos han sido utilizados en la mayoría de las soluciones MPC actuales. Por eso, no es una sorpresa que muchos de los modelos comerciales estén basados en este tipo de modelos. Una gran cantidad de modelos lo implementan pero todos pueden ser derivados de la ecuación 2.9 [CB97, QB03].

2.2 Los Modelos Predictivos de Control a fondo

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}_u\mathbf{u}_k + \mathbf{B}_v\mathbf{v}_k + \mathbf{B}_w\mathbf{w}_k \quad (2.9a)$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{u}_k + \xi_k \quad (2.9b)$$

De forma equivalente, estos modelos pueden ser representados mediante una división de fracciones [Kai80]:

$$\mathbf{y}_k = [\mathbf{I} - \Phi_y(q^{-1})]^{-1}[\Phi_u(q^{-1})\mathbf{u}_k + \Phi_v(q^{-1})\mathbf{v}_k + \Phi_w(q^{-1})\mathbf{w}_k] + \xi_k \quad (2.10)$$

donde q^{-1} es un operador de desplazamiento hacia atrás. Con la aproximación de la identificación del error de la salida [Lju99] se puede minimizar el error $\xi_{k'}$, lo que resulta ser una estimación de un parámetro no lineal. También, multiplicando $[\mathbf{I} - \Phi_y(q^{-1})]$ en ambos lados de la ecuación se obtiene un modelo autorregresivo con entradas exógenas (ARX, acrónimo extraído del nombre inglés “*AutoRegressive model with eXogenous inputs*”),

$$\mathbf{y}_k = \Phi_y(q^{-1})\mathbf{y}_k + \Phi_u(q^{-1})\mathbf{u}_k + \Phi_v(q^{-1})\mathbf{v}_k + \Phi_w(q^{-1})\mathbf{w}_k + \zeta_k \quad (2.11a)$$

donde

$$\zeta_k = [\mathbf{I} - \Phi_y(q^{-1})]\xi_k \quad (2.11b)$$

En esta aproximación se intenta minimizar ζ_k , que se trata de un ruido de color¹ a pesar de que la medición de ξ_k se trata de un ruido blanco².

Sobre este modelo, se puede agregar una nueva opción, emplear el modelo de Bob-Jenkins³ [Van83] al realizar la agrupación de los términos de error en uno único, ε_k :

$$\mathbf{y}_k = [\mathbf{I} - \Phi_y(q^{-1})]^{-1}[\Phi_u(q^{-1})\mathbf{u}_k + \Phi_v(q^{-1})\mathbf{v}_k] + [\Theta_\varepsilon(q^{-1})]^{-1} + \Phi_\varepsilon(q^{-1})\varepsilon_k \quad (2.12)$$

¹Aunque el ruido es una señal aleatoria puede tener ciertas características o propiedades estadísticas. Las variaciones en la densidad espectral hacen que los ruidos puedan tener diferentes tipos de colores, en otras palabras, comportamientos fácilmente clasificables en los que se hace un símil entre su longitud de onda y el espectro visible.

²Es un ruido cuya densidad espectral de potencia es plana, es decir, contiene todas las frecuencias y todas tienen a misma potencia.

³Modelo generado mediante el método de George Box y Gwilym Jenkins que trata de obtener una fórmula que representa los patrones históricos (series temporales) que son considerados como estacionarios.

2. Modelo Predictivo de Control y otras aproximaciones

En caso de tener que tratar con un sistema estable, es decir, un sistema en el que no se produce una elevada cantidad de elementos que obliguen al sistema a tener que estar cambiando de un tipo de monitorización a otro rápidamente, sino que tiene que mantener el control de comportamiento de la planta constante en todo momento, se puede ajustar la ecuación 2.10 de la siguiente forma:

$$\mathbf{y}_k = \sum_{i=1}^{N_u} \mathbf{H}_i^u \mathbf{u}_{k-i} + \sum_{i=1}^{N_v} \mathbf{H}_i^v \mathbf{v}_{k-i} + \sum_{i=1}^{N_w} \mathbf{H}_i^w \mathbf{w}_{k-i} + \xi_k \quad (2.13)$$

Por último, otro de los modelos alternativos es el que se conoce como el modelo de respuesta de paso finito (FSR, de la voz inglesa “*Finite Step Response*”) [CMH83] que queda definido de la siguiente forma:

$$\begin{aligned} \mathbf{y}_k = & \sum_{i=1}^{N_u-1} \mathbf{S}_i^u \Delta \mathbf{u}_{k-i} + \mathbf{S}_{N_u}^u \mathbf{u}_{k-N_u} + \\ & \sum_{i=1}^{N_v-1} \mathbf{S}_i^v \Delta \mathbf{v}_{k-i} + \mathbf{S}_{N_v}^v \mathbf{v}_{k-N_v} + \\ & \sum_{i=1}^{N_w-1} \mathbf{S}_i^w \Delta \mathbf{w}_{k-i} + \mathbf{S}_{N_w}^w \mathbf{w}_{k-N_w} + \xi_k \end{aligned} \quad (2.14)$$

donde $\mathbf{S}_j = \sum_{i=1}^j \mathbf{H}_i$ y $\mathbf{H}_i = \mathbf{S}_i - \mathbf{S}_{i-1}$.

2.2.2.3 Modelos no lineales empíricos

Habitualmente son utilizados dos tipos de modelos no lineales empíricos [Hen98]. El primero de ellos utiliza un modelo lineal con tiempo discreto para cada uno de los estados dinámicos, completándolo con una ecuación de salida que incluye un término lineal sumado a otro no lineal:

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}_u \mathbf{u}_k + \mathbf{B}_v \mathbf{v}_k + \mathbf{B}_w \mathbf{w}_k \quad (2.15a)$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{u}_k + \mathbf{N}(\mathbf{x}_k, \mathbf{u}_k) + \xi_k \quad (2.15b)$$

Con este tipo de modelos únicamente pueden ser controlados los procesos estables, por lo que, los eigenvalores de \mathbf{A} deben estar estrictamente dentro del círculo unitario. Por otra parte, la función no lineal \mathbf{N} es obtenida

2.2 Los Modelos Predictivos de Control a fondo

mediante una red neuronal artificial. Además, ya que el vector de estado \mathbf{x} no está necesariamente limitado a variables físicas, este modelo no lineal se considera más general que la no linealidad de las mediciones. Por ejemplo, en [SBGZ98] se muestra que un modelo dinámico lineal seguido por un mapeo estático no lineal puede representarse de esta forma. Así, se llega a afirmar que este tipo de modelos no lineales puede aproximar, en cualquier momento, los procesos no lineales discretos con un ajuste de memoria con cierto nivel de envejecimiento¹.

En la segunda aproximación, descrita en [SBGZ98, PSRJG00], se utilizan modelos no lineales de entrada-salida. Sin embargo, para que la tarea de identificación de los modelos sea mucho más sencilla se emplea un modelo estático no lineal superpuesto a un modelo dinámico lineal. La descripción recogida por Piche et ál. [PSRJG00] tienen en cuenta una serie de entradas y salidas simples, aunque el concepto es fácilmente aplicable para un modelo de entradas y salidas múltiples. Así, se supone que la entrada del proceso puede descomponerse en dos porciones diferentes: (i) un estado estacionario que obedece a un modelo no lineal estático y (ii) la desviación que sigue el modelo dinámico. Con lo que, para cualquier entrada u_k y salida y_k , las variables que representan la desviación pueden ser calculadas como se muestra en la ecuación 2.16

$$\delta u_k = u_k - u_s \quad (2.16a)$$

$$\delta y_k = y_k - y_s \quad (2.16b)$$

donde u_s e y_s son los valores del estado estacionario para la entrada y la salida respectivamente. Éstos cumplen la siguiente relación no lineal general:

$$y_s = h_s(u) \quad (2.17)$$

Por otra parte, la desviación sigue una relación dinámica lineal de segundo orden como la siguiente:

$$\sum_{i=1}^2 a_i \delta y_{k-i} + b_i \delta u_{k-i} \quad (2.18)$$

¹El envejecimiento de la memoria es una técnica que permite mantener ajustado el modelo al proceso, haciendo que aquellos hechos que sucedieron en el pasado vayan perdiendo importancia dentro del mismo.

2. Modelo Predictivo de Control y otras aproximaciones

El proceso de identificación del modelo (para más información diríjase a la sección 2.2.3) se basa en el hecho de que para el modelo dinámico lineal se va a aplicar un conjunto de datos de prueba recogidos de una serie de pulsos con los que se chequeó el proceso, mientras que para el modelo estático no lineal se construye una red neuronal artificial sobre los datos históricos del proceso productivo.

La utilización de este modelo compuesto en el momento de realizar el control puede describirse como sigue. Basado en la salida deseada del objetivo y_s^d , un programa de optimización no lineal calcula los mejores valores de entrada y salida para u_s^f e y_s^f utilizando para ello el modelo estático no lineal. A lo largo del proceso de cálculo, la ganancia estática no lineal es aproximada mediante una interpolación de las ganancias de los estados estacionarios inicial y final,

$$K_s(u_k) = K_s^i + \frac{K_s^f - K_s^i}{u_s^f - u_s^i} \delta u_k \quad (2.19)$$

donde u_s^i y u_s^f son los valores del estado estacionario en el momento actual y el siguiente respectivamente; y

$$K_s^i = \left. \frac{dy_s}{du_s} \right|_{u_s^i} \quad (2.20a)$$

$$H_s^f = \left. \frac{dy_s}{du_s} \right|_{u_s^f} \quad (2.20b)$$

las cuales son evaluadas utilizando el modelo estático no lineal. Tanto en K_s^i como en K_s^f se pueden aplicar límites mediante la sustitución de la ganancia aproximada en la ecuación 2.19 por los rendimientos del submodelo lineal de la siguiente forma:

$$\delta y_k = \sum_{i=1}^2 a_i \delta y_{k-i} + \bar{b}_i \delta u_{k-i} + g_i \delta u_{k-i}^2 \quad (2.21a)$$

donde

$$\bar{b}_i = \frac{b_i K_s^i (1 - \sum_{j=1}^n a_j)}{\sum_{j=1}^n b_j} \quad (2.21b)$$

$$g_i = \frac{b_i (1 - \sum_{j=1}^n a_j) K_s^f - K_s^f}{\sum_{j=1}^n b_j u_s^f - u_s^i} \quad (2.21c)$$

2.2 Los Modelos Predictivos de Control a fondo

El propósito de esta aproximación es la de reducir la complejidad computacional durante el proceso de cálculo de los valores de control.

Como se puede observar, los valores objetivo del estado estacionario son calculados desde un modelo estático no lineal, mientras que los movimientos para el control dinámico se calculan en base a un modelo de segundo grado que se muestra en la ecuación 2.21a. No obstante, los coeficientes del modelo cuadrático, por ejemplo la ganancia local, cambian de una ejecución a la siguiente, simplemente porque estos son reajustados para que coincidan con la ganancia local del modelo no lineal estático. Esta estrategia de aproximación puede ser interpretada como una linealización sucesiva en los estados inicial y final, seguidos por una interpolación lineal de las ganancias linealizadas. Este modelo asume que la dinámica del proceso sigue siendo lineal en todo el rango de operación. Pero, la dinámica asimétrica (por ejemplo, diferentes constantes de tiempo locales) no puede ser representada por este modelo.

2.2.2.4 Modelos lineales basados en principios fundamentales

Este tipo de procesos no se encuentran representados de ninguna forma en las soluciones comerciales o académicas de los MPC. De esta forma, queda abierta una nueva línea de investigación dentro del desarrollo de modelos lineales basados en los principios fundamentales extraídos del proceso productivo.

2.2.3 Modelado del MPC e identificación de la tecnología

Existen diversos modelos, dependiendo de la tecnología adoptada en cada una de las plantas de producción. Del mismo modo, los diferentes vendedores de sistemas MPC tienen sus propias formas de trabajo a la hora de realizar el proceso de modelado y de identificación (para más información diríjase a la Tabla 3 en [QB03]). No obstante, la mayoría de las ocasiones los modelos son desarrollados utilizando los datos de respuesta obtenidos del proceso. Para ello, se diseñan cuidadosamente una serie de entradas que estimulen al sistema para encontrar las repuestas válidas y útiles en la generación del modelo que se está buscando. A continuación se describirán los procesos que son llevados a cabo: (i) protocolos de pruebas, (ii) identificación de modelos lineales e (iii) identificación de modelos no lineales.

En primer lugar, es necesario la generación de una serie de señales que permitan a los expertos en el dominio construir el modelo que se ajuste al proceso. Para conseguir su objetivo tienen que desarrollar un estudio en el

2. Modelo Predictivo de Control y otras aproximaciones

que determinarán cuáles son las señales a generar para que se estimule el sistema consiguiendo alcanzar tanto situaciones que cuentan con una baja probabilidad de suceder como las de mayor probabilidad. Como bien indican Qin et ál. en su estudio de MPC industriales [QB03], los vendedores de este tipo de sistemas aseguran que ésta es la tarea más importante a la hora de realizar la generación del modelo. Y tanta es su complejidad que, para llegar a desarrollar la prueba que permita obtener el modelo, hay que establecer una prueba inicial que permita ajustar la prueba final. Las razones por las que es imprescindible esta prueba inicial son varias: (i) se necesita buscar la forma de ajustar los instrumentos existentes y los controladores de la planta en relación a las variables que pueden ser modificadas, (ii) se debe obtener el comportamiento (i.e., tiempos, modificaciones que van sufriendo, entre otros) de las variables que únicamente pueden ser monitorizadas pero que sí que tienen su influencia sobre el comportamiento de la planta, y (iii) se obtienen datos para la identificación del comportamiento inicial del proceso.

Así, las pruebas se mantienen en la planta durante 24 horas al día utilizando ingenieros especializados en el dominio para la monitorización de los resultados que se están obteniendo. En el proceso, se fijan los valores de una de las variables sobre las que podemos operar, siempre y cuando el proceso pueda ser determinado fácilmente como lineal, o varias (junto con todas las combinaciones posibles) para procesos que son no lineales mientras van modificándose el resto de las variables, es decir, éstas pasarán por todos los estados posibles. Esto será repetido alrededor de entre 8 y 15 veces para cada una de las variables y se realizará durante un periodo de entre 5 y 15 días. Se llevan a cabo tantas repeticiones con el fin de determinar las tasas de ruido en el proceso productivo y su monitorización. Cabe destacar que durante los procesos de pruebas los operarios únicamente podrán intervenir en el proceso en el momento en el que se alcancen situaciones críticas.

Y en último lugar, seguido a los procesos de pruebas se comienza con el proceso de identificación del modelo. Como ya ha sido comentado hay dos tipos de modelos (i.e., lineales y no lineales) por lo que este paso es el crucial para determinar por qué tipo de estrategia MPC hay que declinarse.

Por una parte, si sobre el proceso productivo en el que se va a trabajar es lineal, hay que llevar a cabo la identificación del modelo que lo representa. En este caso, la estimación de parámetros del modelo tiene que basarse en que el sistema MPC debe minimizar el criterio de los mínimos cuadrados¹ usando

¹Método para la medición de modelos matemáticos, no deterministas pero sí probabilísticos, con el fin de precisar cómo de ajustado está el modelo a la realidad.

2.2 Los Modelos Predictivos de Control a fondo

la aproximación a la ecuación de errores o la aproximación a los errores en la salida [Lju99]. La diferencia entre ambas reside en que mientras que la primera realiza las mediciones en las salidas y las utiliza para realimentar al modelo, la segunda de ellas estima los valores de las salidas y las utiliza para realimentar al modelo. Con el fin de conseguirlo, se realiza una identificación del subespacio en el que se representa el proceso [Lar90, Lju99]. Este método está muy extendido ya que muchas de las soluciones comerciales lo utilizan o basan su nueva aproximación en él. Como alternativa a este último modelo, existe la aproximación de tres pasos utilizada en otros modelos comerciales como el RMPCT (para más información diríjase a la sección 2.3.1.6): (i) la generación de un modelo de Bob-Jenkins [Van83] u otro diferente siguiendo la descomposición de Cholesky¹ [DKA01], (ii) ajustar el modelo identificado mediante cualquiera de los dos métodos para ajustarlo a un orden menor a través del método Gauss-Newton² aplicado a la salida de error y (iii) terminar convirtiendo el modelo a las funciones de la transformada de Laplace.

Por otra parte, en el caso de haber determinado que el proceso sobre el que se va a actuar es de carácter no lineal, se debe partir de la premisa de que la selección de la representación no lineal no es el problema. De hecho, las soluciones existentes utilizan o una representación polinomial o una red neuronal artificial, sino más bien la identificación y selección de un algoritmo robusto y fiable. Una de las soluciones tomadas para solucionarlo es el algoritmo de identificación discutido en [ZGS98], en el que se genera un modelo para cada una de las salidas de forma separada. En otras palabras, para un proceso que tiene m_y variables de salida, el modelo final será construido a través de m_y submodelos dedicados a cada variable. Más concretamente, el procedimiento que se lleva a cabo para identificar cada submodelo es el siguiente:

1. Se especifican una serie de constantes para cada par de entrada-salida. Posteriormente, se pasa a aplicar una serie de filtros de primer orden o construir un modelo de Laguerre³ para cada entrada [ZGS98, SBGZ98]. Los estados de los filtros generarán el vector de estado definido como x .

¹Método descubierto por el matemático André-Louis Cholesky que muestra la posibilidad de descomponer una matriz simétrica definida positiva en el producto de dos matrices: una matriz triangular inferior y la transpuesta de la matriz triangular inferior calculando la primera de ellas a través del triángulo de Cholesky sobre la matriz original. Este método es utilizado para la resolución de sistemas de ecuaciones matriciales.

²Método utilizado para la resolución de problemas no lineales sobre mínimos cuadrados.

³Método que a través de la utilización de una serie de expresiones de tiempo mucho más apropiadas, extraídas de las redes de Laguerre, mejoran el enfoque tradicional de ampliar las funciones de transferencia para la obtención de modelos de predicción con un orden

2. Modelo Predictivo de Control y otras aproximaciones

2. En el segundo de los pasos, es construido un modelo lineal estático para cada salida $\{y_j, j = 1, 2, \dots, m_y\}$ teniendo en cuenta el vector de estado x y utilizando la técnica de mínimos cuadrados parciales (PLS, de la voz inglesa *Partial Least Square*).
3. Posteriormente, se realiza una reducción del modelo identificado en los pasos 1 y 2 mediante el análisis de componentes principales (PCA, del inglés *Principal Component Analysis*) y el equilibrio interno para eliminar variables de estado alineadas.
4. El modelo reducido obtenido del paso anterior es reordenado en un modelo de estado-espacio (A, B) que se utiliza para la generación de una secuencia de estados $\{x_k, k = 1, 2, \dots, K\}$. Entonces, si el modelo converge, es decir, no se pueden llevar a cabo más reducciones del modelo, se pasa al siguiente de los pasos. En caso contrario, se vuelve al punto 2 con el fin de seguir reduciendo el modelo.
5. Un nuevo modelo PLS es construido empleando el vector de estado x y la salida y_j . Así, los coeficientes del modelo PLS pasarán a quedar representados en la matriz C .
6. Finalmente, se construye el modelo mediante una red neuronal artificial utilizando, en primer lugar, los factores latentes del modelo PLS obtenidos en el paso anterior y los residuales de la salida y_j . Por consiguiente, se genera el mapa estático no lineal $g_j(x)$. La utilización de los factores latentes en lugar de los vectores de estado tiene lugar con el fin de mejorar la robustez del método de entrenamiento de la red neuronal artificial y reducir así el tamaño de la misma.

2.2.4 Tecnología de control MPC

Las tecnologías de control MPC son las encargadas de dirigir el proceso desde un estado estacionario a otro estado diferente. Como se puede comprobar, este es el paso final que todo sistema MPC debe llevar a cabo para conseguir que la planta se mantenga produciendo dentro de lo que es considerado como la normalidad. Los objetivos principales de los controladores, según su orden de importancia, son los siguientes [CB04]: (i) prevenir que las entradas y salidas violen las restricciones definidas para el proceso productivo en concreto, (ii)

mucho menor para casos de muestreo rápido y/o dispersión en las constantes temporales [Wah91].

2.2 Los Modelos Predictivos de Control a fondo

monitorizar las variables sobre las que no puede ejercer control con el fin de conocer en todo momento el estado en el que se encuentran, (iii) dirigir las variables sobre las que puede ejercer control con el fin de calcular los valores óptimos, (iv) evitar cambios bruscos en las variables sobre las que puede ejercer el control y (v) cuando las señales que llegan de la planta o los actuadores sobre los que poder realizar las operaciones fallan, debe mantener el control lo máximo posible.

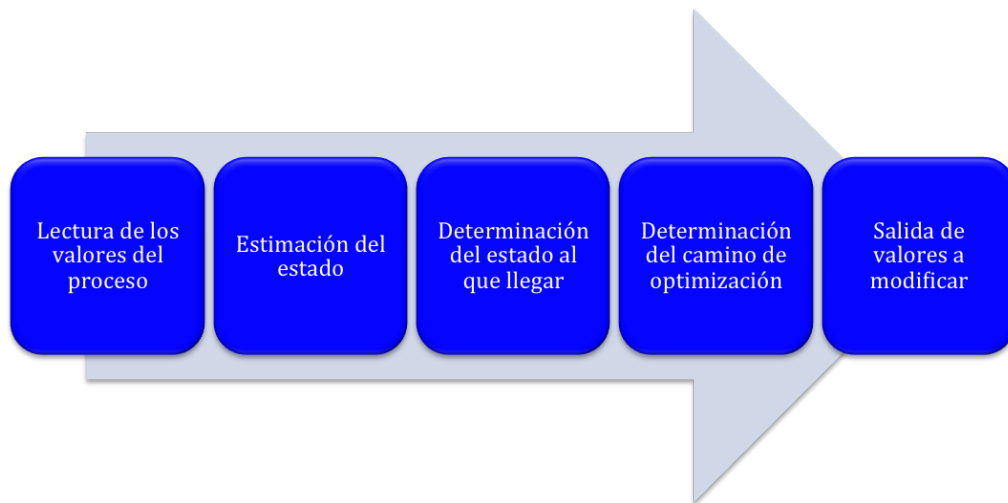


Figura 2.1: Flujo de trabajo general que siguen los controladores MPC para cada una de las ejecuciones que llevan a cabo.

La traducción de los objetivos comentados en una formulación matemática es lo que define el carácter de cada uno de los controladores. Así, como en cualquier problema de diseño, existen diversas soluciones para desarrollar controladores para sistemas MPC. Sin embargo, el flujo de trabajo básico de todos es el siguiente [CB04] (ilustrado en la Figura 2.1). En primer lugar, el controlador realiza las lecturas tanto de las variables sobre las que puede ejercer el control, como de las variables que únicamente puede monitorizar. Normalmente, además de las mediciones (valores numéricos), cada una de las señales lleva una información adicional que intenta comunicar el estado de funcionamiento del sensor que hace las lecturas. Nótese que el envío de toda esta información, concretamente el primero de los pasos, está fuera del alcance del MPC y que dependerá de la tecnología disponible en la planta de producción. Los siguientes pasos del proceso de control responden a las siguientes preguntas: (i) ¿En qué estado se encuentra el proceso ahora?, y de seguir así, ¿dónde se encontraría? (ii) ¿Dónde debería encontrarse el proceso de producción? (iii)

2. Modelo Predictivo de Control y otras aproximaciones

¿Cuál es la mejor manera de dirigir el proceso hasta el lugar en el que debe de encontrarse?

En las próximas secciones se llevará a cabo una descripción mucho más detallada de los diferentes métodos existentes para cada una de las etapas del control de sistemas MPC.

2.2.4.1 Estimación del estado

En esta etapa, el controlador debe tomar las medidas que han sido recogidas de la planta para determinar el estado actual en el que se encuentra la planta y que sucedería en un instante de tiempo $t + 1$. Como bien recoge Qin et ál. [QB03] en este punto es donde la industria de la tecnología MPC ha tenido mayores problemas. Esto se debe a que no se ha realizado la separación del problema general en sus dos componentes naturales, siendo la primera, la estimación del estado; y la segunda, el control del estado. La mayoría de los productos no incorporan la idea de estado y se centran en cambios *ad-hoc* sobre esquemas de polarización para así conseguir dar una retroalimentación. Trabajar de este modo tiene varias implicaciones:

- Otro tipo de medidas que podrían mejorar la procesión del modelo no son fácilmente incorporables a la estructura de control.
- Se necesita de más esfuerzo para controlar las combinaciones lineales de los estados del proceso o para controlar elementos que no han sido medidos.
- Las opciones del modelo para el control de perturbaciones están muy limitados.
- Se deben introducir modificaciones *ad-hoc* para arreglar el modelo en sistemas que son inestables.
- La medida del ruido debe ser llevada a cabo de una manera subóptima.

Para la realización de esta etapa existen modelos que utilizan los filtros de Kalman, ya sean extendidos o no, con el fin de la gestión del posible ruido, o perturbaciones, que se haya producido en la planta [MB88, KSH00], mientras que el resto no repara en solucionar estos temas en los procesos de determinación de los estados actuales y futuro. En los procesos que son más estables,

2.2 Los Modelos Predictivos de Control a fondo

es habitual la utilización de un modelo basado en la comparación de estados, el actual \mathbf{y}_k^m y el predicho \mathbf{y}_k de la siguiente forma:

$$\mathbf{b}_k = \mathbf{y}_k^m - \mathbf{y}_k \quad (2.22)$$

El término \mathbf{b}_k es añadido al modelo con el fin de utilizarlo en las sucesivas predicciones:

$$\mathbf{y}_{k+j} = \mathbf{g}(\mathbf{x}_{k+j}, \mathbf{u}_{k+j}) + \mathbf{b}_k \quad (2.23)$$

Esta forma de trabajo es equivalente a asumir que la perturbación llega a la salida y se mantiene constante durante todo el futuro [ML91, LMG94]. Por otra parte, Muske et ál. analizaron que esta suposición de la correspondencia con los filtros de Kalman se traducía en una ganancia $\mathbf{K}_j = [\mathbf{0I}]$, lo que significaba que no había una realimentación para el estado del proceso estimado, y sí una alimentación de las perturbaciones de salida. [MR93]. También, mostraron que es posible la utilización de otros muchos modelos para la gestión de las perturbaciones. Más concretamente, mostraron que un modelo de perturbaciones de salida constante hace que lentamente se vayan rechazando perturbaciones en la entrada (algo que también queda recogido en [Shi94]). Sin embargo, este problema puede ser solucionado mediante la construcción de un modelo de perturbaciones de entrada que sea similar al interno. Viendo la variedad de modelos existentes para esta etapa, podemos concluir que los MPC tienen características tan dispares que les permiten funcionar en una amplia variedad de escenarios.

2.2.4.2 Determinación del estado al que llegar

Los sistemas MPC suelen disponer de una etapa dedicada únicamente a determinar el estado óptimo al que se pretende llegar en cada ciclo de control para que la producción de la planta sea la correcta. La razón fundamental es que en cada ejecución los objetivos pueden variar debido a las perturbaciones que haya en el proceso. Normalmente, la optimización de este problema suele basarse en mantener los estados lo más cerca posible de las optimizaciones más económicas computacionalmente hablando sin que se lleguen a violar las restricciones de entrada y salida. Las estimaciones de las perturbaciones determinadas en el paso anterior aparecen aquí de forma explícita, pero en este caso pueden llegar a ser eliminadas del proceso. Rao y Rawlings describen una formulación del problema en [RR99], según la cual, para conseguir reali-

2. Modelo Predictivo de Control y otras aproximaciones

zar esta etapa, utilizan un modelo de estado estacionario que puede provenir de linearizar un modelo integral no lineal en cada ejecución del control o, por otra parte, de la versión del modelo estacionario que se utiliza en el paso siguiente (i.e., el paso para determinar cómo llegar al estado final deseado).

Otra de las aproximaciones utiliza la programación lineal para llevar a cabo este proceso. La característica que distingue esta solución es que los objetivos se encuentran en la frontera de las restricciones. Si esa frontera cambia frecuentemente, ya sea por el modelo o por ruido, el cálculo del estado óptimo puede llevar a un bajo rendimiento del controlador. Para solucionarlo, se introduce un fuerte proceso de filtrado de las señales de salida y se desafina el elemento que realiza este cálculo con el fin de minimizar los movimientos.

Esta solución consigue su objetivo, pero también, obstaculiza el rechazo de perturbaciones. Como alternativa, se propuso la incorporación de incertidumbres al modelo. Una solución que no ha llegado a ser implementada posteriormente en ninguno de los modelos comerciales conocidos hasta la fecha. La descripción de estas soluciones fue recogida por Kassmann et ál. en [KBH00a]. Otras de las líneas abiertas para solucionar la programación lineal es la utilización de programación cuadrática. Al utilizar esta técnica, la solución, necesariamente, debe encontrarse en las fronteras de las restricciones con lo que no se tenderá a realizar tantos movimientos como con la programación lineal. Algunos de los MPC comerciales ya han incorporado esta técnica para lograr definir el estado al que debe llegar el proceso [QB03].

Los MPC también disponen de unos procesos para aplicar restricciones a las entradas[CB04]. Normalmente, es sencillo encontrar un estado al que llegar que cumpla con las restricciones de las entradas. No obstante, no sucede lo mismo con las limitaciones en la salida. Esta situación puede darse cuando una gran perturbación entra en el proceso y hace imposible encontrar la salida adecuada, ya que no se puede obtener un resultado exitoso hasta no eliminar la perturbación a la hora de realizar el cálculo del estado deseado.

Por eso, una de las soluciones es permitir las violaciones de las restricciones pero sin sobrepasar una magnitud previamente definida. Y otra de las soluciones adoptadas es la de incorporar penalizaciones a la selección de estados, obteniendo así el que menor sanción tenga. Nótese que estas últimas aproximaciones no eliminan la posibilidad de tener una salida que infrinja las restricciones, únicamente asegura que, de todas las soluciones posibles, va a seleccionarse la que menos restricciones vulnere en base en unos valores introducidos a modo de configuración, ya sean penalizaciones o una desviación máxima.

2.2 Los Modelos Predictivos de Control a fondo

2.2.4.3 Determinación del camino de optimización

En esta etapa del control, un MPC debe computar una serie de ajustes de variables (aquellas que pueden ser modificadas) para, así, dirigir el proceso al estado que fue calculado en la etapa anterior. Para ello, tiene que intentar no violar las restricciones definidas. La mayoría de los MPC pueden ser descritos, de manera aproximada, como una reducción al mínimo de la siguiente función objetivo [CB04]:

$$\begin{aligned}
 J(\mathbf{u}^M) = & \sum_{j=1}^P \{ \|\mathbf{e}_{k+j}^y\|_{\mathbf{Q}_j}^q + \|\mathbf{s}_j\|_{\mathbf{T}}^q \} \\
 & + \sum_{j=0}^{M-1} \{ \|\mathbf{e}_{k+j}^u\|_{\mathbf{R}_j}^q \} + \|\Delta\mathbf{u}_{k+j}\|_{\mathbf{s}_j}^q
 \end{aligned} \tag{2.24a}$$

sujeito al modelo de restricciones:

$$\begin{aligned}
 \mathbf{x}_{k+j} &= \mathbf{f}(\mathbf{x}_{k+j-1}, \mathbf{u}_{k+j-1}) \quad \forall j = 1, P \\
 \mathbf{y}_{k+j} &= \mathbf{g}(\mathbf{x}_{k+j}, \mathbf{u}_{k+j}) \quad \forall j = 1, P
 \end{aligned} \tag{2.24b}$$

y sujeto a restricciones de desigualdad:

$$\begin{aligned}
 \underline{\mathbf{y}} - \mathbf{s}_j &\leq \mathbf{y}_{k+j} \leq \bar{\mathbf{y}} + \mathbf{s}_j \quad \forall j = 1, P \\
 \mathbf{s}_j &\geq 0 \quad \forall j = 1, P \\
 \underline{\mathbf{u}} &\leq \mathbf{u}_{k+j} \leq \bar{\mathbf{u}} \quad \forall j = 0, M-1 \\
 \Delta\underline{\mathbf{u}} &\leq \Delta\mathbf{u}_{k+j} \leq \Delta\bar{\mathbf{u}} \quad \forall j = 0, M-1
 \end{aligned} \tag{2.24c}$$

El objetivo de la función mostrada en la ecuación 2.24a engloba una serie de contribuciones conflictivas. El comportamiento de la futura salida es controlado mediante la penalización de las desviaciones de la trayectoria de salida deseada \mathbf{y}_{k+j}^r , definida como $\mathbf{e}_{k+j}^y \equiv \mathbf{y}_{k+j} - \mathbf{y}_{k+j}^r$, sobre el horizonte de predicción P . Las restricciones de salida son penalizadas minimizando el tamaño de las restricciones de salida para las variables \mathbf{s}_j . Las desviaciones de entradas futuras del estado de entrada deseado \mathbf{u}_s son controladas usando penalizaciones de entrada definidas como $\mathbf{e}_{k+j}^u \equiv \mathbf{u}_{k+j} - \mathbf{u}_s$, sobre el horizonte de control M . Los cambios rápidos en la entrada son penalizados a través de un término que involucra esos movimientos $\Delta\mathbf{u}_{k+j}$. El tamaño de las desviaciones es medido por una norma vectorial, normalmente o la norma L_1 o L_2 ($q = 1, 2$). La importancia relativa de las contribuciones de la función objetivo es controlada

2. Modelo Predictivo de Control y otras aproximaciones

mediante la asignación de pesos dependientes del tiempo a las matrices \mathbf{Q}_j , \mathbf{T}_j , \mathbf{S}_j y \mathbf{R}_j ; siendo elegidas para ser matrices positivas semidefinidas, siendo la solución a la aplicación de un conjunto de M ajustes de entrada:

$$\mathbf{u}^M = (\mathbf{u}_k, \mathbf{u}_{k+1}, \dots, \mathbf{u}_{k+M-1}) \quad (2.25)$$

Muchos de los MPC usan una función objetivo cuadrática similar a la mostrada en la ecuación 2.24a ($q = 2$) para determinar el modo de generar el estado deseado en la planta. En este caso en concreto, toma forma de programación cuadrática que puede resultar confiable y relativamente sencilla. A pesar de ello, para problemas grandes, o procesos que necesitan que el ciclo de control sea muy rápido, no se dispone de tiempo suficiente para solucionarlo. Así que, a modo de solución se utilizan modelos subóptimos capaces de generar una solución aproximada al óptimo. Existen principalmente dos técnicas que permiten conseguir estas soluciones subóptimas: (i) si se sobrepasan las restricciones se iguala el valor a límite marcado por la restricción o (ii) se realizan los cálculos sin las restricciones y, posteriormente, se comprueba si han sido sobrepasadas o no. Ambas técnicas aseguran que no se violarán las restricciones duras, pero, en general, implicarán la pérdida de rendimiento en las predicciones, aunque siempre manteniéndose dentro de unos márgenes aceptables.

2.2.4.4 Formulación de restricciones

Los MPC pueden utilizar diferentes formas para la gestión de las restricciones, normalmente, suelen ser tres [QB03, CB04]. Por una parte, se suelen utilizar restricciones duras que no pueden ser violadas. Así, en la primera de las aproximaciones, todos los valores que sean generados no podrán evadir las limitaciones impuestas.

En segundo lugar, otra aproximación es la de utilizar restricciones débiles. En este caso, sí que está permitido violar la restricción, pero no se podrá realizar en todos los casos y se intentarán minimizar esas violaciones mediante la utilización de penalizaciones cuadráticas en la función objetivo.

Estas últimas restricciones pueden ser gestionadas, también, con la aproximación del punto de referencia. Los puntos de referencia serán definidos para cada una de las restricciones débiles resultando en penalizaciones de la función objetivo cuando los valores se escapen en ambos sentidos (superior e inferior) del punto de referencia. El peso de la salida es ajustado de forma dinámica,

2.2 Los Modelos Predictivos de Control a fondo

sin embargo, este peso solo empieza a ser significativo cuando la variable se comienza a acercar a las restricciones. Cuando una violación de la restricción es predicha, ese peso es utilizado para que el controlador pueda recuperar el valor de la variable, es decir, la pueda colocar de nuevo dentro de los límites establecidos.

Las restricciones fuertes deben ser utilizadas con mucho cuidado, ya que puede llegar a producir inestabilidades del proceso por una indeterminación de valores posibles [Zaf90, MR93]

2.2.4.5 Trayectorias de entrada y salida

Los MPC utilizan básicamente cuatro opciones para especificar el futuro comportamiento de las variables que están siendo controladas en el proceso. Básicamente esos cuatro modelos son [CB04]: (i) un punto de referencia, (ii) una zona, (iii) una trayectoria de referencia o (iv) un embudo.

La primera de ellas, el *punto de referencia*, se toman las mediciones a través del valor constante de un punto de referencia fijo. Todas las desviaciones (superiores o inferiores) a ese punto serán tomadas como penalizaciones dentro la función objetivo. En la práctica, este modelo es muy agresivo y puede conducir a tener que realizar grandes ajustes de las entradas o a desajustar el controlador de alguna manera. Esto es particularmente importante cuando el modelo interno difiere significativamente de los procesos. Para llevar a cabo esta tarea los MPC utilizan factores de supresión.

En segundo lugar, la *zona*, se le proporciona al controlador una zona en la que mantener los valores. Para ello, lo que se hace es fijar los límites superiores e inferiores. La forma de implementarlo es con fronteras e incluso llegar a aplicar la técnica del punto de referencia tanto para los límites superior e inferior. En caso de atravesar esas fronteras, se pasará a imputar el factor de penalización correspondiente.

En tercer lugar, la *trayectoria de referencia* se trata de una curva de primer o segundo orden que servirá como punto de referencia. Al igual que en los dos casos anteriores, las desviaciones de la trayectoria serán penalizadas. La función que representa la trayectoria tiende a convertirse en un punto de referencia cuando el umbral de tiempo sobre el que se trabaja es tendente a infinito. Esta trayectoria de referencia puede ser interpretada matemáticamente como un filtro en la ruta de realimentación, similar al filtro de robustez recomendado en [MZ89].

Y por último pero no menos importante, el *embudo*. Esta solución es similar

2. Modelo Predictivo de Control y otras aproximaciones

a la tomada mediante el ajuste por zona. A pesar de su similitud, su diferencia reside en la forma en la que está definida la superficie de las fronteras superior e inferior. En este caso, en lugar de tratarse de un punto de referencia superior o inferior, se trata de dos curvas que simulan un embudo.

2.2.4.6 Horizonte de salida y parametrización de entradas

Por una parte, los controladores MPC, generalmente, realizan la evaluación de sus variables de control sobre un conjunto finito de intervalos de tiempo futuro, lo que se conoce como *horizonte de predicción* (traducción del nombre inglés “*prediction horizon*”). La solución más ampliamente adoptada es la de una formulación de horizonte finito. Así, la longitud del horizonte P es, básicamente, un parámetro de ajuste de los controladores MPC y suele ser configurado para poder llegar a captar los efectos de todas las variables sobre el resultado final. Otra de las soluciones implementada en muchos de los MPC comerciales [QB03] es la utilización de un horizonte de salida con múltiples puntos. Esto significa que la función de predicción será evaluada en cada punto del futuro. Este proceso, que puede ser muy costoso, podría simplificarse considerando un subconjunto de esos puntos en el horizonte de la predicción, lo que se conoce como *puntos de coincidencia* (traducción literal de la voz inglesa “*coincidence points*”). Para ello, un conjunto de puntos puede ser definido para cada salida, lo que será de gran utilidad cuando una salida responda relativamente rápido a otra de ellas. Debemos destacar que este método es una generalización del anterior, ya que se trata de un superconjunto en el que queda englobado el caso de un único horizonte.

Por otra parte, para la parametrización de las variables que va a manipular el MPC son utilizados tres métodos [CB04]:

- **Movimientos múltiples.** En este primer caso, hasta alcanzar el valor de salida deseado se van dando múltiples saltos durante un horizonte de control M definido como un parámetro de configuración. Una vez finalizado ese periodo de control, el valor de salida será aquel que fue calculado anteriormente por el controlador.
- **Movimiento único.** En este segundo caso, en el momento en el que se detecta la necesidad de la modificación de la salida se realiza un único salto para adaptarla al valor determinado por el controlador. Esta forma de trabajo es un caso especial del modelo anterior en el que el horizonte de control es cero, $M = 0$.

2.3 Una breve historia de los Modelos Predictivos de Control

- **Parametrización en base a una función.** En este tercer caso se trabaja con la misma idea que en el primero de los tres modelos explicados. No obstante, en lugar de ir dando saltos que acerquen el resultado a lo deseado se va a aplicar una función polinomial que va a ir aproximando el resultado. Una vez finalizado el horizonte de control M la función habrá convergido en el resultado calculado por el controlador.

2.3 Una breve historia de los Modelos Predictivos de Control

En los últimos años, la situación de los MPC ha cambiado drásticamente debido al gran aumento en el número de aplicaciones y las mejoras significativas en sus capacidades técnicas. En esta sección se presenta una historia de la tecnología MPC, mostrándose los algoritmos industriales más significativos que fueron surgiendo en los orígenes de esta tecnología. De los algoritmos de control que se habla aquí hay relativamente poca información disponible debido a que son algoritmos que han sido implementados para un proceso productivo en concreto y parte de su información es mantenida en secreto. También, debido a los avances en la materia, se amplía esta visión con la de los algoritmos no lineales, para ello, se dan a conocer dos puntos vista: (i) los algoritmos teóricos y (ii) los algoritmos implantados en la industria.

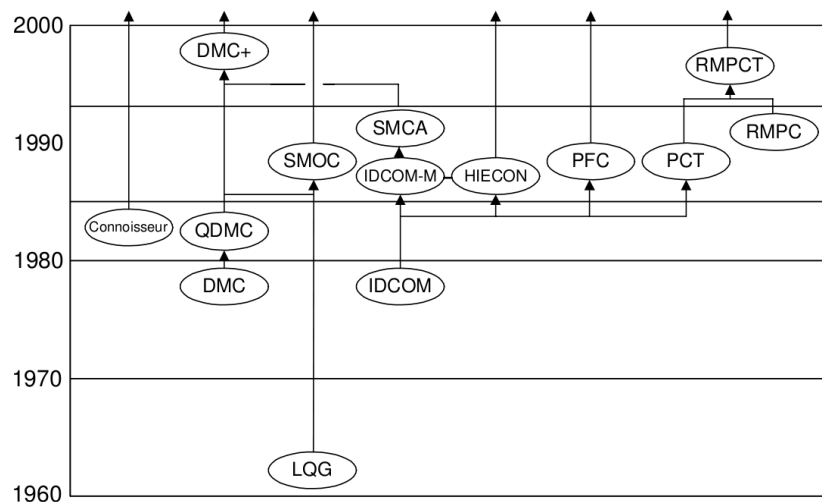


Figura 2.2: Genealogía aproximada de los MPC lineales de la industria.

2.3.1 Modelos Predictivos de Control lineales en la industria, los orígenes

Siguiendo la línea temporal en la que han aparecido los primeros MPC (ver la Figura 2.2) se pueden identificar los siguientes: (i) LQG, (ii) IDCOM, (iii) DMC, (iv) QDMC, (v) IDCOM-M, HIECON, SMCA y SMOC, y (vi) DMC-Plus y RMPCT.

2.3.1.1 LQG

El desarrollo de conceptos de control moderno se originan del trabajo de Kalman et ál. a principios de los 60 [Kal60]. En un contexto de tiempo discreto, el proceso considerado por Kalman y sus compañeros puede ser descrito mediante un modelo de tiempo y un espacio de estado lineal que trabaja con las entradas del proceso o las variables manipuladas, las salidas medidas en el proceso y el vector que representa los estados del proceso a ser controlados. Además, cuenta con un pequeño ajuste de perturbación y ruido (concretamente para el ruido Gaussiano¹) que tiende a ser 0, a pesar de que en la primera de las iteraciones del MPC se supone que su valor no va a ser 0. También, nótese que implícitamente en este modelo se asume que todas las variables se representan como las desviaciones del estado deseado.

Este modelo de MPC, conocido como el LQG, de su nombre en inglés “*Linear Quadratic Gaussian*”, fue el primero en el que en cada uno de los intervalos se utilizaba la medición de la producción para obtener una estimación del estado óptimo. También, para eliminar los problemas de ruido se introdujo el concepto del filtro de Kalman.

La teoría LQG pronto se convirtió en un estándar para solucionar los problemas de control en diversas áreas. Se estimó que existían cientos de aplicaciones reales utilizando LQG y un montante de 400 patentes al año basadas en el filtro de Kalman [GGs01]. Sin embargo, tuvo un pequeño impacto en el desarrollo de tecnologías de control para la industria. Las razones más significativas de que no calase hondo en la industria son las siguientes [RRTP76, GPM89]: (i) sus limitaciones, (ii) que los procesos no eran lineales, (iii) no se tenía en cuenta la robustez (incertidumbre), (iv) sólo disponía de un criterio de actuación y (v) las razones culturales de la época.

¹El ruido Gaussiano es aquel que presenta una distribución de Gauss, independientemente de que exista una correlación del ruido en el tiempo o no.

2.3 Una breve historia de los Modelos Predictivos de Control

2.3.1.2 IDCOM

La primera descripción de las aplicaciones de control MPC fue presentada por Richalet et ál. en 1976 [RRTP76]. Posteriormente, en 1978, resumió el modelo describiendo su enfoque como un modelo de control basado en la predicción heurística (MPHC) [RRTP78]. Su solución fue bautizada con el nombre de IDCOM (como acrónimo de *“Identification and Command”*). Sus características eran las siguientes:

- Se representaba la planta como un modelo de impulsos y respuestas, lineal en las entradas y/o en las variables internas.
- Disponía de un objetivo de rendimiento cuadrático y un número finito de predicciones.
- El comportamiento futuro de la planta es especificado mediante una trayectoria de referencia.
- Las entradas óptimas son computadas usando un algoritmo heurístico iterativo.

Este modelo predecía que la salida en un momento en concreto dependía de una combinación lineal de variables de entradas del pasado y la suma de una serie de pesos para los coeficientes de entradas y salidas. Durante el cálculo dejaba de utilizar ciertos valores del pasado cuando ya no influenciaban la salida. El problema que tiene esta representación es que únicamente es posible para plantas estables.

También, Richalet et ál. hicieron un gran aporte al control de procesos. Concretamente, indicaron que la etapa de control debería estar embebida en una jerarquía de funciones de control dentro de una planta para poder obtener buenos resultados. Describen cuatro niveles de control:

- **Nivel 3:** planificación del tiempo y el espacio en la planta.
- **Nivel 2:** optimización de partes de la producción para minimizar los costes y asegurar la calidad y la cantidad de la producción.
- **Nivel 1:** control dinámico multivariable de la planta.
- **Nivel 0:** control de los sistemas auxiliares y control de válvulas.

2. Modelo Predictivo de Control y otras aproximaciones

2.3.1.3 DMC

Aproximadamente durante la misma época, a principio de los 70, ingenieros de la Shell Oil desarrollaron su propia tecnología MPC independiente. Cutler y Ramaker presentaron los detalles de un algoritmo de control multivariable sin control de las limitaciones de la planta que llamaron DMC [CR80]. Un año después se presentó una contribución que hacía uso de la tecnología DMC en la que el algoritmo había sido modificado para manejar los datos de forma no lineal y, ahora sí, tener en cuenta las limitaciones de la planta [PG80]. Las principales características del algoritmo de control DMC incluyen:

- Modelo de respuesta lineal.
- Su objetivo de rendimiento es cuadrático y con un número finito de predicciones.
- Se especifica el comportamiento futuro de la planta intentando seguir los puntos de ajuste lo más cerca posible.
- Las entradas óptimas son computadas como la solución a un problema de mínimos cuadrados.

El objetivo del controlador DMC es dirigir el proceso tan cerca de los puntos de ajuste marcados como sea posible mediante un cálculo de mínimos cuadrados al que se le añade una penalización en las divergencias. De esta forma, los resultados necesitan menos computación y son menos agresivos que en los modelos anteriores. Al igual que el modelo IDCOM, esta técnica provee un grado de robustez ante los posibles errores.

Los algoritmos IDCOM y DMC representaron la *primera generación* de tecnología MPC. Además, tuvieron un gran impacto en los procesos industriales y sirvieron para definir el paradigma MPC.

2.3.1.4 QDMC

Los algoritmos IDCOM y DMC originales proveyeron un excelente control para procesos multivariables sin restricciones. Sin embargo, el manejo de las restricciones era un tanto *ad-hoc*. Así, los ingenieros de Shell Oil trabajaron contra esta debilidad haciendo que el algoritmo DMC pasara a una programación cuadrática en la que las restricciones aparecerían tanto en las entradas como en las salidas [CMH83, GM86], mejorando los modelos que ya existían. Las características básicas de este nuevo modelo eran las siguientes:

2.3 Una breve historia de los Modelos Predictivos de Control

- Modelo de respuesta lineal.
- Dispone de un objetivo de rendimiento cuadrático en un número finito de predicciones.
- El comportamiento futuro de la planta es especificado intentando apurar los puntos de ajuste y estando sujetos a un plazo de supresión de movimientos.
- Las entradas óptimas son computadas como la solución a un problema cuadrático.

Aunque el algoritmo QDMC es un algoritmo de control algo avanzado, la propia programación cuadrática es uno de los problemas de optimización más simple posible para este tipo de sistemas. Esta forma de perfeccionamiento es positiva para las plantas lineales siendo el problema de optimización resultante convexo. Esto significa que una solución se puede encontrar fácilmente utilizando códigos estándar de optimización.

El algoritmo QDMC puede considerarse como una *segunda generación* de la tecnología MPC, compuesto de algoritmos que proporcionan una manera sistemática para aplicar las restricciones en la entrada y la salida.

2.3.1.5 IDCOM-M, HIECON, SMCA y SMOC

Debido a que la tecnología MPC ganaba aceptación y los problemas con los que debía lidiar crecían y se hacían más complejos, los ingenieros desarrollaron la segunda generación de la tecnología MPC (el algoritmo QDMC). A pesar de la utilización de la programación cuadrática junto a la gestión de las restricciones, su formulación no era satisfactoria debido a que esas restricciones podían ser violadas. Además, la tolerancia al fallo también es interesante ya que es una cuestión que no se había contemplado aún en estos sistemas. También, cada vez era más difícil conseguir expresar en una única función todos los requisitos de control [PG88].

Todas estas razones motivaron a los ingenieros de Adersa, Setpoint Inc. y Shell (Francia) a desarrollar una nueva versión de algoritmos MPC. La versión de Setpoint fue llamada IDCOM-M [GFH88] (la nueva M es para distinguir este algoritmo del anterior, con entradas y salidas simples llamado IDCOM-S) y poco tiempo después pasó a demostrar cómo funcionaba ante un problema en concreto [FM90] con unos resultados francamente satisfactorios.

2. Modelo Predictivo de Control y otras aproximaciones

En la misma época, y de forma similar a Setpoint, la empresa Adersa sacó otro modelo llamado HIECON (de la voz inglesa “*HIErarchical CONs-traint*”) del que no se dispone más información que esta.

Volviendo al algoritmo IDCOM-M, una importante distinción era que utilizaba dos funciones objetivo separadas: (i) una para las entradas y (ii) otra para las salidas. Otras características que también lo distinguían son las siguientes:

- La planta se representa mediante un modelo de impulsos y respuestas.
- Se dispone de un supervisor de control para identificar los conjuntos de la planta que se encuentran en mal estado.
- La formulación se realiza mediante una función multiobjetivo (objetivo de salida y objetivo de entrada cuadrático).
- Controla un subconjunto de puntos en el tiempo futuro para cada salida, llamados puntos de coincidencia, elegidos de una trayectoria de referencia.
- Sólo se calcula un sólo movimiento para cada entrada.
- Las restricciones pueden ser de hardware o de software y pueden estar ordenadas mediante un valor de prioridad.

La empresa Setpoint continuó mejorando el algoritmo IDCOM-M y combinó sus productos de identificación, simulación, configuración y control en un sólo sistema llamado SMCA (de las palabras inglesas “*Setpoint Multivariable Control Architecture*”). En este caso, esta nueva versión se trataba de un motor que permitía solucionar una secuencia de objetivos de optimización separados proveyendo una forma natural para la gestión de múltiples objetivos y restricciones con una valoración para su ordenación.

Por otra parte, a finales de los 80, Shell Research en Francia desarrollo el “*Shell Multivariable Optimizing Controller*” (SMOC) [MB88, YT91]. Este algoritmo incluye muchas características que ahora se consideran esenciales en la formulación de un MPC moderno como:

- Los modelos espacio-estado que son usados para la gama de los MPC dinámicos y no lineales.
- Un modelo de perturbación explícito que describe el efecto de las perturbaciones no medidas.

2.3 Una breve historia de los Modelos Predictivos de Control

- Uso del filtro de Kalman para estimar los estados de la planta y las perturbaciones de la salida de la planta que no han sido medidas.
- Se hace una distinción entre las variables controladas que figuran en el objetivo de control y las que se utilizan para la estimación del estado.
- Las restricciones de entrada y salida están tratadas bajo la programación cuadrática

Estos algoritmos son considerados como la *tercera generación* de la tecnología MPC. Y esta generación se distingue sobre todo por: (i) la gestión de múltiples restricciones, (ii) proveer mecanismos para recuperarse de una solución inviable y (iii) dar mucha más información en la realimentación, entre otros.

2.3.1.6 RMPCT y DMC-plus

Cada vez hay más proveedores comerciales que realizan modificaciones y dan a conocer sus propios sistemas MPC. A finales de 1995, uno de esos proveedores, Honeywell¹, compra la empresa Profimatics Inc. y forma Honeywell Hi-Spec Solutions. El algoritmo que ofrecía la empresa Honeywell fue unido con el algoritmo que ofrecía la empresa Profimatics para crear el algoritmo RMPCT.

Del mismo modo, a principios de 1996, Aspen Technology Inc. compra Setpoint Inc. y DMC Corporation, siguiendo por la adquisición de Treiber Controls en 1998. Las tecnologías de las que disponían estas empresas se unieron para dar paso al producto DMC-plus. Este algoritmo, junto con el anterior pueden ser considerados como la *cuarta generación* de las tecnologías MPC. Y cuentan con las siguientes características:

- Interfaz de usuario basada en ventanas.
- Múltiples niveles de optimización para controlar de forma priorizada diferentes objetivos.
- Flexibilidad adicional en la optimización de objetivos, incluyendo la programación cuadrática y los objetivos económicos.
- Consideraciones directas de la incertidumbre del modelo (diseño robusto del proceso de control).

¹Empresa fabricante de productos de aviónica civil y militar y otros productos aeroespaciales, además de ser integrador y proveedor de servicios. URL: <http://www.honeywell.com>

2. Modelo Predictivo de Control y otras aproximaciones

- Mejora de la tecnología de identificación.

2.3.2 Desarrollos teóricos de la tecnología MPC no lineal

En principio, el método de MPC no lineal (también conocido como NMPC, acrónimo extraído de la voz inglesa “*Non-linear Model Predictive Control*”) se limita a aquellos problemas para los cuales una solución óptima global, en el proceso de optimización dinámica, puede encontrarse entre una ejecución y la siguiente (valores continuos, y no discretos).

Si tratamos con un modelo lineal pero con un objetivo de segundo grado, el problema de optimización resultante tiene la forma de una estructura altamente convexa, para lo que se utilizará la programación cuadrática para la búsqueda de soluciones. Si se modifica el modelo lineal por uno no lineal, en el caso general, se obtiene una pérdida de la convexidad, lo que significa que es mucho más difícil conseguir una solución, y si se llega a encontrar una vez, no puede garantizarse que sea la solución globalmente óptima. Es por ello, que desde el mundo académico se dirigen los esfuerzos para aprovechar la estructura pero mejorando la eficiencia y la fiabilidad de los métodos utilizados [Wri97].

De esta forma, a continuación se pasa a detallar esos avances teóricos que se han dado en el dominio de los MPC no lineales.

2.3.2.1 Estabilidad

La mayor contribución a los sistemas no lineales es la manifestación de Keerthi y Gilbert [KG88] que, para la variedad temporal limitada no lineal y sistemas de tiempo discreto añaden una restricción para la estabilidad $x(k + N|k) = x_s$ al bucle OOP (para más información diríjase a la sección 2.1.2) garantiza que, en condiciones suaves, es decir, no altamente variables, el controlador resultante se mantendrá estabilizado. Básicamente, este resultado es una generalización significativa de los resultados lineales anteriores.

Los esfuerzos en la investigación sobre el problema de la estabilidad de los NMPC con un modelo perfecto han presentado tres soluciones básicas (que se señalan en [ABQ⁺99]). La primera solución, propuesta por Keerthi y Gilbert [KG88], consiste en añadir una restricción de estado terminal al algoritmo NMPC de la forma: $x(k + N|k) = x_s$. Con esta condición forzada, la función objetivo para el controlador se convierte en una función de Lyapunov para el sistema de circuito cerrado, lo que lleva a la estabilidad nominal.

2.3 Una breve historia de los Modelos Predictivos de Control

Lamentablemente esta condición puede ser muy difícil de satisfacer en tiempo real; ya que requiere un número infinito de iteraciones para un código de solución numérica.

Esto motivó a Michalska y Mayne [MM93] a buscar una condición de estabilidad menos estricta. Su idea principal es definir un entorno W alrededor del estado estacionario deseado x_s dentro del cual el sistema puede ser dirigido hacia x_s . Así, les lleva a añadir al algoritmo NMPC una restricción de la forma: $(x(k + N|k) - x_s) \in W$. Si el estado actual $x(k)$ se encuentra fuera de esta región entonces el algoritmo NMPC se resuelve con la restricción anterior. Una vez dentro de la región W se soluciona con el controlador definido. Michalska y Mayne llegan a describirlo como un controlador de modo dual.

Una tercera solución al problema de la estabilidad nominal, descrito por Meadows et ál. [MHER95], consiste en poner el horizonte de predicción y de control en el infinito. Para este caso la función objetivo también sirve como una función de Lyapunov adecuada, lo que lleva a la estabilidad nominal. Ellos demuestran que, si en el cálculo inicial el NMPC tiene una solución factible, entonces existe esa solución para todos y cada uno de los estados temporales posteriores.

2.3.2.2 NMPC de horizonte infinito

En los MPC lineales, los horizontes infinitos (aproximados mediante horizontes extremadamente grandes) son, en muchos casos, una ruta práctica para lograr la estabilidad, ya que existen formas muy eficaces para resolver los enormes problemas de la programación cuadrática. Por contra, para los problemas no lineales, la solución de tales problemas de optimización no es extremadamente difícil, sino imposible de obtener. Por lo tanto, los horizontes finitos son indispensables en los NMPC. Y el trabajo con horizontes infinitos para los MPC no lineales sólo se desarrolla como un método teórico conceptual.

En principio, sería deseable disponer de un procedimiento de diseño del controlador que permita determinar la estabilización de la predicción y el control de horizontes para una configuración NMPC basada en el modelo de la planta y el costo de la fase elegida. Este problema es, sin embargo, muy difícil y aún no ha sido resuelto. Incluso no hay ningún método de análisis disponible que permiten evaluar la estabilidad basada en el conocimiento extraído de la planta, el objetivo funcional y las longitudes del horizonte.

No obstante, existen posibilidades de lograr la estabilidad a pesar de que estas propiedades no se puedan analizar. La idea detrás de estos enfoques es

2. Modelo Predictivo de Control y otras aproximaciones

modificar la configuración del NMPC de tal manera que la estabilidad se pueda garantizar independientemente de la elección de la longitud del horizonte, de la elección de los gastos y para cualquier etapa de la planta. Esto se consigue normalmente mediante la adición de la igualdad adecuada o restricciones de desigualdad en la instalación del modelo. Estas restricciones adicionales no están motivadas por restricciones físicas o los requisitos, pero tienen como única finalidad hacer cumplir la estabilidad del modelo. Por lo tanto, generalmente se denominan restricciones de estabilidad [May97, May00].

2.3.2.3 NMPC con la restricción de igualdad de estado terminal a cero

El esquema NMPC en el que más se trabajó para garantizar la estabilidad utiliza una restricción de la estabilidad en la forma de una restricción de igualdad de estado terminal a cero [KG88, Kle70, MM90]: $x(k + N|k) = 0$, obligando al estado a ser cero en el final del horizonte finito. Keerthi y Gilbert [KG88] fueron los primeros en mostrar la viabilidad de esta formulación implicando la estabilización de una clase de sistemas no lineales limitados. Más tarde, este resultado fue ampliado en [MHER95, RMM94] ya que, incluso, encontraron la formulación para versiones de tiempo continuo [MM90]. En otras palabras, gracias a todos estos trabajos, en condiciones razonables, la estabilidad asintótica del sistema queda probada.

Garantizar la estabilidad mediante la imposición de una restricción de igualdad de estado terminal a cero es, de largo, una de las técnicas más populares. Por una parte, esto se debe sin duda al marco teórico, pero también se debe al hecho de que no se necesita la computación o el ajuste *on-line*. Por otra parte, una restricción de igualdad terminal es una carga artificial que puede requerir importantes gastos adicionales de cómputo (ver [CA98] para una comparación con otros enfoques), y más importante aún, en muchos casos conduce a una grave región de operaciones restringidas debido a problemas de viabilidad.

2.3.2.4 NMPC contractivo

En el NMPC contractivo, como se sugiere en [YP93, DO96], se añade una restricción de la forma $\|x(k + N|k)\|^2 \leq \alpha^2 \|x(k)\|^2$ para controlar el problema. Esta restricción directamente fuerza a que la magnitud del vector de estado se contraiga a un factor pre-especificado cada vez que se realiza el cálculo. Esta restricción se denomina *restricción de contracción de la estabilidad*.

2.3 Una breve historia de los Modelos Predictivos de Control

A diferencia del estándar NMPC, la función de entrada u_k^N sobre el intervalo $[k, k + N - 1]$ es aplicada al sistema no lineal. El siguiente problema de optimización sólo se dará para la instancia $k + N$. En la formulación propuesta por [DO96], se incluye una segunda restricción, como en el caso del modo dual, se utiliza la longitud de horizonte $N \leq n_{max}$ como un reductor adicional.

La viabilidad tiene que ser asumida por todas las instancias temporales durante las que se lleva a cabo una optimización. Por lo tanto, este enfoque no es muy atractivo para las aplicaciones, ya que la viabilidad del problema de optimización en las siguientes instancias no está garantizada.

2.3.2.5 NMPC de modo dual

Desde el punto de vista computacional, una satisfacción de la restricción de igualdad de estado terminal con valor cero es imposible. Si sólo se aplica una aproximación, es decir, sólo requerimos que el estado terminal resida en una pequeña región en torno al origen, la estabilidad está garantizada pero con una gran pérdida, por consiguiente, el sistema no es robusto. Con el fin de relajar la restricción de la estabilidad de la igualdad sin comprometer la estabilidad asintótica, se introduce el llamado NMPC de modo dual [MM93].

El término de modo dual se refiere a dos controladores diferentes que se aplican en diferentes regiones del espacio según el estado, o sea, si esta dentro o fuera de alguna región terminal. Si el estado esta fuera de esta región terminal, un controlador de NMPC con un horizonte variable será aplicado. Si el estado actual se encuentra dentro de la región terminal, se aplica una *realimentación* lineal de la forma $(k) = Kx(k)$.

Computacionalmente hablando, este enfoque es más atractivo que el que se establece en el modelo anterior. Incluso, con una restricción de desigualdad se pueden manejar el proceso con una mayor eficacia. Además, con este modelo se esperan menos problemas de viabilidad y por lo tanto la región de atracción será mayor. También se demostró en [MM93] cómo era la viabilidad del problema de optimización. Las desventajas de este enfoque son la conmutación requerida entre las estrategias de control y el hecho de que la trayectoria predicha y la real, en general, serán diferentes.

2.3.2.6 NMPC de horizonte cuasi-infinito

En el esquema NMPC de horizonte cuasi-infinito [CA98] se utiliza una restricción de estabilidad de desigualdad con la forma $x(k + N|k) \in \Omega$ y una penaliza-

2. Modelo Predictivo de Control y otras aproximaciones

ción terminal cuadrática de la forma $\Phi(x(k+N|k)) = x(k+N|k)^T P x(k+N|k)$. Estas restricciones se agregan a la configuración estándar. La idea básica detrás de este esquema es que el término de penalización terminal no es una especificación de rendimiento que se puede elegir libremente, sino que la matriz P se determina *off-line* de acuerdo con un procedimiento específico, de tal manera que, la función objetivo con Φ es elegida sobre un horizonte finito de coste funcional. Así, se puede lograr la estabilidad, mientras que por otra parte, sólo un problema sobre un horizonte finito tiene que ser solucionado de forma numérica.

En esta configuración, el horizonte de predicción no lineal MPC puede ser pensado como la ampliación hasta casi el infinito. Versiones similares pero de tiempo discreto en sistemas no lineales pueden encontrarse en [FR97]. Al igual que en el enfoque de modo dual, se hace uso de la restricción de desigualdad terminal ya que da ventajas computacionales. En este caso, la aplicación es más simple que el enfoque de modo dual, ya que no hay que cambiar entre las estrategias de control.

2.3.2.7 NMPC robusto

Una serie de resultados han sido publicados para el control de cuán robusto es un Modelo Predictivo de Control lineal (por ejemplo [GM82, Zaf90]). A pesar de ya estar disponibles algunos resultados preliminares, el análisis de las propiedades de robustez en los NMPC aún debe ser considerado como un problema sin resolver.

Algunos de los sistemas anteriormente mencionados tienen algunas propiedades relacionadas con la robustez, o puede hacerse resistente a cambios simples. Sin embargo, sus resultados indican que sólo un grado de incertidumbre lo suficientemente “pequeña” no va a afectar a la estabilidad del modelo. Pero no permiten derivar controladores que garanticen la estabilidad para una descripción de incertidumbre dada. Esto no es sorprendente ya que la definición de las descripciones de lo que es una incertidumbre significativa para los sistemas no lineales es un problema abierto, no sólo en el contexto NMPC, sino también en otras áreas de control. Por una parte, los primeros enfoques NMPC hacen uso de modelos cuantitativos de incertidumbre como los que se describen en [GN95, SF98]. Por otra parte, un análisis más detallado puede encontrarse en [May97, May00].

2.3.3 Implementaciones industriales de la tecnología MPC no lineal

Si bien los aspectos teóricos de los algoritmos NMPC se han discutido con bastante eficacia en varias publicaciones, las descripciones de las aplicaciones industriales NMPC son mucho más difíciles de encontrar. Esto se debe probablemente al hecho de que la actividad industrial en las solicitudes de NMPC ha comenzado a despegar en los últimos años. En la encuesta de la tecnología MPC [QB00] más de 2200 aplicaciones comerciales fueron descubiertas. Sin embargo, casi todas las mejoras se concretaron en los modelos lineales y se agruparon en áreas como las de refinería y procesos petroquímicos. A continuación se va a pasar a indicar los diferentes modelos o implementaciones industriales que existen.

2.3.3.1 Modelos estado-espacio

La primera cuestión que se plantea en las aplicaciones industriales de la teoría NMPC es la derivación de un modelo dinámico no lineal adecuado para el control predictivo del modelo. En la práctica general de los MPC lineales, la mayoría de los modelos dinámicos se derivan de las pruebas de planta o de los sistemas de identificación. Para los NMPC, sin embargo, las pruebas de planta y los sistemas de identificación se vuelven mucho más complicados.

Una clase de modelo de estado-espacio es lo que se adopta en el producto de la compañía Aspen Target, que tiene una ecuación lineal de estado dinámico y una relación no lineal de salida (ver [QB00]). Más concretamente, la no linealidad de la salida se modela con una relación lineal superpuesta a una red neuronal no lineal.

Un tema delicado en el modelado no lineal no es la selección de una relación no lineal, sino, más bien, la selección de un algoritmo de identificación robusta y fiable. El algoritmo de identificación que se discute en [QB00] construye un modelo para cada producto por separado. Además de la identificación del modelo de estado-espacio, un índice de confianza del modelo es calculado de forma *on-line*. Si el sistema indica que la predicción de la red neuronal no es fiable, el mapa neuronal no lineal neto se apaga poco a poco y el modelo de cálculo se basa solamente en el modelo lineal.

Otra característica de este algoritmo de modelado es el uso de filtros de Kalman extendidos (EKF, del inglés “*Extended Kalman Filters*”) para corregir la falta de coincidencia de la planta o del modelo y las perturbaciones no

2. Modelo Predictivo de Control y otras aproximaciones

medidas [QB00]. El EKF proporciona un sesgo y la corrección de ganancia para el modelo durante el funcionamiento del sistema. Esta función sustituye a la salida de error constante que típicamente ha sido empleada en los MPC.

2.3.3.2 Realimentación de la salida

Poniéndonos en el lado de las perturbaciones no medidas y los errores del modelo, es necesario que se dé algún tipo de *realimentación* con una eliminación de la desviación del estado estacionario. El método más común para la incorporación de esa *realimentación* a los algoritmos MPC consiste en comparar la medición y la estimación del proceso. La diferencia entre los dos se añade a las predicciones futuras. Cuando sucede esto, se puede interpretar como que una perturbación de salida que aparece en el sistema y se mantiene constante a lo largo del tiempo. Para el caso de un modelo lineal y sin restricciones activas, Rawlings et ál. [RMM94] han demostrado que esta forma de retroalimentación lleva a un control de compensación libre.

Si el proceso tiene un integrador puro, el supuesto de la perturbación de salida constante ya no da lugar a un control de compensación libre. Para este caso, es común suponer que la integración de una perturbación con un índice de restricción ha llegado a la salida.

De la teoría de control lineal es bien conocido que el conocimiento adicional acerca de las perturbaciones no medidas puede ser aprovechado para ofrecer una mejor retroalimentación mediante el diseño de un filtro de Kalman [KB61]. Muske y Rawlings demostraron cómo esto podía lograrse en el contexto de los MPC [MR93]. Es interesante observar (ver [QB00]) que bastantes algoritmos industriales NMPC ofrecen opciones para la retroalimentación de salida basada en una generalización no lineal del filtro de Kalman conocido como el filtro de Kalman extendido [Ram93]. El modelo Aspen Trajet ofrece un filtro de Kalman extendido para estimar tanto un sesgo como la ganancia de la realimentación.

2.3.3.3 Modelos de entradas / salidas

Los algoritmos MVC y Process Perfecter usan estos modelos de entradas/salidas. En ellos, para simplificar la tarea de identificación del sistema, ambos productos utilizan un modelo estático no lineal superpuesto a un modelo dinámico lineal. Así, Martin et ál. describen los detalles del enfoque para el modelado de Process Perfecter [MBP⁺02]. En él se presenta una sola entrada

2.3 Una breve historia de los Modelos Predictivos de Control

y una sola salida, pero el concepto es aplicable a los modelos de multi-entrada y multi-producto.

Concretamente, en esta aproximación se supone que la entrada del proceso y la producción se puede descomponer en (i) una parte en estado estacionario que obedece a un modelo no lineal estático y (ii) una parte de desviación que sigue un modelo dinámico. Partiendo de esa división, la identificación del modelo dinámico no lineal se basa en los datos de prueba en planta, mientras que el modelo estático no lineal es una red neuronal construida a partir de datos históricos. Se cree que los datos históricos contienen información rica en estado estacionario y las pruebas en plantas son necesarias sólo para el sub-modelo dinámico.

2.3.3.4 Modelo de principios fundamentales

Ya que los métodos de modelización basados en datos empíricos pueden ser poco fiables y requieren una gran cantidad de datos experimentales, algunos productos ofrecen la opción de utilizar los principios fundamentales. Estos productos suelen pedir al usuario que proporcione los modelos de los principios fundamentales con algún tipo de editor de ecuaciones. A continuación, los algoritmos de control pueden utilizar los modelos proporcionados por el usuario para calcular los futuros movimientos de control. El algoritmo NOVA-NLC entra a formar parte en esta categoría. Pero al fin y al cabo, los parámetros del modelo deben ser siempre estimados a partir de los datos de la planta.

2.3.3.5 Optimización de estados estacionarios y dinámicos

Los controladores como el PFC, Aspen Target, MVC, y Process Perfecter dividen el cálculo del control en una optimización local de estados estacionarios, seguida por una optimización dinámica. Los objetivos óptimos de los estados estacionarios se calculan para cada entrada y salida, los cuales pasan al proceso de optimización dinámica para calcular la secuencia de entrada óptima necesaria para avanzar hacia los objetivos marcados. Estos cálculos implican la optimización de un objetivo cuadrático. La excepción de ellos es el controlador NOVA-CNT que realiza ambas optimizaciones al mismo tiempo. Casi todos los productos mencionados en [QB00] utilizan el mismo tipo de objetivos dinámicos, así como las matrices de pesos constantes en la función objetivo.

2.4 Aplicaciones de los Modelos Predictivos de Control

Las Tablas 2.1 y 2.2 resumen las aplicaciones industriales en las que se utiliza la tecnología MPC hoy en día. Ambas tablas muestran tanto la utilización de algoritmos MPC lineales y no lineales, según las encuestas que fueron llevadas a cabo por Qin et ál. [QB03]. Del mismo modo, se ha realizado el recuento de aplicaciones desplegadas en las cada una de las áreas por las diferentes compañías que suministran este tipo de tecnología.

Es importante destacar que los vendedores son libres de definir lo que es una aplicación, por la que hay que tener cuidado al sacar conclusiones de estos datos. En algunos casos, esa aplicación se puede definir como un gran controlador MPC que incluye una planta completa. En otros casos, como el controlador de un motor de un automóvil, el vendedor puede haber informado de una única aplicación incluso existiendo miles de ejemplares que llevan incorporado ese controlador. Téngase en cuenta también que este recuento es el proporcionado por los vendedores, por lo que no incluye las aplicaciones internas realizadas por diversas industrias, ni tampoco aquellas que hayan sido realizadas por centros de investigación con el objetivo de concluir el proceso de transferencia de conocimiento iniciado en las universidades. Sin embargo, sirve para identificar que este tipo de tecnología esta siendo ampliamente utilizado en la industria actual.

Entre las dos tablas (ver Tabla 2.1 y Tabla 2.2) se pueden contar más de 4600 soluciones MPC, más del doble de las realizadas en encuestas anteriores [QB97]. Por consiguiente, es seguro concluir que el uso general de la tecnología MPC crece rápidamente. Todos los vendedores reportan un gran número de solicitudes, lo que hará que siga aumentando el despliegue en la industria de esta tecnología. Debe tenerse en cuenta que las aplicaciones son muy dispares, van desde una única variable a varios cientos de variables.

Además, estas tablas (la Tabla 2.1 y la Tabla 2.2) ilustran la amplia variedad de áreas en las que se aplican los MPC. La industria de las refinerías es la que recoge la mayoría de las soluciones, con un total del 67% de las aplicaciones clasificadas. Como se puede observar, un gran número de aplicaciones también se engloban en áreas tales como las petroquímicas y químicas, a pesar de que han necesitado mayor tiempo para ir aceptando esta tecnología. Comparando con las encuestas anteriores [QB97], otras áreas en las que ha habido un gran crecimiento han sido en la de pulpa y papel, procesamiento

2.4 Aplicaciones de los Modelos Predictivos de Control

Tabla 2.1: Resumen de las aplicaciones lineales MPC por áreas y proveedores industriales (estimaciones basadas en la encuesta de proveedores realizadas por Qin et ál. [QB03]).

Área	Aspen Technology	Honeywell Hi-Spec	Adersa	Ivensys	SGS	Total
Refinería	1200	480	280	25	-	1985
Petroquímicas	450	80	-	20	-	550
Químicas	100	20	3	21	-	144
Pulpa y papel	18	50	-	-	-	68
Aire y gas	-	10	-	-	-	10
Utilidades	-	10	-	4	-	14
Minería	8	6	7	16	-	37
Metalurgia						
Procesamiento alimenticio	-	-	41	10	-	51
Polímeros	17	-	-	-	-	17
Hornos	-	-	42	3	-	45
Aeroespacial						
Defensa	-	-	13	-	-	13
Automoción	-	-	7	-	-	7
Sin clasificar	40	40	1045	26	450	1601
Total	1833	696	1438	125	450	4592

Tabla 2.2: Resumen de las aplicaciones MPC no lineales por áreas y proveedores industriales (estimaciones basadas en la encuesta de proveedores realizadas por Qin et ál. [QB03]).

Área	Adersa	Aspen Tech.	Continental Controls	DOT Controls	Pavilion Tech.	Total
Aire y gas	-	-	18	-	-	18
Químicas	2	-	15	-	5	22
Procesamiento alimenticio	-	-	-	-	9	9
Polímeros	-	1	-	5	15	21
Pulpa y papel	-	-	-	-	1	1
Refinería	-	-	-	-	13	13
Utilidades	-	5	2	-	-	7
Sin clasificar	1	-	1	-	-	2
Total	3	6	36	5	43	93

de alimentos, la aeroespacial y la automovilística.

La Tabla 2.1 ilustra que Aspen Tech. y Honeywell Hi-Spec son dos empresas que enfocan mayoritariamente sus soluciones MPC a la refinería y la industria

2. Modelo Predictivo de Control y otras aproximaciones

petroquímica. Por otra parte, Adersa e Invensys, al parecer, tienen una amplia experiencia con aplicaciones para el procesamiento de alimentos, la minería o la metalurgia, la industria aeroespacial y la automovilística, entre otras. Las aplicaciones reportadas por SGS incluyen una serie de aplicaciones internas con lo que no se ha podido realizar su clasificación. A pesar de ello, y dada la distribución del resto de proveedores, es probable que se dediquen para industrias petroquímicas o la refinería.

La Tabla 2.2, por el contrario, hace referencia a las aplicaciones no lineales. Este tipo de sistemas MPC se reparte más equitativamente entre las áreas de aplicación. Aquellas con el mayor número de casos están relacionadas con productos químicos, polímeros y procesamiento de aire y gas. Sin embargo, como ya se indica en [MJ98], se ha comprobado que el tamaño y el alcance de las aplicaciones no lineales suele ser mucho menor que el de las lineales, debido a la complejidad computacional de los algoritmos.

2.5 Limitaciones de la tecnología existente

Muchos de los algoritmos MPC que existen en la actualidad sufren de limitaciones heredadas de aquellos modelos originales como el DMC y la tecnología IDCOM (ambos comentados en la sección 2.3). Problemas que ya han sido discutidos por otros [ML91, MR93], entre los que se incluyen: (i) limitadas opciones entre las que elegir el modelo, (ii) feedback ineficiente, (iii) la pérdida de la estabilidad nominal y (iv) la solución ineficiente de la optimización dinámica de los modelos.

Los algoritmos lineales MPC que se basan en modelos de convolución (impulso y etapa de respuesta) pueden ser problemáticos cuando se controla un proceso que consta de una amplia variedad de constantes de tiempo. Para este caso es típico el sacrificio del control dinámico del proceso con el fin de mantener el tamaño del modelo dentro de una escala razonable. Un problema potencialmente más significativo, relacionado con los modelos de convolución, es que están limitados a procesos estrictamente estables. Aunque sin duda, es posible modificar los algoritmos para dar cabida a una pequeña solución. A pesar de todo, estas modificaciones pueden conducir a otros problemas, como la adición de ruido en la señal de salida enviada a la planta. Además, no es posible, en general, representar un proceso inestable utilizando un modelo de este tipo. Todos estos problemas pueden superarse mediante el uso de un modelo paramétrico autorregresivo.

2.5 Limitaciones de la tecnología existente

La técnica del sesgo en la retroalimentación utilizada por los controladores MPC (técnica que asume que no hay perturbaciones estocásticas que afectan al sistema y que las mediciones realizadas son perfectas) es, probablemente, la mejor opción para utilizar en plantas estables, es decir, plantas que están en total ausencia de alteraciones y que tienen la posibilidad de recoger múltiple información sobre las mediciones del proceso. Sin embargo, siempre será posible una mejora en la retroalimentación si se tiene en cuenta la caracterización de las perturbaciones, ya que ésta será útil tanto para este tipo de plantas como para las que sí sufren de este problema. En las diferentes aplicaciones, la medición del ruido suele realizarse de forma separada usando técnicas de filtrado *ad hoc*. Estos métodos tienden a funcionar mal en presencia de un alto nivel de ruido. Los profesionales de la industria indican que una de las mejores soluciones para esta problemática es la aplicación del filtro de Kalman [WB95]. Además, ha sido demostrado [MR93] que se puede llegar a obtener mejores resultados utilizando un modelo similar. También han surgido otras aproximaciones, como el algoritmo SMOC [MB88] que se basa en el filtro de Kalman para realizar la realimentación de la planta con el objetivo de solucionar todos los problemas citados. A pesar de todo, esta limitación sigue siendo un reto al que los equipos de investigación tienen que hacer frente.

El ajuste de los controladores MPC para un funcionamiento estable en presencia de restricciones puede ser difícil, incluso cuando el modelo del proceso es perfecto. Por ello, se realiza un gran esfuerzo antes del despliegue del controlador. Hay que reconocer, sin embargo, que en la práctica es imposible simular todas las posibles combinaciones de restricciones activas. Para ello, se toma como solución la construcción de un algoritmo que es inherentemente estable en todos estos casos, lo que en la teoría del control se conoce como un algoritmo de estabilización nominal. Así, en [SR98] se muestra cómo se puede hacer esto para el caso lineal mediante la predicción de horizontes infinitos y su control. También en [CA98] se presenta un método de horizonte cuasi-infinito para sistemas no lineales. Aunque, los resultados de la investigación para el caso no lineal se han cubierto sobradamente por los experimentos llevados a cabo por Mayne et ál. [MRRS00], ninguno de los algoritmos que son utilizados en la industria, comentados en secciones anteriores, disponen de algo similar.

En varios de los modelos existentes se utilizan soluciones subóptimas para la optimización dinámica, presumiblemente con el fin de acelerar el tiempo de generación de la solución. Esto parece difícil de justificar para el caso de aplicaciones petroquímicas, donde cada ciclo de control se ejecuta del orden de una vez por minuto. Sin embargo, para aplicaciones de alta velocidad, es

2. Modelo Predictivo de Control y otras aproximaciones

decir, cuando el controlador debe ejecutarse en unos pocos milisegundos, como en el seguimiento de la posición de un misil, puede que no sea viable otro tipo de soluciones, siendo una solución subóptima la única opción para gestionarlo. Así, en [GGS01] se da un ejemplo de la aplicación de este tipo de soluciones.

La mayoría de los productos que se encuentran desplegados en plantas a lo largo del mundo, como se ha comentado anteriormente, se trata sistemas MPC lineales no explotan la programación cuadrática como técnica de optimización de la dinámica del proceso. A pesar de ello, existen investigaciones, como la de Rao et ál. [RWR98], que indican que gracias a la utilización de este tipo de técnicas se podría manejar procesos mucho más grandes o de una forma mucho más rápida (permitiendo evitar la utilización de sistemas subóptimos).

La incertidumbre del modelo no ha sido abordada adecuadamente por la tecnología MPC. Si bien la mayoría de las soluciones proporcionan estimaciones de la incertidumbre del modelo, sólo dos soluciones comerciales, RMPCT y DMC-plus, proporcionan una manera de utilizar esta información durante el proceso del diseño del controlador. Todos los algoritmos MPC proporcionan una forma de desafinar el controlador con el fin de mejorar la robustez del sistema, pero el equilibrio entre el rendimiento y la robustez general no está muy definido. Hasta que esta conexión se haga, no será posible determinar cuándo un modelo es suficientemente preciso para una aplicación de control en particular. Hay muchos resultados disponibles en la literatura académica, pero la mayoría se centran sólo en la optimización dinámica, considerada como uno de los retos con mayor nivel de complejidad. Aunque se han alcanzado prometedores resultados [VGN95, ?, KBH00b], que con el tiempo pueden entrar en las soluciones comerciales, se necesita una mayor investigación en esta área.

A parte a las limitaciones ya mencionadas se pueden añadir las siguientes:

- Las señales de la planta requieren una atención especial de ingenieros experimentados durante las fases de diseño y prueba del modelo.
- No hay ninguna herramienta para determinar si los datos recogidos son los adecuados para representar el comportamiento del proceso antes de comenzar con diseño de MPC.
- No se aborda la eficiencia estadística y la coherencia de los algoritmos.
- Hay una falta de métodos de validación para comprobar si el modelo identificado es adecuado para el control o si el deterioro del modelo se ha producido después de su puesta en marcha.

- No existe ningún enfoque sistemático para la construcción de modelos dinámicos no lineales. Se necesitan una serie de directrices para realizar las pruebas de la planta con el objetivo de construir un modelo no lineal confiable.

2.6 Modelo Predictivo de Control, la siguiente generación

Como ya se ha podido ver en la sección 2.5, la tecnología MPC no se encuentra optimizada para los procesos de producción. Así, y según la visión de los expertos en el dominio y de los vendedores de tecnologías MPC, unido a las conclusiones extraídas por Morari [ML97] y Rawlings [Raw00], las áreas de investigación en las que los MPC deben evolucionar son:

- **Formulación básica del controlador.** Debido a la dificultad de expresar todos los objetivos de control pertinentes en función de un sólo objetivo, la nueva generación de la tecnología MPC utilizará múltiples funciones objetivo. El horizonte de predicción infinito tiene propiedades teóricas beneficiosas y probablemente se convertirá en una característica estándar. Las opciones de la trayectoria de entrada y salida incluirán los puntos de ajuste, las zonas, las rutas y los redireccionamientos. Del mismo modo, la parametrización de la entrada utilizando funciones base puede hacerse más extensa, y los horizontes de control infinitos con movimientos calculados en cada intervalo de control también pueden llegar a ser posibles. Sin embargo, la investigación aquí detallada resuelve esta problemática al desarrollar diferentes modelos con el fin de ser capaces de gestionar múltiples objetivos, para después, en caso de que uno de ellos difiera de la ruta óptima, sea modificado el proceso de producción.
- **MPC adaptativo.** Unos pocos algoritmos MPC adaptativos como el algoritmo GPC introducido por Clarke, Mohtadi y Tuffs [CMT87] han sido propuestos a lo largo del tiempo. Pero solamente dos de ellos han llegado al mercado (Connoisseur de Invensys y STAR de Dot Products [DMM⁺93]). Esto es así a pesar del fuerte incentivo del mercado para que la comunidad científica desarrolle un controlador MPC que sea capaz de autoajustarse. Viendo que no se han desarrollado tantos modelos, se refleja la dificultad de hacer un sistema de control adaptativo en el mundo real. Sin embargo, debido al nuevo enfoque que se ha dado en

2. Modelo Predictivo de Control y otras aproximaciones

esta investigación se propone un método que no ajusta el sistema de forma automática, sino que proporciona una estrategia de ajuste rápido de los modelos para hacer frente a los cambios en el proceso utilizando para ello la técnica de la compresión bayesiana basada en métodos de Montecarlo.

- **MPC robusto.** Los controladores MPC industriales suelen confiar únicamente en la fuerza bruta de la simulación para evaluar los efectos de la planta. Las garantías de la robustez podrían reducir significativamente el tiempo necesario para ajustar y probar los algoritmos. Es probable que los experimentos realizados [KBH00b, ?, SM98] puedan llegar a convertirse en productos comerciales. Incluso podrían ser combinados con las estimaciones de incertidumbre para simplificar el diseño y la afinación del controlador MPC. De hecho, el nuevo enfoque adoptado en esta tesis doctoral permite medir la robustez de los modelos basados en aprendizaje automático, con lo que de forma transitiva, podemos medir la robustez del sistema MPC.
- **MPC no lineal (NMPC).** Las próximas generaciones de la tecnología MPC trabajarán con modelos no lineales para conseguir la perfección de la combinación entre la representación del proceso y los datos del funcionamiento. Para ello, será definido un modelo fundamental continuo en el tiempo a partir de una representación gráfica de la planta. El procesado de señales de prueba será diseñado de forma automática con el fin de explotar las regiones donde el modelo tiene un funcionamiento inadecuado. Debido a las dificultades del modelado de datos para procesos no lineales se van a necesitar modelos similares a los de Foss et ál. [FLM98]. En el caso que aquí se detalla, se conseguirá realizar un MPC híbrido ya que para el mismo objetivo se van a contar con diferentes modelos trabajando de forma conjunta. Esos modelos serán tanto modelos lineales como no lineales combinando sus resultados, lo que hará que el sistema trabaje con las conclusiones híbridas del modelo.
- **MPC para sistemas híbridos.** Esencialmente, todos los procesos contienen componentes tanto discretas como continuas: válvulas de encendido y apagado, interruptores, entre otros. De este modo, Slupphaung y Foss [SF97, Slu98] y Bemporad y Morari [BM00, BM99a] consideraron la aplicación de los sistemas MPC en este entorno, teniendo siempre en cuenta ambos tipos de variables. En este momento, algunas de las cuestiones interesantes con el desarrollo de esta tecnología para los entornos

2.7 Aproximaciones mediante métodos de aprendizaje automático

híbridos son: (i) hasta dónde puede llegar esta tecnología, (ii) determinar si el desarrollo de modelos que permitan el manejo de este tipo de variables crean cuellos de botella debidos a los procesos de diseño o la computación, (iii) cuáles son los beneficios que se pueden obtener frente a los enfoques tradiciones y (iv) qué nuevos tipos de problemas pueden ser abordados. En esta investigación doctoral se han utilizado modelos que adaptan los datos permitiendo el trabajo con valores continuos y discretos tanto en modelos de representación de conocimiento en forma lineal, como no lineales. Por consiguiente, el modelo que aquí se ilustra es una aproximación a los MPC de entorno híbrido.

2.7 Aproximaciones mediante métodos de aprendizaje automático

La tecnología MPC no contempla el desarrollo de sus fórmulas mediante clasificadores estadísticos basados en aprendizaje automático. En alguna ocasión sí hacen uso de ellos como método para la realimentación de la planta. A pesar de ello, lo que no se contempla es la creación del modelo al completo basado en este tipo de clasificadores. Ésta es la aportación fundamental que se ilustra en esta tesis doctoral (quedando recogida en la sección 1.4.1). Sin embargo, sin entrar a formar parte de la tecnología de los MPC, sí que se han realizado avances desde el punto de vista de las predicciones y de la reducción de ruido utilizando técnicas de aprendizaje automático.

No sólo se utilizan los métodos de control habituales, el aprendizaje automático es cada vez más utilizado para la optimización de los procesos de producción. Los clasificadores con los que más se ha trabajado son las redes neuronales artificiales, por ejemplo, para clasificar piezas de fundición [LSOdM98], para la optimización de los parámetros de calidad [ZXD08], la detección de las causas de defectos de fundición [PK03] y en otros problemas del proceso [Bha99].

Pero la comunidad científica no trabaja únicamente con las redes neuronales artificiales, las redes bayesianas se utilizan como método previo a la utilización de redes neuronales artificiales. Su objetivo es la predicción del número de ferrita en las soldaduras de acero inoxidable [VMB02]. Sorprendentemente, también existen experimentos que tratan de predecir las propiedades de las piezas moldeadas utilizando métodos de aprendizaje automático [Vor01].

Del mismo modo, existen experiencias exitosas con la utilización del al-

2. Modelo Predictivo de Control y otras aproximaciones

goritmo de los k vecinos más próximos (del inglés “*K-nearest neighbour*”) en las que se incluyen la detección de fallos en el proceso de fabricación de semiconductores [HW07] y para la mejora de la calidad del acero.

Pero no solamente son utilizados clasificadores tan habituales como los ya mencionados, sino que modelos normalmente utilizados para trabajar con textos, tales como las máquinas de soporte vectorial (de la voz inglesa “*Support Vector Machines*”) y los árboles de decisión, también se han utilizado en problemas de clasificación relacionados con procesos de producción como en el de la fundición. Ambos se han aplicado en los siguientes ámbitos: el diagnóstico de fallos [YZZ07, WY07], la predicción de series temporales [TVBDW⁺03], la identificación de errores en turbinas de gas [MLP08] y la detección de intrusiones en la red [ZZ06]. Incluso más relacionado con los experimentos que serán presentados más adelante, las máquinas de soporte vectorial se han utilizado con el fin de prever las propiedades mecánicas de materiales de acero [WVG06].

En cuanto a temas de reducción de ruido en las entradas de datos, uno de los experimentos más recientes es el desarrollado por Pham et ál. [PMM⁺07]. Ellos usaban el algoritmo de las abejas, algo similar a los algoritmos genéticos, con el fin de realizar la selección de las características más relevantes de una forma mucho más comprensible.

Como se puede observar no hay demasiadas investigaciones orientadas a los procesos de producción con la utilización de diferentes métodos de clasificación basados en el aprendizaje automático pero, vistas las últimas investigaciones en las que está sumergida la comunidad científica, estas tendencias están cambiando. Así que creemos que es una buena idea realizar la fusión de estos dos métodos.

2.8 Sumario

A lo largo de este capítulo hemos revisado la literatura acerca de los MPC. En concreto, hemos comenzado detallando los conceptos generales que definen los sistemas MPC, así como la formulación matemática para el funcionamiento de un sistema MPC genérico. Del mismo modo, y añadiendo información al concepto general, se ha pasado a detallar cuál es el proceso que deben seguir los ingenieros que pretenden desarrollar un nuevo sistema MPC para una planta.

Más concretamente, hemos identificado las etapas necesarias para el diseño y la afinación, procesos que deben desarrollarse con el fin de ajustar el modelo

al proceso productivo en el que será utilizado. Seguidamente, hemos estudiado la división que puede hacerse en los MPC según la naturaleza del proceso que intentan representar (lineales y no lineales) y el método seguido para su generación (empíricos y basados en principios fundamentales). Asimismo, hemos estudiado la forma en la que un sistema MPC debe ser modelado dependiendo del proceso sobre el que se va a aplicar. Por último, hemos identificado las etapas del ciclo de control de un MPC, así como las diferentes técnicas existentes.

El estudio se complementa con una breve historia de la evolución que han sufrido los MPC. Para ello, partimos desde las primeras generaciones, totalmente lineales y que fueron desarrolladas específicamente para la industria. En segundo lugar, agrupamos las diferentes aproximaciones que fueron desarrolladas en entornos académicos. Y en último lugar, las nuevas generaciones de MPC, las cuales beben de los descubrimientos hechos en el mundo académico, que han sido desplegadas en entornos de producción reales. Además, con el objetivo de demostrar la gran utilidad de estos sistemas se realiza un breve estudio de las áreas de aplicación y cuáles de las casas comerciales se centran más en cada una de las áreas.

Finalmente, y basándonos tanto en aquello comunicado por los expertos en la materia como por los trabajadores que lo utilizan habitualmente, hemos realizado un estudio de las limitaciones inherentes en la tecnología MPC actual. Igualmente, hemos recogido las líneas de trabajo futuras sobre la tecnología MPC y otro tipo de aproximaciones desarrolladas con el fin de mejorar diversos aspectos del proceso productivo y que están basadas en métodos de aprendizaje automático.

*«Alquimista certero, del hierro
pensó hacer oro e hizo del oro
hierro.»*

Proverbio español

3

La fundición, un antiguo proceso rodeado de magia

LOS procesos productivos son parte importante de la sociedad actual. Más concretamente, gracias a ellos, los consumidores pueden disponer de diferentes productos y servicios. Para el desarrollo y comprobación del método que se presenta en esta disertación doctoral, se ha seleccionado como caso de uso el proceso llevado a cabo en fundiciones de alta precisión.

Dentro de los procesos de manufacturación, la fundición es uno de los ejes centrales de la economía. Miles de piezas se crean en las fundiciones repartidas a lo largo de todo el mundo con el fin de crear sistemas más complejos, por ejemplo, pinzas de freno para contribuir a los sistemas de frenado de vehículos motorizados, las hélices que permiten desplazarse a las embarcaciones, los mecanismos que permiten el movimiento de los alerones de un avión o el gatillo de un arma de fuego. Como se puede pensar, en cualquiera de los casos anteriores, el más pequeño de los errores puede tener consecuencias fatales.

A pesar de ser uno de los ejes fundamentales de la sociedad, la fundición se encuentra todavía en un nivel de desarrollo inferior al de industrias de similar importancia. Además, las tendencias actuales fomentan la producción

3. La fundición, un antiguo proceso rodeado de magia

de componentes cada vez más pequeños y precisos, haciendo habitual el descubrimiento de que las piezas resultantes del proceso cuentan con defectos y debemos rechazar el lote al completo.

Por desgracia, aunque hay muchas normas y métodos de comprobación de las piezas obtenidas, éstas se llevan a cabo una vez que la producción ha sido completada. De esta manera, las técnicas más utilizadas para mantener los procesos de producción libres de fallos son (i) el control exhaustivo de la producción y (ii) la utilización de diversas técnicas de simulación [SLS06]. Estas últimas técnicas son extremadamente caras y sólo se llega a lograr buenos resultados mediante las inspecciones a posteriori. Por consiguiente, proporcionar métodos eficaces que permitan detectar este tipo de problemas antes del proceso de producción puede ayudar a aumentar los estándares de calidad y ahorrar recursos en el proceso. Todas estas razones hacen que el proceso de fundición sea el más adecuado para probar la solución que en esta tesis doctoral se detalla.

El resto del capítulo se encuentra organizado como sigue. La sección 3.1 explica el caso de uso sobre el que se ha llevado a cabo los experimentos que prueben la hipótesis planteada. De este modo, se realizará una definición exhaustiva del proceso de fundición, así como las variables del proceso a tener en cuenta a lo largo de los experimentos que hemos desarrollado. La sección 3.2 detalla los defectos y las características sobre las que se ha trabajado. Y finalmente, la sección 3.3 resume los principales aspectos de este capítulo.

3.1 El proceso metalúrgico de la fundición

La fundición es el lugar donde se lleva a cabo el proceso de transformación de los metales en piezas metálicas. A modo de resumen, los operarios introducen los metales ya fundidos, también denominado colada de metal, en un molde para que, una vez terminado su proceso de enfriamiento, se obtenga la pieza final.

En [BSoM95] se ilustran otros procesos que también se encuentran relacionados con la producción de componentes metálicos. De este modo, la tabla 3.1 recoge un breve resumen de esos procesos, indicando su nombre y proporcionando una breve descripción sobre las características de cada uno de ellos. Pero estos procesos no están totalmente inconexos, existen relaciones entre ellos, es decir, las salidas de algunos de ellos sirven como materias primas en otros (para más información diríjase a [BSoM95]).

3.1 El proceso metalúrgico de la fundición

Tabla 3.1: Resumen de los procesos productivos desarrollados sobre metales. En la siguiente tabla se muestra el nombre de los procesos seguidos así como otros sinónimos por los que se conocen. También, se explica el proceso a través de una breve descripción del mismo.

Nombre	Descripción
Producción a granel de metales ferrosos	Se trata de la producción de metal fundido a granel extraído de minerales en bruto y desechos de metales. El metal fundido obtenido puede ser utilizado para: <ul style="list-style-type: none"> • producción continua de colada. • la producción de placas y láminas. • producción de arrabio. • suministro a un proceso de fundición.
Producción de metales no ferrosos a granel	Proceso de producción de metales no ferrosos a partir de minerales en bruto y desechos de metales, sufriendo un proceso de refinamiento y purificación mediante productos químicos, térmicos, electrolíticos y de otros tipos. Los productos suelen consistir en barras extrudidas, varillas, alambres, losas, billetes y lingotes.
Ferroaleaciones y la producción de lingotes	Este proceso se trata de la producción de diferentes aleaciones y adiciones de metales según diferentes grados de certificación, en forma de pepitas, gránulos, etc. para ser utilizadas como materia prima en las fundiciones
Fundición de moldes o piezas	Se trata de la producción de piezas metálicas con una forma concreta. Para ello, se vierte el metal fundido en moldes. Ese metal fundido es producido a partir de chatarra, arrabio, lingotes y aleaciones. Procesos totalmente relacionados con este son el de preparación de moldes o núcleos y el acabado de las piezas manufacturadas.
Hierro forjado	Es el considerado hierro comercialmente puro, con más de un 99% de pureza, trabajado a golpe de martillo, apretado y enrollado a mano para obtener una gran calidad de producto o productos muy complejos. Es una técnica de producción histórica que en la actualidad está limitada a productos muy concretos.
Fabricación	Ensamblado de hojas de metal y chapas, entre otros, para obtener el producto terminado. Puede involucrar tareas de corte, pulido, soldadura, moldeo, etc.
Mecanizado	Proceso de eliminación de metales no deseados mediante la perforación, torneado, etc. con el fin de obtener la forma final deseada y dentro de las tolerancias dimensionales requeridas.

3. La fundición, un antiguo proceso rodeado de magia

3.1.1 Materiales utilizados en los procesos de fundición

En los procesos anteriormente mencionados pueden ser utilizados diferentes metales. Una pequeña clasificación de los más utilizados en estos procesos son los siguientes [Sin89, Bra47]:

- **Ferrosos.** En primer lugar, encontramos los metales ferrosos o férricos. Estos metales son aquellos que contienen como elemento base el hierro, aunque pueden llevar otros elementos en diferentes proporciones.
 - *Hierro.* El hierro es un material maleable, de color gris y que presenta propiedades magnéticas. Dentro de este grupo tenemos (i) el hierro gris que es aquel que contiene más de un 95 % de hierro con carbono, silicio y magnesio; (ii) el dúctil, que se trata de hierro conteniendo carbono, silicio y magnesio; (iii) el maleable, constituido por hierro gris que, cuando la pieza está en frío, se le agrega grafito mediante un tratamiento térmico; y (iv) las aleaciones conteniendo más de un 30 % de cromo y/o níquel para proveer de nuevas características como dureza adicional.
 - *Acero.* Con el nombre que se le da comúnmente en la ingeniería metalúrgica a la aleación de hierro con carbono (entre 0,1 y 2,1 %), este material puede ser dividido en: (i) acero carbono, que contiene niveles de carbono no superiores al 1,7 % (normalmente se encuentra entre 0,1 y 0,5 %) y de manganeso hasta el 1,6 %; (ii) acero de baja aleación, con una baja aleación de carbono (0,35 % máximo) que contiene entre un 3 y un 5 % de cromo y entre un 2 y un 8 % de níquel; y (iii) acero inoxidable, aleación de hierro de base que contiene hasta un 30 % de cromo y/o hasta el 40 % de níquel, aumentando así su resistencia a la corrosión.
- **No ferrosos.** En segundo lugar, agrupamos los metales no ferrosos. Estos comprenden a todos los metales con excepción del hierro. En general son blandos y tienen poca resistencia mecánica. Por ello, y con el fin de mejorar sus propiedades se alean con otros materiales. Los más conocidos son los siguientes.
 - *Aluminio.* Uno de los materiales más comunes en la corteza terrestre. Se desarrollan diferentes aleaciones conteniendo silicio, cobre y/o magnesio.

3.1 El proceso metalúrgico de la fundición

- *Cobre*. Este material normalmente se utiliza de la siguiente forma: (i) cobre puro (más de un 99 % de pureza); (ii) latón, una aleación de cobre y zinc (hasta un 45 %) y a veces con un máximo de 2 % de plomo y 1 % de estaño; (iii) bronce, una aleación de cobre y estaño (máximo un 12 %) a pesar de que existen diversas variaciones como el bronce con plomo (de 9 a 22 % de plomo), *gun metal* (de 3 a 5 % de plomo y entre 2 y 8 % de zinc), bronce de aluminio (de 6 a 9 % de aluminio, de 0 a 5 % de níquel, 0,5 a 5 % de hierro) entre otras.
- *Zinc*. Las aleaciones basadas en zinc cuentan con hasta un 4 % de aluminio y pequeñas cantidades de cobre, magnesio y otros materiales
- *Superaleaciones*. Se tratan de aleaciones basadas en níquel o cobalto en combinación con cromo, hierro, manganeso, molibdeno, entre otras muchas.
- *Magnesio*. Suelen tratarse de aleaciones de magnesio con aluminio, zinc y torio. También pueden añadirse otros metales en pequeñas cantidades.
- *Titanio*. Suele utilizarse como titanio puro, es decir, con más de un 99 % o realizando aleaciones de aluminio, vanadio, zinc, molibdeno y zirconio.

3.1.2 El proceso de fundición y las variables que lo afectan

Particularmente, de los procesos anteriormente comentados, en esta investigación doctoral nos centraremos en los procesos de fundición que se encargan de producir piezas que serán utilizadas para ensamblar el producto final y, además, realizando los experimentos sobre materiales ferrosos como el hierro (fundición de hierro grafito esferoidal) y no ferrosos como el aluminio. Más concretamente, se han seleccionado los datos recogidos en una fundición especializada en el desarrollo de componentes seguros y con un alto nivel de precisión. Los resultados de su proceso productivo son destinados a la industria del automóvil. Esta fundición tiene una producción de alrededor de 45.000 toneladas al año.

Varios factores, como por ejemplo las condiciones extremas en las que se lleva a cabo, hacen que este proceso de producción tenga una gran complejidad. Las piezas, el producto final, se producen mediante el vertido de metal

3. La fundición, un antiguo proceso rodeado de magia

fundido en moldes, utilizando mazarotas para la creación de huecos en las secciones internas. Después de que el metal se haya enfriado lo suficiente, la pieza se separa del molde para someterla a una fase de limpieza y de finalización del producto.

A modo de concretar el proceso de fundición, existen numerosos documentos de dominio público [def03a, def03b] en los que se ofrecen descripciones detalladas del proceso completo. Este proceso según los anteriores documentos puede ser dividido en las etapas que se detallan a continuación.

- **Desarrollo de patrones para los moldes.** En esta etapa, los moldes (exteriores) y mazarotas (interiores) son producidos en madera, metal o resina para ser usados en la creación de los moldes de arena donde será construida la pieza final.
- **Moldes de arena y creación de mazarotas.** El moldeo en arena es el método más extendido para la creación de piezas metálicas. La arena se mezcla con arcilla, agua u otros aglutinantes químicos. Posteriormente, máquinas especializadas crean las dos mitades del molde y las unen para generar el contenedor donde el metal fundido, la colada, será vertido.
- **Fusión y colada.** Los metales se funden y se mezclan para que posteriormente puedan ser vertidos en los moldes de arena que darán la forma a la pieza.
- **Enfriamiento y separación.** La solidificación de las piezas se controla a lo largo de líneas de refrigeración hasta que ésta llega a su fin. Una vez la pieza ya ha alcanzado una temperatura adecuada y está totalmente sólida, pasa a limpiarse, es decir, se retiran todos los restos de la arena del molde. La arena retirada pasa a ser tratada para volver a utilizarla en futuros moldes.
- **Finalización.** Para finalizar el proceso de creación de las piezas, normalmente, se llevan a cabo algunas acciones sobre las piezas, por ejemplo, un tratamiento térmico y la ratificación de defectos de soldaduras, entre otros.
- **Control final.** Una vez que las piezas se encuentran terminadas, son probadas, es decir, se comprueba si hay microrrechupes o si los valores de ciertas características se cumplen. Una vez que se ha determinado el resultado de la pieza, éste se adjunta a la información anterior con

3.1 El proceso metalúrgico de la fundición

el fin de etiquetar los datos y ser útiles en los procesos de aprendizaje automático desarrollados en los experimentos de esta tesis doctoral.

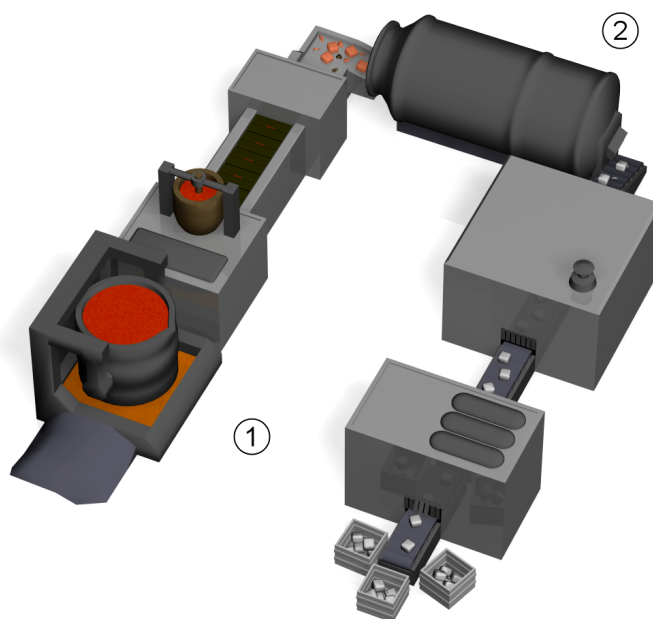


Figura 3.1: Representación del proceso productivo de una fundición.

La figura 3.1 muestra las fases de fusión, moldeo y enfriamiento. Una vez que la materia prima se derrite, se vierte sobre los moldes, marcado con (1). Las líneas de refrigeración (2) aceleran el proceso natural de enfriamiento de la fundición y, cuando estén debidamente solidificadas, los moldes de arena se separan de las piezas y se regresa la arena al molino, de modo que, ésta puede reutilizarse para moldear más piezas de fundición.

La complejidad de la detección de defectos y características de los productos resultantes surge principalmente del elevado número de variables que participan en el proceso de fundición de piezas, y por tanto, pueden influir en el diseño final de los productos. De esta manera, las variables que han sido obtenidas para los experimentos que se detallan en esta tesis doctoral fueron más de 300 diferentes. Sin embargo, y gracias a la ayuda de expertos en el dominio, esta ingente cantidad de variables fue reducido a 24. Más concretamente, las variables a tener en cuenta del proceso al completo pueden ser divididas en (i) variables relacionadas con el metal y (ii) variables relacionadas con el molde. En concreto, las variables a controlar son las siguientes:

3. La fundición, un antiguo proceso rodeado de magia

- **Variables relacionadas con el metal:**

- *Composición:* inoculación, tipo de tratamiento, y los costes.
- *Calidad de la mezcla y potencial de nucleación:* Se obtiene por medio de un programa de análisis térmico [LSS06].
- *Vertido:* Duración del proceso de fundición y la temperatura.

- **Variables relacionadas con el molde:**

- *Arena:* Tipo de aditivos utilizados, las características específicas de arena y realización de pruebas anteriores o no.
- *Moldeado:* la máquina de moldeo utilizada y parámetros

Por lo general, la dimensión y la geometría de la pieza fundida desempeñan un papel muy importante y, por tanto, también se incluyen diferentes variables para mantener el control de estas dos características. De la misma manera, el sistema debe tener en cuenta los parámetros relacionados con la configuración de cada máquina que trabaja en el proceso de fabricación [SSLG07]. Además, se añaden otras variables como la velocidad de enfriamiento y el tratamiento térmico aplicado a la pieza.

3.2 Defectos y características a predecir

Dentro de los procesos productivos, y más concretamente dentro del de la fundición, existe la necesidad de detectar numerosos defectos y características sobre los productos resultantes. Es por ello que, a través de la investigación aquí presentada, se busca la producción de piezas que no desfallezcan durante su utilización, así como reducir el coste del propio proceso de producción, evitando tener que volver a fabricar las piezas detectadas como defectuosas. En concreto, los experimentos presentados se centran en la predicción y solución de dos de las características más complejas de prever en las piezas de fundición: (i) los microrrechupes y (ii) las propiedades mecánicas. En este proceso de investigación se parte de este defecto y esta característica para la validación del sistema desarrollado y con el objetivo de abrir el camino de la optimización de muchos otros defectos y características.

3.2.1 Microrrechupes

Los microrrechupes, también llamados *contracción secundaria*, se tratan de un defecto o irregularidad en las piezas de fundición que suelen aparecer en la fase de enfriamiento de la pieza. Particularmente, este defecto consiste en una forma de rechupe en el que las cavidades son muy pequeñas pero muchas en número y que pueden estar distribuidas sobre un área importante de la pieza. La razón fundamental de la aparición de este defecto es que los metales son menos densos como líquidos que como sólidos, y durante el proceso de enfriamiento, la densidad del metal crece mientras el volumen decrece en paralelo, pudiendo aparecer estas porosidades interdendríticas que reducen la dureza de la pieza, lo que la torna en una pieza no apta para su uso.

La mayor problemática de este defecto es que no puede ser detectado hasta que la producción de la pieza haya finalizado. Los procesos actuales para determinar si una pieza tiene este tipo de defecto se basan en la utilización de rayos X o la emisión de ultrasonidos evitando tener que realizar inspecciones destructivas, en otras palabras, sin la necesidad de acabar con un espécimen para poder hacer las pruebas. La figura 3.2 hace referencia al proceso de análisis de una pieza mediante rayos X para mostrar la zona en la que se encuentra el microrrechupe. Ambos métodos necesitan de personal especializado, maquinaria adecuada y bastante tiempo para poder realizar los análisis. Además, por desgracia, el único momento en el que podemos realizar el examen de la pieza es una vez ya está acabada. Por consiguiente, este tipo de soluciones no las podemos considerar como óptimas para la producción de piezas sin este defecto.

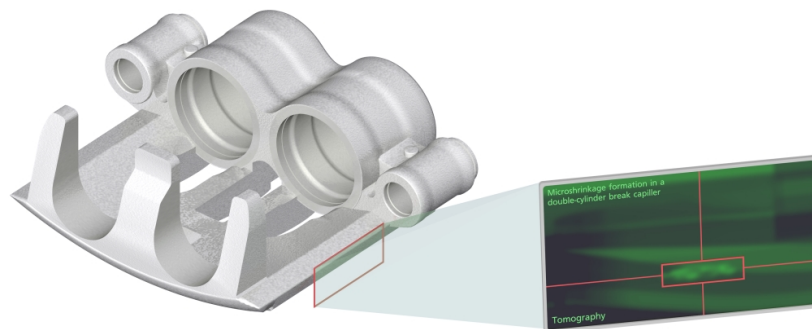


Figura 3.2: Proceso de análisis de una pieza mediante rayos X para la detección de microrrechupes.

Dado el criterio de aceptación o rechazo de los modelos estudiados, los experimentos desarrollados tratan de asemejarse a los tratamientos reales que

3. La fundición, un antiguo proceso rodeado de magia

hacen los clientes sobre los productos manufacturados. En ese sentido, para clasificar los especímenes se han definido una serie de niveles de riesgo posibles. La utilización de estos niveles de riesgo es lo que permite determinar la tendencia que va a seguir la producción en lo que a los defectos de microrechupes se refiere. Concretamente, para este defecto hemos definido cuatro niveles de riesgo. Esos niveles son los siguientes:

- **Riesgo 0:** no se prevé microrrechupes.
- **Riesgo 1:** la probabilidad de que exista microrrechupe es ínfima.
- **Riesgo 2:** existe una probabilidad baja de que aparezca algún microrrechupe en la pieza.
- **Riesgo 3:** la aparición de microrrechupes en la pieza está casi asegurada.

De los cuatro definidos, el *riesgo 0* y el *riesgo 1* no suponen ningún problema para el proceso. Sin embargo, el *riesgo 2* comienza a tenerse en cuenta para realizar modificaciones con el fin de conseguir evitar este defecto. Y finalmente, el *riesgo 3* es con el que obligatoriamente deben modificarse los parámetros de la producción.

3.2.2 Propiedades mecánicas

Una vez que se ha terminado el proceso de producción de las piezas, éstas tienen que estar sujetas a una serie de cargas y fuerzas durante su ciclo de vida. Los ingenieros especializados deben realizar las pertinentes medidas para determinar cómo va a comportarse la pieza. Es decir, si la pieza se va a deformar o romper durante su utilización debido a las fuerzas que se le apliquen. Estas características que cumplen las piezas son las que se conocen como propiedades mecánicas y es de gran importancia saber cómo afectan a las piezas de fundición [GCFC06].

Las propiedades mecánicas se pueden agrupar en las siguientes categorías [LM92]:

- **Fuerza:** es la propiedad que permite a un metal resistir la deformación bajo una carga. Hay muchas clases de fuerza tales como la resistencia a la rotura y la resistencia a la tracción.
- **Dureza:** es la propiedad de resistir la indentación permanente.

3.2 Defectos y características a predecir

- **Tenacidad:** es la propiedad que permite que un material soporte un choque y que lo haga sin deformarse y sin romperse. Esta propiedad es considerada como una combinación de resistencia y plasticidad.
- **Resistencia:** es la propiedad de un material para absorber la energía cuando se deforma elásticamente.
- **Elasticidad:** es la habilidad de un material para volver a su forma original después de retirarle la carga que lo ha deformado.
- **Plasticidad:** es la capacidad de un material para deformarse de forma permanente sin romperse.
- **Fragilidad:** un material quebradizo es aquel que se rompe antes de sufrir una deformación. Por lo general, materiales frágiles tienen un alto valor de resistencia a la compresión pero un valor bajo en la resistencia a la tracción.
- **Ductilidad:** es la propiedad que permite a un material estirarse, doblarse o girar sin agrietarse o romperse.
- **Maleabilidad:** en comparación con la ductilidad, es la propiedad que permite a un material deformarse mediante fuerzas de compresión sin desarrollar defectos en ese material. Un material maleable se puede sellar, martillar, forjar, prensar o laminar en láminas delgadas.

Para establecer los valores de estas propiedades mecánicas, los científicos tienen que probar los materiales en un laboratorio usando procedimientos estándar (por ejemplo, las normas ASTM [fTM08b, fTM08a]).

De todas las propiedades mecánicas, en esta tesis doctoral nos hemos centrado en la carga máxima. Esta propiedad mecánica pertenece a la categoría de fuerza y es la propiedad que determina la fuerza máxima necesaria para romper la pieza al aplicarle una fuerza de tensión sobre la misma. A pesar de esta selección, el método utilizado es fácilmente extrapolable para las demás propiedades mecánicas.

Desafortunadamente, la única manera de estudiar las propiedades mecánicas, y concretamente la carga máxima, es a través de inspecciones destructivas, es decir, mediante la destrucción de especímenes con el fin de determinar los valores. También, y al igual que con el defecto anterior, el proceso requiere de dispositivos adecuados, personal especializado, un buen tiempo para analizar los materiales y que la realización sea llevada a cabo una vez las piezas ya

3. La fundición, un antiguo proceso rodeado de magia

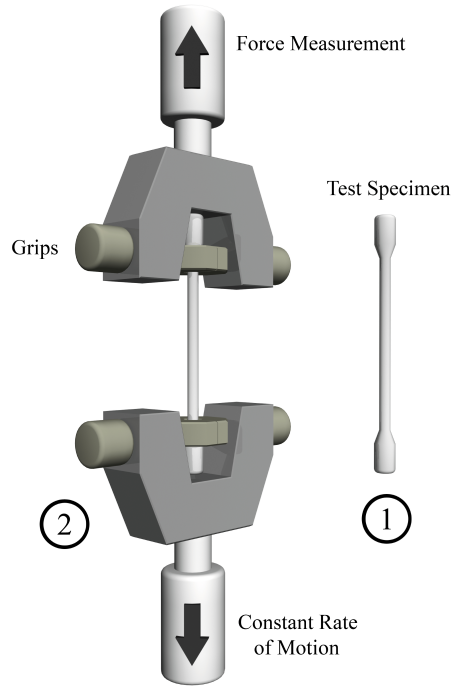


Figura 3.3: Proceso de análisis de la resistencia máxima a la tracción de una pieza.

están finalizadas. Por lo que, incorporar un nuevo método capaz de determinar el resultado final de las propiedades mecánicas antes de realizar las piezas mejorará el proceso.

En cuanto a la carga máxima, su método de prueba se lleva a cabo de la siguiente manera. En primer lugar, un científico prepara una muestra de la pieza original para realizar las pruebas (ver (1) en la figura 3.3). En segundo lugar, la muestra se coloca en la máquina de tracción (2). Y, por último, esta máquina tira de ambos extremos de la muestra para medir la fuerza necesaria para romper el espécimen y cuánto se alarga éste antes de romperse.

En el caso de las propiedades mecánicas, y más concretamente en la carga máxima, el proceso habitual llevado a cabo en la fundición únicamente tiene en cuenta que se sobrepase el umbral de riesgo marcado. De este modo, para esta característica se ha realizado una división simple, y es la siguiente:

- **Riesgo 0:** se sobrepasa el umbral de las 370 MPa ¹.

¹Megapascuales, unidad de presión donde 1 Pascal (Pa) es la presión que ejerce una fuerza de 1 Newton sobre una superficie de 1 metro cuadrado normal a la misma.

- **Riesgo 1:** no se sobrepasa el umbral de las 370 MPa.

En este caso, la producción necesitará ser modificada cuando no se sobrepase el límite indicado, en otras palabras, cuando los elementos de predicción nos indiquen que la producción en el instante de tiempo $t + 1$ se encontrará en el *riesgo 1*.

3.3 Sumario

En este capítulo hemos revisado el caso de uso sobre el que se desarrollan los experimentos de esta tesis doctoral. En concreto, hemos comenzado revisando las actividades que pueden realizarse sobre los diferentes materiales metálicos. Del mismo modo, completamos esta descripción con los diferentes metales que pueden ser utilizados en cada uno de los procesos ya explicados.

Concretamente, para el desarrollo de esta investigación hemos seleccionado la producción de piezas metálicas que son prácticamente el producto terminado y que van a estar hechas tanto con materiales ferrosos (hierro) como con no ferrosos (aluminio).

Posteriormente, dado el proceso seleccionado hemos realizado una breve descripción de las etapas en las que se divide, partiendo de la generación de los patrones para los moldes, pasando por la generación de moldes y del metal fundido que será vertido en los moldes, para llegar a las etapas finales de la producción como son la limpieza y las correcciones finales. Del mismo modo, una vez determinadas las etapas del proceso, hemos indicado cuáles son las variables que afectan a la producción de las piezas y cuáles han sido utilizadas para los procesos de aprendizaje automático.

Finalmente, se ha pasado a determinar los defectos y características de las piezas sobre los que se centrará la investigación. En primer lugar, se define qué son los microrrechupes, pequeñas porosidades que se encuentran en el interior de las piezas. Y en segundo lugar, se explican y se definen las diferentes propiedades mecánicas, las características que determinan el comportamiento de la pieza una vez terminada, indicando que la propiedad sobre la que se va a trabajar es la carga máxima.

«A veces, cuesta mucho más eliminar un sólo defecto que adquirir cien virtudes.»

Jean de la Bruyère (1645-1696)

4

Predicción del estado estacionario $t + 1$

LOS Modelos Predictivos de Control (MPC) son una serie de algoritmos que permiten controlar todas y cada una de las etapas de una planta de producción. Este tipo de sistemas se ha vuelto muy popular para la gestión de los procesos. En sus inicios fueron utilizados para el control de plantas dedicadas a refinar petróleo, pero en la actualidad están implantados en numerosas áreas de producción [AT00].

El gran éxito de los sistemas MPC se debe a tres factores importantes [ML97]. Entre ellos destaca uno, el planteamiento de funcionamiento con el que cuentan, en otras palabras, estos algoritmos consideran el comportamiento de la planta dentro de una dimensión temporal. Tanto es así, que el funcionamiento básico del sistema se basa en la ejecución de un ciclo iterativo en el que la primera de sus tareas es la predicción del estado estacionario en el que se encontrará todo el proceso productivo en un instante de tiempo $t + 1$ localizado en un futuro cercano.

Como ya dejamos recogido en la sección 2.2 existen diferentes MPC dependiendo de su naturaleza (lineales y no lineales) y dependiendo de cómo fueron creados (empíricos y basados en principios fundamentales). La primera de las clasificaciones se basa en el modelo con el que cuenta el MPC, el cual tiene que ajustarse al proceso. No obstante, los más utilizados son los modelos

4. Predicción del estado estacionario $t + 1$

lineales a pesar de que en muchas ocasiones el proceso no lo sea. Normalmente, esto se debe a que las mejoras producidas por los modelos no lineales no son comparables a los esfuerzos necesarios para generarlos y gestionarlos [Hen98]. Además, la complejidad de los modelos y los bajos tiempos de ciclo necesarios nos hacen tener que plantearnos la utilización de sistemas subóptimos [CB04].

Igualmente, el modelo existente dentro de un MPC suele tratarse de un modelo extraído del proceso de producción y que se traduce en una o varias formulas matemáticas que intentan ajustarse al proceso. Sin embargo, para todo su desarrollo ha sido necesario contar con personal experto en el dominio que conoce el funcionamiento a bajo nivel del proceso productivo, expertos en la generación de sistemas MPC para el desarrollo final del modelo; y, en caso de tratarse de un sistema empírico, el control de la planta durante una serie de días para la extracción de la información base que generará el modelo (para más información diríjase a la sección 2.2).

Contra toda esta problemática aparecen los métodos de aprendizaje automático supervisado. Estos algoritmos no han sido introducidos dentro de los sistemas MPC, a excepción de las redes neuronales artificiales utilizadas para realimentar la planta con el fin de solucionar los problemas que sucedían durante el proceso productivo o como la base del modelo no lineal detrás del sistema MPC. Por ende, en este capítulo introducimos la utilización de algoritmos de aprendizaje automático supervisados. Estos algoritmos permiten trabajar tanto en la dimensión lineal como en la no lineal. Además, a través de ellos, el conocimiento es extraído sin necesidad de disponer de un conocimiento experto en el dominio. Así, solucionamos los problemas inherentes a la etapa de predicción que acabamos de comentar.

A pesar de ello, se da la situación de que no siempre podemos seleccionar el mismo algoritmo como el óptimo para la predicción de todos los defectos o características del proceso productivo. Por esa razón, la segunda parte del capítulo se centra en la generación de un sistema de meta-clasificación en el que el conocimiento adquirido por múltiples algoritmos de aprendizaje automático supervisado pueda ser compartido con el objetivo de generar una inteligencia de clasificación colectiva. De esta forma, los sistemas MPC evolucionan al poder generar un sistema de clasificación híbrido basado en la utilización de algoritmos de ambos tipos, lineales y no lineales, que trabajarán de forma conjunta.

A modo de resumen, las contribuciones que se han realizado en este contexto son las siguientes:

4.1 Métodos de aprendizaje supervisado

- Demostramos que el conocimiento de la planta de producción puede ser modelado a través de técnicas de aprendizaje automático.
- Mostramos que la utilización de algoritmos de aprendizaje automático supervisado es factible para la generación del modelo dedicado a la etapa de la predicción de un sistema MPC.
- Proveemos una validación empírica de que nuestro método de predicción basado en un único clasificador es factible para los procesos productivos actuales.
- Evaluamos si es posible la utilización de un único algoritmo de clasificación para las predicciones de diferentes características del proceso.
- Describimos cómo adoptar los métodos de combinación de clasificadores con el fin de obtener un sistema de predicción híbrido.
- Evaluamos de forma empírica los diferentes conjuntos de clasificadores generados para la predicción conjunta de los defectos.

Lo que resta de capítulo está organizado de la siguiente forma. La sección 4.1 analiza los diversos métodos de aprendizaje supervisado que han sido utilizados para la etapa de predicción del estado estacionario $t+1$. Posteriormente, en la sección 4.2 analizamos los métodos de combinación de clasificadores que permiten unir los resultados de un conjunto heterogéneo de clasificadores. Las secciones 4.3 y 4.4 muestran los experimentos desarrollados con el fin de generar un meta-clasificador para la predicción de múltiples características (en nuestro caso, y por tratarse de un proceso de fundición, trabajaremos con los microrrechupes y las propiedades mecánicas) y discuten los resultados obtenidos. Finalmente, en la sección 4.5 se pasa a realizar un resumen de las contribuciones del capítulo.

4.1 Métodos de aprendizaje supervisado

Con el fin de solucionar la problemática definida anteriormente utilizaremos métodos de aprendizaje supervisado, también denominados algoritmos de clasificación. Particularmente, se trata de la búsqueda de algoritmos que pretendan producir una hipótesis general basada en una serie de datos etiquetados que le han sido proporcionados (en nuestro caso el devenir de la producción con el resultado de si existe el defecto o no y la medición de la carga máxima)

4. Predicción del estado estacionario $t + 1$

[KZP07]. En otras palabras, el objetivo del aprendizaje supervisado es el de construir un modelo conciso de la distribución de etiquetas recogidas en los datos para conseguir proporcionar una serie de características de predicción, básicas en la generación de un MPC. Así, en esta sección se revisarán los diferentes algoritmos de aprendizaje supervisado utilizados para generar la etapa de predicción del estado estacionario $t + 1$ del sistema MPC.

4.1.1 Redes bayesianas

La investigación sobre las relaciones causa-consecuencia fue iniciada por el reverendo Thomas Bayes [Bay63], y su trabajo principal es conocido como el “teorema de Bayes” en su honor. De acuerdo a su formulación clásica, dados dos eventos A y B , la probabilidad condicionada de $P(A|B)$ en la que A se produce si se produce B se puede obtener en caso de que se conozca la probabilidad de que A ocurra, $P(A)$, la probabilidad de que B ocurra, $P(B)$, y la probabilidad condicional de B dado A , $P(B|A)$ (como se muestra en la ecuación 4.1)

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (4.1)$$

Partiendo de la idea inicial del reverendo Thomas Bayes nacen las redes bayesianas. Más concretamente, estas redes son modelos probabilísticos que permiten desarrollar un análisis multivariable. Las redes bayesianas pueden ser presentadas gráficamente mediante un grafo dirigido acíclico unido a la distribución de probabilidades asociada a las conexiones representadas en ese grafo (i.e., el conjunto de relaciones probabilísticas entre las variables recogidas a la hora de modelar un problema en particular) [CGH96]. De esta forma, la representación se compone de los siguientes elementos (ver figura 4.1):

- **Nodos.** Representan las variables que influyen en el proceso. Los nodos tienen la capacidad de representar tanto los antecedentes de las condiciones a analizar como los consecuentes a los que tenemos que llegar a través del proceso de inferencia que permite realizar las operaciones de clasificación.
- **Aristas.** Estas representan las relaciones existentes entre los diferentes nodos que componen el modelo. Además, al tratarse de un grafo dirigido, las aristas indican el sentido en el que los nodos influyen a los de su alrededor.

- **Función de probabilidad.** Finalmente, este elemento ilustra la fortaleza de las relaciones definidas a través de las aristas. Esta función es la que almacena la distribución probabilística aprendida a través de un conjunto de datos de aprendizaje.

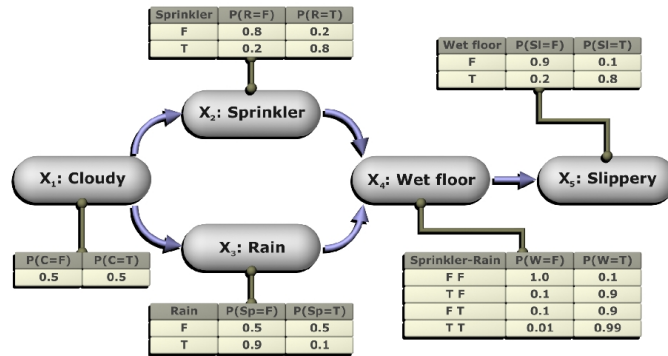


Figura 4.1: Red bayesiana de ejemplo para el control de los aspersores.

Utilizamos este tipo de modelos para varias actividades, por ejemplo, para el desarrollo de una máquina de aprendizaje basada en datos históricos, la búsqueda de patrones sobre datos ambiguos o incompletos, minería de datos para el descubrimiento de las relaciones y la inferencia de variables no observables dado el resto del conjunto [PR00]. En particular, esta capacidad de inferencia se adapta a la necesidad de nuestro MPC. Estas capacidades hacen que las redes bayesianas representen un super-conjunto semántico de los sistemas expertos basados en el encadenamiento de reglas, tanto para adelante como para atrás (de hecho, los modelos bayesianos permiten un tercer tipo más de inferencia, que se conoce como *explicación* o *justificación* [CGH96]). Por otra parte, una red bayesiana puede crecer ampliando su base de conocimientos con nuevas evidencias, sin reducir su nivel de rendimiento [CGH96].

Las redes bayesianas normalmente cuentan con dos etapas durante su proceso de creación: (i) aprendizaje estructural y (ii) aprendizaje paramétrico. De esta forma, el primero de ellos es el encargado de descubrir las relaciones entre las variables que incorpora el modelo. El segundo de ellos, es el encargado de descubrir la distribución estadística que mejor representa los datos que se han tomado como entrenamiento de la red. Para llevar a cabo esta etapa de aprendizaje existen diferentes algoritmos. En esta investigación hemos utilizado el algoritmo *K2* [CH91, CH92], el *escalar colinas* (traducción literal de la voz inglesa *hill climbing*) [Bun96, RN03] y *Tree Augmented Naïve* (TAN) [FGG97],

4. Predicción del estado estacionario $t + 1$

además de la utilización de un clasificador bayesiano ingenuo o *Naïve Bayes* [Lew98].

4.1.1.1 Método de aprendizaje *escalar colinas*

Para este tipo de redes bayesianas se hace uso del algoritmo *escalar colinas*. Particularmente, este algoritmo es una optimización matemática de las técnicas de búsqueda local¹ [RN03]. El algoritmo *escalar colinas* intenta maximizar (o minimizar) una función objetivo $f(\mathbf{x})$, donde \mathbf{x} es un vector de valores continuos y/o discretos. Para ello, parte de una solución generada de forma arbitraria. En cada una de las iteraciones, el algoritmo ajusta un elemento en concreto del vector \mathbf{x} y determina si el cambio mejora el valor actual de la función $f(\mathbf{x})$. El algoritmo *escalar colinas* admite cualquier tipo de cambio siempre y cuando esa modificación haga mejorar la función objetivo. El proceso continuará hasta que no se pueda encontrar ningún cambio que continúe mejorando el resultado de la función $f(\mathbf{x})$. Entonces, en ese momento se dice que los valores contenidos en el vector \mathbf{x} son “localmente óptimos”.

Gracias a este algoritmo estas redes bayesianas serán aprendidas mediante un óptimo local, lo que no garantiza que se haya encontrado la mejor de las soluciones posible (i.e., el óptimo global) de todo el espacio de búsqueda disponible. A pesar de ello, la relativa simplicidad de este algoritmo le ha convertido en la primera de las opciones entre los algoritmos de optimización.

Así, aplicando este algoritmo al aprendizaje bayesiano conseguimos añadir, borrar y direccionar las aristas (más correctamente, revertir su dirección) que forman la estructura de la red bayesiana. También, al realizar esta búsqueda, no se tiene en cuenta el orden de las variables, algo que sí sucede en el algoritmo *K2* [Bun96].

4.1.1.2 Algoritmo de aprendizaje *K2*

El algoritmo *K2* tiene como objetivo la generación de la topología de la red bayesiana. Para ello, intenta encontrar mediante una búsqueda heurística las relaciones más probables dado un conjunto de datos de aprendizaje. Básicamente, se trata de una adaptación del método *escalar colinas* en el que no se

¹Método para la resolución de problemas de optimización computacionalmente complejos. Este tipo de algoritmos se desplaza de una solución posible a otra dentro del espacio de soluciones candidatas (espacio de búsqueda) mediante la aplicación de cambios locales, hasta que una solución considerada como óptima sea encontrada.

4.1 Métodos de aprendizaje supervisado

considera las aristas iniciales generadas a través de un clasificador bayesiano ingenuo para, posteriormente, tener que eliminarlas [FGG97].

De manera más formal, siendo Z el conjunto de n variables discretas, donde la variable x_i en Z tiene r_i valores: $(v_{i1}, \dots, v_{ir_i})$, siendo D el conjunto de datos de entrenamiento con m casos, donde cada caso tiene un valor de asignación para cada una de las variables de Z , siendo B_s la estructura de red más probable para las variables de Z y cada una de las variables x_i en B_s tiene un conjunto de padres π_i ; el algoritmo $K2$ trata de maximizar (o casi maximizar) la función $P(B_s|D)$.

$$\max_{B_s} [P(B_s, D)] = c \prod_{i=1}^n \max_{\pi_i} \left[\prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{i,j} + r_i - 1)!} \cdot \prod_{k=1}^{r_i} \alpha_{i,j,k} \right] \quad (4.2)$$

Para lograr su objetivo, se parte de la ecuación 4.2¹ pero con una serie de supuestos: (i) las variables del dominio se encuentran ordenadas (otra de las diferencias con el método de *escalar colinas*) y (ii) todas las estructuras se consideran igualmente probables. De la misma manera, se modifica la operación de maximización a la derecha de la ecuación 4.2 para realizar una búsqueda mucho más exhaustiva o codiciosa. Concretamente, se utiliza un algoritmo que asume que ningún nodo tiene padres, y al que se le van a ir añadiendo incrementalmente padres siempre que maximicen la función objetivo definida. Cuando ninguna nueva adición pueda incrementar el resultado de la función objetivo, se finalizará con el proceso de búsqueda de padres para ese nodo en concreto, pasando a procesar el siguiente de la lista. Para ello, se utiliza la función:

$$f(i, \pi_i) = \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{i,j} + r_i - 1)!} \cdot \prod_{k=1}^{r_i} \alpha_{i,j,k} \quad (4.3)$$

donde $\alpha_{i,j,k}$ se calcula en relación a π_i , siendo los padres de x_i y relativos al conjunto de datos D . Además, se define u como el máximo número de padres para cualquiera de los nodos.

Al igual que el método anterior, el algoritmo $K2$ suele ser utilizado para el cálculo de resultados preliminares y conocer si el método de las redes bayesianas puede ser el correcto para solucionar nuestro problema de clasificación.

¹Función a maximizar en el aprendizaje estructural de una red bayesiana cuando se utiliza un método de búsqueda exhaustiva [CH91].

4. Predicción del estado estacionario $t + 1$

4.1.1.3 Clasificador bayesiano ingenuo (*Naïve Bayes*)

El clasificador bayesiano ingenuo, también conocido como “*Naïve Bayes*”, se trata de un clasificador que aplica el teorema desarrollado por el reverendo Thomas Bayes [Bay63] pero partiendo de una hipótesis de independencia fuerte entre variables. De forma más simple, un clasificador bayesiano ingenuo asume que la presencia (o la ausencia) de una característica particular de una clase no está relacionada directamente con la presencia (o ausencia) de otra característica para la clase sobre la que realiza las predicciones [JL95, Lew98]. Por consiguiente, todas las variables, o nodos de la red bayesiana, se encuentran directamente vinculados con la variable sobre la que sacar conclusiones.

Dependiendo de la naturaleza del modelo, los clasificadores bayesianos ingenuos pueden ser entrenados de una forma muy eficiente a través del aprendizaje supervisado. Incluso, en muchas aplicaciones prácticas puede realizarse una estimación para los parámetros del modelo utilizando el método de la máxima verosimilitud¹ (traducción del inglés “*maximun likelihood*”). En otras palabras, se puede llegar a trabajar con este tipo de clasificadores sin llegar a utilizar la probabilidad bayesiana u otro tipo de métodos bayesianos. La ventaja que aportan estos clasificadores es que únicamente necesitan de una pequeña cantidad de datos de entrenamiento para estimar los parámetros (las medias y las varianzas de las variables) necesarios para la clasificación. Ya que se parte de la premisa de que las variables son independientes y sólo debe determinarse las varianzas para cada clase y no la matriz de covarianza completa [Lew98].

Para la generación de un clasificador bayesiano ingenuo, se realiza una combinación de un modelo bayesiano de producción con una regla de decisión. La regla más comúnmente utilizada es aquella que toma la hipótesis más probable, conocida como *máximo a posteriori*. De esta forma, la función clasificar(f_1, \dots, f_n) queda definida como se ilustra en la ecuación 4.4.

$$\operatorname{argmax}_c p(C = c) \prod_{i=1}^n p(F_i = f_i | C = c). \quad (4.4)$$

A pesar del hecho de que la suposición de independencia puede ser a menudo imprecisa, los clasificadores bayesianos ingenuos tienen varias propiedades que les hace sorprendentemente útiles en la práctica [JL95]. En particular,

¹En general, para un conjunto de datos fijo y el modelo estadístico subyacente, el método de la máxima verosimilitud selecciona los valores de los parámetros que generan una distribución que da una mayor probabilidad a los datos que han sido observados [Aka73].

el desacoplamiento de las distribuciones de las características condicionadas por la clase significa que puede estimarse independientemente, es decir, como una distribución unidimensional. Esto ayuda a mitigar problemas derivados de la maldición de la dimensionalidad [Tor03]. Además, el clasificador general es lo suficientemente robusto para ignorar las deficiencias de su modelo probabilístico ingenuo [JL95, Lew98].

4.1.1.4 El método *Tree Augmented Naïve* (TAN)

Este tipo de clasificadores bayesianos, *Tree Augmented Naïve Bayes*, es una evolución de los clasificadores bayesianos ingenuos. Particularmente, la variable o nodo clase (sobre el que realizar la predicción final) no va a tener ningún padre, y el resto de las variables que componen el modelo tendrán como padres la variable conclusión y, como mucho, otra variable más. Con lo que cada nodo puede tener una nueva arista que enlace con él [FGG97]. La forma utilizada para la generación de esas nuevas aristas es la que fue presentada en [CL68] con el fin de permitir la creación de redes bayesianas con forma de árbol.

Con esta idea de fondo, un grafo acíclico (es decir, la red bayesiana) en $\{X_1, \dots, X_n\}$ tiene forma de árbol si Π_{X_i} tiene exactamente un padre para todos los X_i , excepto para una de las variables que no tienen ningún padre (esta es la variable que se nombra como la *raíz*). Una red en forma de árbol puede ser descrita identificando los padres de cada variable. Así, una función $\pi : \{1, \dots, n\} \mapsto \{0, \dots, n\}$ se dice que define un árbol para X_1, \dots, X_n , si exactamente hay un i tal que $\pi(i) = 0$ (la raíz del árbol), y no hay una secuencia i_1, \dots, i_k tal que $\pi(i_j) = i_{j+1}$ para $i \leq j < k$ y $\pi(i_k) = i_1$ (en otras palabras, no hay ciclos). De esta forma, se define una red en forma de árbol mediante la ecuación $\Pi_{X_i} = \{X_{\pi(i)}\}$ si $\pi(i) > 0$, y $\Pi_{X_i} = \emptyset$ si $\pi(i) = 0$.

Entonces, la construcción de un clasificador bayesiano a través del método TAN sigue los pasos de la generación de redes en forma de árbol desarrollado por Chow et ál. en [CL68] y los completa con aquellos que permiten finalizar la generación del modelo TAN. Más concretamente los pasos son los siguientes [FGG97]:

1. Se calcula $I_{\hat{p}_D}(A_i; A_j|C)$ para cada pareja de variables que cumplan $i \neq j$. Concretamente, se trata de la función para el cálculo de la información mutua condicional¹ (traducción de la voz inglesa “*conditional mutual*”).

¹La información mutua trata de medir cuanta cantidad de información de \mathbf{Y} proporciona en relación a \mathbf{X} [PLD05].

4. Predicción del estado estacionario $t + 1$

information”), representado en la ecuación 4.5.

$$I_P(\mathbf{X}; \mathbf{Y} | \mathbf{Z}) = \sum_{\mathbf{x}, \mathbf{y}, \mathbf{z}} P(\mathbf{x}, \mathbf{y}, \mathbf{z}) \frac{P(\mathbf{x}, \mathbf{y} | \mathbf{z})}{P(\mathbf{x} | \mathbf{z}) P(\mathbf{y} | \mathbf{z})} \quad (4.5)$$

2. Se construye completamente un grafo no dirigido en el que los vértices son las variables presentes en A . Sobre ese grafo, se realizan unas anotaciones sobre el peso $I_{\hat{p}_D}(A_i; A_j | C)$ asignado a las aristas que conectan A_i con A_j .
3. Partiendo de los resultados anteriores, se construye un árbol que sea el máximo ponderado (traducción del nombre inglés “*weighed maximum spanning tree*”) [FGG97].
4. Se convierte el árbol generado en un árbol dirigido mediante la selección de un nodo raíz y seleccionando la dirección de las aristas para que partan desde el nodo seleccionado.
5. Finalmente se construye la estructura TAN mediante la adición de un nuevo nodo, etiquetado como C y se añaden aristas que enlacen este nodo con cada uno de los nodos A_i .

A pesar de que utilizando el método TAN se parte de un clasificador bayesiano ingenuo, incorporando, en primer lugar, sus ventajas y sus desventajas, la ampliación realizada incorpora un compromiso entre la calidad de las aproximaciones de las correlaciones y la complejidad de la etapa de aprendizaje, con lo que mejora el modelo anterior pero mantiene una estructura sencilla [FGG97].

4.1.2 K vecinos más próximos

El algoritmo “*K-Nearest Neighbour*” (KNN) o de los k vecinos más próximos [FH52] se trata de uno de los algoritmos de aprendizaje automático supervisado para la clasificación de casos más simples. Este algoritmo basa su método de clasificación en la clase a la que corresponden la mayoría de los casos más cercanos a la instancia desconocida, la que va a ser clasificada. En la figura 4.2 se puede visualizar cómo han sido repartidas las evidencias por un espacio bi-dimensional y cómo se han seleccionado las más cercanas a la evidencia de la que no se conoce su clase.

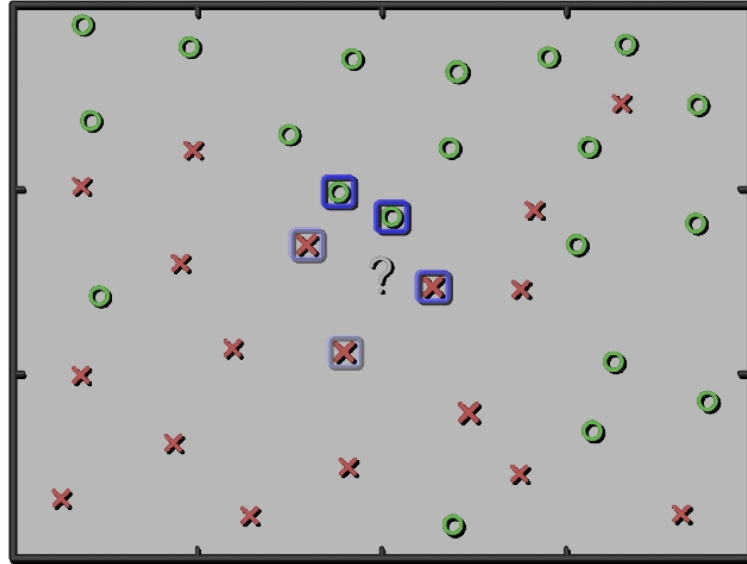


Figura 4.2: Ejemplo de un clasificador KNN en un espacio bi-dimensional.

Por una parte, la fase de entrenamiento de este algoritmo se trata de una representación del conjunto de datos de entrenamiento, las instancias $S = (s_1, s_2, \dots, s_m)$, a lo largo de un espacio n -dimensional en el que n es la cantidad de variables para cada instancia (es decir, en nuestro caso las variables adquiridas del proceso productivo de la fundición).

Por otra parte, la fase de clasificación de una instancia cuya clase es desconocida, se realiza midiendo la distancia entre las instancias que ya han sido representadas en el espacio n -dimensional y la instancia desconocida. De esta manera, para calcular la distancia entre dos puntos X e Y se puede conseguir mediante el uso de cualquier medida de distancia. En nuestro caso hemos utilizado la distancia euclidiana (ver la ecuación 4.6). Durante el proceso de clasificación se pueden seleccionar una cantidad diferente de vecinos para realizar la comprobación de sus tipos. No se puede saber de antemano cuál es la cantidad idónea. Por ello, durante los procesos de pruebas hay que conseguir descubrir ese número. Normalmente, el número de vecinos óptimos suele ser considerado a partir un máximo local¹ para, así, evitar que los experimentos puedan llegar a dilatarse en el tiempo.

¹Como se podrían extender las pruebas demasiado, se considera que cuando se ha encontrado un punto en el que los resultados tienen un mayor nivel de acierto que a sus lados es que se ha encontrado el valor óptimo con el que hay que trabajar.

4. Predicción del estado estacionario $t + 1$

$$\sqrt{\sum_{i=0}^n (X_i - Y_i)^2} \quad (4.6)$$

Por último, hay varias formas de elegir la clase de la instancia desconocida, la técnica más empleada consiste en clasificar la instancia desconocida como la clase más común entre los K -vecinos más próximos. Esta técnica es la que ha sido utilizada para el desarrollo de los experimentos que aquí se detallan.

4.1.3 Redes neuronales artificiales

Las *Redes Neuronales Artificiales* o en inglés “*Artificial Neural Networks*” (ANN) es un modelo matemático o computacional (i.e., aprendizaje automático supervisado) inspirado por la estructura y/o aspectos funcionales de las redes neuronales biológicas. Al fin y al cabo, tratan de simular el comportamiento de las neuronas del cerebro humano [Bis95]. Formalmente, una red neuronal consiste en una serie de neuronas interconectadas que procesan información utilizando un enfoque de la computación basado en conexiones. De esta manera, la activación de una neurona depende de su conjunto de entradas, donde y_i es la activación de la neurona actual, f_i es la función de activación, $W_{j,i}$ es el peso de la neurona y a_j es la activación de la neurona de entrada (así se muestra en la ecuación 4.7). Las redes neuronales artificiales modernas son herramientas de modelado de datos estadísticos no lineales, que normalmente son utilizadas para modelar relaciones complejas entre entradas y salidas o para buscar patrones imbricados en los datos.

$$y_i = f_i \left(\sum_{j=0}^n W_{j,i} \cdot a_j \right) \quad (4.7)$$

Más exactamente, lo que hemos utilizado a la hora de desarrollar los experimentos ha sido un tipo de red neuronal llamado *multilayer perceptron* (MLP) con una alimentación hacia delante de los datos que mapea el conjunto de datos de entrada en un conjunto de salidas apropiadas. Las MLP consisten en un conjunto de capas de nodos en un grafo dirigido, cada capa se encuentra totalmente conectada con la siguiente (ver la Figura 4.3).

A excepción de los nodos de entrada, el resto de los nodos son neuronas (o elementos de procesamiento) con una función de activación no lineal. Las neuronas de las MLP son conocidas como *perceptrones*. Para la creación del mo-

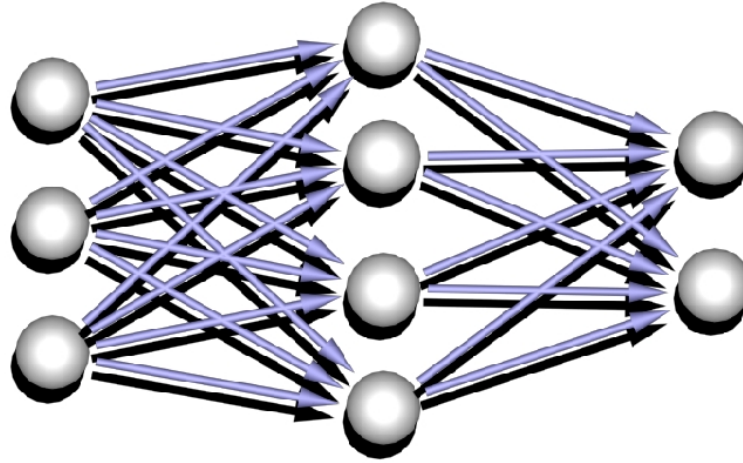


Figura 4.3: Representación gráfica de una red neuronal de tipo MLP.

delo se utiliza el algoritmo de *retropropagación* o *backpropagation* [MSTC94] que calcula los pesos de la función de activación para cada neurona. Las diferentes capas en las que se organizan este tipo de redes neuronales artificiales se pueden clasificar como:

- **Capa de entrada:** Esta capa tiene el mismo número de nodos que de variables definidas en el modelo. Por consiguiente, los valores que se dan a las variables del modelo son los que van a entrar a través de esos nodos.
- **Capas ocultas:** Se crean estas capas para adaptar el aprendizaje y van a ser utilizadas para conectar las neuronas de la entrada con las de la salida.
- **Capa de salida:** En la capa de salida va a haber tantas neuronas como grupos de clasificación. Según la excitación de estas neuronas se decidirá el resultado de la clasificación para la evidencia que entró al modelo.

En caso de que una MLP tuviera una función de activación lineal en todas sus neuronas se trataría de un simple mecanismo de encendido-apagado que determinaría cuándo se ha disparado una neurona, algo que podría hacerse fácilmente mediante álgebra lineal, viéndose reducido al mínimo el número de capas de estas redes, concretamente, a dos capas que formarían un modelo de entrada-salida básico. No obstante, lo que hace a las MLP diferentes es que cada neurona utiliza una función de activación no lineal que está desarrollada

4. Predicción del estado estacionario $t + 1$

para modelar la frecuencia de las potenciales acciones, o disparos, de las neuronas biológicas de un cerebro. Esta función se puede modelar de muchas formas diferentes. Las dos funciones de activación más utilizadas en las aplicaciones actuales son sigmoides [Bis95] y se describen en 4.8

$$\phi(y_i) = \tanh(v_i) \quad (4.8a)$$

$$\phi(y_i) = (1 + e^{-v_i})^{-1} \quad (4.8b)$$

en las que la primera de ellas es una tangente hiperbólica en el rango de -1 a 1 , y la segunda, una función logística, similar en forma pero en rangos que van de 0 a 1 . Aquí y_i es la salida del i -ésimo nodo (neurona) y v_i es la suma ponderada de las sinapsis de la entrada. Otras funciones de activación más especializadas utilizan funciones de base radial (traducción literal del termino inglés *radial basis functions*) y que son utilizadas en otros modelos de aprendizaje supervisado como las máquinas de soporte vectorial.

4.1.4 Máquinas de soporte vectorial

Las *máquinas de soporte vectorial* o *Support Vector Machines* (SVM) [Vap00] son un concepto estadístico relacionado con el aprendizaje supervisado. Este concepto está propiamente enfocado a la solución de problemas de clasificación y regresión. Más concretamente, dado un conjunto de datos de entrenamiento (muestras) podemos etiquetar los datos y entrenar un SVM con el fin de obtener un modelo con la capacidad de predecir la clase de una nueva instancia [Bis95].

De forma intuitiva, podemos pensar que un SVM es un modelo que representa los puntos de las muestras en el espacio. Una vez repartidos, trata de realizar una separación de las clases por un espacio lo más amplio posible. Cuando llegan nuevas muestras que necesitan de un proceso de clasificación se ponen en correspondencia con ese modelo y teniendo en cuenta la proximidad a esos conjuntos se puede clasificar en una u otra clase.

De forma más formal, un SVM construye un hiperplano o conjunto de hiperplanos en un espacio de alta dimensionalidad o infinito-dimensional, el cual es utilizado como frontera entre las diferentes clases sobre las que clasificar. Este hiperplano es el que maximiza la distancia entre los ejemplos o evidencias de las clases, como bien se ha comentado anteriormente. Esta separación entre estas regiones se denomina *margen funcional*. En general, cuanto

mayor sea el margen funcional, menor será el error de generalización del clasificador. Un ejemplo de esta representación se puede ver en la figura 4.4. Los *vectores de soporte* son las instancias que se encuentran cerca del hiperplano [Pla98, KSBM01, HT98].

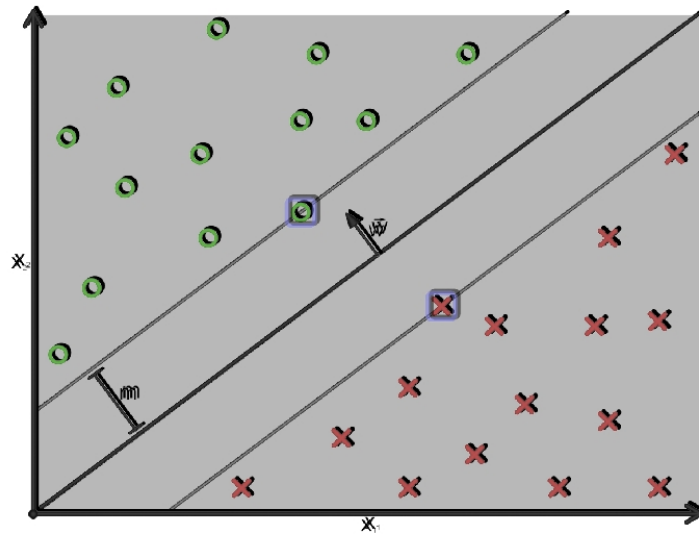


Figura 4.4: Representación gráfica de la división de un espacio bi-dimensional mediante SVM.

Nótese que si los datos de entrenamiento son linealmente separables, podremos seleccionar los dos hiperplanos del margen, de tal forma que no haya puntos entre ellos y se pueda maximizar esa distancia [Bis95]. Para conseguirlo utilizamos la geometría, concretamente, buscamos la distancia entre esos dos hiperplanos ($\frac{2}{\|\mathbf{w}\|}$) en la que tratamos de minimizar $\|\mathbf{w}\|$. También, con el fin de prevenir la aparición de los puntos en la zona intermedia se añaden unas restricciones, con lo que obtenemos que $y_i(\mathbf{w} \cdot \mathbf{x}_i - b) \geq 1$, para todos los i que cumplan $1 \leq i \leq n$. En este caso, \mathbf{w} es el vector normal al hiperplano, cada \mathbf{x}_i es un vector real p -dimensional, e y_i es -1 ó 1 indicando a que clase pertenece \mathbf{x}_i .

Pese a ello, la solución mostrada es compleja de resolver ya que depende de $\|\mathbf{w}\|$, lo que implica el cálculo de una raíz cuadrada. Esto puede ser sustituido por $\frac{1}{2}\|\mathbf{w}\|^2$, quedando representado como se muestra en la ecuación 4.9, es decir, que ahora en lo que nos centramos es en buscar un punto de referencia.

4. Predicción del estado estacionario $t + 1$

$$\min_{\mathbf{w}, b} \max_{\alpha \geq 0} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i [y_i (\mathbf{w} \cdot \mathbf{x}_i - b) - 1] \right\} \quad (4.9)$$

Este problema ahora puede ser solucionado a través de la programación cuadrática, un caso particular de la matemática no lineal, en la que la expresión anterior puede tratarse como una combinación lineal de vectores de entrenamiento (ver ecuación 4.10).

$$\mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i \quad (4.10)$$

Así, por una parte, solo unos pocos α_i serán mayores que cero. Y por otra parte, los correspondientes \mathbf{x}_i son los vectores de soporte. Partiendo de esta base, somos capaces de determinar el desplazamiento b . De todas formas, en la práctica es mejor realizar el cálculo de la media sobre todos los N_{SV} vectores de soporte, obteniendo la ecuación 4.11.

$$b = \frac{1}{N_{SV}} \sum_{i=1}^{N_{SV}} (\mathbf{w} \cdot \mathbf{x}_i - y_i) \quad (4.11)$$

A pesar de los cálculos y las optimizaciones realizadas, el espacio no puede ser dividido siempre por un hiperplano. En esos casos se hace uso de una función de núcleo k (o función de *kernel*). Esta función estudia las relaciones entre los datos y crea divisiones más complejas en el espacio que las que se pueden generar a través de los hiperplanos.

Escribiendo la regla de clasificación en su forma dual¹ revela que el máximo margen del hiperplano, y en su defecto, el problema de clasificación, es una función de los vectores de soporte. Por consiguiente, utilizando el hecho de que $\|\mathbf{w}\|^2 = \mathbf{w} \cdot \mathbf{w}$ y sustituyendo $\mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i$ se reduce la operación a la siguiente ecuación a maximizar en α_i .

$$\tilde{L}(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j) \quad (4.12)$$

¹En la teoría de la optimización, el principio de la dualidad manifiesta que los problemas pueden ser vistos desde dos perspectivas, el problema principal y el problema dual. Ambos son complementarios y encontrando una solución a uno de ellos se determinará la solución a ambos.

Por lo que la función de núcleo queda definida como $k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j$ (una función de núcleo lineal). Para los experimentos desarrollados en esta tesis doctoral la función de núcleo ha sido sustituida por las siguientes funciones:

- **Kernel polinomial.** [AW99] el producto escalar es sustituido por la siguiente ecuación:

$$k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i, \mathbf{x}_j)^p \quad (4.13)$$

- **Kernel polinomial normalizado** [AW99] el producto escalar es sustituido con:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \frac{(\mathbf{x}_i, \mathbf{x}_j)^p}{\sqrt{(\mathbf{x}_i, \mathbf{x}_j)^p}} \quad (4.14)$$

- **Kernel Radial Basis Function RBF.** [AW99] la ecuación utilizada para sustituir el producto escalar es:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \cdot \|\mathbf{x}_i, \mathbf{x}_j\|^2) \text{ para } \gamma > 0 \quad (4.15)$$

- **Kernel Pearson VII.** [ÜMB06] se realiza una sustitución del producto escalar por la siguiente ecuación:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{\left[1 + \left(\frac{\sqrt{\|\mathbf{x}_i - \mathbf{x}_j\|^2} \cdot \sqrt{2^{\frac{1}{\omega}} - 1}}{\delta}\right)^2\right]^\omega} \quad (4.16)$$

Desde su creación, los SVM se utilizan ampliamente y han conseguido lograr muy buenos resultados por ejemplo, en la categorización de textos [PZH08].

4.1.5 Árboles de decisión

Estos clasificadores [Qui86] constituyen una herramienta de soporte a la toma de decisiones que es representada como un árbol (ver la Figura 4.5). También son utilizados para modelar las decisiones y las posibles consecuencias que pueden surgir una vez hayan sido tomadas. En este sentido, los árboles de decisión se componen de nodos de diferentes tipos:

4. Predicción del estado estacionario $t + 1$

- **Nodos de decisión:** Son los nodos en los que hay que tomar una decisión basada en unos valores que se han obtenido. Es decir, evaluar condiciones para poder continuar a delante.
- **Nodos finales:** Los nodos finales son los que marcan el resultado. En el caso de la clasificación indicarán el grupo al que pertenece la evidencia que está siendo clasificada.

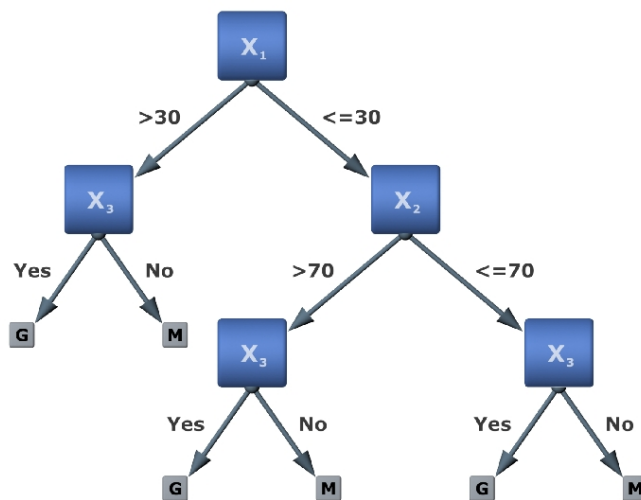


Figura 4.5: Representación gráfica de un árbol de decisión.

Los árboles pueden ser aprendidos mediante la división del conjunto de datos de prueba en diferentes subconjuntos basados en la prueba de valor de un atributo. El proceso es repetido en cada subconjunto derivado de una forma recursiva, conocido con el nombre de particionamiento recursivo (del inglés “*recursive partitioning*”). Este proceso llegará a su fin cuando el subconjunto de un nodo sea el mismo que la variable objetivo, o cuando la división no aporte valor a las predicciones. Normalmente, los algoritmos utilizados trabajan de una forma *top-down* eligiendo la variable que se va a usar en la división del conjunto de datos. El concepto de “mejor” es definido por cómo de buena es la división de la variable para obtener una serie de subconjuntos homogéneos que representen lo mismo con respecto a la variable objetivo. Dependiendo de los algoritmos utilizados, el cálculo será realizado según diferentes fórmulas.

Para la minería de datos pueden ser descritos como la combinación de técnicas matemáticas y computacionales para ayudar a la descripción, categorización y generalización del conjunto de datos dado. De esta forma, los datos

llegan como registros en la forma $(\mathbf{x}, Y) = (x_1, x_2, x_3, \dots, x_k, Y)$. La variable dependiente, Y , es la variable objetivo que se está intentando comprender, clasificar o generalizar. El vector \mathbf{x} está compuesto por las variables de entrada x_k utilizadas en la tarea con la que estamos trabajando.

Los árboles de decisión tienen una serie de ventajas y de inconvenientes [Qui86]. En primer lugar, entre sus ventajas se cuentan (i) la simplicidad de entendimiento y de interpretación, (ii) los datos no requieren de grandes preparaciones, (iii) son capaces de gestionar tanto datos numéricos como nominales, (iv) es posible validar el modelo, con lo que se convierte en mucho más confiable y robusto y (v) el tiempo necesario para el análisis de grandes cantidades de datos es muy bajo. Por el contrario, entre sus limitaciones se encuentran aspectos tan reseñables como: (i) el problema del aprendizaje de un árbol de decisión es NP-complejo, con lo que los métodos utilizados se basan en heurísticas, obteniendo unos resultados no totalmente óptimos; (ii) los métodos de aprendizaje pueden crear árboles demasiado complejos que no se acaban ajustando perfectamente a los datos y (iii) existen formas que son difíciles de representar con los árboles de decisión, como podrían ser los condicionamientos basados en la operación XOR.

De entre todos los algoritmos de aprendizaje supervisado existentes, para la realización de los experimentos de esta tesis doctoral, hemos utilizado el *J48* (la implementación del *Weka* [Gar95] del algoritmo *C4.5* desarrollado por Ross Quinlan [Qui93]). Igualmente, hemos empleado los bosques aleatorios (traducción del inglés “*random forests*”), que es una combinación de varios árboles generados al azar con el fin de buscar la creación de un clasificador mucho más fuerte [Bre01].

4.1.5.1 El algoritmo C4.5

Utilizando el algoritmo *C4.5* [Qui93] se realiza una construcción de árboles de decisión de la misma forma que con el método *ID3*¹ (de la voz inglesa “*Iterative Dichotomiser 3*”) [Qui86], en el que para realizar la división de los nodos se utiliza el concepto de la entropía de la información representada en la ecuación 4.17:

¹El algoritmo de aprendizaje *ID3*, desarrollado por Ross Quinlan, es el precursor del algoritmo *C4.5*. De forma resumida, el algoritmo *ID3* toma los atributos no utilizados y calcula los valores de la entropía concernientes a las muestras de prueba. Posteriormente, elige el atributo para el que la entropía es mínima y genera un nodo conteniendo ese atributo.

4. Predicción del estado estacionario $t + 1$

$$E(S) = - \sum_{j=1}^n f_S(j) \log_2 f_S(j) \quad (4.17)$$

donde $E(S)$ es la entropía del conjunto de datos S , n se trata del número de diferentes valores que tiene un atributo del conjunto S (la entropía se calcula para un atributo dado), $f_S(j)$ es la frecuencia del valor j dentro del conjunto de datos S . Así, en el cálculo de la entropía, un valor de 0 identifica un conjunto clasificado perfectamente.

Partiendo de esta premisa de cálculo, en los datos de entrenamiento, el conjunto $S = \{s_1, s_2, \dots\}$ de instancias ya clasificadas, cada muestra $s_i = \{x_1, x_2, \dots\}$ es un vector en el que cada x_j representa los atributos o las características de la muestra. Estos datos pasarán a ser aumentados a través de un vector $C = \{c_1, c_2, \dots\}$ donde cada c_i representa la clase a la que pertenece cada muestra.

Entonces, el algoritmo C4.5, en cada nodo selecciona el atributo que divide de forma más efectiva ese conjunto de muestras en los consecuentes subconjuntos de una clase o de la otra. Su criterio es la normalización de la ganancia de información resultante de elegir un atributo en ese proceso de división. El atributo con el mayor valor será el elegido para realizar la decisión.

El algoritmo C4.5 tiene una serie de casos base que utiliza para realizar su proceso de aprendizaje [Qui93]. El primero de ellos es “*todas las muestras del conjunto de datos de entrenamiento S pertenecen a la misma clase*”. Cuando se da esta situación, el algoritmo simplemente crea un nodo hoja indicando que ese es el resultado de la clasificación. El segundo caso base es “*ninguna de las características proporciona valor en el cálculo de la ganancia de la información*”. Para esta situación, el algoritmo crea un nodo de decisión en la parte alta del árbol utilizando el valor esperado de la clase. Y finalmente, el tercero de ellos es “*encontrase una instancia perteneciente a una clase que desconocía*”. Al igual que en el caso anterior, el algoritmo C4.5 crea un nodo de decisión en la parte alta del árbol utilizando el valor esperado de la clase.

4.1.5.2 El método de bosques aleatorios

Los bosques aleatorios, o en inglés “*random forests*”, es un clasificador que consiste en un conjunto de varios árboles de decisión pero que realiza la salida de la clasificación como si se tratase de un único árbol de decisión. El algoritmo para inducir el bosque aleatorio fue presentado por Leo Breiman en [Bre01].

4.2 Combinación de métodos de aprendizaje automático

Varias aproximaciones han trabajado en la creación de un conjunto de clasificadores (un *ensemble*¹) con árboles de decisión [Bre96, AG97, Ho98, Die00b]. Pero todas esas aproximaciones tienen un elemento común tras ellas, y es que para el k -ésimo árbol se genera un vector aleatorio Θ_k , independiente de los vectores que ya fueron creados anteriormente, $\Theta_1, \dots, \Theta_{k-1}$, pero con la misma distribución. Este árbol se genera utilizando el conjunto de datos de entrenamiento y Θ_k , construyendo un clasificador $h(\mathbf{x}, \Theta_k)$ donde \mathbf{x} es el vector de entrada. Entonces, una vez generados diversos árboles, ellos realizan una votación para determinar cuál es la clase más popular.

Más formalmente, los bosques aleatorios son un clasificador consistente en una colección de árboles de decisión de la forma $\{h(\mathbf{x}, \theta_k), k = 1, \dots, n\}$ donde $\{\theta_k\}$ son vectores aleatorios independientes distribuidos de forma idéntica. En ellos cada árbol indica la salida de la clase resultante y a través de un proceso de voto se selecciona la clase \mathbf{x} más popular [Bre01].

Este tipo de clasificadores tienen una serie de ventajas como [Bre01]: (i) es uno de los algoritmos de aprendizaje más precisos [CKY08], (ii) tiene una ejecución eficiente para grandes conjuntos de datos y (iii) dispone de un método muy efectivo para la generación de los clasificadores cuando los datos tienen una gran parte de datos sin informar, entre otros.

Por otra parte, algunas de las desventajas de este método de clasificación son las siguientes: (i) se ha observado que en algunas ocasiones, y dependiendo del conjunto de datos, el clasificador puede sobreajustarse [Seg04] y, a diferencia de los árboles de decisión, (ii) este tipo de clasificación no es tan fácil de seguir o de interpretar [BBHK10].

4.2 Combinación de métodos de aprendizaje automático

Los clasificadores unitarios son capaces de obtener buenos resultados por sí mismos, pero como se comentó anteriormente, no podemos asegurar que un clasificador en concreto sea el idóneo para la predicción del estado estacionario de la planta. Además, en la actualidad, todos los modelos de los sistemas MPC son lineales y no lineales (para más información diríjase al capítulo 2). Frente a esta problemática, diversos estudios han sido desarrollados para permitir una combinación de clasificadores [Kun04]. Estas técnicas nacen con el

¹Es la combinación de varios clasificadores débiles en uno más fuerte.

4. Predicción del estado estacionario $t + 1$

objetivo de alcanzar mejores decisiones en los problemas de clasificación a pesar de incorporar un mayor grado de complejidad al proceso.

Desde un punto de vista estadístico [Die00a], suponiendo que el conjunto de datos etiquetados \mathbf{Z} y un número n de diferentes clasificadores con un relativo buen comportamiento en las clasificaciones para \mathbf{Z} , podemos seleccionar uno de ellos para solucionar los problemas de clasificación, pero corremos el riesgo de escoger una mala opción para resolver este problema. Si aprendemos múltiples clasificadores para el conjunto de datos \mathbf{Z} que tienen bajas tasas de errores en sus procesos de clasificación, a pesar de ser indistinguibles por su tasa de error, cada uno de los clasificadores puede tener representaciones diferentes de la generalización del conjunto de datos \mathbf{Z} . Así que, en lugar de optar por un único clasificador, la opción más segura es la de utilizar todos ellos y realizar una “media” de sus salidas. Por consiguiente, el clasificador resultante no tienen por qué ser mejor que el clasificador que mejor se ajustaba al proceso de clasificación, pero disminuirá o eliminará el riesgo de la utilización de un clasificador no adecuado.

Desde un punto de vista computacional [Kun04], algunos de los algoritmos de aprendizaje supervisado utilizan métodos que detectan soluciones basadas en máximos locales. Por lo que, el problema de clasificación puede representarse como una trayectoria que se inicia en algún lugar del espacio de clasificación y acaba convergiendo en una zona cercana a la del clasificador óptimo. De esta forma, una agregación de clasificadores se aproxima mucho más al clasificador óptimo que lo que lo hace uno de ellos.

Desde un punto de vista figurativo [Kun04], es posible que en el espacio de clasificación no se encuentre el clasificador óptimo. Por ejemplo, la solución óptima para la generalización del conjunto de datos \mathbf{Z} es un clasificador no lineal. Sin embargo, las soluciones adoptadas son lineales. Ninguno de los clasificadores lineales se aproximarán al espacio en el que se encuentra el clasificador óptimo (no hay posibilidad de llegar a conseguirlo). Pero por otra parte, la utilización de muchos de esos clasificadores pueden llegar a aproximar la frontera de las predicciones a la del clasificador no lineal. Además, los clasificadores no lineales siendo mucho más complejos de configurar podrían ser simulados a través de un conjunto de clasificadores mucho más simples.

Igualmente, en el caso que nos ocupa, podemos añadir que debido a la naturaleza de los procesos de producción estos pueden ser divididos en lineales o no lineales. A través de la utilización de estos métodos de combinación somos capaces de generar un sistema de inteligencia colectiva para la clasificación que incorpora tanto clasificadores lineales como no lineales. De esta forma,

4.2 Combinación de métodos de aprendizaje automático

el resultado de la clasificación del estado estacionario $t + 1$ en la planta de producción se realizará a través de un sistema de clasificación híbrido.

Existen diversos métodos que permiten la combinación de clasificadores. Muchos de ellos quedan recogidos en [Kun04]. A pesar de que las soluciones pueden ser óptimas para la gestión del proceso productivo, algunos únicamente permiten la unión de clasificadores del mismo tipo. Para la investigación de esta tesis doctoral, ese tipo de combinaciones han sido desechadas debido a que su utilización elimina la ventaja de la generación de un sistema de clasificación híbrido. Por consiguiente, de los métodos que permiten la unión de clasificadores heterogéneos han sido utilizados: (i) la combinación por voto [Kun04], (ii) *grading* [SF01], (iii) *stacking* [Wol92] y (iv) multiesquema (una implementación de Weka [Gar95] que permite la selección de los resultados de un clasificador de entre todos los del conjunto y que basa su selección en el porcentaje de acierto y las tasas de error). A continuación pasaremos a revisar cada uno de los métodos empleados en esta tesis doctoral.

4.2.1 Combinación por voto

La democracia en la clasificación de elementos es una de las estrategias más antiguas para la toma de decisiones. En [Day87] se realiza una revisión de los conceptos de consenso en un contexto de teoría electoral. Las raíces de este sistema son trazadas desde la antigua Grecia y en el senado de Roma. El criterio de la votación por mayoría viene a establecerse en 1356 en la elección de los reyes alemanes, para 1450 fue adoptado en la Cámara de los Comunes Británica. Sin embargo, extendiendo la teoría electoral, existen otros métodos que permiten realizar esa combinación de clasificadores. Estos pasarán a ser comentados a continuación.

4.2.1.1 La mayoría decide

La aproximación más simple en concepto es la de la *mayoría decide*. Sin embargo, para esta aproximación existen tres patrones de funcionamiento diferentes [Kun04]: (i) unanimidad, (ii) la mayoría simple y (iii) pluralidad. A pesar de contar, con las tres formas de combinación, para los experimentos de esta tesis doctoral se ha optado por el método de pluralidad. Entonces, asumiendo que las salidas etiquetadas de los clasificadores son dadas como vectores binarios c -dimensionales $[d_{i,1}, \dots, d_{i,c}]^T \in \{0, 1\}^c$, $i = 1, \dots, L$ donde $d_{i,j} = 1$ si el clasificador D_i etiqueta \mathbf{x} en ω_j , ó 0 en caso contrario. La pluralidad de voto resulta

4. Predicción del estado estacionario $t + 1$

en un conjunto de clasificación para la clase ω_k tal que

$$\sum_{i=1}^L d_{i,k} = \max_{j=1}^c \sum_{i=1}^L d_{i,j} \quad (4.18)$$

Frente al problema de los empates, estos son resueltos arbitrariamente. Esta regla normalmente es llamada en la literatura como *la mayoría decide*. De hecho, también coincide con la regla que anteriormente hemos nombrado como la mayoría simple (el 50 % de los votos más uno) en el caso de trabajar sobre 2 clases. En [KBD01] se sugiere una pluralidad de voto basada en un umbral. Por lo que, se aumenta el conjunto de etiquetas de clase Ω con una nueva, ω_{c+1} , para todos los objetos en los que la combinación falló al determinar la clase con suficiente confianza o se produjeron empates. Así, la decisión es

$$\begin{cases} \omega_k, & \text{Si } \sum_{i=1}^L d_{i,k} \geq \alpha \cdot L \\ \omega_{c+1}, & \text{en caso contrario} \end{cases} \quad (4.19)$$

donde $0 < \alpha \leq 1$. Para la mayoría simple, podemos simplemente asignar α con el valor $\frac{1}{2} + \varepsilon$, donde $0 < \varepsilon < 1/L$. En el caso de que $\alpha = 1$, nos encontramos tratando con una combinación de clasificadores basada en la *unanimidad* en el voto. En esta última, la decisión es tomada para una clase en concreto si todos y cada uno de los clasificadores han llegado a la misma conclusión. En caso contrario, el conjunto se niega a decidir y le asigna la etiqueta ω_{c+1} a \mathbf{x} .

4.2.1.2 Producto de probabilidades

En este segundo método de unión de clasificadores por voto se tienen en cuenta las probabilidades [KHDM98]. Así, para la *regla del producto* tenemos que $p(x_1, \dots, x_R | \omega_k)$ representa la distribución de probabilidad conjunta de las mediciones extraídas de los clasificadores. Asumimos que esas representaciones son estadísticamente y condicionalmente independientes. Dada la suposición anterior, obtenemos la siguiente deducción.

$$p(x_1, \dots, x_R | \omega_k) = \prod_{i=1}^R p(x_i | \omega_k) \quad (4.20)$$

donde $p(x_i | \omega_k)$ es la medida del modelo del proceso para la i -ésima representación. Aplicando una serie de sustituciones en la ecuación 4.20 basadas en la *regla de decisión* bayesiana [KHDM98] podemos obtener la ecuación 4.21

4.2 Combinación de métodos de aprendizaje automático

$$P(\omega_k|x_1, \dots, x_R) = \frac{P(\omega_k) \prod_{i=1}^R p(x_i|\omega_k)}{\sum_j^m P(\omega_j) \prod_{i=1}^R p(x_i|\omega_j)} \quad (4.21)$$

y mezclando de nuevo la ecuación 4.21 con *regla de decisión bayesiana* [KHDM98] obtenemos la siguiente regla de decisión, es decir, realizaremos la asignación $Z \rightarrow \omega_j$ si

$$P(\omega_j) \prod_{i=1}^R P(x_i|\omega_j) = \max_{k=1}^m P(\omega_k) \prod_{i=1}^R P(x_i|\omega_k) \quad (4.22)$$

o en términos de probabilidades *a posteriori* producidas por los respectivos clasificadores asignaremos $Z \rightarrow \omega_j$ si

$$P^{-(R-1)}(\omega_j) \prod_{i=1}^R P(\omega_j|x_i) = \max_{k=1}^m P^{-(R-1)}(\omega_k) \prod_{i=1}^R P(\omega_k|x_i) \quad (4.23)$$

La regla de decisión 4.23 cuantifica la probabilidad de una hipótesis mediante la combinación *a posteriori* de las probabilidades generadas por los clasificadores a través de la *regla del producto*. Efectivamente, es una regla de fusión muy severa. La razón es que este tipo de fusión puede inhibir una de las salidas cuando la probabilidad es muy próxima a 0.

4.2.1.3 Media de probabilidades

La *regla de la media* parte de la generación de la *regla de la suma* para, posteriormente, realizar una división entre la cantidad de participantes en la generación del resultado final R [KHDM98]. En más detalle, considerando la regla de decisión 4.23, en algunas aplicaciones podría ser apropiado considerar que las probabilidades *a posteriori* computadas por cada uno de los clasificadores no serán derivadas de las probabilidades *a priori*. Esta es una suposición algo fuerte, pero que puede ser realmente satisfecha cuando la información discriminativa disponible es altamente ambigua debido a altos niveles de ruido. En ese caso, podemos suponer que las probabilidades *a posteriori* pueden ser representadas como

$$P(\omega_k|x_i) = P(\omega_k)(1 + \delta_{ki}) \quad (4.24)$$

4. Predicción del estado estacionario $t + 1$

donde δ_{ki} satisface $\delta_{ki} \ll 1$. Entonces, sustituyendo el valor de $P(\omega_k|x_i)$ 4.24 en la ecuación 4.23 por las probabilidades *a posteriori*, encontramos

$$P^{-(R-1)}(\omega_k) \prod_{i=1}^R P(\omega_k|x_i) = P(\omega_k) \prod_{i=1}^R (1 + \delta_{ki}) \quad (4.25)$$

Si expandimos el producto y descuidamos cualquier término de segundo orden o mayor, podemos aproximar la parte derecha de 4.25 como

$$P(\omega_k) \prod_{i=1}^R (1 + \delta_{ki}) = P(\omega_k) + P(\omega_k) \sum_{i=1}^R \delta_{ki} \quad (4.26)$$

Por lo que, sustituyendo los valores obtenidos en las ecuaciones 4.26 y 4.24 dentro de la ecuación 4.23 obtenemos la regla de la decisión de la suma (para la media debería realizarse una división entre R) en la que se hará la asignación $Z \rightarrow \omega_j$ si

$$(1 - R)P(\omega_j) + \sum_{i=1}^R P(\omega_j|x_i) = \max_{k=1}^m \left[(1 - R)P(\omega_k) + \sum_{i=1}^R P(\omega_k|x_i) \right] \quad (4.27)$$

4.2.1.4 Probabilidad máxima

Para la *regla de la probabilidad máxima* [KHDM98], comenzamos con la *regla de la suma* representada en 4.27 y aproximamos la suma mediante el máximo de las probabilidades *a posteriori*, así, obtenemos que asignaremos $Z \rightarrow \omega_j$ si

$$(1 - R)P(\omega_j) + R \max_{i=1}^R P(\omega_j|x_i) = \max_{k=1}^m \left[(1 - R)P(\omega_k) + R \max_{i=1}^R P(\omega_k|x_i) \right] \quad (4.28)$$

que bajo la consideración de igualdades en las precedencias se simplifica en una asignación de $Z \rightarrow \omega_j$ si

$$\max_{i=1}^R P(\omega_k|x_i) = \max_{k=1}^m \max_{i=1}^R P(\omega_k|x_i) \quad (4.29)$$

4.2 Combinación de métodos de aprendizaje automático

4.2.1.5 Probabilidad mínima

Para la *regla de la probabilidad mínima* [KHDM98], comenzamos con la *regla del producto*, definida en la ecuación 4.23, en la que dejando de lado el producto de las probabilidades *a posteriori* se obtiene que asignaremos $Z \rightarrow \omega_j$ si

$$P^{-(R-1)}(\omega_j) \min_{i=1}^R P(\omega_j|x_i) = \max_{k=1}^m P^{-(R-1)}(\omega_k) \min_{i=1}^R P(\omega_k|x_i) \quad (4.30)$$

lo que bajo el supuesto de igualdades en las precedencias simplificamos a la asignación de $Z \rightarrow \omega_j$ si

$$\text{med}_{i=1}^R P(\omega_j|x_i) = \max_{k=1}^m \text{med}_{i=1}^R P(\omega_k|x_i) \quad (4.31)$$

4.2.2 Grading

En un problema de aprendizaje con n_e muestras para el entrenamiento, cada una compuesta por n_a atributos x_{ij} y una etiqueta de clase cl_i , asumimos que tenemos un conjunto de n_c clasificadores base C_k . Los clasificadores base forman el conjunto de clasificadores que pretendemos combinar a través de la técnica de *grading* [SF01], los cuales han sido evaluados utilizando algún tipo de los esquemas de validación cruzada (en inglés “*k-fold cross validation*”) [Bis95]. Con la aplicación de esta técnica aseguramos que todas y cada una de las muestras han sido utilizadas para su aprendizaje. Entonces, el método de combinación de clasificadores conocido como *grading* hace uso de esas predicciones con el fin de entrenar un conjunto de meta-clasificadores que aprenderán cómo realizar las predicciones cuando se crea que el clasificador base obtendrá un resultado correcto en su pronóstico.

El conjunto de datos para el entrenamiento de cada uno de esos meta-clasificadores se construye con las predicciones de los correspondientes clasificadores base C_k al que se le añade una nueva clase. Es decir, los meta-clasificadores utilizan n_c conjuntos de entrenamiento con los datos originales y un nuevo atributo que califica los resultados de los clasificadores base. Los valores que puede adoptar este nuevo atributo son correcto (+) o incorrecto (-). Posteriormente, se realiza un entrenamiento de n_c clasificadores de primer nivel, los meta-clasificadores. Todos y cada uno de ellos aprenderán el comportamiento de los clasificadores base utilizando uno de estos conjuntos de datos. Esta idea se basa en que los diferentes clasificadores base cometen

4. Predicción del estado estacionario $t + 1$

diferentes errores, y cada uno de los clasificadores de este primer nivel intenta determinar cómo se comportarán los clasificadores base. Nótese que las muestras negativas en estos conjuntos de datos son principalmente las tasas de error correspondientes al clasificador base y que han sido estimadas a través del procedimiento de validación cruzada. Así, utilizando la técnica de *grading* se va a tomar la decisión para cada muestra de forma separada, centrándose únicamente en aquellos clasificadores base cuya predicción obtuvo una calificación positiva. Con lo que el método de *grading* puede ser visto como un método de generalización de la selección por validación cruzada.

El funcionamiento en tiempo de clasificación es como sigue [SF01]:

1. Cada clasificador base realiza una predicción para la instancia que quiere clasificarse.
2. Los clasificadores de primer nivel, los meta-clasificadores califican el resultado obtenido por los clasificadores base para la instancia que se está tratando de clasificar.
3. Se deriva la clasificación utilizando únicamente los resultados calificados como positivos.

Los conflictos (i.e., múltiples clasificadores con diferentes predicciones base han sido predichos como correctos) pueden ser resueltos mediante la utilización del método de votación (ver sección 4.2.1) o mediante la utilización de la confianza estimada en el clasificador.

Para el segundo de los métodos de resolución de conflictos, la confianza perteneciente a los clasificadores de primer nivel será sumada por clase y después normalizada para producir una distribución probabilística apropiada. En el caso de que no se califique ninguno de los clasificadores base de forma positiva, todos ellos serán utilizados con un nuevo valor de confianza fruto del cálculo de $(1 - \text{confianza})$, prefiriendo la utilización de estos clasificadores base en lugar de los de primer nivel ya que los meta-clasificadores son más inciertos. La clase más probable de la distribución final de clases será elegida como la predicción final. En este caso los empates son eliminados al elegir la clase que más frecuentemente ocurre de entre las clases más probables.

Más formalmente, tomamos p_{ikl} como la probabilidad de la clase calculada para cada clasificador base k para la clase l y la muestra i . Para simplificar las ecuaciones, escribimos P_{ik} para hacer referencia al vector $(p_{ik1}, p_{ik2}, \dots, p_{ikn_l})$ de todas las probabilidades de las clases para la muestra i y el clasificador k .

4.2 Combinación de métodos de aprendizaje automático

Asimismo, la predicción del clasificador base k dada la muestra i es la clase l con la máxima probabilidad p_{ikl} , es decir, $c_{ik} = \operatorname{argmax}_l \{p_{ikl}\}$.

Posteriormente, el método *grading* construye n_c conjuntos de datos de entrenamiento, uno para cada clasificador base k , añadiendo las predicciones g_{ik} como la nueva clase al conjunto de datos original (g_{ik} es 1 si la predicción del clasificador base k para la muestra i es correcta (predicción positiva) y 0 en los casos contrarios). $prMeta_{ik}$ es la probabilidad de que el clasificador de base k vaya a predecir bien la muestra i y que fue estimada por el meta-clasificador k . Con esta información, la probabilidad final estimada para la clase l y la muestra i , en el caso de que al menos un meta-clasificador indique que su clasificador va a tener un resultado positivo (i.e., $prMeta_{ik} > 0,5$), utilizamos la siguiente fórmula:

$$prGrading_{il} = \sum \{prMeta_{ik} | c_{ik} = l \wedge prMeta_{ik} > 0,5\} \quad (4.32)$$

4.2.3 Stacking

El método *stacking* [Wol92], al igual que el anteriormente explicado (ver la sección 4.2.2), se trata de un método para la combinación de clasificadores que intenta obtener mejores resultados que los basados en la idea de combinar los resultados obtenidos mediante el método de la validación cruzada.

Así, para el método de *stacking* serán utilizados una serie de clasificadores o generalizadores que son definidos como elementos que realizan un mapeo de un conjunto de m pares de aprendizaje $\{x_k \in \mathbf{R}^n, y_k \in \mathbf{R}, 1 \leq k \leq m$ con una pregunta $\in \mathbf{R}_n\}$ y una conjetura $z \in \mathbf{R}$. Aunque ampliamente queda generalizada la conjetura $z \in \mathbf{R}^m$ [Wol89]. Y para el primero de sus pasos, se seleccionarán un conjunto de r particiones, cada una de ellas divide θ (conjunto de entrenamiento) en dos conjuntos, normalmente disjuntos. Etiquetamos al conjunto de particiones como θ_{ij} , donde $1 \leq i \leq r$ y $j \in \{1, 2\}$. Estas particiones es lo que se conoce como *conjunto de particiones*.

Entonces, definimos el espacio inhabitado \mathbf{R}^{n+1} del original conjunto de datos θ como el espacio de nivel 0. Cualquier clasificador que realiza una generalización en este nivel es denominado *generalizador de nivel 0*, y el conjunto de datos originales es conocido como *conjunto de aprendizaje de nivel 0*. Para cada una de las particiones r de θ , $\{\theta_{i1}, \theta_{i2}\}$, vemos un grupo de k números determinados por un subconjunto de N $\{G_j\}$. Típicamente, estos k números pueden ser elementos como (i) las suposiciones realizadas por el clasificador o

4. Predicción del estado estacionario $t + 1$

generalizador $\{G_j\}$ cuando fue enseñado con θ_{i1} y presentada una pregunta de clasificación a través del elemento de entrada de θ_{i2} , (ii) la componente de entrada de θ_{i2} o (iii) el vector en el espacio de entrada conectando la componente θ_{i2} a su θ_{i1} vecino más próximo. Posteriormente, se toma cada agrupación de k números como la componente de entrada en un espacio \mathbf{R}^{k+1} . El correspondiente valor de salida de cada punto es calculado de la componente de salida correspondiente a θ_{i2} como función del resultado obtenido en el generalizador G_j . El espacio \mathbf{R}^{k+1} es llamado *espacio de nivel 1*. Debido a que disponemos de r particiones de θ , tenemos r puntos en el espacio de nivel 1. Esos puntos son conocidos como el conjunto de aprendizaje *reducido* o de *nivel 1*.

A la hora de realizar una clasificación mediante el cómputo de un clasificador en el nivel 1, la técnica de *stacking* nos deja hacerlo de múltiples maneras. La forma común, y la utilizada para los experimentos de esta tesis doctoral, es tomar la pregunta en el nivel 0, pasar las transformaciones que producen las componentes de entrada para adaptarlo al conjunto de datos aprendido en el nivel 1, y así, obtener la pregunta de clasificación adaptada al espacio de entrada del nivel 1. Finalmente, se realiza la respuesta con la generalización obtenida en el nivel 1. Una vez ha sido obtenida la solución de este nivel 1, se transforma la respuesta de nuevo al espacio de nivel 0. Este proceso es lo que se conoce con el nombre de “*stacked generalization*”. Este proceso puede complicarse determinando que el número de niveles sea mayor que 1 (i.e., utilización de apilamientos múltiples).

Debe tenerse en cuenta que aspectos relacionados con esta técnica se encuentran rodeados de un halo de misticismo. Por ejemplo, actualmente no existen reglas que definan cuáles son los clasificadores que deben ser utilizados en el nivel 0, cuáles en el nivel 1 y qué k números deben ser empleados para generar el espacio de entrada del nivel 1 [Wol92].

4.2.4 Multiesquema

Este método es una implementación de Weka [Gar95] que permite realizar la combinación de los clasificadores de la forma más simple posible. Así, frente a los últimos métodos que han sido comentados (*grading* y *satching*), este método hace uso de una regla de combinación basada únicamente en la validación cruzada de los clasificadores y su tasa de error medida como el error cuadrado medio.

Partiendo de la misma definición del espacio de clasificación y los generalizadores que se hace en la sección 4.2.3, así como la del conjunto de peticio-

nes realizadas; para esas particiones hechas en base a la validación cruzada (CPVC), $r = m$, para todos los i , θ_{i2} consiste en un único elemento de θ ; el correspondiente θ_{i1} se compone del resto de θ , y $\theta_{i2} \neq \theta_{j2}$ cuando $i \neq j$.

La técnica de “el ganador se lleva todo” adoptada por la validación cruzada es una forma sencilla de realizar el CPVC para mapear un generalizador G con un conjunto de datos de aprendizaje θ y estimar la tasa de error de G cuando se generaliza θ . Instintivamente, se estima la precisión de la generalización mirando cómo de bien se ajusta el clasificador o generalizador a una parte del conjunto de datos cuando se le enseña con el resto de él. De forma más rigurosa, se realiza calculando la media, sobre todo i , del error de G al predecir que salida corresponde a la entrada θ_{i2} cuando se realiza el aprendizaje con el resto de θ , θ_{i1} . De esta forma, la estimación del error realizada a través de la validación cruzada es la siguiente (nótese que los valores utilizados para los cálculos son los valores obtenidos en el error medio cuadrado para cada clasificador G):

$$CV(G, \theta) \equiv \frac{\sum_i [G(\theta_{i1}, \text{entrada de } \theta_{i2}) - (\text{salida de } \theta_{i2})]^2}{m} \quad (4.33)$$

De esta forma, a través de la utilización de esta métrica, el método multiesquema es capaz de determinar la solución de que clasificador debe tenerse en cuenta con el fin de realizar la generalización más precisa.

4.3 Validación empírica

A lo largo de esta sección se van a recoger la experimentación que hemos realizado para validar la predicción del estado estacionario en el instante de tiempo $t + 1$ mediante dos enfoques diferentes: (i) la predicción de múltiples características utilizando un único clasificador y (ii) la predicción de múltiples características utilizando la combinación de múltiples clasificadores.

4.3.1 Metodología general

Para realizar la validación de ambas aproximaciones, hemos recogido datos de una fundición especializada en el desarrollo de componentes seguros y con un alto nivel de precisión. Las piezas resultantes del proceso productivo son enviadas posteriormente a la industria del automóvil. Esta fundición tiene una producción de alrededor de 45.000 toneladas al año. Como ya se ha comen-

4. Predicción del estado estacionario $t + 1$

tado, estos experimentos se centran en la predicción del estado de la planta en un instante de tiempo localizado en un futuro próximo, comprobando la aparición de defectos de microrrechupes o el cumplimiento de ciertas características, en otras palabras, si las propiedades mecánicas (más concretamente la carga máxima) cumplen con los valores de seguridad necesarios. Para más información sobre el proceso desarrollado en una fundición o qué estados han sido utilizados en cada una de las clases de predicción diríjase al capítulo 3.

Utilizando este conjunto de datos, queremos responder las siguientes preguntas relacionadas con la investigación:

¿Cuál de los clasificadores seleccionados es el óptimo para la predicción de cada una de las características seleccionadas?

¿Diferentes características del proceso de fabricación pueden ser predichas utilizando el mismo clasificador con la misma configuración?

¿Es posible la generación de un sistema de clasificación híbrido para la predicción de características del proceso de producción?

¿Cuál de los sistemas de combinación de clasificadores es el óptimo para la predicción de cada una de las características seleccionadas?

Como se puede comprobar, las dos de las primeras preguntas hacen referencia a la primera de las aproximaciones, la predicción del estado estacionario $t + 1$ con un único clasificador, mientras que las dos últimas hacen referencia a la segunda de ellas, la predicción del estado utilizando técnicas de combinación de clasificadores. En concreto, con el fin de realizar la evaluación de cada una de las propuestas realizadas, hemos conducido la siguiente metodología de experimentación:

- **Adquisición de datos.** Proceso mediante el cual se va a extraer el conocimiento de la fundición y que serán los datos que representan el proceso productivo junto con la etiqueta del resultado que obtuvo ese proceso. Estos datos son la base fundamental para que, en las etapas posteriores, seamos capaces de desarrollar el proceso de aprendizaje automático supervisado.
- **Validación cruzada.** Se ha realizado una validación cruzada de k partes [Bis95] donde $k = 10$. De esta manera, nuestro conjunto de datos

para el aprendizaje se divide 10 veces en 10 diferentes partes. En cada posible división se seleccionan 9 de esas partes para la fase de aprendizaje (90 % del conjunto de datos total) y sólo una de las partes para la fase de pruebas (10 % del total de los datos). Al repetirse este proceso durante 10 veces, se acaba utilizando todas las partes tanto para la etapa de aprendizaje como para la etapa de pruebas.

- **Aprendizaje del modelo.** En esta fase se ha realizado el proceso de aprendizaje de los diferentes métodos de aprendizaje supervisado y de combinación de clasificadores mencionados anteriormente. A pesar de que la metodología es la misma para ambos experimentos, en este momento hemos hecho la división del aprendizaje del clasificador único o de la combinación de clasificadores.

– *Predicción con clasificador único.* Para el método de la clasificación con un único clasificador se han aplicado los siguientes modelos (para más información diríjase a las sección 4.1):

- * Redes bayesianas (BN del vocablo “*Bayesian Networks*”): para el aprendizaje de las redes bayesianas se han utilizado diferentes algoritmos de aprendizaje estructural (aprendizaje que permite descubrir las relaciones presentes entre las variables): *K2* [FGG97], *escalar colinas* [RN03] y el aprendizaje *bayesiano ingenuo aumentado mediante un árbol* (TAN, del inglés “*Tree Augmented Naïve*”) [FGG97]. Por otra parte, se han realizado experimentos con el clasificador *bayesiano ingenuo* [JL95], que aunque es un algoritmo muy sencillo, en el que se asumen que todas las variables son independientes, suele obtener unos resultados aceptables.
- * *K* vecinos más próximos (KNN, del inglés “*K-Nearest Neighbours*”): para el clasificador KNN [FH52] se han realizado experimentos con $K = 1$, $K = 2$, $K = 3$, $K = 4$ y $K = 5$. No se han continuado las pruebas con más valores de K porque el máximo local de precisión para cada uno de los experimentos posibles se encuentra dentro de este rango de valores.
- * Redes neuronales artificiales (ANN, de la voz inglesa “*Artificial Neural Networks*”): se ha utilizado una red neuronal compuesta de *múltiples capas de perceptrones* (MLP, *MultiLayer Perceptron*) [Bis95] con el algoritmo de *retropropagación*, o en inglés

4. Predicción del estado estacionario $t + 1$

backpropagation [MSTC94], para su aprendizaje. Se ha seleccionado únicamente este modelo debido a que es uno de los más utilizados en experimentos relacionados con la clasificación de defectos (ver sección 2.7).

- * Máquinas de soporte vectorial (SVM, *Support Vector Machines*): Para entrenar los SVM se han utilizado diferentes funciones de núcleo [Vap00]: *núcleo polinomial* [AW99], *núcleo polinomial normalizado* [AW99], *núcleo universal Peason VII* [ÜMB06] y *núcleo de función de base radial* (RBF, de *Radial Basis Function*) [AW99].
 - * Árboles de decisión (DT, de la voz inglesa “*Decisión Trees*”): se han llevado a cabo experimentos con *bosques aleatorios* [Bre01], una agrupación de árboles de decisión generados al azar que se construye utilizando diferentes cantidades de estos árboles para, de este modo, encontrar el máximo local, igual que sucede con el algoritmo KNN. Igualmente, se ha utilizado el algoritmo C4.5 [Qui93] para la generación de un árbol de decisión.
- *Predicción con combinación de clasificadores.* Teniendo en cuenta los resultados anteriores, para el método de clasificación mediante la utilización de procedimientos que permiten la combinación de resultados se han aplicado los siguientes modelos que permiten la unión de un conjunto de clasificadores heterogéneos (para más información diríjase a las sección 4.2). Nótese que los clasificadores utilizados para la generación del conjunto básico son los empleados en el experimento anterior:
- * *Combinación por voto:* siendo uno de los métodos más antiguos para la combinación de decisiones, este método es el primero a la hora de realizar las pruebas de combinación de clasificadores. Sin embargo, los métodos de combinación por voto pueden hacerse a través de diferentes reglas que permiten modificar el comportamiento de la clasificación. Más concretamente, hemos utilizado la regla de la decisión de la mayoría [Kun04]. Esta regla es la que más se acerca a un proceso de votación tal y como lo conocemos hoy en día, es decir, utiliza las nociones de la teoría electoral. También hemos probado las reglas del producto de probabilidades [KHDM98] y la media de las probabilidades [KHDM98]. Estas dos reglas plantan las bases necesarias para la deducción del resto de reglas utilizadas. Esas

últimas reglas son la probabilidad máxima [KHDM98] y probabilidad mínima [KHDM98].

- * **Grading:** el método se encarga de realizar una clasificación en un primer nivel para identificar si los clasificadores van a obtener una predicción correcta. En caso de ser positiva, esas son las que se tendrán en cuenta para derivar el resultado final [SF01]. Los clasificadores que han sido utilizados como clasificadores de primer nivel han sido un clasificador bayesiano ingenuo, una red bayesiana aprendida con TAN, un KNN donde $1 \leq K \leq 5$ y un árbol de decisión basado en el algoritmo C4.5.
- * **Stacking:** aunque similar al anterior, el método de *stacking* [Wol92] realiza una definición de una serie de clasificadores de nivel 0, los clasificadores que van a ser combinados y, posteriormente, genera una serie de transformaciones en los datos y unos segundos clasificadores que derivarán la salida según el conocimiento que han adquirido. Para finalizar, realizarán las transformaciones a la inversa para, así, obtener la clasificación final. Para generar esos clasificadores de apoyo hemos seleccionado un clasificador bayesiano ingenuo, una red bayesiana aprendida con TAN, un KNN con valores de K que van de 1 a 5 y un árbol de decisión creado con el algoritmo C4.5.
- * **Multiesquema:** se basa en realizar un ranking del funcionamiento de los clasificadores en base a los resultados obtenidos al realizar una validación cruzada y a su tasa de error. Por lo que para este caso de combinación no hemos tenido que realizar ningún tipo de configuración extra.

- **Pruebas del modelo.** La primera de las medidas sobre las que se ha trabajado es el nivel de precisión alcanzado por el clasificador. El segundo es la evaluación de las tasas de error. Se ha evaluado la tasa de error entre el conjunto de valores predichos X y el conjunto de valores reales Y (ambos con el tamaño del conjunto de datos de pruebas m) mediante el error medio absoluto (MAE, *Mean Absolute Error*) (se muestra en la ecuación 4.34).

$$MAE(X, Y) = \sum_{i=1}^m \frac{|X_i - Y_i|}{m} \quad (4.34)$$

Del mismo modo, se ha utilizado la medida de la raíz del error cuadrado

4. Predicción del estado estacionario $t + 1$

medio (*Root Mean Square Error*, RMSE). (Se muestra en la ecuación 4.35).

$$RMSE(X, Y) = \frac{1}{m} \cdot \sqrt{\sum_{i=1}^m (X_i - Y_i)^2} \quad (4.35)$$

4.3.2 Predicción mediante clasificador único

Una vez aplicada la metodología mencionada en la sección 4.3.1, hemos obtenido los resultados que se muestran a continuación. Para facilitar la legibilidad de los mismos, hemos dividido los resultados según el objetivo de clasificación seleccionado para cada predicción.

4.3.2.1 Microrrechupes

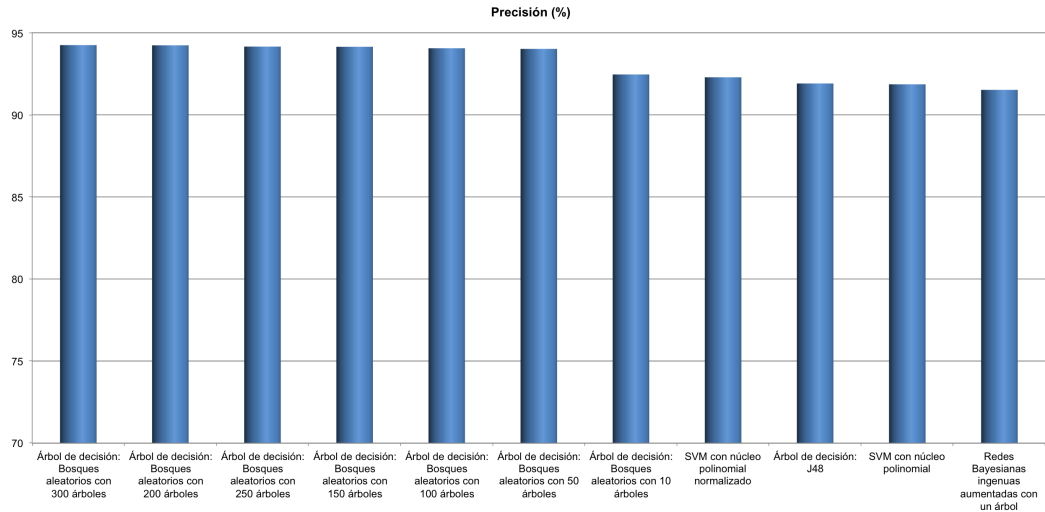
Anteriormente, ya se había experimentado la utilización de clasificadores para la predicción de microrrechupes [PBZ08]. Gracias a ese experimento se puede afirmar que este tipo de aproximaciones mejoran el proceso productivo de la fundición. Sin embargo, hasta donde abarca nuestro conocimiento, no existe ningún enfoque que sea capaz de predecir este problema utilizando un clasificador diferente a las redes bayesianas. Por ello, estos experimentos se desarrollan para comprobar el comportamiento de otros clasificadores con este defecto y determinar si las redes bayesianas eran los clasificadores más adecuados para la detección de este problema.

En relación a los experimentos, las Figuras 4.6a y 4.6b muestran los resultados obtenidos en términos de exactitud de la predicción y las Figuras 4.7a y 4.7b muestran las tasas de error (MAE y RMSE) de los clasificadores.

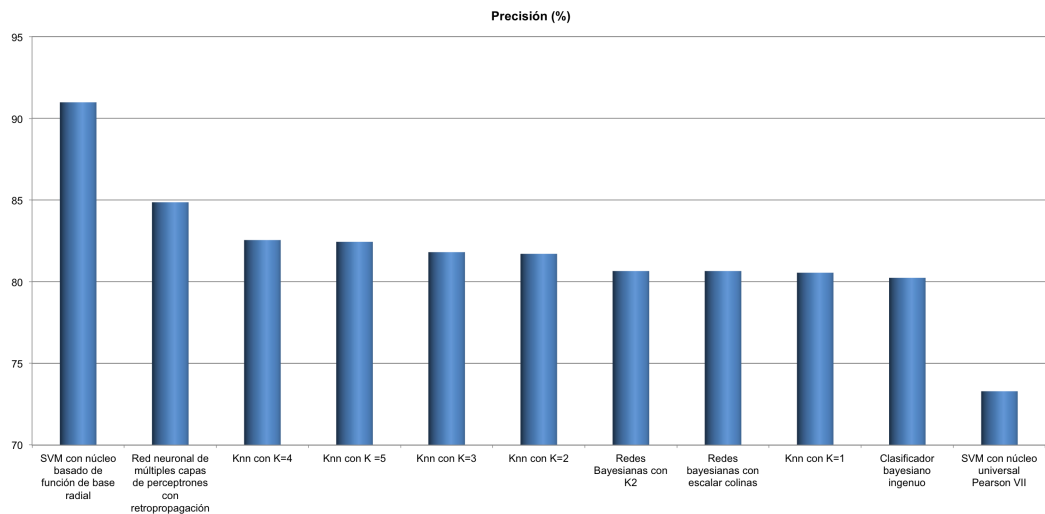
Como se puede observar, casi todos los algoritmos de clasificación obtuvieron buenos resultados, manteniéndose en un porcentaje de acierto que va desde el 73,29% hasta el 94,25%. El mejor de todos ellos ha sido uno de los clasificadores lineales, los árboles de decisión, y más concretamente, la variante de *bosques aleatorios*. Además, 12 de los 22 clasificadores superan el 90% de nivel de precisión, y 21 de los 22 superan un 80% de acierto.

De esta manera, casi todos los algoritmos obtuvieron buenos resultados, sin embargo, tanto los árboles de decisión (utilizando *bosques aleatorios* y el método C4.5) como las máquinas de soporte vectorial (utilizando el *núcleo*

4.3 Validación empírica



(a) La figura muestra cuáles son los clasificadores que mejores resultados han obtenido a la hora de realizar la predicción de los microrrechupes. Puede observarse que todos ellos han tenido un comportamiento bastante similar.



(b) La figura ilustra cuáles han sido los clasificadores que peores resultados han obtenido. Como era de esperar, uno de los peores ha sido el clasificador bayesiano ingenuo debido a la simplicidad del modelo.

Figura 4.6: Resultados obtenidos en la predicción del estado estacionario $t + 1$ con el fin de detectar de posibles problemas de microrrechupes. Se muestran los datos en términos de porcentaje de precisión. Los *bosques aleatorios creados a partir de 250 árboles de decisión* fueron el mejor clasificador unitario para la detección del problema, llegando a alcanzar un 94,25 % de precisión.

4. Predicción del estado estacionario $t + 1$

polinomial normalizado y núcleo polinomial) y las redes bayesianas aprendidas mediante el algoritmo TAN parecen los clasificadores más adecuados si tenemos únicamente en cuenta la precisión.

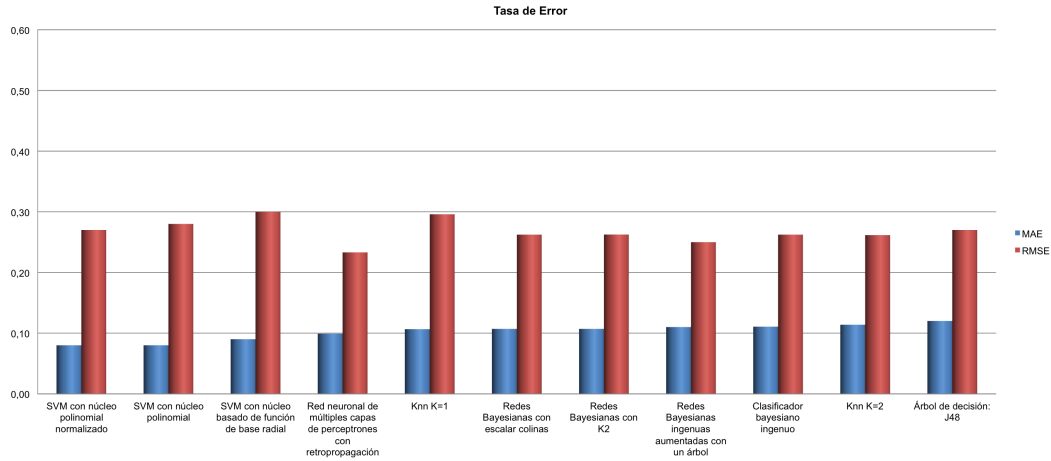
Por otra parte, el clasificador *bayesiano ingenuo* se comportó como uno de los peores clasificadores. Hay que tener en cuenta que este clasificador es un tipo de red bayesiana en la que todas las variables de entrada se suponen estadísticamente independientes. Por ello, se salta la dependencia causal que pueda surgir entre las variables, y hace que no pueda lograr tan buenos resultados como los otros clasificadores. Y sorprendentemente, el peor de los clasificadores es el que utiliza las máquinas de soporte vectorial con *núcleo universal Pearson VII*. Este resultado nos sorprende debido a que los otros SVM, utilizando diferentes núcleos, se ajustaron mucho mejor al proceso productivo.

También, el algoritmo de los K vecinos más próximos (KNN), un clasificador no lineal, logró mejores resultados de los que uno puede pensar de antemano. KNN no tiene una fase de entrenamiento o de aprendizaje como tal, únicamente dispone de una pequeña etapa de procesado de datos, sólo se centra en la semejanza entre las instancias. Se comporta relativamente bien, aunque los métodos con un proceso de aprendizaje mucho más robusto y que pueden extraer mucha más información de los datos acaban comportándose mucho mejor que este clasificador.

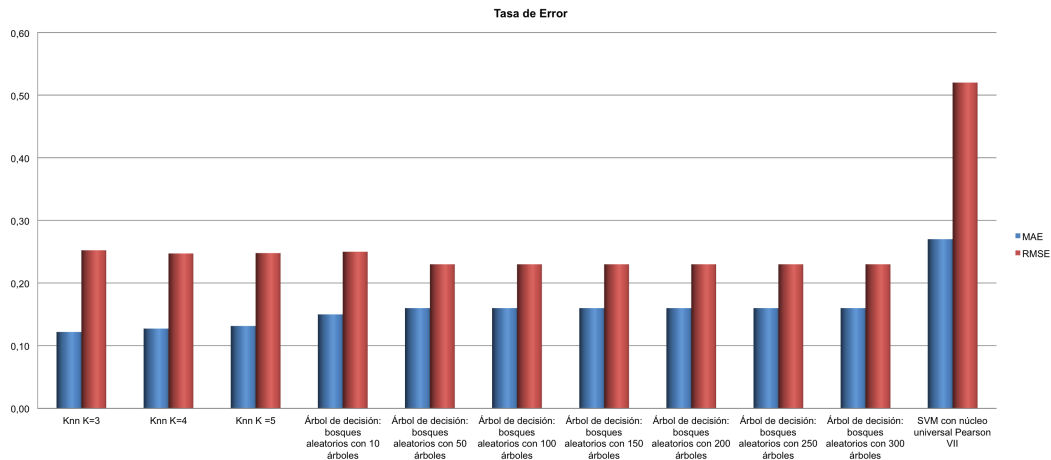
En lo que respecta a las tasas de error, se puede observar que para el MAE (Figuras 4.7a y 4.7b) no se mantiene el mismo comportamiento de los clasificadores que en el nivel de precisión. Mientras que las máquinas de soporte vectorial eclipsan al resto de los clasificadores (con excepción del que utiliza el *núcleo universal Pearson VII*), los árboles de decisión generados mediante *bosques aleatorios* son los que peores resultados obtienen. A pesar de todo, la variación entre los valores obtenidos por los primeros y los segundos es de 0,008 unidades, un valor relativamente pequeño para preocuparnos. En el resto de los casos, también se observan variaciones pero que se mantienen dentro del rango ya comentado. En definitiva, en lo referente a la tasa de error, todos los clasificadores obtienen unos resultados muy parejos.

Teniendo en cuenta la tasa de error RMSE (Figuras 4.7a y 4.7b), se vuelve a ver que los árboles de decisión generados con *bosques aleatorios* son los clasificadores que han alcanzado mejores resultados (al igual que en el nivel de precisión). Por eso, este clasificador es considerado como el mejor para la predicción de los microrrechupes ya que son los más equilibrados. En el caso que nos ocupa, la variación entre todos los clasificadores es mínima ya que la mayoría ronda el rango de valores que van desde 0,23 a 0,26 unidades, lo

4.3 Validación empírica



(a) La figura muestra cuáles son los clasificadores que mejores tasas de error han obtenido a la hora de realizar la predicción de los microrrechupes. Nótese que el clasificador que mejores resultados en términos de MAE ha alcanzado no es el mismo que obtuvo la mejor precisión. Sin embargo, en términos de RMSE, los bosques aleatorios vuelven a ser los mejores clasificadores.



(b) Esta segunda figura completa la lista de las tasas de error para los clasificadores utilizados en la predicción de los microrrechupes. Sorprendentemente, a pesar del comportamiento similar de todos los clasificadores, los SVM con núcleo universal Pearson VII disparan sus errores.

Figura 4.7: Los valores relativos a las tasas de error que han alcanzado los clasificadores a la hora de realizar la predicción del estado estacionario $t + 1$ para el defecto de microrrechupes. La medida se ha hecho a través de los cálculos del “*Mean Absolute Error*” y “*Root Mean Square Error*”. Destaca que el comportamiento de todos los clasificadores ha sido muy similar.

4. Predicción del estado estacionario $t + 1$

que muestra que todos ellos comparten un comportamiento extremadamente parejo en lo que a las tasas de errores se refiere. Sin embargo, y al igual que en todos los casos anteriores, las máquinas de soporte vectorial aprendidas con el *núcleo universal Pearson VII* son las que peores resultados obtienen. Por consiguiente, este clasificador se podría descartar para la predicción de los microrrechupes como clasificador único. Por el contrario, y pensando en la posterior generación de un meta-clasificador, o conjunto de clasificadores que trabajan unidos para alcanzar un objetivo común, sí que es interesante que forme parte de la comunidad y que se tenga en cuenta su opinión para intentar realizar la predicción de resultados.

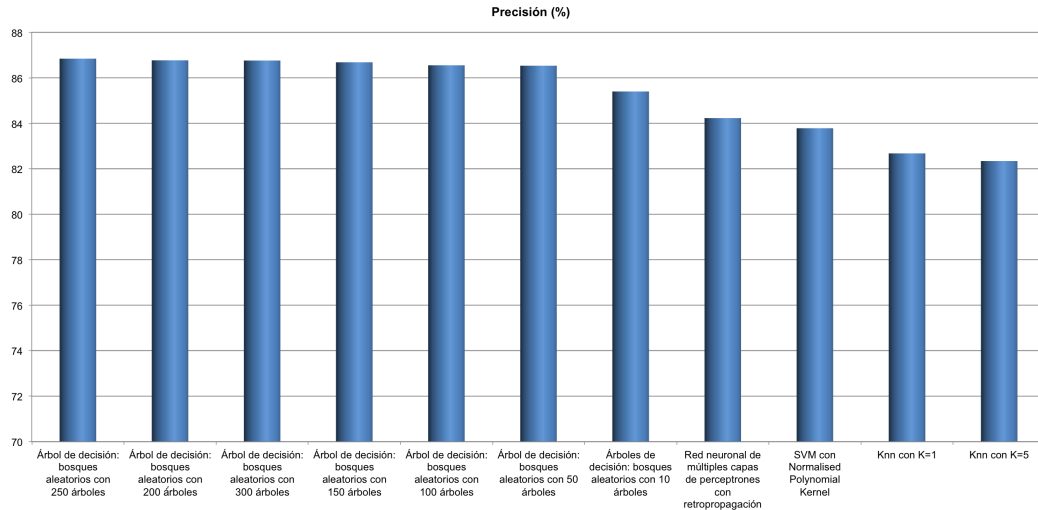
En resumen, este tipo de clasificación basada en técnicas de aprendizaje automático supervisado (que en la mayoría de los casos ha alcanzado más de un 90 % de precisión) es una interesante solución a adoptar para la predicción de microrrechupes. De esta manera, se reduce de forma significativa el coste y la duración de los métodos de pruebas actuales, así como se asegura la reducción del número de piezas que pueden desfallecer mientras están siendo utilizadas. Cabe destacar que los *bosques aleatorios* pueden ser utilizados de la misma forma que se utilizaban las redes bayesianas en [PBZ08]. Además, nótese que aunque se buscan nuevos modelos de MPC no lineales, el clasificador que mejor resultado ha obtenido es inherentemente lineal. La razón fundamental, como se explica en 4.2.1, es que los *bosques aleatorios* son una agrupación de clasificadores que por sí solos se sabe que van a funcionar con peor rendimiento, pero que al trabajar en grupo obtienen mejores resultados. Incluso, a través de la combinación de varios clasificadores lineales, se puede llegar a simular un comportamiento no lineal. A pesar de todo, la variación entre este clasificador lineal y el primero de los no lineales es de 2 unidades. Por lo tanto, se deja la puerta abierta a la combinación de clasificadores como hacen los *bosques aleatorios*.

4.3.2.2 Propiedades mecánicas

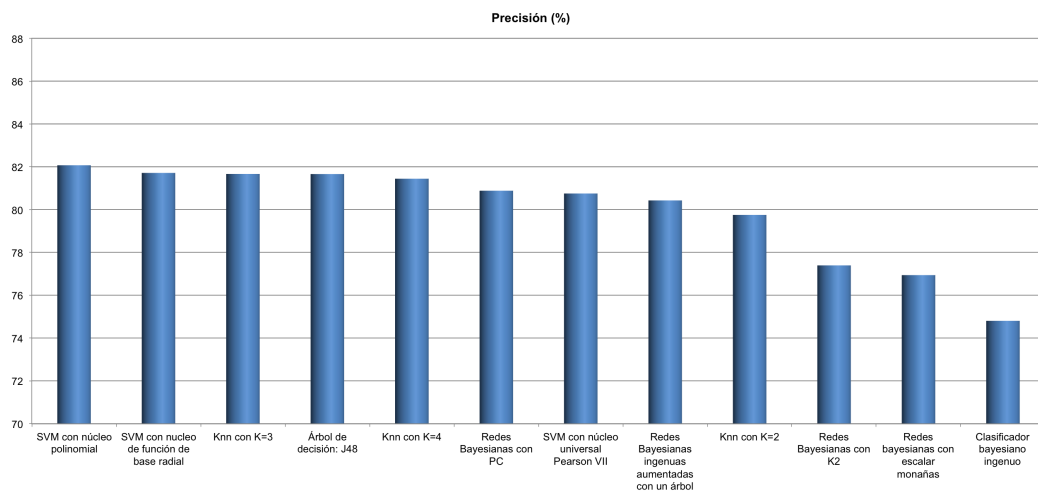
Sobre las propiedades mecánicas, hasta donde llega nuestro conocimiento, no existe ningún modelo que permita predecir su estado antes de llegar a realizar las piezas. Por eso, se decidió dar los mismos pasos que con los microrrechupes.

Uno de los aspectos que merecen ser comentados es la inexistencia de un modelo previo de redes bayesianas, por consiguiente, el primero de los pasos fue replicar el experimento con el fin de comparar los resultados. Es por ello, que en las Figuras 4.8a, 4.8b, 4.9a y 4.9b, se puede ver un clasificador más

4.3 Validación empírica



(a) Los mejores clasificadores tienen un comportamiento similar. Las diferencias entre ellos son mínimas, llegando a una variación máxima de 4,5 unidades. El método de árboles aleatorios eclipsa al resto de los clasificadores ya que, independientemente de su configuración, todos ellos se colocan a la cabeza de los clasificadores.



(b) La segunda tanda de clasificadores se encuentran también muy cerca de los resultados generales de la predicción. Podemos observar que el clasificador con el que se inició la investigación de los microrrechupes, redes bayesianas con el aprendidas con PC, es sobrepasado por muchos de los otros.

Figura 4.8: Resultados alcanzados al realizar la predicción de la carga máxima en un estado de tiempo $t+1$. De todos los clasificadores el que mejores resultados ha obtenido ha sido el generado a partir de *bosques aleatorios con 250 árboles*. La precisión a la que se ha llegado ha sido de un 86,84 %.

4. Predicción del estado estacionario $t + 1$

que la de los experimentos de los microrrechupes. En este caso el clasificador que ha sido añadido es la red bayesiana en la que para realizar el aprendizaje estructural se utiliza el algoritmo PC o NPC. En particular, el algoritmo PC [SGS00] se utiliza aquí para conseguir la estructura de causalidad y/o relaciones correlativas entre las variables que afectan al proceso de fundición. En otras palabras, el algoritmo PC utiliza los datos de la muestra de entrenamiento para definir un modelo bayesiano, lo que representa el conjunto de las dependencias y las relaciones de independencia entre los parámetros. Si se conoce de antemano alguna de estas relaciones entre las variables y éstas deben aparecer en el modelo bayesiano, podemos hacer uso del algoritmo NPC [KM08]. El algoritmo NPC permite definir estas relaciones iniciales que son conocidas como *condiciones de ruta necesarias*.

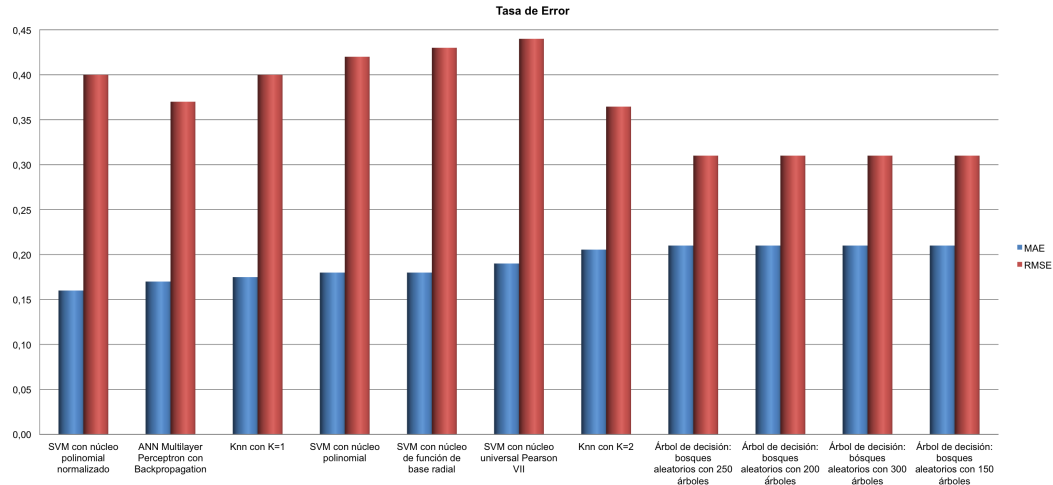
En el caso de la carga máxima, los árboles de decisión generados con *bosques aleatorios* obtuvieron los mejores resultados en términos de precisión. Para la predicción de la carga máxima se consiguió alcanzar un nivel de precisión del 86,84 % utilizando 250 árboles. Este es el número de árboles seleccionado como óptimo, ya que como se ilustra en la Figura 4.8a se trata de un máximo local. A pesar de ser el mismo clasificador que obtuvo los mejores resultados para los microrrechupes, esta vez fue generado con menos árboles aleatorios. Por lo que no podemos utilizar los mismos parámetros de configuración para este clasificador. Por consiguiente, para cada uno de los problemas hay que generar un modelo específico que se encarga de la predicción de cada uno de los defectos.

Muy de cerca se encuentran las redes neuronales artificiales que obtuvieron un buen nivel de precisión (un 84,23 %). La diferencia es mínima, únicamente se reduce el nivel de precisión en 2,61 unidades. Con lo que podemos considerar que también se obtienen buenos resultados mediante la utilización de clasificadores no lineales.

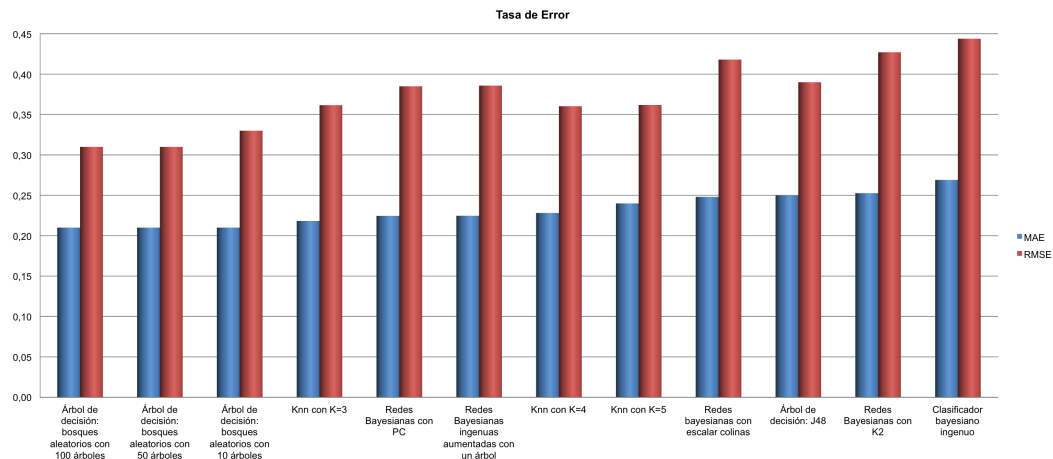
Es sorprendente el comportamiento alcanzado por el algoritmo de los K vecinos más próximos. A pesar de sus simplicidad, los valores de $K = 1$ y $K = 5$ alcanzaron unos resultados muy similares a los de los árboles de decisión y redes neuronales artificiales. Sin embargo, el resto de los valores de K no lograron unos valores tan interesantes. Así, este modelo es una buena opción a tener en cuenta a la hora de desarrollar un meta-clasificador. No obstante, en caso de querer trabajar con un único clasificador para la predicción de las propiedades mecánicas no sería la opción más adecuada.

En el caso de la predicción de la carga máxima, cabe destacar el mal resultado que lograron los modelos basados en clasificadores bayesianos. La

4.3 Validación empírica



(a) A continuación observamos los primeros clasificadores y sus tasas de errores. A pesar de que pueda parecer que los comportamientos han sido muy dispares, no hay grandes diferencias entre los resultados alcanzados por cada clasificador.



(b) Se muestra la segunda tanda de clasificadores y las tasas de error que han alcanzado. Los clasificadores que habían copado las primeras posiciones mantienen un comportamiento similar en sus tasas de errores, que en términos de MAE son mejorados, y en términos de RMSE vuelven al cauce de las predicciones.

Figura 4.9: Tasas de error alcanzadas al realizar la predicción de la carga máxima, una de las propiedades mecánicas de las piezas de fundición. Estas tasas han sido medidas en términos de “Mean Absolute Error” y “Root Mean Square Error”. El comportamiento de las tasas de errores no sigue los mismo patrones que la precisión en el acierto.

4. Predicción del estado estacionario $t + 1$

mayoría de las aproximaciones basadas en este tipo se encuentran entre los clasificadores con peor comportamiento. Destacamos que pese a que la naturaleza del problema es similar y que muchas de las variables con las que se trabaja son las mismas que utilizamos para la predicción de los microrrechupes, los clasificadores no son capaces de alcanzar el mismo nivel de resultados. Aspecto harto interesante debido a que nos indica claramente que no se puede presuponer que un mismo clasificador se comporte de forma similar a pesar de que la gran mayoría del conocimiento con el que se ha generado sea análogo.

En lo referente a las máquinas de soporte vectorial, su comportamiento parece desigual, ya que unos se postulan como la mejor alternativa a los árboles de decisión y las redes neuronales artificiales (aquellos que son generados con el *núcleo polinomial normalizado*), mientras que otros no son capaces de superar clasificadores tan simples como el algoritmo de los K vecinos más próximos. También, al igual que para los microrrechupes, el peor de los clasificadores basados en máquinas de soporte vectorial son los que se crearon utilizando el *núcleo universal Pearson VII*, lo que hace pensar que no es capaz de adaptarse a los datos que son extraídos de la fundición. De todas formas, al igual que ya se comentó para la predicción de los microrrechupes, es interesante la utilización de este clasificador en el conjunto de clasificadores que formen parte de un meta-clasificador, ya que puede aportar y compartir el conocimiento alcanzado con el resto de los clasificadores.

En cuanto a las tasas de error, para el MAE (ver Figuras 4.9a y 4.9b) nos encontramos en la misma situación que en la predicción de microrrechupes. Ninguno de los clasificadores mostró el mismo comportamiento alcanzado en la precisión. A pesar de todo, los valores de las tasas de error van desde 0,16 unidades a 0,27 (una variación de 0,11 unidades) siendo de poca importancia la diferencia del comportamiento que lograron cada uno de ellos. Sin embargo hablando de la tasa de error RMSE (ver Figuras 4.9a y 4.9b), los resultados obtenidos se acercan más a lo sucedido con la precisión en la predicción. Al igual que en el caso del MAE, la diferencia entre los valores es muy baja (los valores se encuentran en el rango que va desde las 0,31 unidades a las 0,44 unidades). Pero en este caso, los árboles de decisión vuelven a ser los mejor clasificados, Así, los podemos considerar como el método más equilibrado.

Como conclusión, aunque la predicción de la carga máxima no alcanza el mismo nivel de precisión que en la predicción de los microrrechupes, se puede considerar que este tipo de solución es adecuada, ya que con el nivel alcanzado (un 86,84 %) los beneficios que puede aportar al sistema de producción de una fundición son elevados. También, se ha detectado que existen problemas

en los datos extraídos de la fundición, fallos introducidos por los operarios que se encargan de realizar las anotaciones. Este ruido en la muestra de entrenamiento hace que el comportamiento de los clasificadores no sea el más adecuado. Esto es así, porque al no estar trabajando con todas las medidas tomadas correctamente, se está distorsionando la realidad de los clasificadores. Una vez que se haya solucionado este problema, los resultados obtenidos serán mejores debido a que la representación será más fiel al modelo real de la fundición.

En el caso de tener que utilizar un clasificador para la predicción de este tipo de características, lo más idóneo sería la utilización de los árboles de decisión generados con *bosques aleatorios*, ya que con los experimentos aquí realizados se demuestra que esta aproximación obtiene buenos resultados y que podría ser utilizado para sustituir los actuales sistemas de predicción empleados en los MPC.

4.3.3 Predicción mediante meta-clasificador

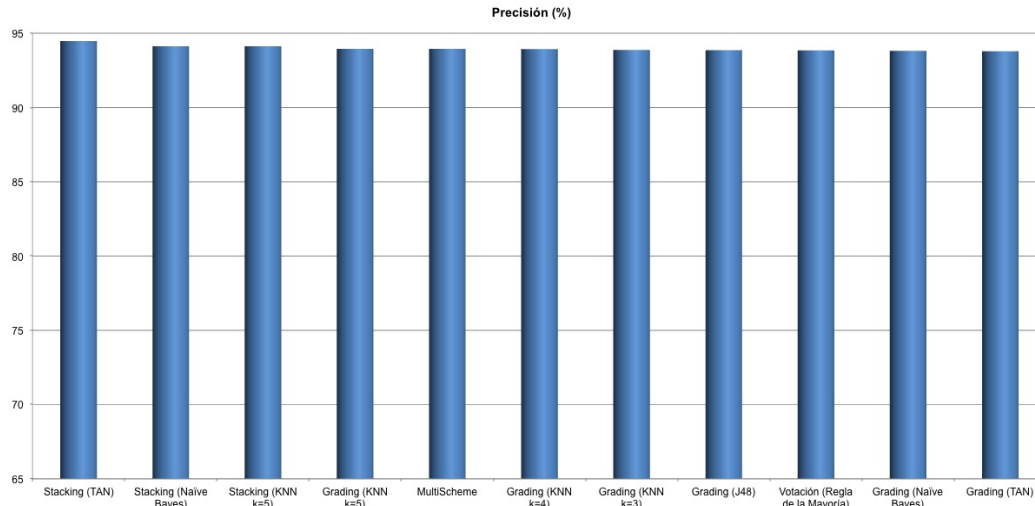
Una vez descubiertos los resultados que tienen los clasificadores trabajando de forma separada, y siguiendo lo que se indica en la metodología presentada en la sección 4.3.1, hemos obtenido los siguientes resultados. Al igual que en los experimentos anteriores, y con el objetivo de hacer mucho más legibles los resultados, hemos realizado una división según el objetivo de clasificación seleccionado.

4.3.3.1 Microrrechupes

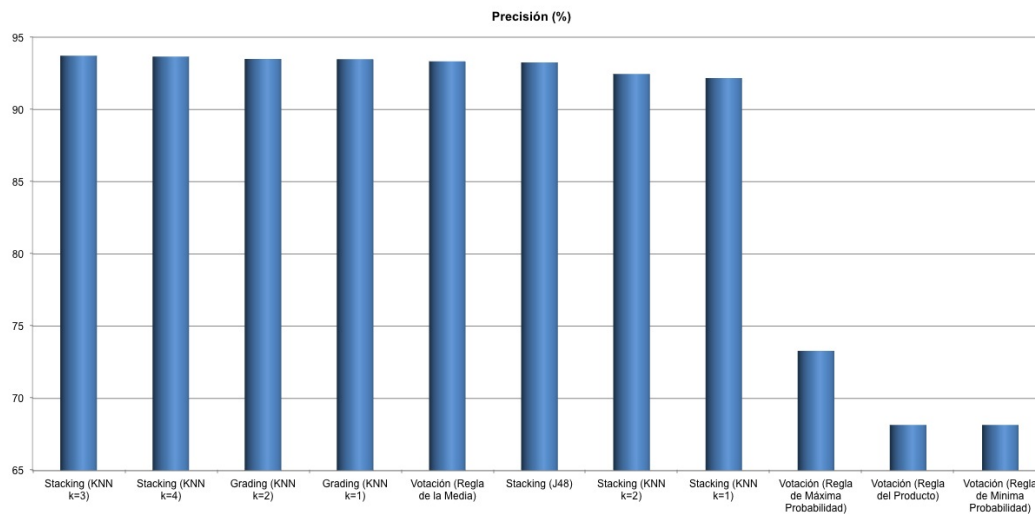
Los métodos de aprendizaje automático cada vez se utilizan más dentro de sistemas de monitorización y predicción. Pero todos esos experimentos hacen uso de clasificadores que solucionan el problema trabajando ellos solos. Sin embargo, y a nuestro entender, no existe ninguna investigación en la que se utilicen sistemas de meta-clasificación, o más concretamente, métodos de combinación de clasificadores simples y que su objetivo sea la mejora de los procesos productivos. Así, en esta tesis doctoral se ha llevado a cabo una investigación pionera en su campo.

Con el ánimo de poder comparar los resultados con los de los clasificadores unitarios, hemos realizado las mediciones en los mismos términos, la precisión en el acierto (ver las Figuras 4.10a y 4.10b) y las tasas de error MAE y RMSE (ver las Figuras 4.11a y 4.11b).

4. Predicción del estado estacionario $t + 1$



(a) Los primeros métodos de combinación han tenido un comportamiento bastante similar. Sin embargo, el método *stacking* es el que copa las primeras posiciones para la generación de un sistema de meta-clasificación.



(b) La figura ilustra la segunda parte de métodos de clasificación. La gran mayoría de ellos tiene un comportamiento similar a los mejores, sin embargo, los tres últimos, basados en la técnica de la *votación*, que parecen no ajustarse al proceso.

Figura 4.10: Resultados obtenidos en el proceso de predicción del estado estacionario $t + 1$ para detectar problemas de microrrechupes a través de la utilización de meta-clasificadores. Se muestran los datos en términos de porcentaje de precisión. La técnica de *stacking basada en un clasificador bayesiano TAN* ha sido la mejor de todas ellas, consiguiendo obtener un 94,47% de precisión.

Entonces, en términos de precisión, podemos observar que el mejor de los clasificadores ha sido el que desarrolla su proceso mediante el método *stacking* en el que los clasificadores del segundo espacio han sido creados a través de un clasificador bayesiano ingenuo ampliado a través de un árbol. La precisión a la que se ha conseguido llegar con las técnicas de meta-clasificación ha sido un 94,47%. Pero los comportamientos de todos los meta-clasificadores han sido similares. Así, el porcentaje de acierto se mueve en un rango que va de los 68,17% a los 94,47% ya comentados. A pesar de que parece que hay una gran diferencia entre los clasificadores, podemos indicar que 19 de los 22 meta-clasificadores probados han alcanzado una precisión superior al 92%. Con lo que el rendimiento, como ya hemos comentado, es similar y no tan dispar como esas cifras podrían hacer pensar.

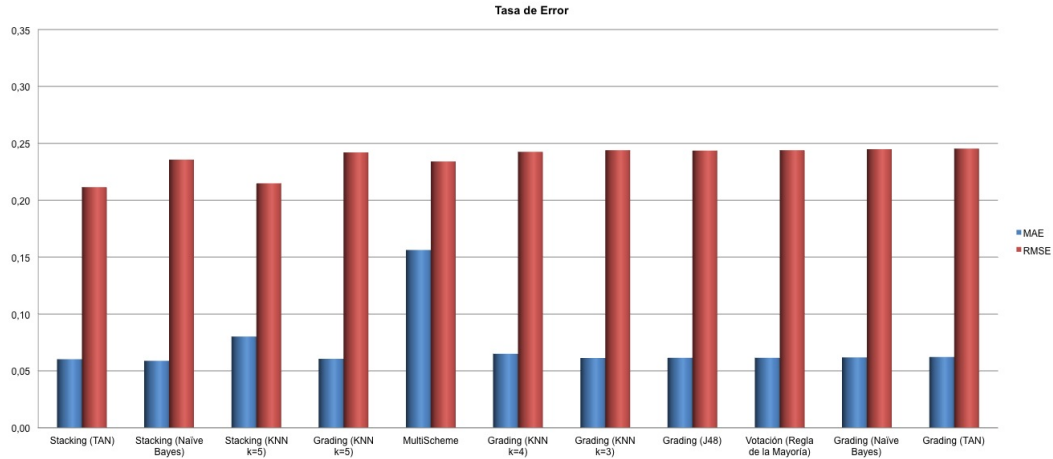
Sorprendentemente, el método *stacking*, además de ser el que mejor resultado ha obtenido, ocupa las tres primeras posiciones. Con lo que este tipo de combinación de clasificadores es capaz de ajustarse mucho mejor al proceso de detección de microrrechupes que el resto de los métodos. Pero, por otra parte, a pesar de esa supremacía, los métodos basados en *grading* también obtienen unos buenos resultados. No debemos olvidar que la variación entre la mayoría de los meta-clasificadores únicamente es de 2,29 unidades, con lo que podemos considerar que su comportamiento ha sido satisfactorio.

Otro interesante aspecto es el buen resultado que se ha obtenido con métodos de combinación muy simples como son el Multiesquema y la combinación por voto a través de la regla de la mayoría. Dado que los resultados alcanzados son buenos, 93,94% y 93,85% respectivamente, estos métodos pueden ser utilizados con el objetivo de minimizar los cálculos y la complejidad computacional que tienen los otros métodos de combinación de clasificadores.

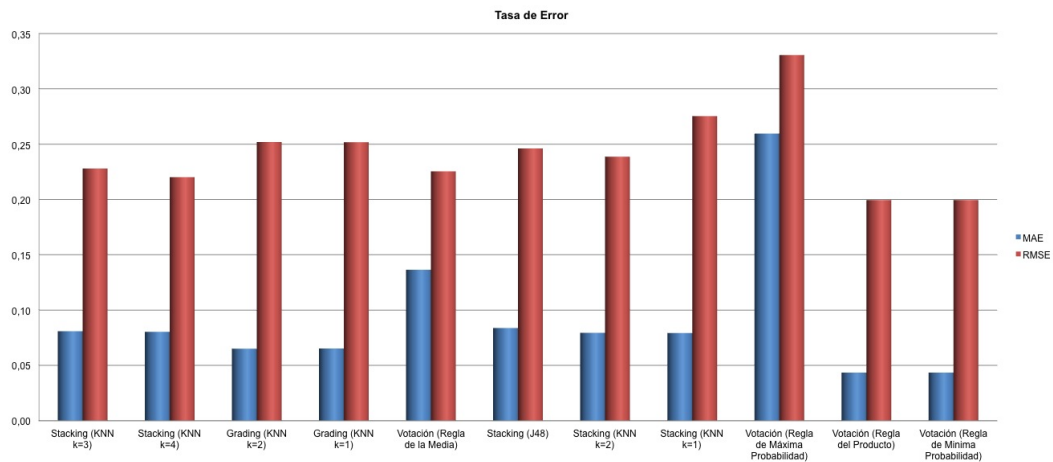
Algo que requiere ser destacado es que al hacer una comparación de los métodos de combinación y de los métodos de clasificación unitarios hemos descubierto que la afirmación de que los métodos de combinación aproximan los resultados es correcta. De esta forma, aproximadamente todos los clasificadores con sus porcentajes superiores al 92% quedan relativamente cerca de todos aquellos clasificadores que fueron capaces de obtener más de un 90% de acierto. Sin embargo, en los clasificadores unitarios encontramos casi tantos resultados en la decena de los 80 como en la de los 90, con lo que este tipo de sistemas parece más preciso. Y no sólo eso, en este caso hemos sido capaces de superar el porcentaje de precisión que tenían los clasificadores unitarios.

Pero dejando de lado aquellos clasificadores que obtuvieron mejor rendimiento, encontramos tres que no lo hicieron así. Más concretamente, son los

4. Predicción del estado estacionario $t + 1$



(a) Los primeros métodos de combinación han tenido un comportamiento bastante similar. Exceptuando el método *Multiesquema* que su tasa de error MAE se dispara, los demás, tanto en MAE como en RMSE, no consiguen más de 0,05 unidades de diferencia.



(b) El comportamiento de este segundo grupo ha sido similar. Exceptuando dos de los métodos de combinación por voto, sus valores no llegan a diferenciarse en más de 0,05 unidades. Nótese que los clasificadores con mejores tasas de error son los que peor precisión tienen.

Figura 4.11: Resultados obtenidos en relación a las tasas de error para el proceso de predicción del estado estacionario $t + 1$ con el fin de la detección de posibles problemas de microrrechupes a través de la utilización de metaclasificadores. Se muestran los datos en términos de “*Mean Absolute Error*” y “*Root Mean Square Error*”. Quitando casos especiales, el comportamiento de los metaclasificadores ha sido similar.

clasificadores que combinan los resultados a través de voto y utilizan las reglas de la probabilidad máxima, el producto y la probabilidad mínima. Estos resultados nos dan a entender que el proceso no es tan simple como para suponer de antemano cómo se relacionan las probabilidades. De esta forma, estos modelos no llegan a ajustarse al proceso y obtienen malos resultados. Por ello, parece que las mejores opciones son aquellas capaces de obtener el conocimiento de una forma mucho más razonada o no tienen en cuenta las probabilidades para realizar las uniones como es el método de voto que utiliza la regla de la mayoría.

En lo que respecta a las tasas de error, podemos observar que de la misma forma que para los clasificadores unitarios, no obtenemos el mismo comportamiento que el alcanzado por los meta-clasificadores en relación a la precisión. Mientras que el método *stacking* creado a través de un clasificador TAN fue el que alcanzó los mejores resultados de precisión, sorprendentemente, en lo que a MAE se refiere, los métodos de combinación que mejor tasa de errores han alcanzado han sido aquellos que obtuvieron peores resultados en la precisión, más concretamente, los métodos de combinación por voto que utilizan la regla del producto y la de la probabilidad mínima. Ambos alcanzaron un valor de 0,0435. A pesar de estos resultados, no debemos obviar que el mejor meta-clasificador en términos de precisión no se encuentra muy lejos de estos dos. Así, el método *stacking* que utiliza los clasificadores bayesianos ingeniosos ampliados con un árbol como clasificadores del segundo espacio obtuvo un valor de 0,0602 para el MAE. Por lo que podemos mantenerlo como la opción adecuada para realizar las predicciones gracias a la estabilidad alcanzada entre el porcentaje de precisión y su tasa de error.

No obstante, a pesar de poder identificar los mejores meta-clasificadores según su tasa de error, el comportamiento alcanzado por la gran mayoría ha sido similar. De esta forma, eliminando excepciones, los meta-clasificadores han alcanzado unos valores de MAE que van desde las 0,0435 unidades a las 0,0838, lo que no supone una variación de ni siquiera una décima. Las excepciones de comportamiento han sido los métodos de combinación a través del voto y que aplican las reglas de la media (alcanzando un 0,1364) y la máxima probabilidad (con un valor de 0,2595 convirtiéndolo en el que peor tasa de MAE tiene), y el método Multiesquema que tan buenos resultados había obtenido en la precisión. Estos resultados nos continúan indicando que el método de votación, en general, parece no adaptarse al proceso de fundición sobre el que estamos trabajando. Por otra parte, nos sorprendemos de la variación con la que cuenta el método Multiesquema (con un valor de 0,1562), con lo que

4. Predicción del estado estacionario $t + 1$

para minimizar la complejidad computacional deberemos seleccionar únicamente el método de combinación por voto que aplica la regla de la mayoría. En este caso, este último método ha alcanzado un valor de 0,0615 dejándolo en la zona media de los clasificadores. En otras palabras, se trata de un método de complejidad reducida con una buena tasa de acierto y una gestión de los errores más que aceptable.

En cuanto a la segunda de las tasas de error que estamos midiendo, el RMSE, el comportamiento es bastante parejo. No se posicionan los meta-clasificadores en las mismas posiciones, pero los resultados que obtienen hacen pensar en que el comportamiento de todos ellos para la gestión de los errores es similar. Como podemos observar, los valores alcanzados van desde las 0,1995 unidades (alcanzadas por los métodos de votación con las reglas del producto y de la mínima probabilidad) hasta las 0,3306 (alcanzadas por el método de votación con la regla de máxima probabilidad). La diferencia entre ellos no supera las 0,1311 unidades. Sin embargo, eliminando el último de ellos que es el que más se desvía del funcionamiento general, volvemos a obtener que los comportamientos de los clasificadores no se distancian de forma significativa (i.e., no llegan a desviarse más de una décima, concretamente, 0,0759 unidades). Por consiguiente, las deducciones llevadas a cabo a través de la tasa de error MAE son aplicables para la tasa RMSE.

En resumen, estas técnicas de clasificación mediante métodos de combinación de clasificadores unitarios, en las que la mayoría de ellos han alcanzado más de un 92 %, no solo han aproximado los resultados individuales, sino que han conseguido superarlos. En definitiva, consideramos que son una interesante solución para incorporar una nueva forma de predicción de los estados estacionarios $t + 1$.

Por otra parte, a través de la utilización de los meta-clasificadores conseguimos desarrollar un método de predicción que elimina ciertas de las limitaciones con las que cuentan los MPC: (i) se amplían las opciones para la selección de modelos, (ii) no se necesita de personal experto ya que las relaciones son aprendidas de forma automática y (iii) los métodos de aprendizaje automático y de combinación de clasificadores tienen bien definidas las directrices para realizar las pruebas de los modelos, con lo que se soluciona el problema de determinar si el modelo seleccionado es el adecuado.

También, desarrollando este tipo de modelos se trabaja sobre diferentes aspectos esperados para las siguientes generaciones de los sistemas MPC: (i) la combinación de múltiples meta-clasificadores permite realizar un sistema de clasificación multiobjetivo, (ii) debido a que los sistemas de aprendizaje

automático son conocidos dentro de la comunidad científica, es fácil aplicar los métodos que demuestran cómo de robustas son las soluciones adoptadas y (iii) en lugar de seleccionar un modelo lineal o mejorar un modelo no lineal, a través de la combinación de clasificadores disponemos de un modelo híbrido que hace uso de ambas aproximaciones para tomar la decisión de clasificación.

4.3.3.2 Propiedades mecánicas

Para realizar las mediciones nos hemos basado en los mismos parámetros que utilizamos para los microrrechupes, particularmente, se tratan del grado de precisión alcanzado por diferentes métodos de combinación de clasificadores, y las mediciones de las tasas de error en términos de MAE y RMSE. Así, en las Figuras 4.12a y 4.12b podemos visualizar los resultados obtenidos en términos de precisión. En el caso que nos ocupa, el método de meta-clasificación que mejores resultados ha llegado a obtener es el método de *grading* que hace uso de clasificadores de primer nivel generados con clasificadores bayesianos ingenuos ampliados mediante un árbol. La máxima precisión alcanzada ha sido de un 86,63%. Aunque el meta-clasificador que mejores resultados obtuvo no es el mismo que el que lo consiguió para los microrrechupes, el método de *stacking* con clasificadores TAN sigue estando dentro de el top 10 de los meta-clasificadores con un resultado muy similar a su vencedor, un 85,73% de precisión.

Muy cerca se encuentran los métodos de combinación de clasificadores basados en *Multiesquema* y *grading* construido a través de un árbol de decisión aprendido con el algoritmo C4.5. En términos de precisión, la diferencia entre estos tres métodos no es superior a las 0,29 unidades. Otro aspecto a destacar es que para la predicción de la carga máxima no contamos con un método que, independientemente de las reglas o clasificadores utilizados, se posicione en las primeros lugares. La distribución de los métodos es mucho más arbitraria. También, sorprendentemente, igual que en el caso de los microrrechupes, dos clasificadores sencillos como son el método multiesquema y el de combinación de resultados por votación (usando la regla de la mayoría) se encuentran entre los mejores métodos. Los resultados que han alcanzado son de una precisión de 86,37% y 85,95% respectivamente. Lo que vuelve a hacernos pensar en que son los métodos idóneos para la reducción de la complejidad computacional que añaden estos métodos de meta-clasificación al sistema MPC.

Aquí también tenemos unos clasificadores que se desvían de la tendencia marcada por la mayoría, específicamente, volvemos a encontrarnos de nuevo

4. Predicción del estado estacionario $t + 1$

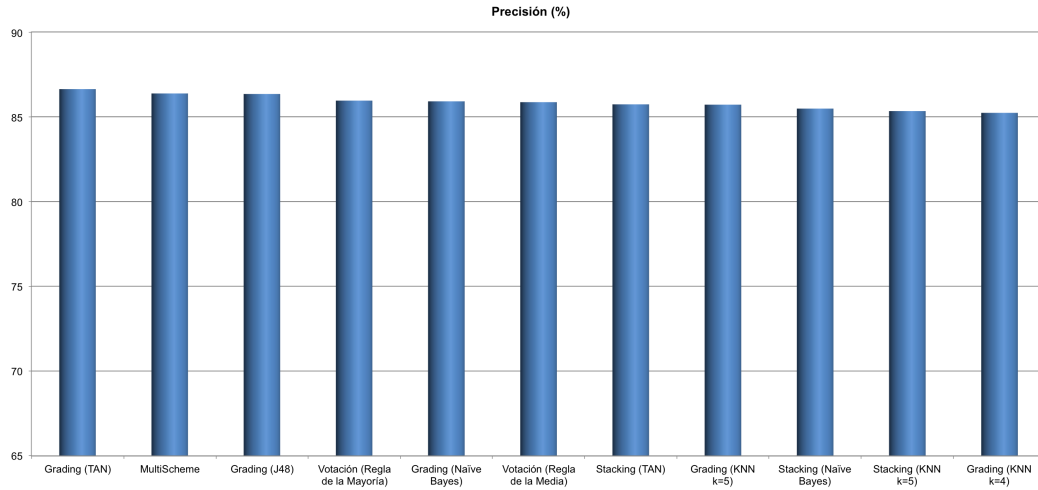
con los clasificadores que combinan los resultados a través de las técnicas de votación. Esta vez, las reglas que no se adaptan al proceso de producción de la fundición son las del producto y la de la probabilidad mínima. A parte, encontramos uno de los métodos de *stacking* generados a través de clasificadores KNN. En este último caso, al igual que los meta-clasificadores basados en *grading* que hacen uso de los KNN, los resultados son bajos debido a que el valor óptimo a adoptar en la k es el de 5. Por eso, éstos acaban desviándose del comportamiento que tienen la mayoría de los clasificadores.

Teniendo en cuenta estos resultados, podemos deducir que el comportamiento de los métodos de combinación de clasificadores también es dependiente del defecto, por lo que hemos solucionado el problema de la selección de los clasificadores, pero hemos añadido el problema de la selección de un método de meta-clasificación que se ajuste al proceso. A pesar de este hecho, la precisión de todos los métodos es muy pareja, ya que 19 de los 22 meta-clasificadores han alcanzado una precisión superior al 80 %. Parece que los resultados obtenidos no son realmente buenos, sin embargo, los valores de precisión obtenidos mediante los clasificadores unitarios son similares, con la excepción de que aquí los resultados tienden a acercarse a los valores altos de la decena, mientras que la mayoría de los clasificadores unitarios tienden hacia los valores bajos de la decena. Sin embargo, en comparación con el experimento realizado para los microrrechupes, en esta ocasión no hemos sido capaces de superar la precisión alcanzada por los clasificadores unitarios. No obstante, la aproximación es buena, ya que únicamente contamos con una desviación de 0,21 unidades. Sin embargo, el comportamiento de los meta-clasificadores es mucho más estable que el conseguido a través de clasificadores unitarios.

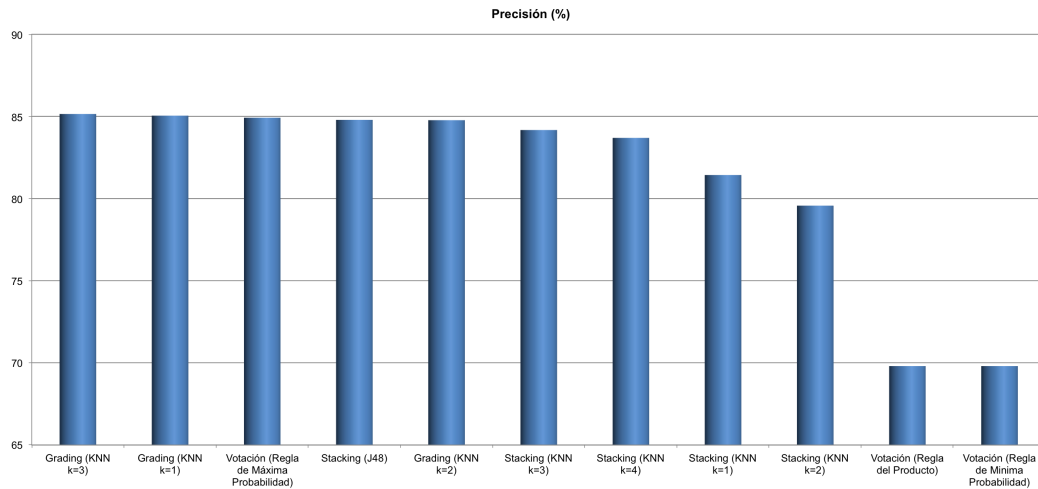
En cuanto a las tasas de error (ver las Figuras 4.13a y 4.13b), observamos que el comportamiento que se recoge es similar al que se viene dando a lo largo de todos los experimentos, es decir, no se mantiene el mismo comportamiento que marcan los niveles de precisión, pero la mayoría de los métodos se encuentran muy parejos.

Así, en lo que respecta las medidas relativas a MAE, de la misma forma que sucedía en la predicción de los microrrechupes, los mejores métodos han sido los que obtuvieron la peor precisión. Ambos métodos comparten tanto el porcentaje de aciertos como los valores alcanzados en las tasas de error. En este caso, su valor es de 0,0936. Pero como ya hemos comentado, y algo que parece una tendencia marcada en los experimentos de predicción de esta tesis doctoral, el comportamiento de la mayoría de los meta-clasificadores es similar. El rango en el que se mueven los clasificadores va desde las 0,3137 unidades a

4.3 Validación empírica



(a) Los primeros métodos de combinación han tenido un comportamiento bastante similar. Además, se puede observar que no hay un método que podamos designar como el mejor de ellos.



(b) La figura ilustra la segunda parte de métodos de clasificación. La gran mayoría de ellos tiene un comportamiento similar a los mejores, sin embargo, destaca que al igual que en los microrrechupes, algunas reglas del método de *votación* no se ajustan al proceso.

Figura 4.12: Resultados obtenidos en el proceso de predicción del estado estacionario $t+1$ con el fin de la detección de posibles problemas en la carga máxima a través de la utilización de meta-clasificadores. Se muestran los datos en términos de porcentaje de precisión. La técnica de *grading basada clasificadores TAN* ha sido la mejor, consiguiendo un 86,63 %.

4. Predicción del estado estacionario $t + 1$

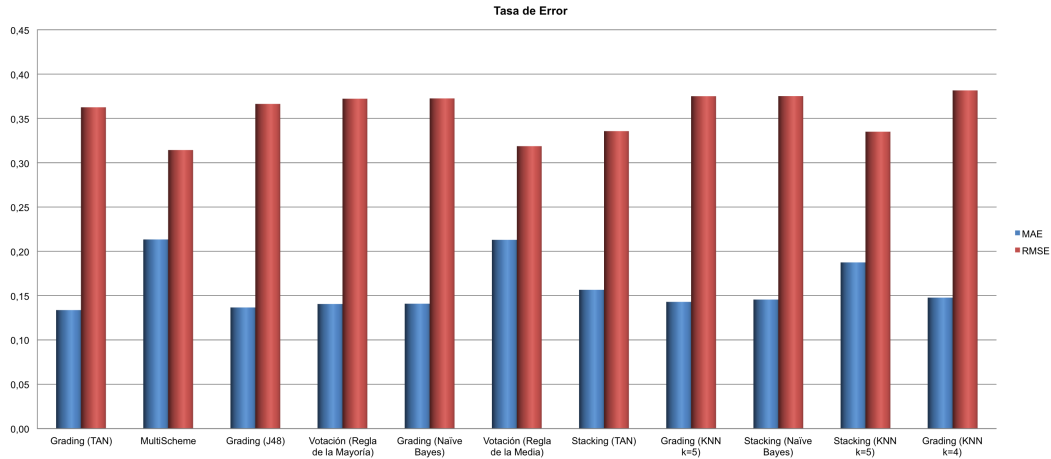
las 0,0936, es decir, 0,2201 unidades. Si eliminamos las excepciones como ya lo hicimos anteriormente, observamos que 19 de los métodos de combinación tienen una desviación de una décima.

A pesar de esta peculiaridad que ya observamos en la predicción de los microrrechupes, el meta-clasificador basado en *grading* construido a base de clasificadores TAN para su primer nivel de predicción se encuentra el siguiente en el ranking. Este clasificador es el que mejor precisión obtuvo, con lo que se postula como el mejor de los métodos para la predicción de los estados estacionarios $t + 1$ para el objetivo de la carga máxima. Por otra parte, y algo sorprendente, el segundo de los meta-clasificadores no es capaz de manejar los errores en la misma forma que la precisión, ya que nos lo encontramos con uno de los peores valores de MAE de todos los clasificadores. De esta forma, nos da pie a dejarlo de lado a la hora de crear un sistema que reduzca la complejidad computacional. Aun así, quedaría la posibilidad de la utilización del método de votación basado en la regla de la mayoría, ya que este meta-clasificador sí que es capaz de manejar los errores de una forma correcta. Únicamente se desvía en menos 5 centésimas del resto de los meta-clasificadores que se encuentran por encima de él.

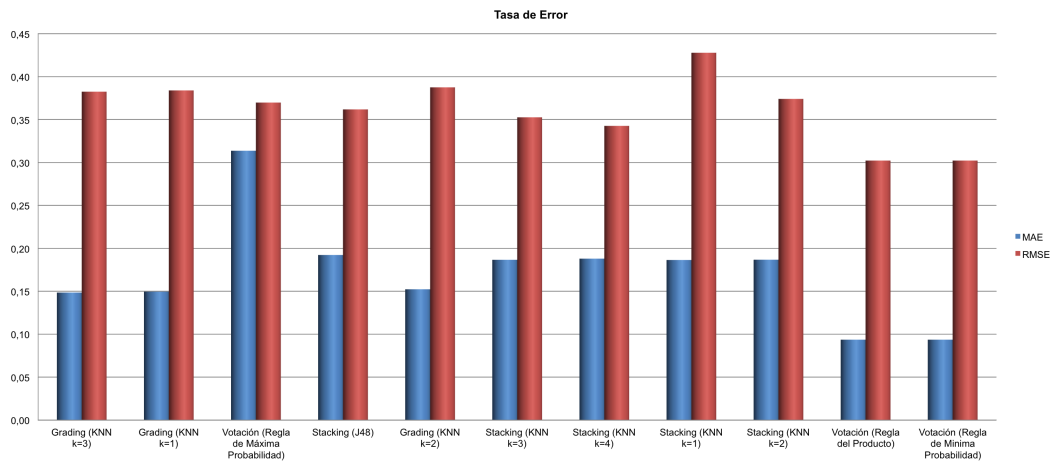
En lo referente a la tasa de errores RMSE, volvemos a encontrarnos que los peores clasificadores en términos de precisión son los mejores en el control de errores. Obviando esta circunstancia, y centrándonos en el resto de los clasificadores que sí obtuvieron buenos resultados en la predicción, podemos observar que los comportamientos de todos ellos vuelven a ser muy similares. La desviación entre todos ellos ronda la décima, y en caso de centrarnos en el comportamiento más repetido, se encuentran todos dentro de la misma décima. Por consiguiente, el meta-clasificador más adecuado será aquel que nos dio un mejor porcentaje de precisión debido a que las desviaciones en los errores no son significativas. Sí queremos destacar que, al comprobar esta tasa de error, el clasificador Multiesquema, uno de los que consiguieron la mejor precisión, vuelve a colocarse entre los primeros.

A modo de conclusiones, este experimento nos muestra que la utilización de estos métodos para la predicción de la situación de la planta es factible. En el caso de la carga máxima no hemos logrado superar los resultados obtenidos por el clasificador unitario óptimo. A pesar de este hecho, los resultados generales sí que superan a los clasificadores unitarios. En primer lugar, la tendencia de los meta-clasificadores les lleva a alcanzar la tasa más alta de predicción, algo que no sucedía con los clasificadores unitarios ya que una parte conseguía acercarse a la franja de los 85% pero otra gran mayoría se centraba en la

4.3 Validación empírica



(a) A continuación observamos los primeros clasificadores y sus tasas de errores. Los comportamientos han sido muy dispares y los resultados no siguen el mismo patrón que pudimos ver para la precisión. Sin embargo, la variación entre todos ellos no es muy elevada.



(b) La segunda tanda de clasificadores mantiene un comportamiento similar a los primeros. Se detectan ciertos casos especiales como los dos primeros con un bajo valor de MAE o el *stacking con KNN k=1* con un alto valor de RMSE.

Figura 4.13: Tasas de error alcanzadas al realizar la predicción de la carga máxima mediante la utilización de métodos para la combinación de clasificadores. Estas tasas han sido medidas en términos de “*Mean Absolute Error*” y “*Root Mean Square Error*”. La tasa de error MAE está optimizada por los métodos que cuentan con más complejidad, mientras que la tasa RMSE es optimizada por los métodos más simples.

4. Predicción del estado estacionario $t + 1$

franja de los 80 %. En segundo lugar, las tasas de error han sido reducidas, de tal forma que, la peor de las tasas de error de los meta-clasificadores es menor que la peor de las tasas de error alcanzada por los clasificadores unitarios. Asimismo, si realizamos la comparación del clasificador con más precisión frente al meta-clasificador con mejores resultados, también obtenemos unas tasas de errores más bajas.

Además, al igual que en los microrrechupes, este método aporta un avance en el estado de la técnica aplicada en las etapas de predicción de los sistemas MPC, al solucionar algunas de las limitaciones con las que contaban estos sistemas, como la aparición de nuevos modelos de predicción, la posibilidad del aprendizaje a través de métodos automáticos permitiendo portar la solución a otros dominios o que este tipo de técnicas tienen muy bien definidas los métodos de evaluación. Del mismo modo, se da una serie de pasos para comenzar a conseguir la siguiente generación de los sistemas MPC al incorporar la robustez de los métodos de aprendizaje automático y la posibilidad de crear un modelo híbrido.

4.4 Discusión de los resultados

Los resultados obtenidos validan nuestra hipótesis inicial de que el proceso de fundición puede ser modelado como una nube de conocimiento experto que permite realizar la predicción del estado estacionario en el que se encontrará la planta en un instante de tiempo futuro $t + 1$. También, estos experimentos muestran la validez de realizar esta tarea a través de la posibilidad de crear un meta-clasificador que combina métodos de clasificación heterogéneos. De esta forma, y a pesar del hecho de que la combinación de clasificadores ha obtenido un buen rendimiento cuando se ha enfrentado a la predicción de los microrrechupes y la carga máxima, hay una serie de aspectos a discutir en relación a la viabilidad de esta solución.

En primer lugar, tal y como hemos visto, con los métodos para la combinación de clasificadores solucionamos el problema de la selección de un clasificador para la predicción de un defecto. Sin embargo, aparece un nuevo problema, los métodos de combinación también se comportan de forma diferente para la predicción de cada defecto. Por lo que hemos desplazado el problema desde el primer nivel de clasificación al segundo de ellos. Nótese que a pesar de esta desventaja, los beneficios que aporta la utilización de este tipo de clasificaciones a los sistemas MPC son mayores, en otras palabras, somos capaces de

generar una nueva generación de MPC capaz de ajustarse al proceso mediante la utilización de un sistema de predicción híbrido manteniendo, o incluso, mejorando la precisión y reduciendo las tasas de error.

En segundo lugar, debido al hecho de que la adquisición de datos se realiza de forma manual, encontramos una serie de irregularidades en los datos que hacen que los resultados obtenidos no sean tan buenos como deberían. Más concretamente, los datos con los que trabajamos parece que tienen varios registros incorrectos. Una de las soluciones para este problema es la instalación de sistemas de captura de datos mucho más precisos. Esta tarea también es de gran complejidad debido a las duras condiciones en las que se desarrolla el proceso productivo de la fundición. Otra de las soluciones es aplicar técnicas que permitan reducir el ruido, por ejemplo la *descomposición en valores singulares* (del inglés “*Singular Value Decomposition*”) o el *análisis de la semántica latente* (de la voz inglesa “*Latent Semantic Analysis*”) han sido utilizados para los procesos de recuperación de información de documentos [SFS08] con grandes resultados en la reducción de ruido y ambigüedades.

En tercer lugar, la adquisición del conocimiento puede considerarse como un cuello de botella. Así como hemos mencionado anteriormente, el proceso de captura de datos se lleva a cabo de forma manual. Las inspecciones para la extracción de los datos y su etiquetado son un proceso de larga duración. Los sistemas SCADA (del inglés “*Supervisory Control And Data Acquisition*”) [DS99] automatizan la función de adquisición de información, lo que para nosotros es conocimiento. Por consiguiente, este tipo de sistemas puede ser empleado con el fin de reducir el tiempo necesario para la recuperación de la información desde las diferentes etapas por las que pasa el proceso de producción. Nótese, también, que el sistema que se está presentando en esta tesis doctoral no va a sustituir por completo a los análisis realizados en laboratorio. La información que allí se obtiene es la base para la readaptación de los modelos que componen el sistema MPC.

Para concluir, la utilización de algoritmos de aprendizaje automático supervisado puede ser un problema en sí mismo. Aunque nuestro conjunto de datos es reducido, puede ir creciendo con el paso del tiempo según se vayan obteniendo nuevas inspecciones. Este crecimiento hace que sea necesario incrementar las posibilidades de almacenamiento de información, y que va a acabar traducándose en una mayor complejidad computacional para el proceso de generación de nuevos modelos, afectando a la precisión de los mismos [CHL06]. Por consiguiente, la solución a este problema es la reducción del conjunto original de entrenamiento [CJ06]. De esta forma, el método que podemos

4. Predicción del estado estacionario $t + 1$

desarrollar para solventarlo es conocido como *reducción de datos* o en inglés “*Data Reduction*”. A través de estas técnicas, seremos capaces de reducir las mediciones, el almacenamiento y la transferencia de la información. También, la reducción de datos lucha contra la *maldición de la dimensionalidad* para intentar mejorar el comportamiento de los modelos (i.e., mejorando la velocidad, la precisión y reduciendo la complejidad computacional en relación a la visualización y entendimiento de los datos) [Tor03]. Específicamente, esta técnica puede gestionar los siguientes aspectos. Primero, la selección de instancias (IS, de su nombre en inglés “*Instance Selection*”) trata de determinar las instancias más relevantes o ser capaz de generar nuevas inferencias [LM01]. Segundo, la selección de características (FS, del inglés “*Feature Selection*”) ajusta el número de atributos o características de entrenamiento [LM08]. Más concretamente, ya estamos pensando en la aplicación de estos métodos para reducir el futuro conjunto de entrenamiento y, también, readaptar los modelos con las nuevas evidencias recogidas de la planta. De hecho, este método puede dar más importancia a las nuevas evidencias que las antiguas, haciendo que el sistema MPC pueda ser mucho más sensible a los cambios del proceso.

4.5 Sumario

En este capítulo hemos hecho dos contribuciones en la mejora de los procesos productivos a través de la utilización de sistemas MPC. Más concretamente, las mejoras han estado dirigidas a lograr un proceso de predicción óptimo. Así, en primer lugar, nos hemos centrado en realizar este proceso a través de métodos de aprendizaje automático supervisado y, en segundo lugar, mediante la utilización de sistemas de meta-clasificación que permiten la unión de múltiples clasificadores (tanto homogéneos como heterogéneos). Para llevar a cabo esta investigación, hemos utilizado la información extraída de un proceso real de fundición en la que contamos con las diferentes variables y niveles de riesgo que fueron explicados en el capítulo 3.

Dado que los sistemas MPC actuales se basan en la generación de una fórmula matemática, lo que implica tener un alto conocimiento del funcionamiento de la planta, proponemos la utilización de algoritmos de aprendizaje automático que solucionan este problema. Hemos evaluado este enfoque centándonos en dos aspectos: la precisión del modelo y las tasas de error alcanzadas. Más concretamente, se ha conseguido demostrar que este enfoque es capaz de alcanzar unas tasas de acierto cercanas al 95 % en las predicciones de microrrechupes y del 87 % para la carga máxima.

Una vez presentados los modelos anteriores, surge la necesidad de unirlos en un único clasificador que los integre. La razón es que tenemos el problema de la selección del método adecuado. La utilización de técnicas de combinación de clasificadores permite obviarlo y, así, conseguir una aproximación al clasificador óptimo. Pero además, los métodos actuales únicamente cuentan con dos visiones del modelo (la lineal y la no lineal) y mediante la utilización de estas técnicas proponemos la generación de un sistema de predicción híbrido no existente hasta la fecha. Para validarlo, hemos partido de los clasificadores obtenidos en el paso anterior y hemos testado diferentes métodos de combinación. Los resultados alcanzados demuestran que estos métodos no sólo permiten aproximar los clasificadores unitarios, sino mejorarlos y reducir las tasas de error.

Creemos que el enfoque alcanzado puede suponer un cambio importante en los sistemas MPC ya que permite evolucionar la perspectiva de estos sistemas consiguiendo un proceso de predicción basado en un aprendizaje automático, y genera un nuevo tipo de modelos híbridos que pretenden ser aquellos capaces de ajustarse tanto a procesos productivos lineales, como a los no lineales.

«Las especies que sobreviven no son las más fuertes, ni las más rápidas, ni las más inteligentes; sino aquellas que se adaptan mejor al cambio.»

Charles Darwin (1809-1882)

5

Adaptando los modelos al proceso cambiante

UN sistema MPC es una representación algorítmica, y también matemática, del proceso de producción sobre el que trabaja. Es por ello, que durante su etapa de creación (ya sea a través de principios fundamentales o empíricos) se busca la representación que más se ajuste al proceso [QB03]. No obstante, esa búsqueda por conseguir el mejor sistema, i.e., el mejor modelo, plantea un problema de base. Los procesos suelen cambiar a lo largo de su vida, con lo que el modelo generado inicialmente, con el paso del tiempo, no será capaz de obtener los mismos resultados que al principio. Cabe destacar, que en caso de utilizar los principios fundamentales conocidos para su generación, el sistema es capaz de adaptarse a esos cambios mucho mejor [QB03], aunque, esto no soluciona el problema.

El grado de importancia de este inconveniente dependerá del proceso que estamos intentando representar. Esto es así debido a que existen procesos productivos que no se sufren de grandes cambios, como puede ser la refinería de petróleo. Pero por otra parte, muchos de los procesos tienen un alto grado de variabilidad, con lo que este factor es mucho más importante. Particularmente, en el caso que nos ocupa, la fundición pertenece al segundo de los grupos [SS05], ya que el sistema se debe encontrar preparado para rápidamente saltar

5. Adaptando los modelos al proceso cambiante

del control de un tipo de piezas a otro totalmente diferente, mientras debe ser capaz de adaptarse a la generación de nuevas piezas.

Otro factor de interés con el que debemos lidiar es la inclusión de nuevas formas para la predicción del estado estacionario $t + 1$. Así, en el capítulo 4 mostramos la utilización de diferentes métodos de aprendizaje automático para realizar nuestras predicciones. La incorporación de éstos hace que debamos disponer de una nueva manera que permita adaptarlos. Nótese, que este proceso es complicado debido a que los clasificadores seleccionados no son capaces de autodaptarse según reciben nueva información como sí lo son otros, por ejemplo, un modelo bayesiano ingenuo actualizable [JL95] o las redes neuronales artificiales con capacidad de evolución [YAO93]. Por consiguiente, la única forma de adaptación que podemos llevar a cabo se centra en el aprendizaje de nuevos modelos de clasificación que sustituirán los antiguos.

Para conseguir la adaptación debemos actuar sobre el proceso que se encarga del aprendizaje. Otro factor a tener en cuenta es la importancia estadística que reciben las nuevas evidencias con respecto a las antiguas. Una nueva evidencia frente a todo el conjunto anterior no tiene una gran importancia ya que, simplificando ideas y basándonos en la estadística básica, no es un valor representativo del conjunto. Además, dentro de ese conjunto general, las evidencias producidas hace años no representan tan bien el actual proceso productivo como lo hacen las nuevas. Para combatir este tipo de problemática disponemos de la técnica para el envejecimiento de muestras mediante la utilización de factores de desvanecimiento (en inglés “*fading factors*”) [FW95].

De esta forma, en este capítulo proponemos un nuevo método que combina diversas técnicas con el fin de aprender los nuevos modelos que realizarán las predicciones de los estados estacionarios. Concretamente, proponemos un método que trabaja sobre los conjuntos de datos utilizados en las etapas de aprendizaje. Para ello, y a través de la compresión bayesiana basada en métodos de Montecarlo, generamos un nuevo conjunto de datos reducido (con un menor número de instancias y, así, envejecer la muestra a lo largo del tiempo), pero que mantiene la misma distribución estadística (aprendida a través de las probabilidades condicionales inherentes en una red bayesiana), al que unimos las nuevas evidencias (el conjunto original está reducido por lo que, ahora, las nuevas evidencias son más importantes). Este nuevo método nos permite ir realizando un aprendizaje acumulativo, en otras palabras, un aprendizaje iterativo que alterna procesos de reducción de datos y agregación de nuevas evidencias provocando una aceleración en el proceso de adaptación y puesta al día de los modelos dedicados a la predicción de estados estacionarios.

5.1 Capacidades estadísticas para la adaptación de modelos

En resumen, las contribuciones llevadas a cabo a través de la investigación presentada en este capítulo han sido las siguientes:

- Demostramos que existen técnicas estadísticas con la capacidad de proporcionar métodos para la generación de nuevos modelos que se vayan adaptando al proceso de una forma mucho más rápida.
- Proponemos un método de adaptación basado en el envejecimiento de muestras.
- Evaluamos el comportamiento de nuestro método para la generación de nuevos conjuntos de datos.
- Demostramos que los modelos reducidos, los cuales son generados a través de la compresión bayesiana, son una gran aproximación estadística de los modelos completos a pesar de construirse con un conjunto de datos que cuenta con un número menor de evidencias.
- Evaluamos de forma empírica el proceso de aprendizaje utilizando conjuntos de datos acumulados mediante la técnica de compresión bayesiana.

Lo que resta de capítulo queda organizado como sigue. La sección 5.1 plantea las bases estadísticas que fundamentan el método de adaptación desarrollado. La sección 5.2 presenta la aproximación desarrollada para la adaptación de los modelos mediante la compresión bayesiana basada en métodos de Montecarlo. Posteriormente, la sección 5.3 detalla la metodología seguida con el fin de demostrar la viabilidad de nuestro método. Del mismo modo, se visualizan los resultados alcanzados durante el proceso de validación. La sección 5.4 pone el punto crítico al modelo que aquí presentamos. Y finalmente, la sección 5.5 recoge un breve resumen de las contribuciones realizadas en el proceso de adaptación de los modelos de aprendizaje supervisado.

5.1 Capacidades estadísticas para la adaptación de modelos

Esta sección muestra las herramientas estadísticas en las que se basa el método de adaptación desarrollado en esta tesis doctoral. Así, en primer lugar, introducimos la capacidad que disponen las redes bayesianas para la compresión de datos mediante técnicas estadísticas.

5. Adaptando los modelos al proceso cambiante

En segundo lugar, realizaremos una breve introducción al método de Montecarlo, un método estadístico numérico normalmente utilizado para realizar la aproximación de expresiones matemáticas complejas y costosas de evaluar con exactitud, y que, particularmente, permitirá realizar la generación de un nuevo conjunto de datos que mantenga la distribución estadística inicial.

5.1.1 Redes bayesianas para comprimir datos

Los métodos de aprendizaje automático siguen contando con una gran aceptación en investigaciones dirigidas a diferentes campos. Uno de estos métodos son las redes bayesianas. Históricamente, las soluciones que se han tratado de obtener con este tipo de algoritmos han estado relacionadas con diagnósticos, predicciones o la inferencia eficiente. Un claro ejemplo es el capítulo 4 en el que las utilizamos como elemento de predicción. A pesar de este hecho, las redes bayesianas pueden ser utilizadas para otros propósitos, como la compresión de conjuntos de datos sin pérdida. Ejemplos de este tipo de aproximaciones los encontramos en [Fre97, DM99].

De esta forma, la idea fundamental que promueve la investigación de estos métodos es aquella que nos indica que dado un fichero f compuesto por un conjunto de datos \mathcal{D} cuya cantidad de información l es muy grande, en un análisis estadístico, y según la teoría de información [BA02], la cantidad de datos realmente significativos es mucho menor a l . Esta aproximación es utilizada por los analistas de información con el fin de reducir grandes conjuntos de datos con una alta dimensionalidad. El método que se utiliza habitualmente es el análisis de los componentes principales (PCA, del la voz inglesa “*Principal Component Analysis*”) [THM02]. Esta técnica permite reducir la dimensionalidad, sin embargo, cuenta con un inconveniente, y es que hacemos la reducción o compresión con una pequeña pérdida de información.

Con el fin de solucionar este problema, y partiendo de la función conocida como el Criterio de Información Bayesiana (CIB) [Sch78]

$$C(B, D) = \log P(D|B) - |B| \cdot 0,5 \log k \quad (5.1)$$

donde $|B|$ es el número de parámetros (probabilidades) almacenadas en la red y k es el número de registros del conjunto de datos. La maximización de la ecuación 5.1 corresponde directamente a la minimización de los bits requeridos para el almacenamiento de (i) los parámetros de la red B obteniendo un nivel de precisión razonable y (ii) una codificación eficiente (codificación aritmética)

5.1 Capacidades estadísticas para la adaptación de modelos

de \mathcal{D} utilizando la distribución de probabilidad interna de B .

Así, mientras las aproximaciones tradicionales para la compresión se centran en producir codificadores en los que la longitud de sus palabras codificadas fuese casi óptima, donde la longitud óptima para el patrón de entrada \mathbf{v} es $\log_2 P_r(\mathbf{v})$, donde $P_r(\mathbf{v})$ es la distribución de probabilidades; la codificación aritmética [RL79, WNC87] es un algoritmo para producir codificaciones óptimas cuando la distribución de probabilidad $P_r(\mathbf{v})$ es conocida. Por consiguiente, si \mathbf{v} es un valor binario, $P_r(\mathbf{v})$ puede ser fácilmente estimada y la codificación aritmética puede ser utilizada para la producción de una solución cercana al óptimo. Si \mathbf{v} es de una gran dimensionalidad y la distribución es muy compleja, sería deseable la estimación a través de un modelo de probabilidades flexible $P(\mathbf{v})$. Pero para este modelo necesitamos tanto su estimación $P(\mathbf{v}) \approx P_r(\mathbf{v})$, como la manera en la que descomponer $P(\mathbf{v})$, de tal forma que el codificador pueda realizar su trabajo con una variable cada vez (o en pequeños grupos).

Para lograr este propósito disponemos de los modelos gráficos. Estos permiten dar una descripción estructurada de $P(\mathbf{v})$. Pero las estructuras que mejores resultados obtienen son aquellas que cuentan con aristas dirigidas [HS86]. Por ende, las redes bayesianas son unos modelos que claramente se ajustan para la compresión de la información.

Así, este nuevo dominio de utilización de las redes bayesianas promovió nuestra investigación para el envejecimiento de muestras basado en la representación de la distribución estadística $P_r(\mathbf{v})$ a través de la aproximación de una red bayesiana. Este modelo será el que nos permitirá transformar un conjunto de datos en otro totalmente diferente (con un número menor de instancias pero la misma distribución probabilística), el cual nos permitirá crear los nuevos conjuntos de datos para el aprendizaje (para más detalles diríjase a la sección 5.2).

5.1.2 Generación de información a través del método de Montecarlo

Muchos de los problemas que nos encontramos en el mundo científico son muy complejos, por ejemplo, la estadística termodinámica considera que las propiedades térmicas de la materia son el resultado de la interacción entre un gran número de partículas elementales. De esta forma, la descripción determinista en términos de una ecuación de los movimientos de todas esas partículas no

5. Adaptando los modelos al proceso cambiante

tiene ningún sentido, sin embargo, se necesita una descripción probabilística. Esta descripción puede estar incluso intrínsecamente implícita en la naturaleza cuántica de los procesos básicos o, dado que el problema no puede ser caracterizado por completo, sólo podemos considerar una serie de grados de libertad mientras que el resto actúan de fondo causando ruido aleatorio. Este concepto de las distribuciones de las probabilidades muchas veces no puede ser computado analíticamente en una forma explícita. Así, nacen los métodos de Montecarlo con el objetivo de realizar estimaciones numéricas de una distribución de probabilidad haciendo uso de *números pseudo-aleatorios*¹.

No hay un consenso de cómo definir los métodos de Montecarlo. De este modo, por ejemplo, Ripley en [Rip87] define un modelo mucho más probabilístico a través de una simulación estocástica, siendo reservado el método de Montecarlo para la integración y las pruebas estadísticas. Por otra parte, Sawilowsky [Saw03] distingue entre la simulación, el método Montecarlo y la simulación Montecarlo. Para él, mientras el método Montecarlo es una técnica que puede utilizarse para solucionar problemas matemáticos o estadísticos, la simulación Montecarlo es una representación ficticia de la realidad que realiza una toma de muestras repetida para el descubrimiento de las propiedades de un fenómeno o su comportamiento. Por otra parte, Kalos y Whitlock [KW09] indican que estas distinciones no son tan fáciles de mantener en todas las ocasiones. De hecho, la emisión de la radiación de los átomos es un proceso estocástico por naturaleza. Este puede ser simulado directamente o descrito su comportamiento medio mediante ecuaciones estocásticas que para ser resueltas debe hacerse uso de los propios métodos de Montecarlo.

Los métodos de Montecarlo están muy unidos a la generación de números pseudo-aleatorios. En la mayoría de las aplicaciones se hace uso de este tipo de números permitiendo que los cálculos sean más simples y que las simulaciones pueden volverse a ejecutar. No obstante, hay ciertas ocasiones en las que la imprevisibilidad es vital [Dav92]. Pero dejando estos casos de lado, para obtener una buena calidad en las simulaciones, lo único que se necesita es que las secuencias de números sean lo suficientemente aleatorias. Normalmente, los generadores de números pseudo-aleatorios deben cumplir una serie de características [Bin79]. Principalmente, necesitan estar uniformemente distribuidos en el intervalo $[0, 1]$ y no estar correlacionados. No existen algoritmos que cumplan estas necesidades, por lo que adoptamos las características que describió Sawilowsky [Saw03] para alcanzar unas simulaciones de calidad:

¹Secuencia de números producidos por computador mediante la utilización de un procedimiento determinístico, partiendo desde una semilla en concreto.

5.1 Capacidades estadísticas para la adaptación de modelos

- El generador de números pseudo-aleatorios tiene una serie de características, como por ejemplo, la necesidad de que transcurra un largo periodo antes de volver a repetir la secuencia de números.
- Los números generados pasan las pruebas de aleatoriedad que intentan aproximar las características definidas en [Bin79].
- Existen muestras suficientes para garantizar resultados precisos.
- Se utiliza la técnica de muestreo óptima para la solución en concreto.
- El algoritmo utilizado es válido para aquello que estamos modelando.
- Los números simulan el fenómeno en cuestión.

En el caso que nos ocupa, nos interesa la capacidad de generar estimaciones numéricas para distribuciones de probabilidad conocidas. Así, para una distribución de probabilidad conocida p_i , en la que un estado (discreto) i ocurre con $1 \leq i \leq n$ y $\sum_{i=1}^n p_i = 1$, puede ser generada a través de métodos numéricos basados en la generación de números pseudo-aleatorios uniformemente distribuidos en el intervalo que va de 0 a la unidad. La forma de hacerlo es la siguiente, definiendo $P_i = \sum_{j=1}^i p_j$, elegimos un estado i si el número aleatorio ζ satisface $P_{i-1} \leq \zeta < P_i$, asumiendo que $P_0 = 0$. Con la generación de un gran número de ensayos (M) podemos aproximar la distribución p_i con una tasa de error del orden de $1/\sqrt{M}$.

En muchas ocasiones, existen métodos en los que el muestreo no es totalmente claro, por ejemplo, esto sucede con la probabilidad de Boltzmann (dentro del campo de la mecánica estadística):

$$P_{eq}(\mathbf{X}) = \left(\frac{1}{Z}\right) \exp\left\{\frac{-\mathcal{H}(\mathbf{X})}{k_b T}\right\} \quad (5.2)$$

donde k_b es la constante de Boltzmann¹, T es la temperatura absoluta y $Z = \sum_{\mathbf{X}} \exp\{-\mathcal{H}(\mathbf{X})/K_b T\}$ es la función de partición. Entonces, para realizar el muestreo para este tipo de situaciones no debemos seleccionar los puntos \mathbf{X} en el espacio de forma completamente aleatoria. La solución es un muestreo de puntos importantes, o en inglés “*important sampling*”, generando puntos \mathbf{X} , preferiblemente, de la parte importante del espacio en el que se representa dicha función.

¹Constante física que relaciona temperatura absoluta y energía. Su valor es $k \approx 1,3806504 \times 10^{-23}$ J/K = $1,3806504 \times 10^{-16}$ ergios/K.

5. Adaptando los modelos al proceso cambiante

Los métodos de Montecarlo también pueden ser utilizados para otros propósitos [Bin79]:

- **Ciencias físicas.** Algunas de sus aplicaciones son el cálculo de estadísticas en las teorías de partículas y polímeros [Bae09], y en astrofísica se han utilizado para modelar la evolución de las galaxias [MDM⁺82] y la transmisión de las microondas a través de una superficie rugosa planetaria [Gol79].
- **Ingeniería.** Dentro de este campo suelen ser utilizados para análisis de sensibilidades¹ y en el análisis probabilístico cuantitativo para el diseño de procesos. Más concretamente, para los análisis de riesgo y poluciones [IPRDNT02, IPDNCT04].
- **Biología computacional.** Se utilizan estos métodos para la generación de simulaciones por ordenador con el fin de estudiar las proteínas [OGLC09] y las membranas [MS93] entre otros.
- **Estadística aplicada.** Normalmente son utilizados con dos propósitos: (i) para la comparación de estadísticas utilizando muestras pequeñas sobre datos realistas [Saw03] y (ii) para proporcionar pruebas de hipótesis que son más eficientes que las pruebas exactas, tales como las pruebas de permutación.
- **Juegos.** Se incorporan los métodos de Montecarlo dentro de la teoría de juegos con el fin de optimizar los algoritmos de búsqueda y las simulaciones [SV10].
- **Finanzas y negocios.** Habitualmente son utilizados para simular el valor de las empresas, evaluar las inversiones o para evaluar los productos financieros derivados.
- **Telecomunicaciones.** En los procesos de diseño de redes, estos métodos nos ayudan a simular los usuarios que las están utilizando.
- **Soluciones matemáticas.** Los métodos de Montecarlo suelen ser utilizados para solucionar problemas como la integración, la simulación y la optimización.

¹Es el estudio de cómo la incertidumbre en la salida de un modelo (numérico o de otra forma) se puede distribuir a diferentes fuentes de incertidumbre en la entrada del modelo.

5.2 Adaptación de los modelos que forman un MPC

Una vez presentados los métodos que hemos identificado como útiles para la generación del proceso de adaptación de los modelos capaces de predecir los estados estacionarios, en esta sección pasamos a describir el método propuesto. Para ello, ofrecemos una visión general del proceso y de cada una de sus etapas, así como la forma de introducirlo en un sistema MPC con el fin de conseguir que el propio sistema sea autoadaptativo.

5.2.1 Compresión bayesiana basada en métodos de Montecarlo

La compresión bayesiana basada en métodos de Montecarlo trabaja sobre los procesos de aprendizaje de los modelos que predicen el estado estacionario $t + 1$. La metodología de predicción seleccionada en el capítulo 4 nos hace tener que trabajar sobre los conjuntos de datos que generan los modelos. La razón es que las selecciones iniciales dadas en la investigación no nos llevaron a la utilización de modelos evolutivos capaces de irse adaptando por sí solos al proceso. De esta forma, la solución es la generación de nuevos modelos que sustituirán los antiguos y, por consiguiente, trabajar sobre los datos que utilizamos para su generación.

El método desarrollado en esta tesis doctoral hace uso de los métodos estadísticos presentados en la sección 5.1. Para poder conseguir el objetivo de la adaptación, este método trabaja en tres aspectos fundamentales: (i) en el descubrimiento de la distribución de probabilidad existente en los datos, (ii) la simulación de unos nuevos datos de aprendizaje acordes a la distribución de probabilidad de la información y (iii) la reducción del conjunto de datos original para envejecer las muestras y que las evidencias de nueva incorporación sean más representativas.

Para el descubrimiento de la distribución de probabilidad, y tomando las ideas de [Fre97] y [DM99], realizamos una aproximación del comportamiento de la planta a través de un modelo en forma de grafo con aristas dirigidas. En nuestro caso, y tal como se hacía en las investigaciones ya citadas, emplearemos una red bayesiana. Este tipo de red almacena la estructura de relaciones entre las variables, así como la fortaleza de esas relaciones y la distribución probabilística de que se den (para más información diríjase a la sección 4.1.1).

5. Adaptando los modelos al proceso cambiante

De esta forma, y haciendo uso de la propia naturaleza de la red, seremos capaces de aproximar la distribución probabilística inherente en los datos.

En cuanto a la simulación de los datos, los métodos de Montecarlo y los números pseudo-aleatorios nos permiten generar puntos en el espacio representado por la distribución de probabilidad [Bin79]. En nuestro caso, esta viene dada por lo que aprendió la red bayesiana. Mediante los métodos de Montecarlo iremos generando valores para cada una de las variables que compone la nueva evidencia y que será utilizada en la adaptación.

Y por último, y muy unido a la simulación de los datos, la reducción del conjunto de datos para conseguir envejecer las muestras viene dada por la cantidad de nuevas evidencias que van a generarse. Así, a mayor grado de compresión, mayor grado de envejecimiento obtendremos. Sin embargo, deberemos tener cuidado ya que una compresión elevada podría hacernos perder información valiosa.

Estos tres aspectos se desglosan en una serie de operaciones que hay que realizar para la generación del modelo. El proceso consta de una serie de operaciones de inicialización (esta inicialización es la base de todo el método, con lo que los resultados obtenidos en esta etapa influirán a los procesos ulteriores) y de un ciclo iterativo determinado por los parámetros de configuración iniciales (i.e., el porcentaje de compresión que va a tener el nuevo conjunto de datos). Más concretamente, las etapas que hay que atravesar para generar ese conjunto de datos son las siguientes (ver Figura 5.1):

- **Aprendizaje de la red bayesiana.** En el primero de los pasos descubriremos la distribución probabilística de los datos. Para ello, aprenderemos una red bayesiana con el fin de tener acceso a su estructura de relaciones y a sus tablas de probabilidades.
- **Ordenación de los nodos para la generación de la información.** Dado que las probabilidades aprendidas por la red bayesiana son condicionales debemos ordenar los nodos para que la selección que hagamos en la tabla de probabilidades esté adecuada a las condiciones precedentes ya escogidas.
- **Para cada evidencia a generar y cada nodo de la red.**
 - *Selección de la distribución de probabilidades y su representación.* En esta etapa determinaremos que parte de la distribución de probabilidades, en nuestro caso la tabla de probabilidades del nodo

5.2 Adaptación de los modelos que forman un MPC

con el que estamos trabajando, es relevante dadas las condiciones previas seleccionadas. Posteriormente, se realizará una representación espacial de esa distribución para permitirnos hacer la selección final.

- *Generación del nuevo estado.* Mediante la generación de un número pseudo-aleatorio seleccionaremos un punto en el espacio de representación que estará asociado al estado que tendrá el nodo sobre el que estamos trabajando.

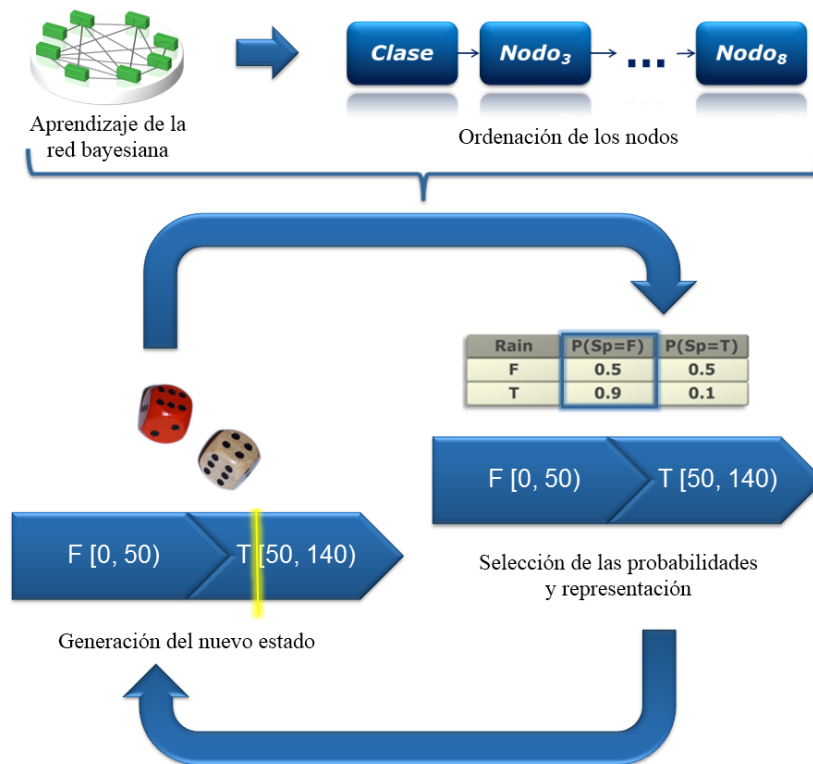


Figura 5.1: Procedimiento general para la generación de un conjunto de datos reducido a través de la compresión bayesiana basada en métodos de Montecarlo.

5.2.1.1 Aprendizaje de la red bayesiana

Este primer paso es fundamental, ya que en él obtendremos la aproximación del modelo de la planta. En otras palabras, obtendremos la distribución

5. Adaptando los modelos al proceso cambiante

probabilística inherente en los datos. Para ello, utilizaremos una red bayesiana, ya que entre las características de la red se encuentra la aproximación probabilística del modelo aprendido a través de sus datos de entrenamiento. Además, estos modelos permiten representar la forma en la que las variables que lo componen se influyen entre ellas.

Como bien pudimos observar en la sección 4.1.1, una red bayesiana cuenta con dos procesos de aprendizaje. El primero de ellos es el que se encarga de aprender la estructura de la misma (i.e., las relaciones) y el segundo de ellos es el que obtiene la distribución probabilística (i.e., la fortaleza de los enlaces) para cada uno de los nodos. En particular, el proceso de aprendizaje de la estructura puede realizarse con diferentes métodos como el algoritmo de escalar colinas [RN03], el K2 [FGG97] o a través de una red bayesiana ingenua aumentada mediante un árbol (en inglés “*Tree Augmented Naïve*”) [FGG97]. Cada uno de ellos obtiene estructuras diferentes, con lo que la representación probabilística que se aprenderá en un paso posterior será diferente debido a cómo se haya determinado las influencias de las variables.

Para la reducción de nuestro conjunto de datos hemos seleccionado la última de las tres, “*Tree Augmented Naïve*”. La razón de su selección es muy clara. Por una parte, a la hora de hacer los experimentos de predicción para ambos defectos, este tipo de red es la que mejores resultados ha obtenido. Así, consiguieron una mejor precisión en sus predicciones y una tasa de error muy similar entre todas ellas, lo que nos hace pensar que se ajusta mejor al proceso de fundición. Y por otra parte, los algoritmos K2 y escalar colinas obtienen sus resultados a través de máximos locales y, normalmente, son utilizadas para hacer pruebas previas de la clasificación bayesiana [FGG97], lo que no nos asegura que sean los algoritmos óptimos para la representación del proceso.

5.2.1.2 Ordenación de los nodos para la generación de información

En los siguientes pasos se van a ir generando nuevas instancias basadas en la distribución probabilística de la red bayesiana. Pero para conseguir una instancia completa deberemos generar los valores para todos y cada uno de los nodos (variables) que componen la red. En el proceso de generación de sus estados, las probabilidades cambian dependiendo de las influencias que tienen sobre el nodo de trabajo el resto de los nodos. De esta forma, este algoritmo de ordenación busca identificar la secuencia de nodos sobre la que trabajar.

En particular, el algoritmo de ordenación obtiene una secuencia de nodos $\mathfrak{N} = \{cl, n_i, \dots, n_j\}$ tal que el primero de ellos, cl , siempre será la clase sobre

5.2 Adaptación de los modelos que forman un MPC

la que precedimos y vendrá seguida por el resto de los nodos, de tal forma que tomando un punto cualquiera de la secuencia, x , para el nodo n_x , en caso de tener padres, todos ellos se encontrarán en la subsecuencia de nodos $[cl, n_{x-1}]$; y todos sus hijos se encontrarán en la secuencia de nodos $[n_{x+1}, n_j]$. De esta forma, a la hora de generar el estado para el nodo n_{x+1} ya habrán sido asignados los estados de todos sus padres y podremos ir directamente a la sección relevante de su tabla de probabilidades.

El algoritmo seguido para la ordenación es el siguiente:

Algoritmo 2: Algoritmo para la ordenación de los nodos.

Input: Lista de nodos de la red bayesiana, \mathbf{N}

Output: Lista de nodos ordenados, \mathfrak{N}

$\mathfrak{N} \leftarrow \emptyset$

colaNodos $\leftarrow \emptyset$

foreach $nodo \in \mathbf{N}$ **do**

if $Padres(nodo) = \emptyset$ **then**

 | Anexar(\mathfrak{N} , nodo)

end

else

 | Encolar(colaNodo, nodo)

end

end

while $Tamaño(colaNodo) > 0$ **do**

 nodoTrabajo \leftarrow Desencolar(colaNodo)

if $Padres(nodoTrabajo) \in \mathfrak{N}$ **then**

 | Anexar(\mathfrak{N} , nodoTrabajo)

end

else

 | Encolar(colaNodo, nodoTrabajo)

end

end

return \mathfrak{N}

5.2.1.3 Selección de la distribución de probabilidades y su representación

Una vez ordenados los nodos de la red bayesiana comenzaremos a recorrerlos para generar los estados que tendrán cada uno de ellos. No obstante, antes

5. Adaptando los modelos al proceso cambiante

de ello, debemos conocer las probabilidades actuales para el nodo con el que trabajamos y realizar una representación que utilizaremos en la selección del estado.

De esa forma, en primer lugar, tendremos en cuenta los estados que han sido seleccionados en los padres del actual nodo de trabajo. Debemos tener presente que dependiendo de los estados alcanzados por los padres, las probabilidades del estado del hijo varían. Además, las tablas de probabilidades de cada nodo disponen de dos entradas posibles, (i) la secuencia de estados de los padres, permitiendo seleccionar la columna de probabilidades óptima y (ii) el estado del actual nodo, permitiendo seleccionar la probabilidad concreta de ese estado.

Haciendo uso de las dos entradas, proporcionaremos a la tabla de probabilidades la secuencia de estados de los padres y cada uno de los estados de los nodos. Al realizar esta operación recogeremos una lista de bi-tuplas de la forma (i, p_i) , donde i hace referencia al estado del nodo y p_i a la probabilidad asociada a ese estado y que cumple $0 \leq p_i \leq 1$. Esta lista es la que marca la distribución estadística que debemos seguir dadas las probabilidades condicionales almacenadas en la red.

Para poder realizar la posterior elección del estado para el nodo actual, obtendremos todas las bi-tuplas asociadas al estado y les aplicaremos una función de redondeo, obteniendo valores que únicamente cuenten con dos decimales. Posteriormente, convertiremos esos valores fraccionarios en valores naturales, aplicando una multiplicación que eliminará sus decimales. Con estos nuevos valores, la representación de las probabilidades será realizada de forma lineal desde el valor 0 hasta el valor total de la suma de los números naturales asociados a cada estado. Este segmento quedará separado en diferentes bloques de desigual tamaño que marcan las secciones asociadas a cada uno de los estados posibles para el nodo de trabajo.

5.2.1.4 Generación del nuevo estado

El último de los procesos que llevamos a cabo con el fin de generar la nueva instancia es aquel que se encarga de determinar el estado para el nodo actual. En este momento, nos encontramos con una representación lineal de los estados posibles para el nodo. Además, ese conjunto lineal de valores se encuentra dividido en secciones ponderadas según la distribución de probabilidad que nos ofreció la red bayesiana, y más concretamente, la selección realizada en la tabla de probabilidades del nodo en cuestión.

5.2 Adaptación de los modelos que forman un MPC

A partir de ahora vamos a aplicar la teoría de muestreo explicada en los métodos de Montecarlo. El paso anterior nos ha permitido seleccionar la parte importante del espacio, con lo que el muestreo que vamos a realizar es del tipo “*important sampling*” (ver sección 5.1.2). De la misma forma, estos métodos permiten la generación de nueva información basada en una distribución de probabilidad conocida [Bin79], y esta ha sido recogida en el mismo momento que identificamos el espacio importante para el muestreo. Por consiguiente, a través de la generación de un número pseudo-aleatorio, que cumpla las condiciones establecidas por los métodos de Montecarlo [Saw03] y que se encuentre entre los valores máximo y mínimo de la representación realizada, determinaremos el punto de selección. Específicamente, ese punto queda dentro de una de las secciones realizadas en la representación lineal y que, por ende, lleva asociado un estado, el estado que elegiremos para el nodo con el que trabajamos.

5.2.2 Consiguiendo un sistema MPC autoadaptativo

Una vez presentado el modelo desarrollado, podemos comprobar que el proceso de funcionamiento está directamente relacionado con las etapas de aprendizaje de los diferentes clasificadores que se encargan de las predicciones del sistema. En el capítulo 4 presentamos diferentes métodos de aprendizaje supervisado con el fin de conseguir el objetivo de la predicción, del mismo modo, que realizamos una combinación de estos métodos para la solución de diversos problemas. Sin embargo, los modelos seleccionados no cuentan con la capacidad de evolución. Por consiguiente, la solución adoptada para la adaptación de los modelos es la sustitución de los antiguos por unos nuevos que se ajusten mucho más al proceso.

Las etapas de aprendizaje de los nuevos modelos suelen ser tediosas, por lo que optamos por una solución desconectada. En otras palabras, el sistema MPC cuenta con dos procesos concurrentes, el primero, que se encarga de realizar la gestión del proceso de producción, y el segundo, que se encarga de realizar los aprendizajes de los nuevos modelos adaptados al proceso. El lanzamiento de este segundo proceso dependerá de la parametrización inicial con la que cuente el sistema. En este caso deberíamos indicar: (i) la periodicidad de adaptación del proceso y (ii) el grado de compresión que deben alcanzar los conjuntos de datos utilizados.

Por lo tanto, configurado un sistema MPC con un ciclo iterativo de adaptación t y un grado de compresión p , cada t se lanzará el proceso de adaptación.

5. Adaptando los modelos al proceso cambiante

Para ello, partiremos del conjunto de datos de entrenamiento actual \mathfrak{D}_1 compuesto por m muestras de la forma $m_i = \{n_1, n_2, \dots, n_a, cl_j\}$, siendo a el número de diferentes variables de control y cl_j el estado j de la clase sobre la que hay que realizar la clasificación. Por otra parte, obtendrá un segundo conjunto de datos, \mathfrak{D}_2 , con n nuevas evidencias del proceso productivo. Posteriormente, se realizará la aplicación de la función de compresión a \mathfrak{D}_1 dado el porcentaje de compresión p , obteniendo como resultado un nuevo conjunto de datos \mathfrak{D}_3 que cuenta con $0,01pm$ evidencias. Una vez obtenido el resultado, se aplicará una operación de unión de conjuntos ($\mathfrak{D}_3 \cup \mathfrak{D}_2$) resultando en el conjunto de datos que será utilizado para el aprendizaje de todos y cada uno de los modelos que componen el sistema, \mathfrak{D}_a . Una vez los modelos han sido construidos, el último de los pasos es la sustitución en el hilo de monitorización principal, pasando este hilo de adaptación a esperar un nuevo ciclo t para así, repetir el proceso de adaptación.

Este procedimiento puede ser arrancado de forma automática, haciendo que los modelos se mantengan actualizados a lo largo del ciclo de vida completo del sistema MPC. La ventaja proporcionada por la utilización de métodos predictivos basados en un aprendizaje supervisado hace que seamos capaces de no necesitar la ayuda de terceros (i.e., ingenieros, matemáticos o expertos en el dominio) en el proceso de adaptación, logrando el grado de automatismo solicitado por los usuarios de este tipo de sistemas. Este avance en el estado de la técnica supone una solución a otra de las limitaciones detalladas en la sección 2.5, evolucionando los sistemas MPC hacia la nueva generación detallada en la sección 2.6.

5.3 Validación empírica

En esta sección quedan recogidos los experimentos realizados con el fin de validar el método de compresión bayesiana basado en métodos de Montecarlo y la adaptación de los modelos incorporados en el sistema MPC. De esta forma, se ha llevado a cabo un primer experimento con el fin de demostrar que los modelos reducidos mantienen la misma distribución probabilística que los datos. Y posteriormente, se ha realizado un segundo experimento en el que se ha obtenido un conjunto de datos acumulado, simulando el proceso de adaptación a lo largo de dos años en un planta de producción, comprobando los resultados obtenidos por los modelos del sistema MPC.

Para ambos experimentos, hemos testado los modelos generados a través

de los métodos habituales y, posteriormente, hemos enfrentado los modelos aprendidos a través de esos nuevos conjuntos reducidos y acumulados al conjunto de datos original.

Nótese que con estos experimentos no buscamos mejorar los clasificadores, sino que pretendemos demostrar que después de realizar la adaptación del sistema nos encontramos aproximadamente en el mismo lugar que realizando el proceso con los datos al completo.

5.3.1 Metodología general

Dado que el objetivo es probar el método desarrollado aplicándolo a los mismos modelos que se detallaron en el capítulo 4, para la realización de estos experimentos vamos a partir del mismo conjunto de datos. Como ya explicamos anteriormente, los datos son relativos a una fundición que obtiene piezas dedicadas a la industria de la automoción (alrededor de unas 45.000 toneladas anuales), con lo que estas piezas deben pasar por unos rigurosos controles de seguridad y debe garantizarse, ante todo, un nivel de calidad que no ponga en peligro las vidas de aquellas personas que las utilicen (para más información diríjase al capítulo 3).

Al igual que en el capítulo 4, el objetivo de los modelos será la predicción de los estados estacionarios $t + 1$ para los defectos de microrrechupe y la característica de la carga máxima (una de las propiedades mecánicas que pueden ser englobadas dentro de la *fuerza*). Con motivo de no extendernos, indicamos al lector que en el capítulo 3, y más concretamente en la sección 3.2, podrá ampliar esta información.

Bajo este contexto, con la investigación aquí detallada queremos llegar a responder las siguientes preguntas:

¿El método desarrollado cumple con los principios estadísticos que rigen a la compresión bayesiana y los métodos de Montecarlo?

¿Existe algún grado o porcentaje de compresión que sea el óptimo para realizar las reducciones del conjunto de datos original?

¿El comportamiento de los modelos generados se aproxima al de los modelos originales probados en el capítulo 4?

5. Adaptando los modelos al proceso cambiante

¿Los modelos generados son capaces de comportarse de manera similar ante el conjunto de datos original a pesar de haber realizado su aprendizaje con unos conjuntos de datos adaptados a través de nuestro método?

¿La acumulación de conjuntos de datos reducidos permite envejecer las muestras y obtener resultados similares a los que podríamos adquirir al realizar el proceso de aprendizaje con el conjunto de datos al completo?

Como se puede observar, las cuatro primeras preguntas están tanto dirigidas al proceso de investigación de los modelos creados con conjuntos reducidos como acumulados, mientras que la última de ellas se centra en la segunda de las investigaciones que simula el entorno real de adaptación. Específicamente, para descubrir las respuestas a estas preguntas, hemos utilizado la metodología de investigación que detallamos a continuación.

- **Adquisición de datos.** Compartiendo el proceso ya desarrollado para los experimentos del capítulo 4, realizamos el proceso de adquisición de información en la planta seleccionada. Estos datos serán la entrada para todo el proceso determinando, así, la validez del procedimiento de compresión bayesiana basado en métodos de Montecarlo.
- **Adecuación de los conjuntos de datos.** Los datos recogidos de la fundición necesitan ser adaptados para el proceso que vamos a realizar. En primer lugar, para los experimentos en los que utilizaremos conjuntos de datos reducidos, hemos generado los conjuntos de datos necesarios para cubrir todos los porcentajes de compresión posibles. Para ello, aplicaremos el método de compresión bayesiana a los conjuntos de datos originales, para ambos defectos, con un rango de porcentajes de compresión [1, 99). Nótese que no utilizamos el grado de compresión 99, es decir, utilizar únicamente un 1 % del tamaño de los datos originales ya que se da la situación de que no conseguimos las evidencias suficientes que permitan realizar una validación cruzada para el valor de $k = 10$.

En segundo lugar, para el experimento de conjuntos de datos acumulados, hemos dividido los conjuntos de datos originales en cuatro nuevos ficheros. La división la hemos realizado considerando que el tiempo de adaptación no debe llegar a un año (es demasiado tiempo sin realizar una actualización y puede que el modelo lleve mucho tiempo trabajando

con un rendimiento bajo) ni a nivel mensual (podríamos alcanzar un sobreajustamiento de los modelos). Por eso, hemos seleccionado la opción de actualización para cada medio año, que en el caso de la fundición con la que trabajamos es más que adecuado debido a la variabilidad de su proceso. Una vez obtenidos estos cuatro ficheros, hemos tomado el primero de ellos y le hemos aplicado todos los posibles porcentajes de compresión, de 1 a 98; hemos añadido el segundo de los ficheros para, posteriormente, aplicar de nuevo todos los porcentajes de compresión posibles; en tercer lugar, hemos repetido el proceso con el tercero de los ficheros y, finalmente, hemos agregado el último de ellos. Los conjuntos de datos acumulados resultantes son los que hemos utilizado para el resto de los experimentos.

- **Validación cruzada.** El método de la validación cruzada ha sido utilizado para los experimentos en los que nos hemos centrado en realizar las pruebas de los modelos. Más concretamente, y como ya explicamos en la sección 4.3.1, el método de validación cruzada [Bis95] divide k veces el conjunto de datos en k partes, seleccionando $k - 1$ para el aprendizaje y una para las pruebas. Hemos seleccionado un valor de $k = 10$. En el caso en el que probamos los modelos frente al fichero original, hemos conducido el proceso de aprendizaje empleando los ficheros completos para, finalmente, enfrentarlos a los conjuntos de datos originales.
- **Aprendizaje del modelo.** Para ambos experimentos, hemos aprendido los modelos utilizando los conjuntos de datos generados anteriormente y aplicando diferentes parámetros de configuración a cada uno de ellos. Concretamente, hemos utilizado los mismos modelos que en los anteriores experimentos: *redes bayesianas* (utilizando los algoritmos K2, escalar colinas y “*Tree Augmented Naïve*” para el aprendizaje estructural y también un clasificador bayesiano ingenuo), *K vecinos más próximos* (con valores de K que se encuentran entre 1 y 5), *máquinas de soporte vectorial* (utilizando diferentes funciones de núcleo: polinomial, polinomial normalizado, “*Radial Basis Function*” (RBF) y el basado en la función Pearson VII), *árboles de decisión* (recurriendo a los bosques aleatorios configurados con diferentes cantidades de árboles, n , $n = 50$, $n = 100$, $n = 150$, $n = 200$, $n = 250$, $n = 300$ y $n = 350$; y un árbol de decisión construido con el algoritmo C4.5) y *redes neuronales artificiales* (específicamente una red de múltiples capas de perceptrones aprendida con el algoritmo de retropropagación).

5. Adaptando los modelos al proceso cambiante

- **Pruebas de los modelos.** Hemos evaluado cada uno de los métodos generados a través del porcentaje de precisión que han alcanzado en sus predicciones y el área formado por la curva ROC (del inglés “*Receiver Operating Characteristic*”). Este cálculo establece la relación entre los falsos negativos y los falsos positivos [SKM09]. Hemos decidido utilizar un segundo valor para las comparaciones porque, como se dice en [PF97], la precisión suele ser un métrica pobre para medir el comportamiento de los clasificadores.

5.3.2 Predicción sobre conjuntos de datos reducidos

Una vez aplicada la metodología mencionada en la sección 4.3.1, hemos obtenido los resultados que se muestran a continuación. Para facilitar la legibilidad de los mismos, hemos dividido los resultados según el objetivo de clasificación seleccionado para cada predicción.

5.3.2.1 Microrrechupes

Los métodos de aprendizaje automático han sido utilizados con el fin de mejorar alguna parte del proceso productivo. Sin embargo, y a nuestro entender, no conocemos de ninguna solución que permita realizar la adaptación de esos modelos. Conocemos las diversas técnicas presentes en la literatura pero no hay una aplicación directa al proceso.

Así, a continuación vamos a pasar a presentar los resultados obtenidos en el proceso de investigación. En lo que al experimento se refiere, éste va a estar dividido en dos partes. La primera realiza la prueba de los modelos aprendidos sobre un conjunto de datos al que únicamente se le ha aplicado un único proceso de compresión. Su finalidad es la de permitirnos valorar si los modelos que generamos cumplen con las bases estadísticas de partida y, a su vez, comprobar si las tasas de aciertos y errores de estos modelos son aceptables para que los posteriores procesos de acumulación de conjuntos sean lo más precisos posible. La segunda comprueba esos modelos con el conjunto de datos original. Este experimento nos permitirá comprobar cómo de generalizado se encuentra el conjunto de datos reducido y cómo, con menos instancias que el conjunto original, es capaz de ajustarse al proceso.

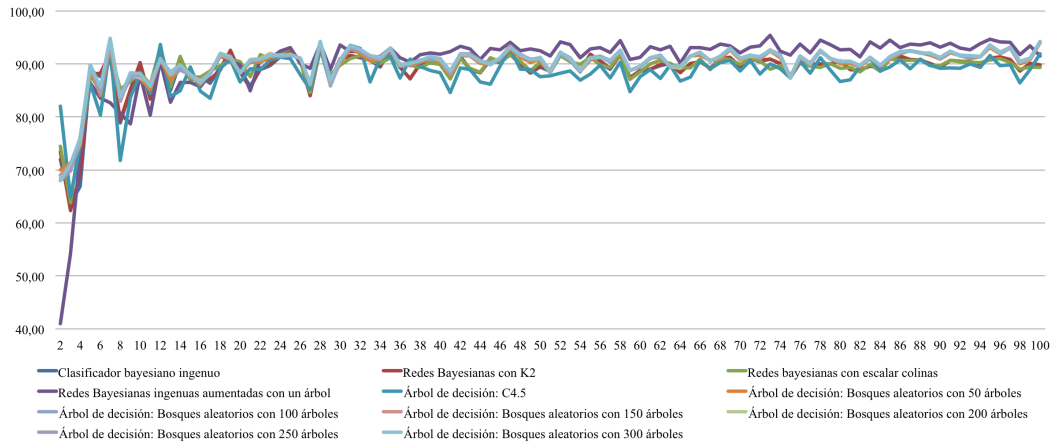
Entonces, comenzando por el primero de ellos, las Figuras 5.2a y 5.2b hacen referencia a la precisión alcanzada en los procesos de predicción del defecto de microrrechupes y, las Figuras 5.3a y 5.3b hacen referencia a las tasas de error

alcanzadas por los clasificadores medidas como el área bajo la curva ROC. Como se puede observar, las gráficas indican el valor que ha sido logrado por cada porcentaje de compresión aplicado. Nótese, que a mayor número en el eje de abscisas, representamos un número menor de compresión. En otras palabras, la numeración que se muestra en el eje de abscisas hace referencia al porcentaje de instancias utilizado para la generación de ese conjunto de datos.

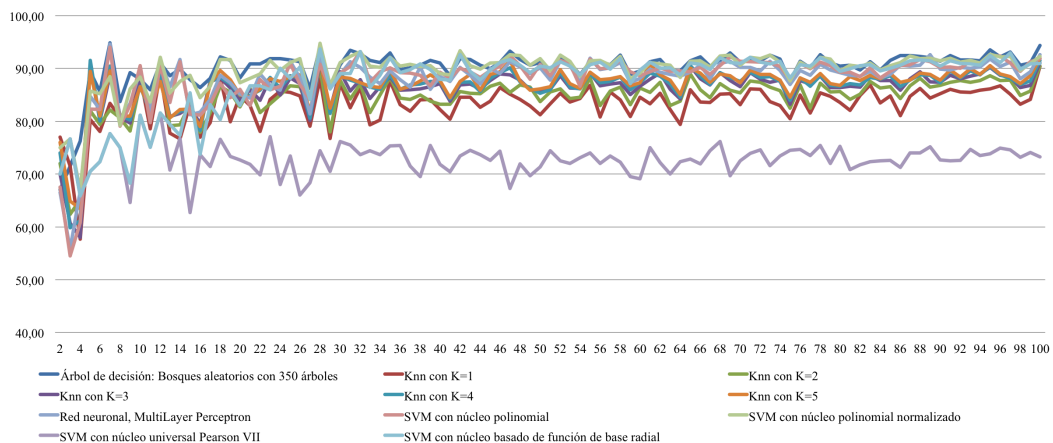
Por lo tanto, en lo que representa los porcentajes de acierto podemos comprobar que el comportamiento de todos los clasificadores es prácticamente el mismo, lo único que cada uno desplazado en el eje de ordenadas debido a las tasas de precisión alcanzadas. Entonces, esta situación nos hace pensar que el método de compresión bayesiana basado en métodos de Montecarlo acaba afectando a todos los modelos de una forma similar. La razón de notar este tipo de interacción entre el método desarrollado y los resultados alcanzados en todos los modelos es que la curva que representa el porcentaje de precisión de cada clasificador encaja perfectamente en la de otro de ellos. Es decir, los picos y los valles los encontramos en los mismos puntos para todos los clasificadores.

También, a la hora de comprobar el comportamiento de los clasificadores podemos observar que hay una serie de perturbaciones en la curva de resultados. Así, nos encontramos posiciones más altas seguidas de posiciones más bajas a lo largo de todos los porcentajes de aciertos. La explicación de las perturbaciones es sencilla. En primer lugar son producidas por el no determinismo de los métodos de Montecarlo. Este no determinismo significa que no podemos determinar que el conjunto de datos generado para un porcentaje de compresión es el óptimo, ya que los puntos generados en el espacio, a pesar de ser muestreados en base a una distribución de probabilidad, pueden no ser aquellos que representen mejor al conjunto inicial. Por consiguiente, para ciertos grados de compresión los conjuntos de datos generados son mejores que para otros. Además, en caso de volver a lanzar el proceso de compresión podríamos generar un nuevo conjunto de datos que ahora sí que es más preciso para los puntos que en este caso fueron bajos y menos preciso para los que aquí fueron altos. En segundo lugar, las perturbaciones pueden estar creadas por una baja calidad de los datos que gracias al posible ruido con el que cuentan (idea que ya fue citada en la sección 4.4) perturben el devenir de nuestro método de compresión, al no generar una distribución estadística que se ajuste perfectamente al modelo, i.e., la generación de la red bayesiana no es todo lo correcta y óptima posible. De esta forma, y bajo la luz de estos hechos, es más interesante la tendencia general llevada a cabo por el clasificador que la clasificación puntual en un porcentaje de compresión x .

5. Adaptando los modelos al proceso cambiante



(a) La figura muestra el primer grupo de clasificadores basado en un conjunto de datos reducido prediciendo microrrechupes. A excepción de los niveles en los que hemos generado un número reducido de instancias, los clasificadores mantienen un comportamiento similar al inicial.



(b) Mostramos el segundo conjunto de clasificadores que utilizan un conjunto de datos reducido para la predicción de los microrrechupes. Las tendencias de los clasificadores son las mismas que vimos en el primero de los grupos.

Figura 5.2: Resultados obtenidos durante la comprobación de los modelos generados a través de un conjunto de datos reducido con el fin de la detección de posibles problemas de microrrechupes. Se muestran los datos en términos de precisión para cada uno de los porcentajes de compresión. Aunque los resultados pueden fluctuar en alrededor de 10 unidades, se comprueba que la tendencia en la clasificación es similar a la inicial y el proceso de compresión les afecta a todos por igual.

Específicamente, podemos comprobar como a pesar de las fluctuaciones, los resultados obtenidos para cada clasificador se mantienen relativamente constantes a lo largo de los porcentajes de compresión. Más concretamente, mantienen una tendencia similar desde un porcentaje de generación de instancias del 16 % (i.e., una compresión del 84 %) hasta un porcentaje de generación del 99 % (una compresión del 1 %). Sin embargo, para altos niveles de compresión, la cantidad de datos generada es tan pequeña que no es capaz de mantener la distribución probabilística inicial. Por ende, los resultados que se obtienen son mucho peores.

Los clasificadores de la Figura 5.2a se encuentran obteniendo unos resultados que rondan en torno al 90 % (las fluctuaciones se mueven desde el 85 a el 95 %), con lo que se mantiene aproximadamente la distribución probabilística del modelo inherente del proceso y la precisión que se alcanza se aproxima a la lograda al utilizar el conjunto inicial al completo. En lo que respecta a los clasificadores del segundo de los grupos (ver Figura 5.2b), estos obtienen unos resultados sensiblemente más bajos. La razón es que esos clasificadores tampoco fueron capaces de aportar mejores resultados anteriormente. Sin lugar a duda, esta gráfica nos ilustra con el clasificador que peor rendimiento ha obtenido, las máquinas de soporte vectorial construidas con una función de núcleo Pearson VII. Sin embargo, su baja tasa de acierto no se debe a que el método de compresión lo haya perjudicado, sino que se debe a que el modelo no era capaz de adaptarse al proceso ya en las predicciones con el conjunto de datos al completo. Otro punto a destacar, es que clasificadores que obtuvieron peor precisión que este método al aplicar el método de compresión, se han visto mejorados. De esta forma, nos hace pensar que al realizar el proceso de compresión eliminamos cierto ruido, ocurrencias estadísticamente anecdóticas, que hace que los modelos puedan adaptarse mucho mejor.

En lo que respecta a los errores, podemos observar en las Figuras 5.3a y 5.3b que las tendencias visualizadas en los porcentajes de acierto se vuelven a repetir. De esta forma podemos indicar que estos clasificadores, a parte de tener un interesante porcentaje de aciertos, llevan a cabo una buena gestión de los errores.

La figura 5.3a nos ilustra con los clasificadores del primer grupo. Estos, que fueron los que mejor precisión alcanzaron, obtienen unas tasas muy cercanas al valor óptimo a lo largo de toda la secuencia de porcentajes de compresión posible. Particularmente, el clasificador que mejor trata los errores es la red bayesiana ingenua ampliada con un árbol. No debemos dejar de lado que este modelo es el que ha sido utilizado para la representación de la distribu-

5. Adaptando los modelos al proceso cambiante

ción probabilística, con lo que su ajuste será mucho mayor. No obstante, los bosques aleatorios siguen muy de cerca a este modelo. De tal forma, que los clasificadores que mejores tasas de acierto adquirieron en este experimento (siguiendo a las a las redes bayesianas aprendidas con TAN) y en el experimento descrito en 4.3.2.1 (i.e., los bosques aleatorios) vuelven a ser los que mejor se adaptan al proceso. Particularmente, en este caso, el clasificador que peores resultados ha obtenido es el árbol C4.5. Este clasificador se desvía bastante de la tendencia normal del grupo, consiguiendo superar únicamente la barrera de los 0,90 en contadas ocasiones. Específicamente, a pesar de mantener unos porcentajes de precisión similares a sus compañeros, este clasificador también fue el que obtuvo una tendencia de acierto más baja.

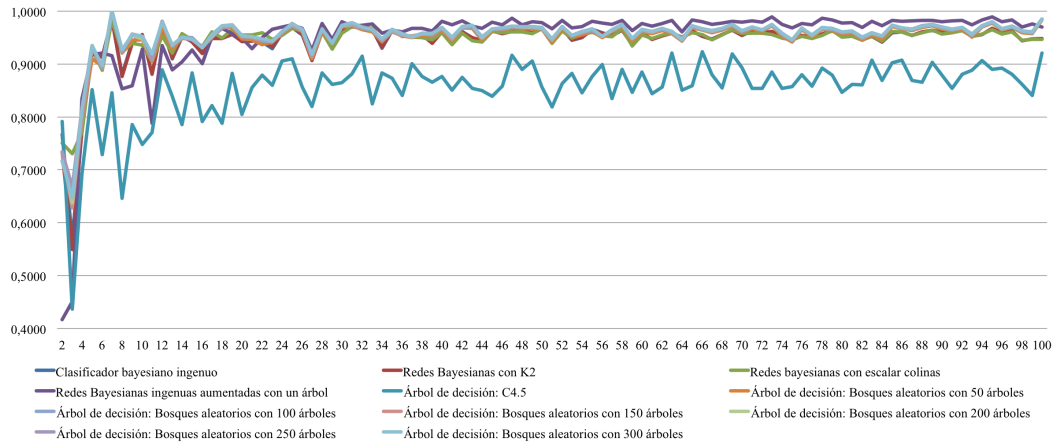
Al igual que sucedía en el porcentaje de precisión, a lo largo de la curva generada para los errores encontramos una sección en la que el rendimiento de los clasificadores desciende. Esa sección es la que hace referencia a unos porcentajes muy altos de compresión en la que se pierda la capacidad de representación del proceso productivo. A pesar de este hecho, el rango en el que sucede es el mismo que ya notamos en los porcentajes de acierto (entre el 2 y el 16 % de instancias generadas). Pero este descenso no se hace muy pronunciado hasta alcanzar un valor de instancias creadas del 8 %.

El segundo de los grupos (ver Figura 5.3b), y al igual que sucedió en los porcentajes de precisión, obtiene un comportamiento mucho más heterogéneo. Pero esa heterogeneidad se basa en el grado de gestión de errores, ya que la tendencia general de todos los clasificadores es la misma, alcanzando máximos y mínimos en los mismos lugares. Las tendencias en la gestión de los errores para este segundo grupo de clasificadores va desde los 0,85 a los 0,95, valores que se encuentran muy cercanos al óptimo.

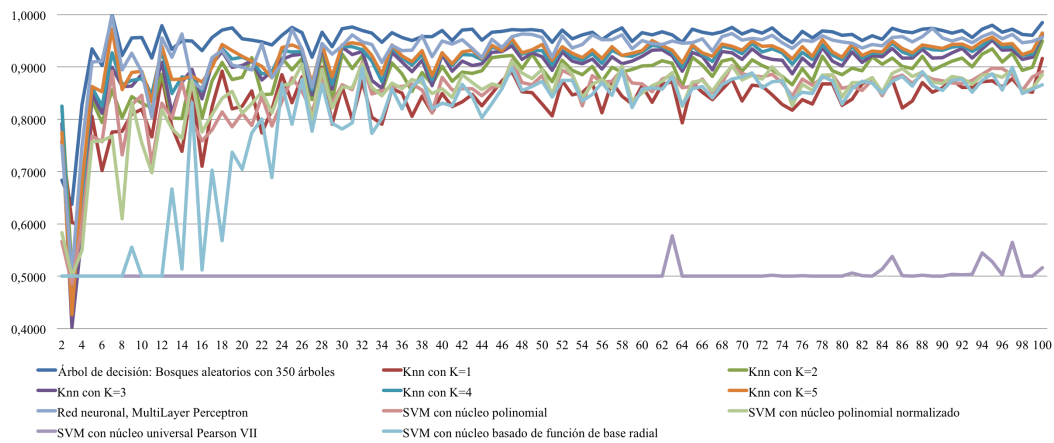
Para este grupo de clasificadores también encontramos un rango de compresión que decrementa el rendimiento de los modelos en la gestión de los errores. Este rango coincide con el que ya ha sido identificado en los porcentajes de acierto y la gestión de errores en el primero de los grupos.

A pesar de que todo el conjunto de clasificadores (tanto los de la Figura 5.3a como los de la Figura 5.3b) se comportan de forma similar, podemos destacar que el peor de los clasificadores es el SVM creado con una función de núcleo Pearson VII. No solo obtuvo la peor precisión, sino que la gestión de errores es bastante mala. A lo largo de todos los porcentajes de compresión mantiene una tendencia que le lleva a obtener un valor de 0,5 unidades en la mayoría de los casos. El desplazamiento que sufre con respecto a sus compañeros es de más de 3 décimas. Por otra parte, los SVM generados con una función de núcleo

5.3 Validación empírica



(a) La figura muestra las tasas de error alcanzadas por un primer grupo de clasificadores basado en un conjunto de datos reducido. Observamos que para compresiones muy altas los resultados empeoran y que el árbol de decisión construido con C4.5 no puede igualar al resto de los clasificadores.



(b) En este segundo grupo encontramos que la variación es mucho mayor. Sin embargo, las tendencias de los clasificadores siguen siendo las mismas. Las máquinas de soporte vectorial con función de núcleo Pearson VII no pueden igualar al resto de los clasificadores.

Figura 5.3: Tasas de error obtenidas durante la comprobación de los modelos generados a través de un conjunto de datos reducido con el fin de la detección de posibles problemas de microrrechupes. Se muestran los datos en términos de área bajo la curva ROC para cada uno de los porcentajes de compresión. En general, los resultados se acercan mucho al valor óptimo del área excepto para compresiones muy grandes.

5. Adaptando los modelos al proceso cambiante

RBF acaban convergiendo con el anterior. A pesar de que este clasificador no es uno de los que mejor precisión y tasas de error tiene, durante la mayoría del rango de compresiones mantiene una tendencia similar a la del resto de clasificadores. Sin embargo, a la hora de llegar a la sección que hace empeorar el comportamiento, este clasificador reduce su rendimiento tanto que acaba igualándose al peor de los clasificadores.

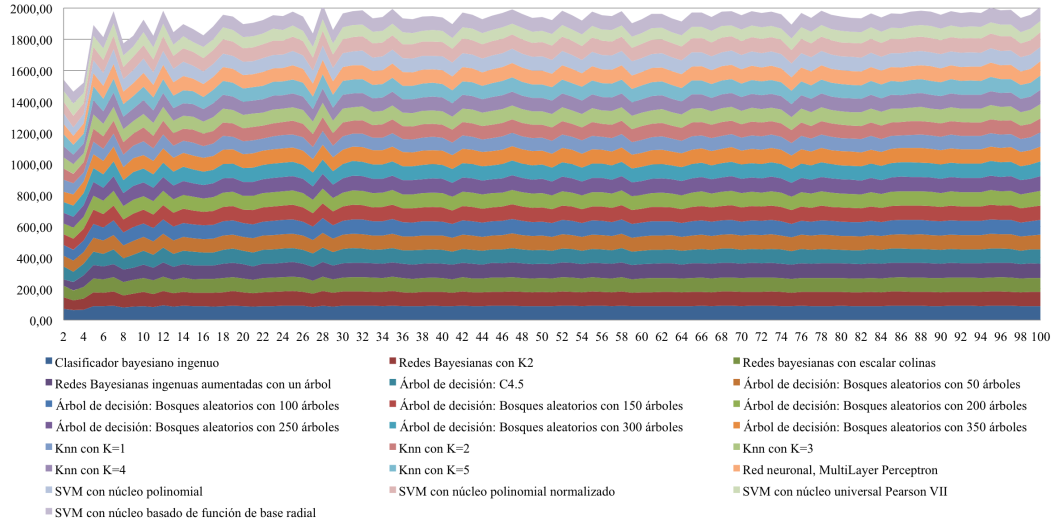
Dada la dificultad a la hora de vislumbrar los resultados alcanzados por todos y cada uno de los clasificadores en las gráficas presentadas, hemos optado por la opción de mostrar un resumen de los resultados a través de unas gráficas de áreas apiladas.

De esta forma, la Figura 5.4a hace referencia a las tasas de precisión alcanzadas por los modelos. Ahora sí, con mayor claridad observamos cómo la tendencia se mantiene. Los máximos y mínimos se encuentran para cada uno de los clasificadores en los mismos lugares. Esta gráfica tiene un gran interés ya que nos indica que los métodos estadísticos aplicados para la generación de los nuevos modelos reducidos afectan del mismo modo a todos los clasificadores, por lo que, como ya podíamos pensar anteriormente, los resultados no tenían por qué ser mejores, pero sí deberían mantener una tendencia similar. Por otra parte, la gestión de los errores se muestra en la Figura 5.4b. Igualmente, se observa que la tendencia general seguida por cada uno de los clasificadores es similar, así, llegamos a las mismas conclusiones que en los resultados de la precisión.

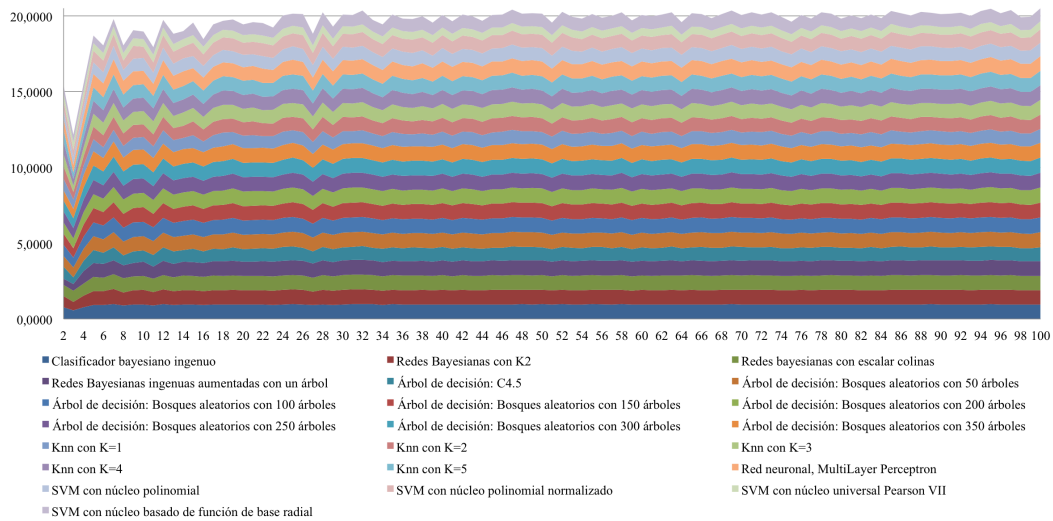
Con este primer experimento, hemos podido comprobar que el comportamiento de los clasificadores que han utilizado un conjunto de datos reducido es realmente bueno. No solo eso, sino que, a pesar de que se empeoren los resultados cuando los porcentajes de precisión son muy altos y quitando algún caso especial, la diferencia para un clasificador es de aproximadamente tan solo un 30%. Algo similar sucede con las tasas de error, habiendo una diferencia de alrededor de las 3 decimas. Un aspecto harto interesante, ya que utilizando muy pocas evidencias hemos conseguido representar el proceso, por lo tanto, el modelo inherente es suficientemente representativo desde un punto de vista estadístico. De esta forma, la reducción de datos para el defecto de los microrrechupes se ha realizado de forma exitosa plantando una buena base para que el posterior experimento sobre los conjuntos de datos acumulados pueda demostrar la viabilidad del método desarrollado.

Una vez determinado el comportamiento de los modelos, como segunda parte, hemos querido testarlos con el conjunto de datos original comprobando, de esta forma, si los modelos reducidos son capaces de ajustarse a lo que la

5.3 Validación empírica



(a) A modo de simplificación, la primera figura muestra las tendencias seguidas por todos los clasificadores en términos de precisión en la predicción de la aparición de microrrechupes analizada en este experimento. Como se puede observar, las tendencias seguidas por todos ellos son muy similares.



(b) En el segundo de los gráficos apilamos todos los resultados del área bajo la curva ROC alcanzados por los clasificadores a la hora de prever la aparición de microrrechupes.

Figura 5.4: Resultados apilados obtenidos en los experimentos orientados a la prueba de los modelos reducidos para la predicción de defectos de microrrechupe en las piezas extraídas de la fundición. La figura recoge tanto los valores alcanzados en la precisión como los alcanzados en el área bajo la curva ROC.

5. Adaptando los modelos al proceso cambiante

producción nos indica.

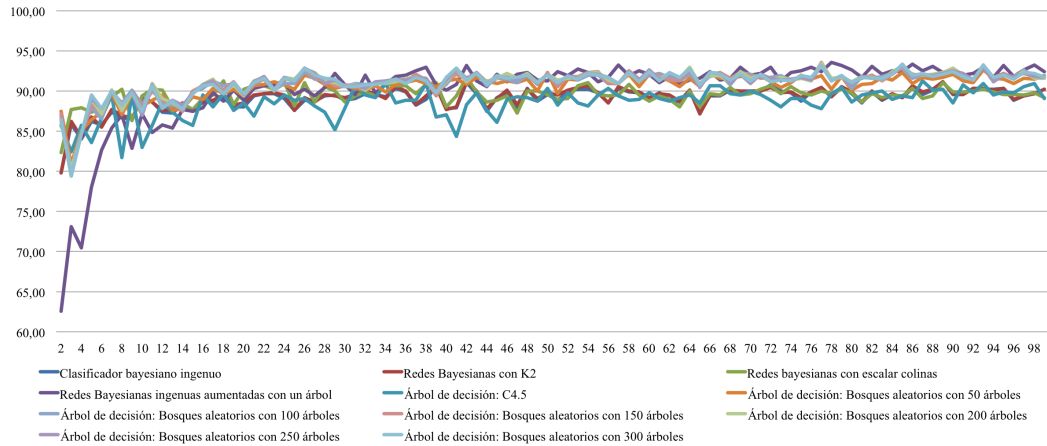
Así, para este experimento hemos tomado los mismos clasificadores que llevamos utilizando a lo largo de esta tesis doctoral y hemos aplicado la etapa de aprendizaje con el conjunto reducido al completo. Posteriormente, le hemos aplicado el conjunto original comparando si los resultados obtenidos son los mismos que se detectaron a través de los análisis realizados en planta. También, y debido al no determinismo de los métodos de Montecarlo, hemos querido utilizar los mismos conjuntos de datos reducidos para así poder comparar el comportamiento obtenido con los anteriores resultados. Debemos indicar que este proceso no cuenta con la validación cruzada pero nos permite comprobar el comportamiento bajo una situación lo más real posible.

En las Figuras 5.5a y 5.5b se pueden observar los resultados en términos de precisión obtenidos al probar los modelos reducidos con los datos originales. Como se puede observar a primera vista, la tendencia de comportamiento seguida por todos los clasificadores concuerda claramente con los experimentos anteriores (ver Figura 5.2). Igualmente, se siguen presentando las fluctuaciones a lo largo de la curva de resultados. Pero como ya se comentó antes, son debidas al no determinismo de los métodos de Montecarlo y/o ruido en los datos. Del mismo modo, en este caso también se observa un decremento en la precisión cuanto mayor es la tasa de compresión. Con lo que, este experimento sobre el conjunto de datos original mantiene las generalidades descubiertas al probar los modelos por sí mismos.

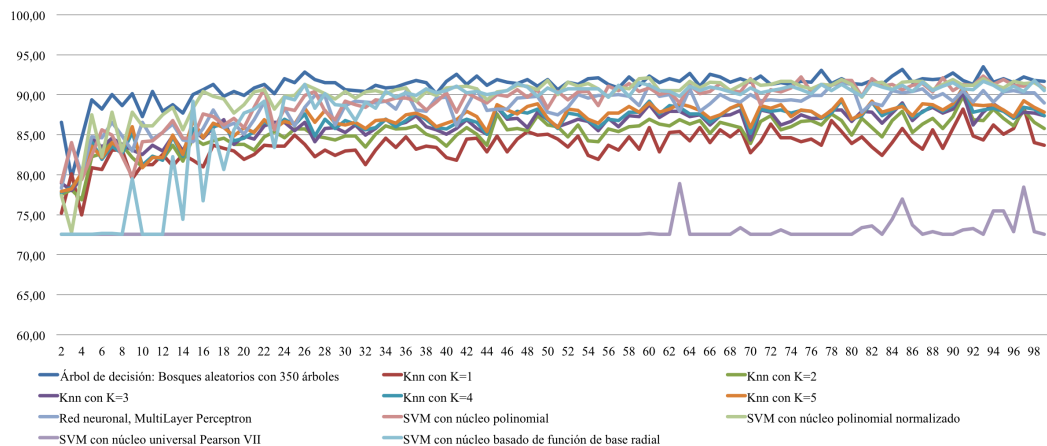
Más concretamente, la Figura 5.5a muestra los resultados alcanzados por el primer grupo de clasificadores. Como bien se puede observar, su tendencia es la habitual. Sin embargo, sorprendentemente, en este caso el clasificador bayesiano ingenuo aumentado con un árbol no es el que mejor se ajusta al proceso. Todos los demás clasificadores se encuentran a su mismo nivel, encontrando diversos niveles de compresión en los que es superado. Pero, además, este clasificador también se convierte en el peor de todos cuando el grado de compresión es muy alto. De tal forma, que el clasificador bayesiano TAN únicamente llega a obtener un porcentaje de acierto del 62,57%.

Para este primer grupo, el descenso pronunciado de los porcentajes de precisión se vuelve a dar dentro del mismo rango, el que va entre la creación de un 2 y un 16% de las instancias. Claramente puede ser identificado en esos puntos porque es cuando todos los clasificadores comienzan a encontrarse bajo la barrera del 90% de precisión. A pesar de este hecho, en este experimento podemos comprobar que la caída obtenida es menor a lo obtenida con los modelos reducidos. Esta caída es de unas 20 unidades, aspecto chocante ya

5.3 Validación empírica



(a) El primer grupo de clasificadores que hemos testado contra el conjunto de datos original obtienen unos resultados similares (algo más bajos porque no se ha ponderado con una validación cruzada). Todos los clasificadores se ven afectados de forma similar al método de compresión.



(b) En esta segunda figura comprobamos el comportamiento que ha tenido el segundo grupo de clasificadores. La precisión es menor, al igual que sucedía en las pruebas con los clasificadores unitarios. De la misma forma que en las pruebas de los datos reducidos los SVM con función de núcleo Pearson VII se distancian del comportamiento normal.

Figura 5.5: Resultados obtenidos durante la comprobación de los modelos generados a través de un conjunto de datos reducido frente al conjunto de datos original con el fin de la detección de posibles problemas de microrrechupes. Se muestran los datos en términos de precisión para cada uno de los porcentajes de compresión. El comportamiento de cada uno de los clasificadores fluctúa. A pesar de ello, estos mantienen una tendencia constante en su predicción.

5. Adaptando los modelos al proceso cambiante

que las tasas de acierto que somos capaces de alcanzar con los modelos que han sido aprendidos únicamente con un 2 % de instancias supera el 80 %.

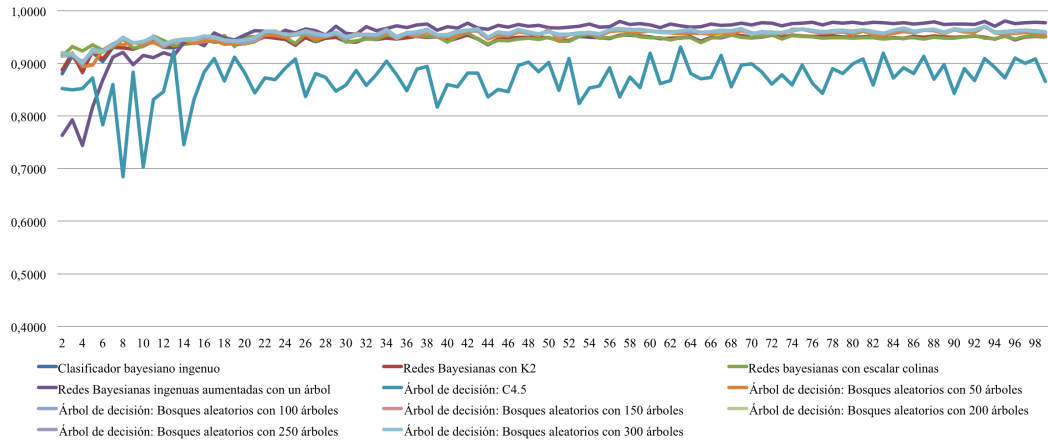
Como observamos en el experimento sobre los modelos, en general, el peor de los clasificadores vuelve a ser el árbol construido con el algoritmo C4.5. Su tendencia a lo largo de la mayoría del rango de compresiones lo convierten en el peor de este grupo. Pero, por otra parte, es el método de aprendizaje automático supervisado que menos se ve afectado a la hora de utilizar un alto grado de compresión en los datos. En otras palabras, en el rango de generación de instancias [2, 18] es el clasificador que mantiene un comportamiento más estable.

Por otra parte, la Figura 5.5a nos ilustra con los resultados logrados por el segundo grupo de clasificadores. En esta ocasión volvemos a ver como este grupo tiene unos resultados mucho más variados, pero únicamente en lo que a valores de precisión se refiere, porque las curvas que presentan los clasificadores son similares entre todos ellos (algo que también se producía en el caso estudiado anteriormente). Como ya fue explicado, las variaciones vienen marcadas porque los modelos no son capaces de obtener el mismo grado de aciertos. A pesar de ello, la tendencia que siguen vuelve a ser la misma que ya observamos en el primero de los grupos, en otras palabras, a mayor grado de compresión obtenemos una menor precisión.

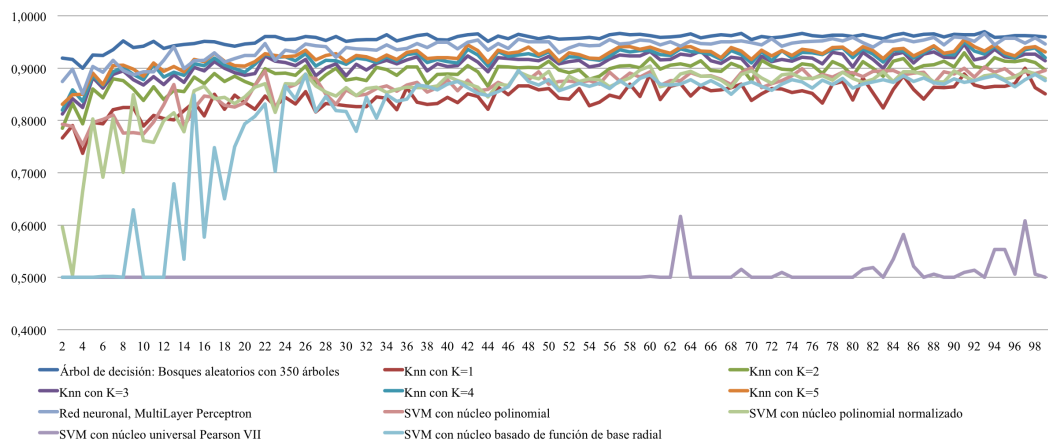
En este segundo grupo, la reducción máxima en los porcentajes de acierto ronda el 75 % (dejando de lado los resultados cosechados por las máquinas de soporte vectorial). Así, igual de sorprendente que en el caso del grupo anterior, una gran compresión conlleva una pérdida de precisión pero no tan grande como nos podríamos imaginar de antemano. Con lo que para ambos grupos mantenemos unas tasas de precisión muy interesantes para cuando deban generarse los modelos mediante un conjunto de datos acumulado.

También, como ya venimos observando en estos experimentos, las máquinas de soporte vectorial que utilizan una función de núcleo Pearson VII vuelven a no adaptarse al proceso, pudiendo denominarlos como los peores clasificadores en la predicción de los microrrechupes, tanto si utilizamos un conjunto de datos reducido, como el original. Pero, algo interesante es que este clasificador no se ve afectado por el método de compresión bayesiana basado en técnicas de Montecarlo, ya que su comportamiento a lo largo de todos los porcentajes de compresión es prácticamente constante. También, otro de los clasificadores SVM, esta vez utilizando una función de núcleo RBF, llega a alcanzar unos niveles de precisión muy bajos. Particularmente, este último clasificador se trata de uno de los mejores del segundo grupo, sin embargo,

5.3 Validación empírica



(a) La figura muestra las tasas de error alcanzadas por el primer grupo de clasificadores. Se ilustra que las tendencias de comportamiento de los clasificadores son similares, afectados en igual medida por el método de compresión. La excepción es el árbol de decisión construido con el algoritmo C4.5.



(b) La segunda figura muestra el comportamiento del resto de los clasificadores. Como se puede observar, se mantienen las mismas tendencias, aunque obteniendo peores resultados (al igual que en los casos anteriores). El comportamiento del modelo SVM generado con una función de núcleo Pearson VII se aparta del resto de clasificadores.

Figura 5.6: Tasas de error obtenidas al enfrentar los modelos aprendidos con el conjunto reducido a los datos originales dada la intención de detectar micro-rechupes. Se muestran los datos en términos de área bajo la curva ROC para cada uno de los porcentajes de compresión. Por lo general se adquieren buenos resultados y los clasificadores replican lo que hicieron al ser testado únicamente su modelo reducido.

5. Adaptando los modelos al proceso cambiante

cuando comienza el proceso de decadencia que todos los clasificadores sufren, este se ve afectado en mayor manera igualando su comportamiento con el del SVM con Pearson VII.

En lo que respecta a los errores, las Figuras 5.6a y 5.6b muestran el rendimiento de gestión en base al área bajo la curva ROC. Sorprendentemente, la gestión de errores que se ha alcanzado en este experimento, en el que se testea cada uno de los modelos con el conjunto de datos original, es incluso mejor que la obtenida al probar el rendimiento de cada uno de los clasificadores en sí mismos. De la misma manera, como podemos observar, y evitando algunos de los clasificadores, las tendencias de comportamiento vuelven a ser las mismas que llevamos analizando a lo largo de todo el experimento; determinando que hay un decremento en la gestión de los errores cuantas menos instancias son producidas en el proceso de reducción. Tampoco podemos dejar de lado que las perturbaciones que se producían sobre las curvas de evaluación, aquí vuelven a producirse. Sin embargo, estas han visto reducida su amplitud considerablemente. Por lo tanto, el grado de ajuste y adaptación que se alcanza con los modelos reducidos es realmente interesante, con lo que puede hacernos prever que los resultados de los aprendizajes acumulados también van a ser satisfactorios.

En particular, la Figura 5.6a muestra los resultados alcanzados por el primer grupo de clasificadores. Como bien queda ilustrado en la gráfica, la mayoría de los clasificadores, a excepción del árbol construido con el método C4.5, mantienen una tendencia casi constante a lo largo de todo el rango de compresiones posibles. Además, es importante destacar que dentro del rango que consideramos como el de rendimiento óptimo (el que comprende los porcentajes de creación de variables que van desde 19 al 99%) el área bajo la curva ROC es de más de un 0,95 (nótese que el valor óptimo a alcanzar en esta medida es de 1). A pesar de que en el rango [2,18] sigue existiendo un descenso del rendimiento, esta vez es mucho menor ya que la mayoría de los clasificadores llegan a descender como máximo unas 5 o 6 centésimas. Si para realizar esta medición tenemos en cuenta las excepciones que marcan los extremos, se acaba decrementando el rendimiento entre unas 0,7 y 0,8 unidades. Comparado con lo conseguido por este mismo grupo (ver Figura 5.3a) podemos observar que la adaptación es mejor que la que nos ofrecieron las pruebas de los modelos por sí solos al realizar la medición con el conjunto de datos original.

Al igual que en la precisión, contamos con dos casos particulares que merecen ser comentados. El primero de ellos es el de los resultados alcanzados por

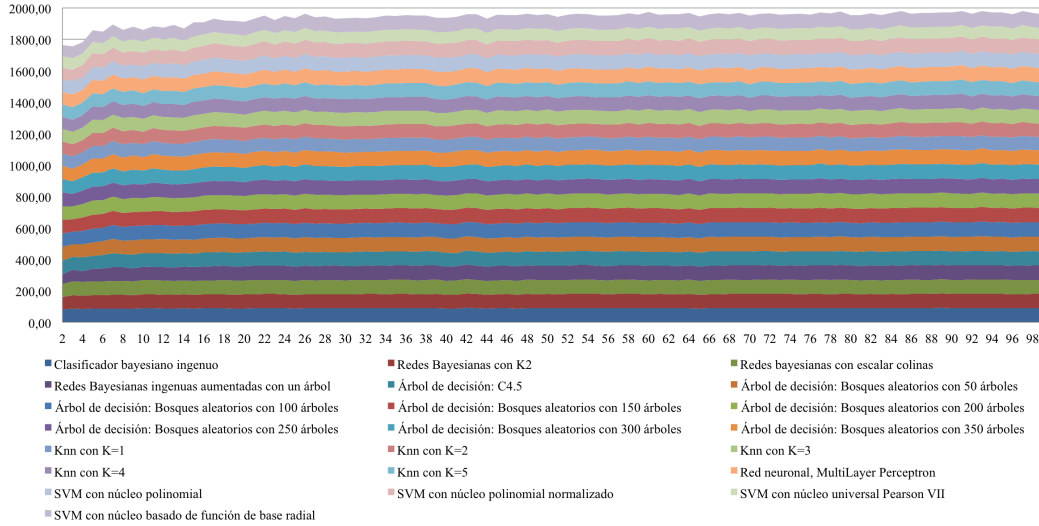
la red bayesiana ingenua aumentada con un árbol. Esta mantiene uno de los mejores resultados alcanzados por los clasificadores. No obstante, son seguidas muy de cerca por los bosques aleatorios y otros clasificadores bayesianos que utilizaron otros métodos de aprendizaje estructural. Pero al igual que en el caso de la precisión en el acierto, en el rango de compresiones mayores sufre una pérdida de la efectividad en la gestión de los errores. El segundo de los casos que merece ser comentado es el de los árboles C4.5. Como hasta ahora, este modelo no se acerca de forma clara al rendimiento de sus compañeros de grupo. También, a pesar de las reducciones en las fluctuaciones, este clasificador las sigue manteniendo con una amplitud muy similar a la de los clasificadores basados en los conjuntos reducidos (ver Figura 5.6a). A pesar de esto, el árbol C4.5 no puede ser considerado como el peor de los clasificadores, ya que otros de los que se encuentran en el segundo de los grupos obtienen peores resultados. Asimismo, aunque este modelo nos sorprendía por la estabilidad alcanzada en la precisión de acierto, aquí no es capaz de mantenerla. Esto es así debido a que a pesar de que el descenso sea muy pequeño, las fluctuaciones durante el rango [2, 18] son muy pronunciadas.

En el segundo de los grupos de clasificación (ver Figura 5.6b) podemos comprobar que los clasificadores acaban teniendo el mismo comportamiento que los del primer grupo. En general, la tendencia que siguen los clasificadores es casi constante pero a diferentes niveles. Esto, como ya se ha comentado, está relacionado con el comportamiento de cada clasificador, en otras palabras, los clasificadores se ajustan de forma diferente al proceso. De esta forma, el hecho de que el área bajo la curva ROC sea superior para algunos métodos e inferior para otros no está tan relacionado con el método de compresión como lo está con el rendimiento de cada clasificador.

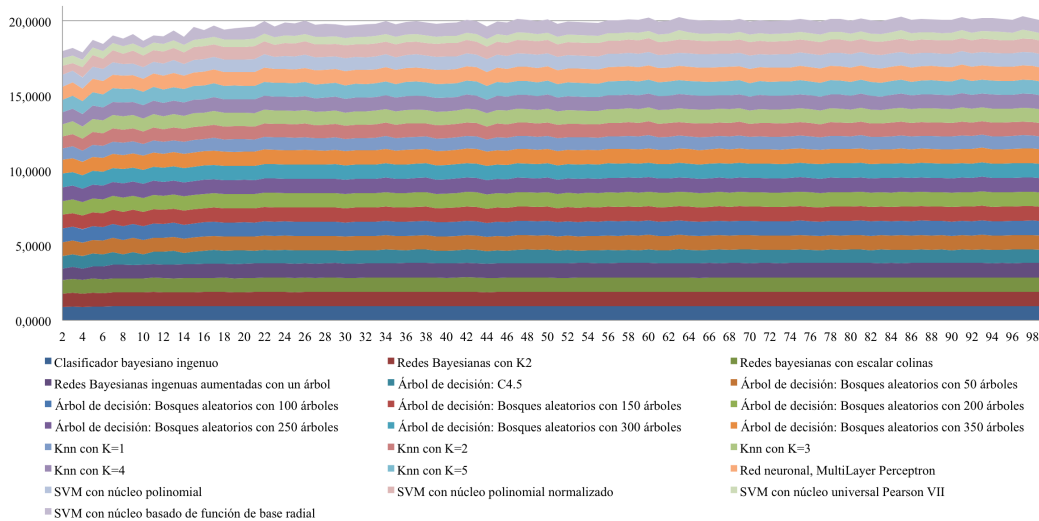
Para este segundo grupo también detectamos un punto en el que el rendimiento de los clasificadores para la gestión de los errores desciende. Este punto es el mismo que ya identificamos en los experimentos anteriores (la generación del 18% de las instancias). Y como ya sucedía con el primer grupo de clasificadores, el descenso relativo de cada clasificador es muy bajo. Eliminando excepciones, el rango de valores para el área bajo la curva ROC en el que se mueven los clasificadores es el de 0,75 a 0,95. Pero tomando únicamente al zona de confort que va de (18, 99] los valores no descienden de la barrera del 0,80.

A parte, en este segundo grupo nos encontramos tres clasificadores que abandonan la normalidad, y todos ellos son máquinas de soporte vectorial construidas con diferentes funciones de núcleo. El primero de ellos, construi-

5. Adaptando los modelos al proceso cambiante



(a) La primera figura muestra los diferentes resultados de precisión alcanzados por los clasificadores mostrando las tendencias seguidas por todos ellos de una forma mucho más legible. Como se puede observar, la línea que trazan todos ellos es muy similar.



(b) La segunda gráfica muestra los resultados apilados relativos a las tasas de error, más concretamente, el área bajo la curva ROC que alcanzaron los clasificadores en su proceso de predicción de microrrechupes.

Figura 5.7: Resultados apilados obtenidos en los experimentos orientados a la prueba de los modelos reducidos, probados frente al conjunto de datos original, para la predicción de defectos de microrrechupes. Se recogen tanto los valores alcanzados en la precisión como aquellos relativos al área bajo la curva ROC.

do con una función Pearson VII, mantiene el comportamiento constante que llevamos observando en todos estos experimentos. Además, el rendimiento alcanzado queda alineado con el rendimiento medio posible que un clasificador puede tener en la gestión de errores mediante el área bajo la curva ROC, más específicamente, un valor de 0,5. Los otros dos clasificadores, generados con una función de núcleo RBF y una polinomial normalizada, destacan porque, en comparación con el resto, el decremento del rendimiento es mucho más vertiginoso. Bajan tanto que su comportamiento llega a alinearse al del SVM con Pearson VII. La diferencia que hay entre ellos dos es el punto en el que empieza su vertiginosa bajada de rendimiento, entre el 22 y el 24 % de instancias generadas para el primero de ellos y entre el 14 y 16 % de instancias para el segundo.

Al igual que ya hicimos anteriormente, con el fin de obtener una mejor visualización de los resultados y las tendencias que han tenido los clasificadores, apilamos sus resultados, tanto los relativos a las tasas de acierto, como a la gestión de los errores, para descubrir todo aquello que ya se había citado tras un minucioso análisis de los resultados. Así, la Figura 5.7a hace referencia a la precisión de los modelos basados en un conjunto de datos reducido al probarlo frente al conjunto de datos original. La gráfica deja bien claro que la tendencia seguida por todos los clasificadores es casi constante hasta que se llega a un punto de descenso del comportamiento de clasificación. Sin embargo, como se puede ver los valores apilados para el peor de los casos no difieren tanto del mejor de los casos. Por otra parte, la Figura 5.7b muestra los valores apilados para el área bajo la curva ROC. Igual que en el caso de la precisión, comprobamos que las variaciones entre todos los clasificadores han sido muy pequeñas, y que el peor de los puntos apenas varía del mejor de ellos. Además, ambas imágenes dejan más clara cuál es la zona de confort en la que mejor se desenvuelven los clasificadores y dónde empieza el descenso del rendimiento para ambos casos.

A modo de breve conclusión, este experimento nos ha permitido visualizar el comportamiento del sistema de compresión bayesiana basado en métodos de Montecarlo. En primer lugar, hemos podido comprobar que los modelos generados dan unos buenos resultados por sí mismos y que estos modelos cuentan con una distribución probabilística muy pareja a la inherente en los datos extraídos del proceso productivo. En segundo lugar, hemos comprobado que los datos reducidos son capaces de tener un buen rendimiento frente al conjunto de datos original, consiguiendo comprobar cómo funciona el método en un caso muy cercano a la realidad. Ambos experimentos abren la puerta a que

5. Adaptando los modelos al proceso cambiante

el método de comprensión desarrollado sea el adecuado para las adaptaciones acumulativas comprobadas más adelante.

5.3.2.2 Propiedades mecánicas

Al igual que ya hicimos en el capítulo 4, una vez analizado el primero de los defectos pasamos a aplicar el método de comprensión con el fin de comprobar el comportamiento alcanzado en el segundo de los defectos. Más concretamente, trabajamos con las propiedades mecánicas, especialmente con la carga máxima que soportan las piezas de fundición. A nuestro entender, al igual que no existía ningún experimento relacionado con las predicciones de las propiedades mecánicas en piezas de fundición, no conocemos estudio alguno que, en el caso de las propiedades mecánicas, se encargue de realizar una adaptación de los modelos de predicción. De esta forma, estos experimentos son pioneros al determinar el funcionamiento de un nuevo método de adaptación al proceso productivo para modelos construidos mediante aprendizaje automático supervisado.

Para llevar a cabo los experimentos relacionados con la carga máxima hemos seguido la misma metodología aplicada al defecto de los microrrechupes en la sección 5.3.2.1. Específicamente, hemos dividido el experimento en dos partes. En primer lugar, realizamos un análisis de los clasificadores aprendidos mediante conjuntos de datos reducidos. En segundo lugar, probaremos el comportamiento de los modelos reducidos con el conjunto de datos original. Ambos experimentos permiten comprobar si los nuevos modelos reducidos representan correctamente el proceso y son capaces de obtener unos buenos resultados en la predicción del estado estacionario $t + 1$. Del mismo modo, estos experimentos nos permiten evaluar el grado de ajuste de los modelos al proceso real cuando se realiza la prueba con los datos originales y determinar cómo afecta el método a cada uno de los clasificadores.

Comenzando por el primero de ellos, las Figuras 5.8a y 5.8b nos muestran los resultados alcanzados por los clasificadores para cada uno de los porcentajes de comprensión posibles. Por otra parte, las Figuras 5.9a y 5.9b hacen referencia a las tasas de error medidas como el área que se encuentra bajo la curva ROC. Recordamos que en las gráficas, el eje de abscisas marca el porcentaje de variables que han sido generadas para el conjunto de datos reducido. Por consiguiente, a mayor número, un menor grado de comprensión ha sido aplicado al conjunto de datos. Del mismo modo que en los experimentos anteriores, no hemos llegado a realizar una comprensión del 99 % (i.e., generar

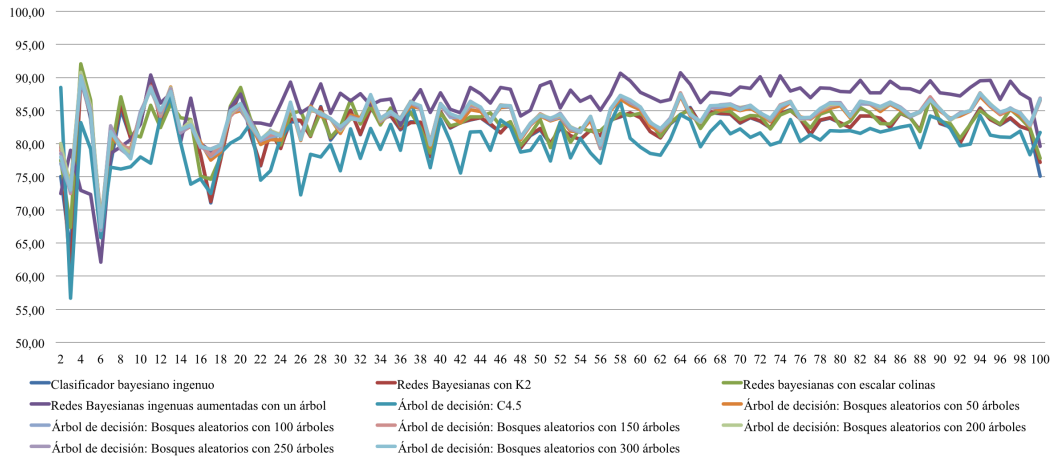
únicamente un 1 % de instancias) porque el conjunto de datos llega a ser tan pequeño que no nos permite realizar una validación cruzada con un valor de $k = 10$.

En lo que respecta a los porcentajes de acierto alcanzados por los clasificadores, se muestra que el funcionamiento de la compresión bayesiana basada en métodos de Montecarlo es muy similar al obtenido en los experimentos de los microrrechupes. De esta forma, volvemos a observar que todos los clasificadores disponen de una serie de fluctuaciones provocadas, en primer lugar, por el no determinismo en los muestreos de los métodos de Montecarlo y, en segundo lugar, por la mala calidad de los datos (i.e., disponen de ruido introducido por los operarios de la planta al realizar la captura de los datos a través de un proceso manual). Sobre todo, el problema de la calidad de los datos hace que en este caso la variabilidad en la precisión entre los diferentes clasificadores sea mucho mayor. Pero este es otro de los puntos que ya fueron observados en los experimentos desarrollados en la sección 4.3.2.2. Sin embargo, dejando de lado las fluctuaciones y visualizando las tendencias obtenidas por los clasificadores, observamos que el método afecta a todos por igual y que volvemos a encontrarnos en una situación en la que los máximos y los mínimos de las curvas de resultados coinciden en la mayoría de los porcentajes de compresión. Además, de forma general, se puede comprobar que a lo largo de la curva de resultados, muchos de los clasificadores mejoran en sus porcentajes de acierto, con lo que determinamos que la generación de los modelos reducidos acaba con la existencia de ocurrencias anecdóticas que pueden perturbar su comportamiento.

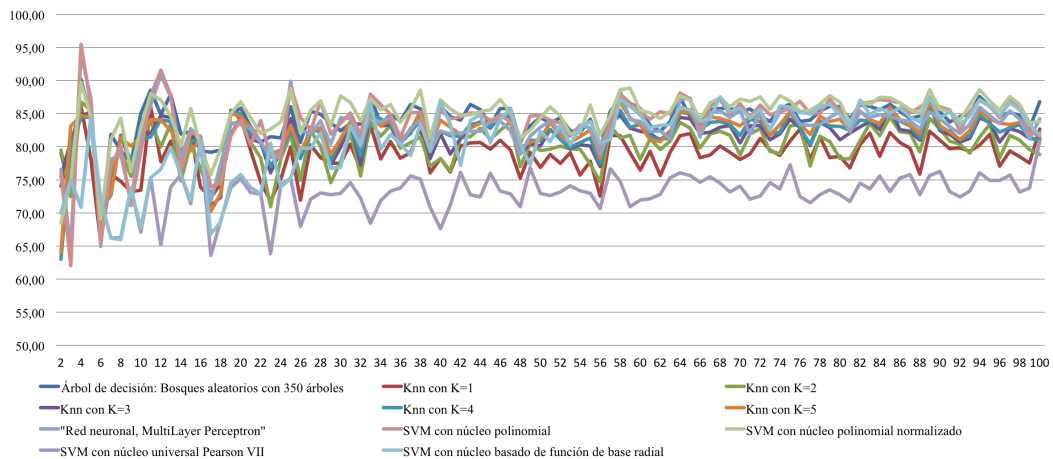
Más concretamente, la Figura 5.8a muestra los resultados alcanzados por el primer grupo de clasificadores. Como hemos comentado, observamos que las fluctuaciones son mayores y de mayor amplitud que en los experimentos anteriores. Pero la tendencia general de los clasificadores es la misma, con lo que el método de compresión no los entorpece en su trabajo. Incluso, podemos observar, que exceptuando el caso de los bosques aleatorios que aproximadamente mantienen todos el mismo comportamiento y una tendencia de un 85 % de acierto, el resto de los clasificadores, una vez se ha comenzado a aplicar la compresión hacen que su tendencia sea superior al porcentaje de acierto basado en el conjunto de datos general. Esto es lo que podemos identificar como una reducción del ruido.

En los experimentos desarrollados para los microrrechupes pudimos determinar que hay una cierta zona de confort en la que los clasificadores obtienen los mejores resultados y las mejores tendencias. Asimismo, contábamos con

5. Adaptando los modelos al proceso cambiante



(a) El primer grupo de clasificadores se comporta de forma similar. Los defectos en los datos hacen que las fluctuaciones sean mucho mayores. Sin embargo, destaca que al comenzar a comprimir los conjuntos de datos, muchos clasificadores mejoraron su precisión, lo que indica que el método incluso puede reducir ruido.



(b) Al igual que en los casos anteriores, este segundo grupo de clasificadores tiene un comportamiento más disperso, pero, cada clasificador es afectado de la misma forma por el método de compresión, con lo que sus tendencias son similares.

Figura 5.8: Resultados obtenidos durante la comprobación de los modelos generados a través de un conjunto de datos reducido con el fin de la detección de posibles problemas de carga máxima. Se muestran los datos en términos de precisión para cada uno de los porcentajes de compresión. A pesar de la fluctuación, las tendencias para cada clasificador también son similares para la predicción de este defecto.

una zona en la que el rendimiento comenzaba a descender pero sin llegar nunca a ofrecernos unos valores bajísimos de precisión. Este hecho nos hacía determinar que la representación de los modelos era buena. En el caso de las propiedades mecánicas, y más específicamente, en el de la carga máxima, también encontramos este comportamiento. A parte de determinar una zona de descenso de la precisión, observamos que las fluctuaciones crecen en amplitud haciendo que sea difícil marcar el comportamiento de los clasificadores. Aproximadamente, la zona de confort es la que ocupa el rango de generación de instancias [24, 99]. En este caso, el descenso no es muy pronunciado, pero los picos de las fluctuaciones hacen que en ciertos puntos no se supere un porcentaje del 60 %. Además, es interesante que todos los clasificadores, tanto los que mejores resultados han tenido como los peores del grupo, llegan a descender hasta el mismo nivel. De esta forma, este grado de compresión no llega a ser excesivamente representativo. Este inconveniente viene dado por la incapacidad de obtener una buena representación a través de la red bayesiana que genera la distribución probabilística. Así, mientras que para valores bajos de compresión sí que es capaz de aproximar el comportamiento de la planta, con valores muy altos no lo hace tan bien.

Por otra parte, en los experimentos de la carga máxima también observamos ciertos clasificadores que se desplazan de la normalidad marcada por la mayoría. Así, por ejemplo, el árbol de decisión generado con el método C4.5. Como viene siendo habitual, este es el peor de los clasificadores del grupo. Es también uno de los que mayor amplitud tiene en sus fluctuaciones. Sin embargo, al aplicar el método conseguimos obtener mejores resultados que los que se obtenían con el modelo original. El segundo de ellos es la red bayesiana ingenua aumentada con un árbol. Como ya sucedía anteriormente, este clasificador es el que mejor se adapta una vez realizada la reducción ya que este mismo método fue utilizado para la generación de los datos. No obstante, a pesar de ser el mejor del grupo, todos los bosques aleatorios y los demás clasificadores bayesianos se encuentran muy cerca.

En lo que respecta al segundo grupo, la Figura 5.8b nos ilustra con los resultados en términos de precisión. De la misma forma que sucedía en los microrrechupes, este segundo grupo cuenta con una variación de precisiones mayor. No obstante, para el caso de la carga máxima se encuentran todos ellos mucho más cercanos entre sí. Fácilmente se puede vislumbrar que las tendencias que están siguiendo todos los clasificadores vuelven a ser similares, tanto que hay curvas que durante varios porcentajes coinciden totalmente. Del mismo modo que en el primero de los grupos, y a pesar de que en los primeros

5. Adaptando los modelos al proceso cambiante

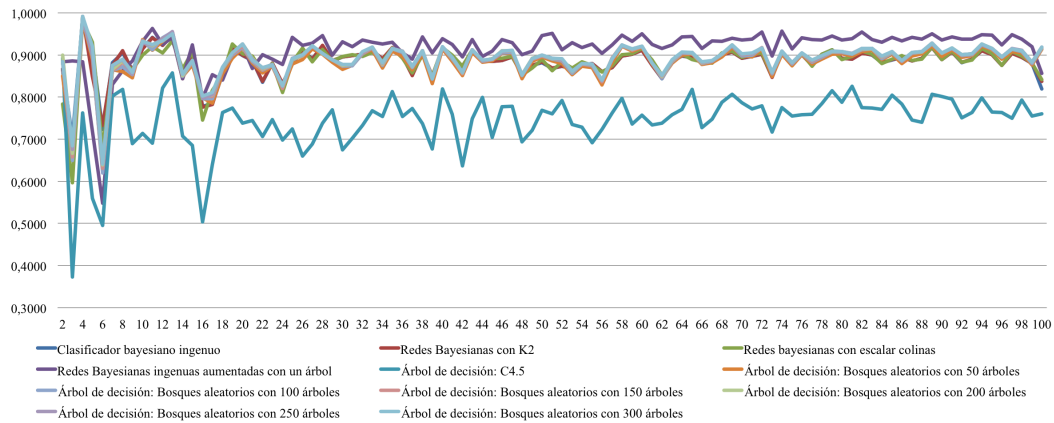
porcentajes de compresión el rendimiento baja entre un 2 y un 5 %, posteriormente, los clasificadores vuelven a marcar una tendencia de predicción superior a la precisión original.

Este segundo grupo también muestra dos zonas diferenciadas, (i) una en la que se alcanza un porcentaje de aciertos considerable y (ii) otra en la que hay un descenso en el rendimiento de los modelos. En este grupo se reproduce el comportamiento del primer grupo. La reducción no es excesivamente grande, pero es la etapa en la que se comienzan a obtener unas fluctuaciones de mayor amplitud y con una frecuencia de aparición mucho mayor. Además, ciertos picos de las fluctuaciones muestran que los clasificadores no son capaces de obtener más de un 62 % de acierto. El punto que divide las dos áreas es el mismo que se detectó en el primero de los grupos, la generación del 24 % de las evidencias.

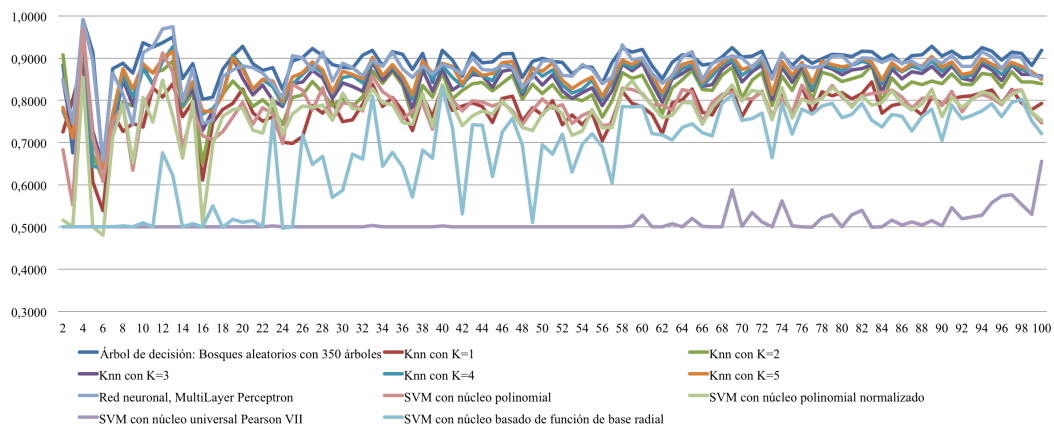
Debemos destacar sobre todo el comportamiento de las máquinas de soporte vectorial con función de núcleo Pearson VII, convirtiéndose de nuevo en el peor de los clasificadores en todos los experimentos desarrollados. A pesar de este hecho, y debido a que los resultados alcanzados por el clasificador con el conjunto de datos original son mejores que todos los que pueden ser obtenidos a lo largo de todos los valores de compresión, consideramos que el método de compresión le ha afectado negativamente. Aunque debemos recordar que los resultados que obtuvo este clasificador en los procesos de predicción del estado estacionario $t + 1$ no fueron excesivamente buenos (se encontraba entre los seis peores resultados).

Centrándonos ahora en la gestión de los errores que han tenido cada uno de los clasificadores, la Figura 5.9a muestra los resultados obtenidos por el primero de los grupos de clasificación. Concretamente, nos muestra los valores obtenidos en el área bajo la curva ROC. Básicamente, las conclusiones a las que podemos llegar tras analizar los resultados son similares a las que ya se obtuvieron al analizar la precisión de los modelos. Así, identificamos que las tendencias de todos los clasificadores son muy similares. Dentro de todo el conjunto de clasificadores reconocemos tres grupos claramente diferenciados. El primero de ellos se compone del clasificador bayesiano ingenuo aumentado con un árbol. Este es el que mejores resultados consigue, pero como ya comentamos anteriormente, este clasificador es el que ha sido utilizado en el proceso de compresión, por lo que la distribución de la información está adaptada a él. El segundo de ellos es el que marca la normalidad, compuesto por los bosques aleatorios y otros algoritmos bayesianos (utilizando algoritmos de aprendizaje estructural como K2 o escalar colinas, y la red bayesiana ingenua)

5.3 Validación empírica



(a) La figura muestra las tasas de error alcanzadas por un primer grupo de clasificadores basado en un conjunto de datos reducido. Casi todos los clasificadores se comportan de una forma similar ante el método de compresión y realizan una buena gestión de los errores. El árbol con C4.5 sigue siendo el modelo cuyo comportamiento no se ajusta al del resto de los clasificadores.



(b) Las variaciones del segundo grupo son mayores. A pesar de este hecho, las tendencias siguen siendo muy parecidas. En este caso, los SVM con funciones de núcleo Pearson VII y RBF no consiguen seguir al resto de los clasificadores.

Figura 5.9: Tasas de error obtenidas durante la comprobación de los modelos generados a través de un conjunto de datos reducido con el fin de la detección de posibles problemas en la carga máxima. Se muestran los datos en términos de área bajo la curva ROC para cada uno de los porcentajes de compresión. De forma similar a los microrrechupes, las tendencias que siguen los clasificadores son similares. También, observamos que a grandes porcentajes de compresión las variaciones en el comportamiento son mayores.

5. Adaptando los modelos al proceso cambiante

que se encuentra muy cerca de la red bayesiana TAN. Y por último, el grupo del árbol de decisión construido con el método C4.5 que es el que más se aparta de la normalidad. Como siempre, no se convierte en el peor de los clasificadores en términos de error, pero su adaptación al proceso de producción no es tan buena como la de otros.

La tendencia de este primer grupo sigue la que ya fue marcada por los porcentajes de aciertos. Se trata de una tendencia muy similar, que a partir de un grado de creación de variables (el 24 %, el mismo que en la precisión) comienza a mantener una peor gestión de los errores. Además, bastante en concordancia con los resultados obtenidos en los porcentajes de aciertos, la reducción que se experimenta se trata de la aparición de una mayor amplitud en las perturbaciones. En el peor de los casos se llega a bajar hasta un valor de 0,37 unidades (alcanzado por el árbol de decisión C4.5). Sin embargo, los valores alcanzados por la mayoría de los clasificadores durante el rango de rendimiento óptimo acaba rondando los 0,90, un valor más que interesante a la hora de trabajar en la predicción de las propiedades mecánicas.

También, comprobamos cómo todos los clasificadores sufren una gran mejora al realizar el proceso de compresión. De esta forma, se llega a realizar una mejora de aproximadamente una décima, con lo que se consigue pasar de valores cercanos a las 0,80 ó 0,85 unidades a los 0,90 ó 0,95. Por otra parte, hay un par de casos en los que en los primeros valores de compresión se reduce la gestión de errores pero, posteriormente, se recuperan y vuelven a mantener una tendencia con resultados superiores a los obtenidos con el conjunto de datos original. Particularmente, esos casos especiales son el de los bosques aleatorios construidos con 300 árboles y el del árbol de decisión C4.5.

La Figura 5.9a recopila la información sobre la gestión de los errores alcanzada por el segundo de los grupos de clasificación. Tal como se puede observar, y como ya vimos en la precisión, en este caso los valores obtenidos son más dispares. Básicamente, la razón fundamental para disponer de líneas de comportamiento que no coinciden tanto es el propio rendimiento de los clasificadores. A pesar de este hecho, las tendencias de la mayoría de ellos sigue siendo la misma. Muchos de ellos marcan máximos y mínimos en las mismas zonas.

Básicamente, dentro de este segundo grupo de clasificadores también pueden observarse las diferentes áreas de trabajo que hay en los grados de compresión. Por un lado, en el rango (24, 99] está la zona de confort en la que los clasificadores realizan la mejor gestión de los errores. Y no solo eso, sino que como ya pudimos comprobar en el primero de los grupos de clasificación, la

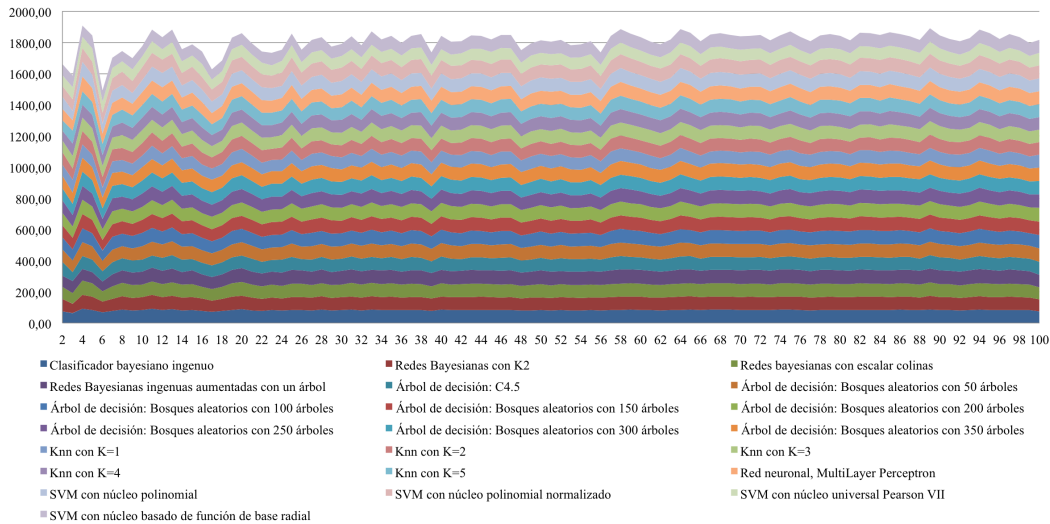
mayoría de los clasificadores sufren una mejora en cuanto comienza el proceso de compresión. Esto es debido a la eliminación de ruido en el conjunto de datos original. No obstante, existen algunos clasificadores como el KNN con $K = 1$ o los bosques aleatorios de 350 árboles que sufren un descenso en el rendimiento durante las primeras compresiones para, posteriormente, volverse a igualar a lo que eran capaces de obtener sin aplicar nuestro método. Por otro lado, el rango [2, 24] es donde encontramos los peores resultados de los clasificadores. De nuevo, el descenso está marcado no por una curva, sino por la variación en la amplitud de las fluctuaciones que hacen que los resultados tengan una gran desviación. En este caso, el peor de los clasificadores no llega a superar la barrera de las 0,5 unidades.

Por último, hay dos clasificadores que claramente no son capaces de realizar una buena gestión de los errores. Ambos son máquinas de soporte vectorial (SVM) que parecen no adaptarse al proceso de producción o al proceso de compresión. El primero de ellos está construido con una función de núcleo Pearson VII. Este clasificador es el que está alcanzando los peores resultados en todos los experimentos. También, observamos que el comportamiento vuelve a ser similar al de los microrrechupes. Su tendencia de comportamiento es casi lineal y se ajusta al valor medio que puede tener el área bajo la curva ROC. Por otra parte, el segundo de ellos lleva la función de núcleo RBF. Este, es el único de los clasificadores que mantienen una tendencia descendente desde el comienzo de las compresiones. Además, a partir de un grado de generación de evidencias del 58 % comienza a obtener una amplitud de perturbación muy alta. Finalmente, este clasificador acaba igualándose a su otro compañero para obtener valores de área bajo la curva ROC que no sobrepasan las 0,5 unidades.

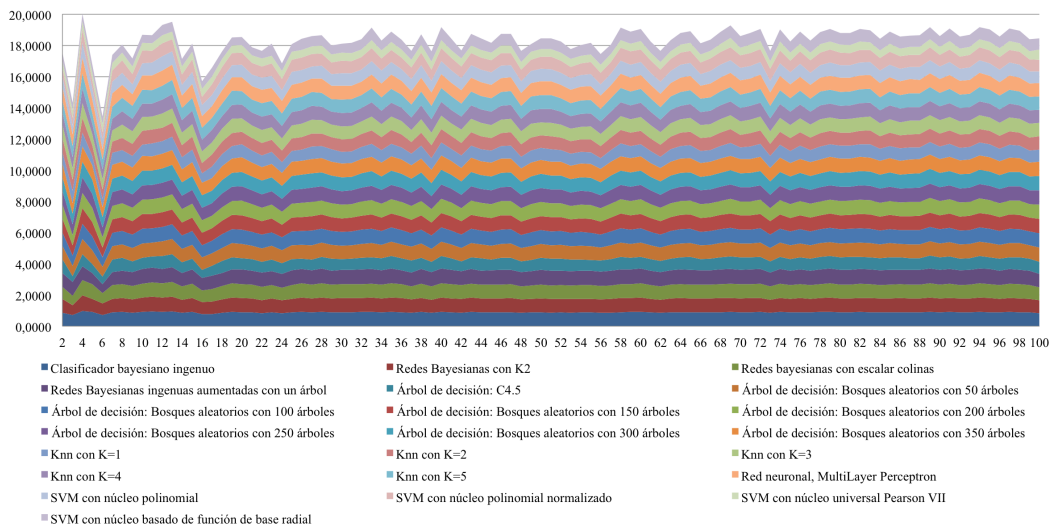
Como ya hemos realizado con anterioridad y con el fin de mostrar los resultados en una forma mucho más legible hemos agrupado todos los resultados en dos gráficas de áreas apiladas. Concretamente, la Figura 5.10a hace referencia a los valores de precisión apilados. Así, podemos observar cómo las tendencias de clasificación son muy similares entre sí. Además, se comprueba fácilmente el lugar en el que se encuentran la mayoría de las variaciones, dejando más claro las zonas de trabajo con las que contamos. Igualmente, en la Figura 5.10b recogemos los valores apilados del área sobre la curva ROC. Prácticamente observamos situaciones muy similares para todos los clasificadores. De esta forma, las tendencias en la gestión de los errores quedan bien marcadas, al igual que la zona de trabajo óptima y la que sufre un descenso de rendimiento.

Las pruebas de los modelos en sí mismos, realizadas para determinar el

5. Adaptando los modelos al proceso cambiante



(a) A modo de simplificación, la primera figura muestra las tendencias seguidas por todos los clasificadores en términos de precisión en su proceso de predicción. Como se puede observar, las tendencias seguidas por todos ellos son muy similares.



(b) En el segundo de los gráficos apilamos todos los resultados del área bajo la curva ROC alcanzados por los clasificadores a la hora de prever la aparición de un valor no válido en la propiedad mecánica de carga máxima.

Figura 5.10: Resultados apilados obtenidos en los experimentos orientados a la prueba de los modelos reducidos para la predicción de defectos de carga máxima en las piezas extraídas de la fundición. La figura recoge tanto los valores alcanzados en la precisión como los alcanzados en el área bajo la curva ROC.

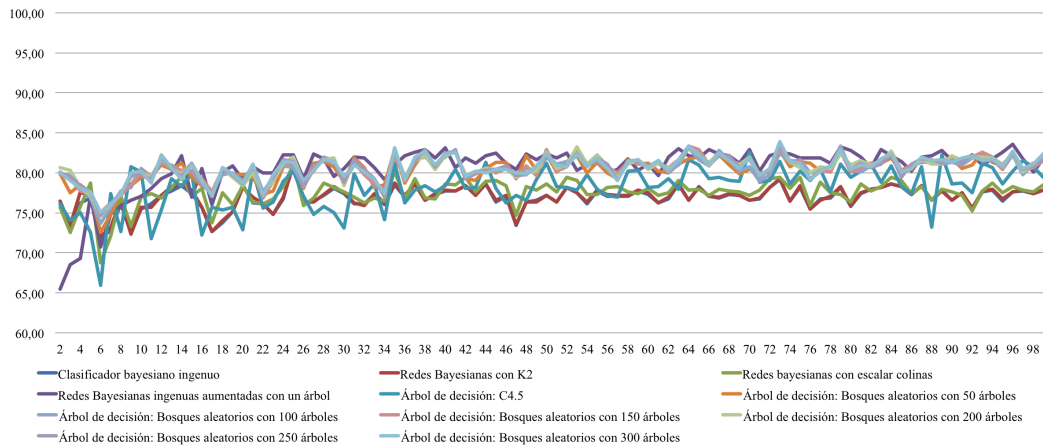
valor que alcanzará la carga máxima en un estado localizado en un futuro próximo, nos ha permitido comprobar que el método de compresión mantiene la representación estadística extraída de los datos originales. Así, los resultados que se obtienen son más que prometedores (aproximadamente se mantiene la precisión y la gestión de errores de estos clasificadores) para que el método de compresión bayesiana basado en métodos de Montecarlo sea adecuado para un aprendizaje acumulado que adaptaría los modelos al nuevo proceso productivo.

En el segundo de los experimentos, y con el fin de poder realizar una serie de comparaciones, hemos llevado a cabo el mismo tipo de mediciones. Por lo que, las Figuras 5.11a y 5.11b muestran los resultados obtenidos por los clasificadores aprendidos con el conjunto de datos reducido al completo y testados con el conjunto de datos original en términos de precisión. Por otra parte, las Figuras 5.12a y 5.12b muestran toda la información relativa a las tasas de error medidas como el área bajo la curva ROC. A modo de que todos los experimentos puedan ser comparados y fácilmente legibles, hemos continuado con el mismo tipo de notación en el eje de abscisas, por consiguiente, indicamos el porcentaje de instancias que han sido generadas para crear los conjuntos de datos. Nótese que estos experimentos hacen referencia a las pruebas de los diferentes conjuntos de datos reducidos como datos de aprendizaje y el conjunto de datos original como el conjunto de datos de prueba. Así, en este experimento no será aplicada la validación cruzada. Y además, para ello hemos utilizado los mismos conjuntos de datos reducidos (debido al no determinismo del muestreo de los métodos de Montecarlo) y los mismos clasificadores que llevamos utilizando en esta tesis doctoral.

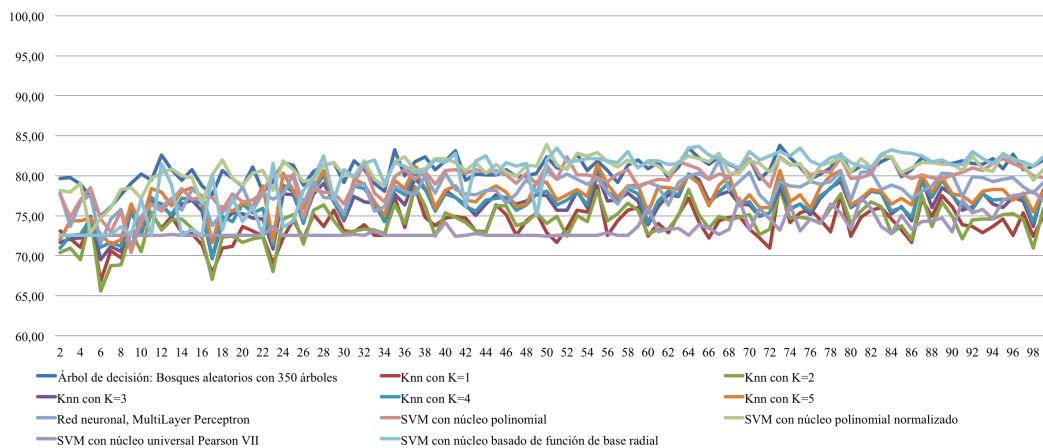
Específicamente, la Figura 5.11a hace referencia a los valores obtenidos por un primer grupo de clasificación. Los clasificadores que aquí encontramos son los mismos que hemos podido analizar dentro en el primer grupo del resto de los experimentos. En este caso, podemos observar que el porcentaje de precisión se ha reducido sensiblemente, pero tanto para los diferentes porcentajes de compresión, como para el original. Esto se debe a que se ha eliminado la ponderación introducida por la validación cruzada. No obstante, los resultados obtenidos se encuentran muy cerca de los valores obtenidos por los clasificadores sin ningún tipo de compresión (ver la sección 4.3.2.2) dándose una pequeña reducción de 5 unidades en el mejor de los casos y 10 en el peor de ellos.

Aunque se ha producido esa pequeña reducción en la precisión, podemos observar que la forma en la que trabaja nuestro método nos es familiar, es decir, nos encontramos las curvas de precisión que cuentan con una serie de imperfecciones o fluctuaciones generadas por las mismas razones que en los

5. Adaptando los modelos al proceso cambiante



(a) La primera figura muestra el comportamiento obtenido en lo que niveles de precisión se refiere por el primer grupo de clasificadores. En este caso, debido a la calidad inferior de los datos comprobamos que las fluctuaciones aumentan en número y en desplazamiento. Sin embargo, el comportamiento similar se mantiene.



(b) La segunda figura muestra cómo se han comportado el segundo grupo de clasificadores. Observamos que, como en todas las gráficas anteriores, el comportamiento es similar destacando los resultados casi lineales del clasificador SVM con función de núcleo Pearson VII para gran parte de los porcentajes de compresión.

Figura 5.11: Resultados obtenidos durante la comprobación de los modelos generados a través de un conjunto de datos reducido frente al conjunto de datos original con el fin de la detección de posibles problemas de carga máxima. Se muestran los datos en términos de precisión para cada uno de los porcentajes de compresión. Al no utilizar la validación cruzada los resultados no superan los anteriores.

casos anteriores. No obstante, teniendo en cuenta la tendencia seguida por esas curvas y ajustando su recorrido, podemos observar que la predisposición al funcionamiento de los clasificadores vuelve a ser muy pareja. De este modo, a pesar del desplazamiento en la precisión (es decir, un desplazamiento vertical en la curva), los máximos y los mínimos obtenidos a lo largo de todo el proceso de compresión prácticamente coinciden.

Como ya pudimos comprobar en los experimentos sobre los conjuntos de datos reducidos y los modelos que hicieron uso de ellos, obtenemos dos secciones a lo largo de todo el rango de compresiones. La sección que permite obtener el mejor rendimiento en la precisión y la que muestra un decremento en los niveles de acierto de los modelos aprendidos. Particularmente, este caso combina lo que ya descubrimos con el defecto de los microrrechupes, es decir, un descenso paulatino del rendimiento para grandes valores de compresión; junto al caso descubierto en el experimento anterior, en otras palabras, una ampliación de la amplitud de las fluctuaciones. De este manera, la mayoría de los clasificadores tienen un descenso de unas 5 u 8 unidades, exceptuando algunos casos especiales, como los árboles de decisión que utilizaron el método C4.5 para su aprendizaje o la propia red bayesiana ingenua aumentada con un árbol, que es mayor. El segundo de los clasificadores, siendo el generador de la compresión tiene un buen rendimiento siempre que se encuentra dentro de la zona de confort. Sin embargo, al enfrentarlo al conjunto de datos original no ha sido capaz de representar tan bien el modelo como parecen hacerlo otros clasificadores. Cabe destacar que en este experimento el árbol de decisión C4.5 únicamente se convierte en el peor clasificador del grupo fuera de la zona de trabajo óptimo. Mientras tanto, dentro de la zona de confort, este clasificador es uno más del grupo.

En cuanto al segundo grupo (ver Figura 5.11b) obtenemos unos resultados muy similares a los del primer grupo. En este caso, el porcentaje de precisión es más reducido que en el primero de ellos. Pero como ya venimos comentando, estos clasificadores no fueron capaces de obtener mejores resultados que sus compañeros cuando estaban siendo probados para la predicción del estado estacionario $t + 1$.

No obstante, y de igual modo que el primero de los grupos, estos modelos o clasificadores también han sufrido una reducción en el porcentaje de precisión. Pero, a pesar de este hecho, las tendencias de ejecución presentes en los comportamientos de cada uno de los clasificadores vuelven a ser similares a lo que estamos viendo. De esta forma, cuanto mas incrementamos el grado de compresión, peores representaciones del proceso obtenemos y los clasificadores

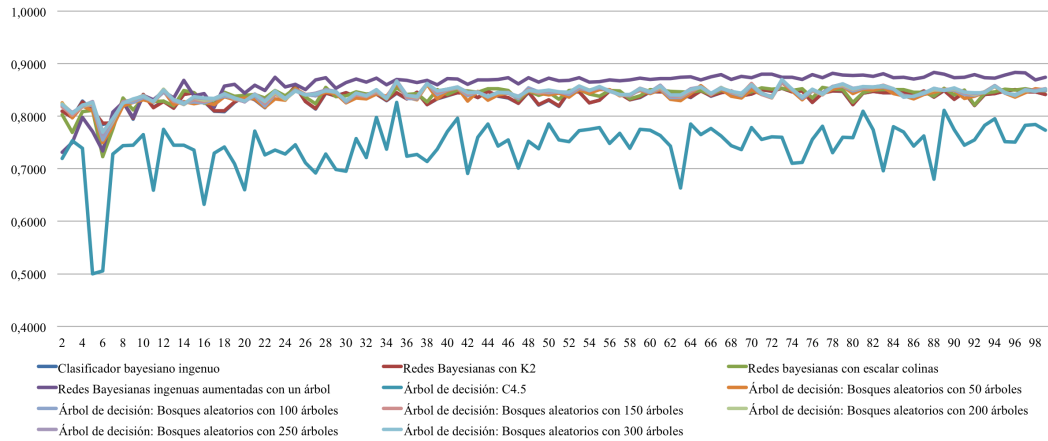
5. Adaptando los modelos al proceso cambiante

no son capaces de adaptarse tan bien al mismo. A pesar de disponer de ese grado de descenso, el porcentaje de precisión no decrece significativamente. Así, éstos sufren una pérdida de unas 5 unidades en su tasa de acierto.

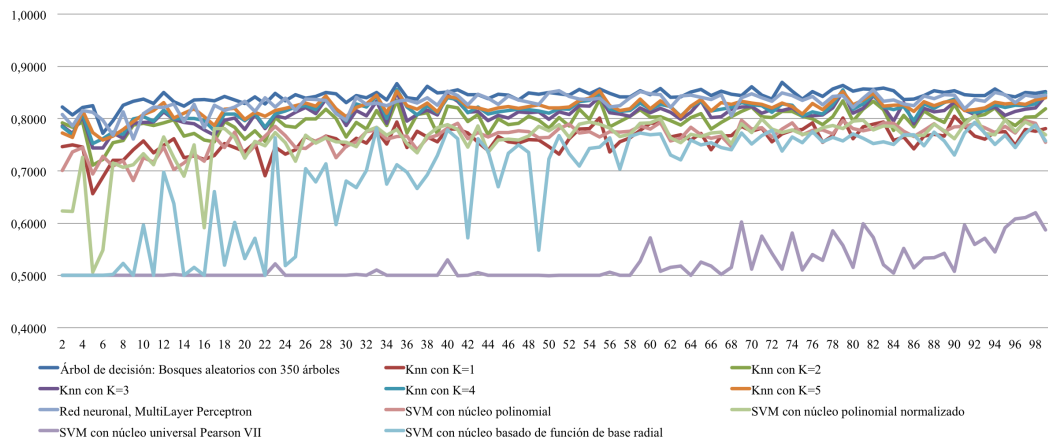
En el caso que nos ocupa, volvemos a determinar fácilmente las dos áreas de comportamiento de los clasificadores. Claramente, cerca de los valores que indican que el conjunto de datos ha sido generando utilizando el 24 % de las instancias, comenzamos a ver cómo la horizontalidad de las tendencias pasa a inclinarse hacia abajo. Además, en este caso, al igual que en el primer grupo de clasificación, observamos que la amplitud de las perturbaciones crece pero no de una manera tan descontrolada como en los experimentos realizados al probar los clasificadores por sí mismos. También, destacamos que el clasificador SVM con función Pearson VII no se trata del peor de los clasificadores. Hay otros que en ciertos porcentajes de compresión consiguen descender de los valores que logra este SVM. Algunos ejemplos de esta situación son ofrecidos por los clasificadores KNN con diferentes valores de K . Pese a ello, la tendencia general que sigue el clasificador SVM con función de núcleo Pearson VII en el proceso de predicción de la carga máxima es mayoritariamente lineal haciéndole mantenerse en las últimas posiciones a lo largo de todos los porcentajes de compresión posibles.

Pasando ahora a la gestión de errores, la Figura 5.12a nos indica los resultados obtenidos por el primer grupo de clasificadores en relación a la gestión de los errores. Como ya se ha comentado, los errores los medimos a través del área generada bajo la curva ROC. Así, en el caso de los errores, los comportamientos de todos los clasificadores de este grupo son mucho más similares entre sí, mucho más que cuando realizamos la medición de la precisión. Tanto es así, que volvemos a ver los tres grupos que ya pudimos identificar en el experimento de los modelos por sí solos. Así, vemos que a pesar de que la red bayesiana ingenua aumentada con un árbol (TAN) tiene un descenso significativo en el rendimiento fuera de la zona de confort, esta vez, para los errores, su comportamiento vuelve a ser uno de los mejores, manteniéndose bastante constante y teniendo una bajada de rendimiento (la más grande de este grupo) de menos de dos décimas. Después, podemos incluir a los bosques aleatorios y a los demás clasificadores bayesianos en el grupo de la normalidad. Estos clasificadores tienen un comportamiento muy constante a lo largo de todo el espectro de compresiones. Su comportamiento es tan constante que sólo comienza a desfallecer cuando se está llegando a un grado de compresión de un 94 % (generación del nuevo conjunto de datos utilizando únicamente el 6 %). Sorprendentemente, la desviación máxima alcanzada en el punto mencionado

5.3 Validación empírica



(a) En la primera de las figuras se muestra que todos los clasificadores han obtenido unas tasas de gestión de errores muy similares (rondan el 0,85 de área bajo la curva ROC). Como en el resto de casos, el árbol de decisión construido a través del algoritmo C4.5 no alcanza a sus compañeros.



(b) La segunda figura muestra que las tendencias en los comportamientos son similares pero obteniendo resultados que varían entre un área del 0,75 y el 0,85. Sin embargo, como viene pasando en el resto de los experimentos, el clasificador SVM creado con funciones de núcleo RBF y Pearson VII no se ajusta tan bien como el resto.

Figura 5.12: Tasas de error obtenidas al enfrentar los modelos aprendidos con el conjunto reducido a los datos originales manteniendo el objetivo de predecir la carga máxima. Se muestran los datos en términos de área bajo la curva ROC para cada uno de los porcentajes de compresión. Por lo general se adquieren buenos resultados a excepción de alguno de los clasificadores que se separa de sus compañeros.

5. Adaptando los modelos al proceso cambiante

es de una décima. Por último, tenemos el árbol de decisión C4.5 que para la gestión de los errores vuelve a ser el peor clasificador de este grupo. Igual que en los casos anteriores no puede ser identificado como el peor de todos los clasificadores, ya que en el segundo grupo se encuentran otros con peores resultados. La diferencia de este clasificador con el resto de los que conforman el primero de los grupos es el alto grado de variaciones con gran amplitud que sufre su curva.

Para este caso, es complejo descubrir las zonas de confort y de descenso de los resultados. El comportamiento de la mayoría de los clasificadores ha sido muy lineal. Sí que podemos asegurar que una compresión muy grande perjudica el rendimiento en la gestión de los errores. No obstante, se puede observar que la zona comprendida entre una generación con el 4 % de los datos y con el 8 % sufre una fuerte caída del rendimiento. De esta forma, estudiando las curvas que más descienden (i.e., la relacionada con la red bayesiana TAN y el árbol de decisión generado con el método C4.5) observamos que existe una zona en la que la horizontalidad se torna en una pendiente descendente. El punto identificado se encuentra en el rango de valores que van desde la generación del 24 al 28 % de las instancias. De esta forma, conseguimos realizar aproximadamente las mismas dos divisiones con las que ya contábamos en el experimento anterior y en los resultados de precisión de este experimento.

Por otra parte, la Figura 5.12b muestra los resultados relacionados con las tasas de error que obtuvieron los clasificadores del segundo grupo a la hora de realizar la predicción del estado estacionario $t + 1$, y así, determinar posibles fallos en los valores de la carga máxima. Algo habitual que vamos observando a lo largo de los experimentos es que este segundo grupo nunca tiene los valores tan unificados como el primero de los grupos. A pesar de ello, sorprendentemente, esta vez las curvas que representan el comportamiento para todos y cada uno de los clasificadores de este grupo son más cercanas que en ninguno de los otros experimentos. Además, las tendencias mostradas no sufren de tantas fluctuaciones como en ocasiones anteriores. Los resultados de este grupo no consiguen superar a los alcanzados por los clasificadores del grupo anterior, sin embargo, la diferencia no es tan grande, encontrándose estos clasificadores en el rango que va desde las 0,75 unidades a las 0,86 unidades aproximadamente.

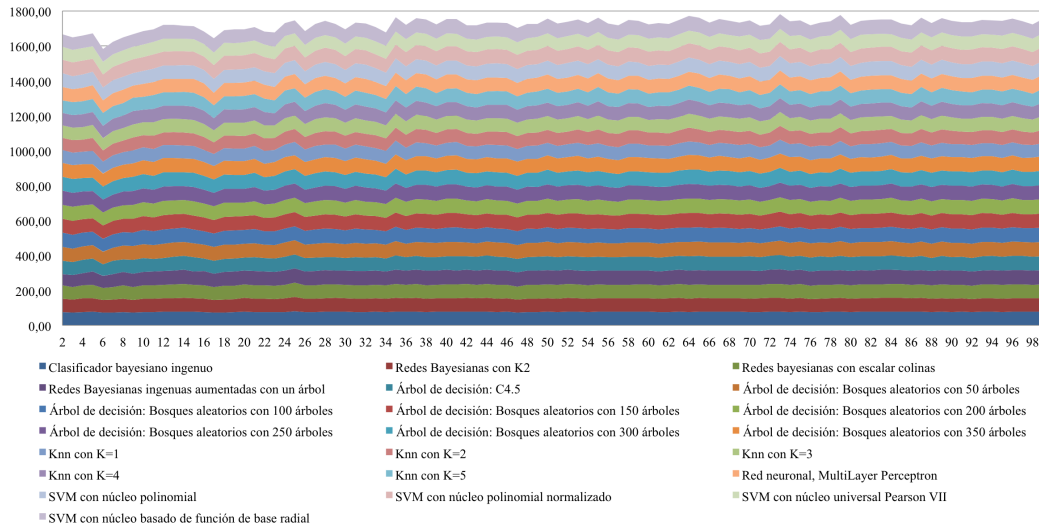
En este grupo, debido a la paridad que hay entre los resultados obtenidos y que las tendencias para el manejo de los errores son mayoritariamente constantes, es realmente complicado determinar un punto en el que dividir el rendimiento de los clasificadores. A primera vista percibimos que hay un des-

censo de rendimiento. De esta forma, y aplicando la misma técnica del grupo anterior, nos fijaremos en aquellos clasificadores que tienen una variación mayor. Concretamente, tomando algunos de los KNN se observa que a lo largo de una gran parte de los porcentajes de compresión se mantienen muy cercanos a la línea de las 0,8 unidades y en un momento en concreto descienden bajo ella. Igualmente, otro clasificador como las redes neuronales artificiales también sufren ese descenso. Para todos estos casos, el punto aproximado en el que se realiza ese cambio es el mismo que el que fue detectado para los clasificadores del primer grupo y en los experimentos que probaron el funcionamiento de los clasificadores por sí mismos.

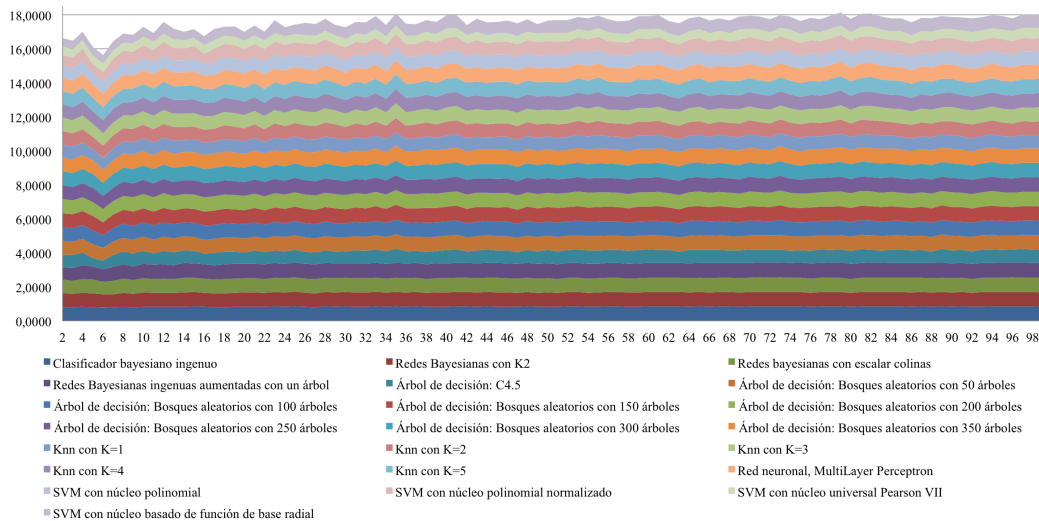
Como ya pudimos comprobar en ocasiones anteriores, en este grupo siempre tenemos una serie de excepciones de comportamiento que se encuentran relacionadas con los métodos SVM (máquinas de soporte vectorial). En especial, uno de ellos siempre se ajusta al rango medio de valores posibles a alcanzar como área bajo la curva ROC (0,5 unidades). Además, la tendencia que sigue este clasificador es totalmente lineal. Y como en anteriores experimentos, es el SVM creado con una función de núcleo Pearson VII. Otro de los clasificadores de la misma familia, el generado con una función de núcleo RBF, emula el comportamiento de otros experimentos. En otras palabras, comienza con un rendimiento en la gestión de errores similar al que disponen el resto de los clasificadores pero paulatinamente va reduciéndose el control sobre los errores. Para este clasificador, cuanto mayor grado de compresión se utiliza para la generación de los modelos, más acerca su comportamiento al del SVM con la función Pearson VII. Por último, otro de los clasificadores SVM que utiliza una función de núcleo polinomial normalizado también tiene un descenso con grandes fluctuaciones en los niveles de compresión máxima. De esta forma, en algunos momentos concretos, los resultados de este clasificador llegan a equipararse con los de sus dos compañeros.

La complejidad para la identificación de los clasificadores dentro de las gráficas de resultados, así como la legibilidad y la posibilidad de determinar correctamente las tendencias de cada uno de ellos; han hecho que decidamos realizar un pequeño resumen de los resultados a través de un gráfico de áreas apiladas. De esta forma, la Figura 5.13a ilustra los resultados de precisión apilados alcanzados por los diferentes clasificadores utilizados. Como se puede observar, la tendencia seguida por todos los clasificadores es muy similar. También, podemos comprobar cómo las fluctuaciones no disponen de una amplitud tan grande como en otros experimentos. Del mismo modo, también podemos identificar que hay una bajada en el rendimiento cerca de la zona

5. Adaptando los modelos al proceso cambiante



(a) Esta primera figura ilustra los diferentes resultados de precisión alcanzados por los clasificadores mostrando las tendencias seguidas por todos ellos de una forma mucho más legible. Como se puede observar, las tendencias son muy similares entre sí.



(b) La segunda gráfica muestra los resultados apilados relativos a las tasas de error, más concretamente, el área bajo la curva ROC que alcanzaron los clasificadores en su proceso de predicción de la carga máxima.

Figura 5.13: Resultados apilados obtenidos en los experimentos orientados a la prueba de los modelos reducidos, probados frente al conjunto de datos original, para la predicción de defectos en la carga máxima. Se recogen tanto los valores alcanzados en la precisión como aquellos relativos al área bajo la curva ROC.

de compresión máxima, y como bien indicamos anteriormente, es la cercana a los porcentajes de generación de instancias que se encuentra entre el 24 y el 28 % (la misma sección identificada en experimentos anteriores).

Por otra parte, la Figura 5.13b hace referencia a las tasas de error alcanzadas por los diferentes clasificadores y apiladas con el fin de permitir visualizar mejor los detalles ya comentados en las anteriores figuras. De esta forma, podemos ver cómo la tendencia en el manejo de errores se encuentra mucho más ajustada en este caso. Todos ellos prácticamente se comportan de forma similar. Además, la forma en la que el método de compresión les afecta es muy pareja entre todos ellos, aunque cada clasificador obtiene diferentes valores para el área bajo la curva ROC dependiendo de cómo se ajusta al proceso. Por otra parte, la difícil identificación de las zonas de trabajo queda mucho más clara a través de esta gráfica, determinando que los valores se encuentran entre el 72 y 76 % de compresión, es decir, en el mismo lugar que para la precisión.

Como conclusión, y al igual que el experimento sobre los microrrechupes, estos dos experimentos nos han permitido tomar un contacto con el modelo para la adaptación del proceso a través de la compresión Bayesiana basada en métodos de Montecarlo. En un primer lugar, hemos podido comprobar que la compresión de los conjuntos de datos es adecuada para el proceso de adaptación al mantener la distribución probabilística inherente en el proceso de fundición. Además, el primero de los experimentos nos ha permitido comprobar que la generación de los modelos a un alto grado de compresión es realmente interesante, ya que las pérdidas son sorprendentemente bajas, mucho menos de lo que podíamos esperar de antemano.

Y el segundo de los experimentos, el que nos ha hecho enfrentar esos conjuntos reducidos al conjunto de datos original, nos ha dado una visión de una actuación más real del método. En este caso, los resultados son muy esperanzadores. De esta forma, el método de compresión parece un método candidato a conseguir la adaptación de los modelos con un envejecimiento de muestras muy útil a la hora de mantener el sistema MPC actualizado a través de un aprendizaje acumulativo.

5.3.3 Predicción sobre conjuntos de datos acumulados

Una vez obtenidos los resultados relacionados con el aprendizaje de los modelos mediante la utilización de conjuntos de datos reducidos, pasaremos a presentar los resultados alcanzados en el segundo de los experimentos (det-

5. Adaptando los modelos al proceso cambiante

llado en la sección 5.3.1). Al igual que en los ensayos anteriores, realizaremos una prueba de los modelos con el conjunto de datos acumulado y posteriormente testaremos el comportamiento que tienen frente al conjunto de datos original. Como ya venimos haciendo, con el fin de aportar legibilidad a los experimentos, separaremos los resultados por el objetivo de predicción sobre el que se ha trabajado.

5.3.3.1 Microrrechupes

Hasta ahora hemos desarrollado una serie de experimentos para comprobar si los modelos generados a través de un conjunto de datos reducido son significativamente representativos, es decir, si se ajustan al proceso de producción sobre el que estamos trabajando. Sin embargo, este proceso no es la adaptación completa de los modelos. Por ello, en esta segunda parte del capítulo llevaremos a cabo una simulación de una adaptación real del proceso productivo. En el caso de los microrrechupes, el conjunto de datos originales son los que hemos utilizado a lo largo de esta tesis doctoral y hemos realizado un proceso de aprendizaje acumulado como el descrito en la sección 5.2.2. Como ya fue comentado anteriormente, hemos seleccionado el tiempo t de adaptación con una duración de medio año, ya que para la cantidad de datos que disponemos y el periodo al que pertenecen, es la opción más adecuada. Por consiguiente, el proceso de reducción ha sido realizado tres veces (i.e., reducción del conjunto de datos y anexión de las nuevas instancias).

A continuación pasamos a comentar los resultados obtenidos al realizar la predicción sobre la posibilidad de que aparezcan microrrechupes en las piezas fundidas. En estos experimentos seguiremos la misma metodología desarrollada en los anteriores. Concretamente, comprobaremos: (i) el comportamiento de los clasificadores con el fin de validar si los modelos generados cumplen con las mismas bases que un modelo aprendido con un único conjunto de datos reducido y (ii) enfrentaremos esos modelos con el conjunto de datos original para comprobar la generalización del proceso de adaptación desarrollado.

Para el primero de los experimentos, las Figuras 5.14a y 5.14b muestran los resultados alcanzados por los diferentes clasificadores una vez aplicado el proceso de adaptación. Aquí, al igual que en los experimentos previos, realizamos la medición de la precisión alcanzada por cada uno de estos clasificadores a lo largo de todos los posibles porcentajes de compresión. Posteriormente, las Figuras 5.15a y 5.15b hacen referencia a la gestión de errores que han tenido los clasificadores. Para ello, efectuamos las mediciones a través del área que se

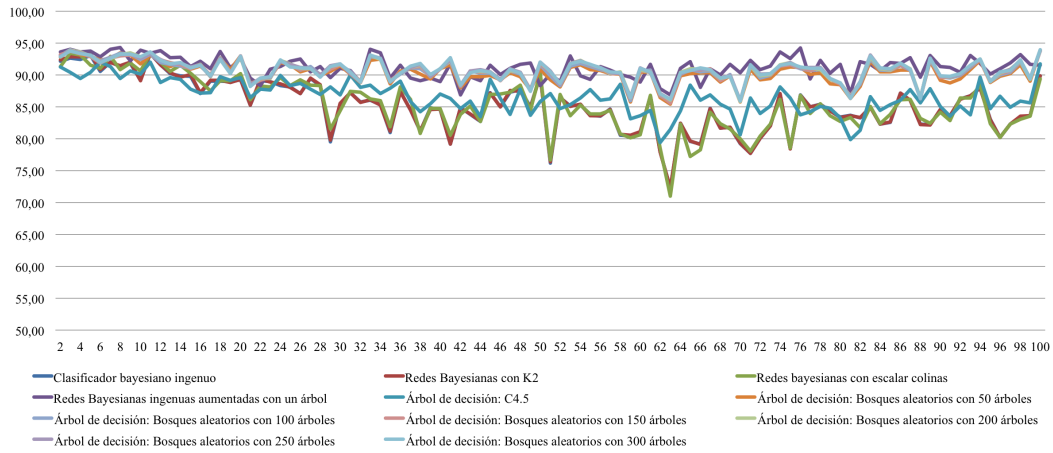
encuentra bajo la curva ROC. Del mismo modo que en anteriores experimentos, el eje de las abscisas nos muestra el porcentaje de instancias que han sido generadas en cada una de las reducciones del proceso de adaptación. Nótese que en el proceso de adaptación, siempre que se ha realizado una reducción del conjunto de datos le hemos aplicado el mismo grado de reducción.

En particular, la Figura 5.14a ilustra los resultados del primero de los grupos de clasificación. La separación de grupos que se ha realizado es la misma que en experimentos anteriores con el fin de poder llevar a cabo las pertinentes comparaciones. A primera vista, se puede distinguir una clara tendencia que afecta a todos los clasificadores. Estos, son capaces de adaptarse mejor al proceso cuando los grados de compresión son muy bajos o muy altos. No obstante, los grados de compresión medios no obtienen malos resultados, pero en comparación, los mejores porcentajes de precisión se alcanzan siempre en los extremos. Este tipo de resultados nos ilustran que el proceso de adaptación con el envejecimiento de muestras desarrollado a través de la compresión de los datos es viable. También, nos indica que el proceso de envejecimiento es correcto al conseguir unos porcentajes de acierto muy similares a lo que cada uno de los clasificadores pudieron conseguir con el conjunto de datos al completo.

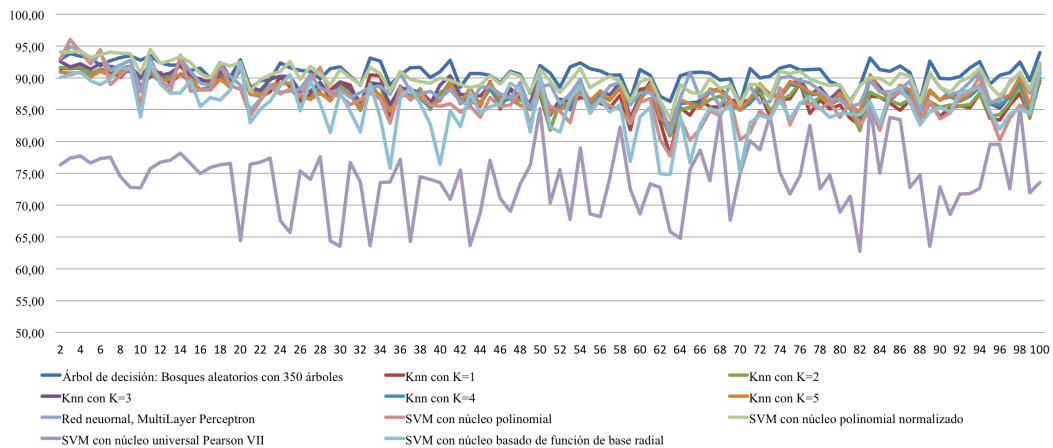
Tal y como viene sucediendo en los experimentos de compresión, las tendencias generales que siguen los clasificadores son muy parecidas. Así, la tendencia general de este grupo es la de mantener una precisión un poco más baja a la que se obtiene con el aprendizaje del conjunto de datos original para, posteriormente, acercarse a un grado de compresión del 74 % (i.e., una generación del 26 % de las instancias) y comenzar a incrementar su rendimiento. Sorprendentemente, los grados de compresión que mejores resultados dan son aquellos que fueron identificados en los experimentos anteriores como la franja de funcionamiento con la precisión más baja. Pero recordemos que aquí el conjunto resultante no es un conjunto totalmente reducido, sino que se han ido acumulando instancias y hace que este grupo sí mantenga mucho mejor la representación del proceso.

En cuanto a los clasificadores, en general, dentro de este último rango de compresiones, el máximo, todos acaban acercándose entre ellos consiguiendo valores que van del 85 al 95 % de acierto. Y si reducimos mucho más esta franja, [2, 10], los resultados de precisión obtenidos no bajan del 90 %. Nótese, que muchos de los clasificadores llegan a mejorar los resultados que obtuvieron al utilizar el conjunto completo de datos. En el otro rango de compresiones, los clasificadores pueden ser agrupados en tres grupos. El primero de ellos es

5. Adaptando los modelos al proceso cambiante



(a) La figura muestra el primer grupo de clasificadores, en el que cuantas menos instancias hemos creado para la acumulación, indica que se ha alcanzado una mejor adaptación. Los clasificadores que se desvían un poco de la normalidad son las redes bayesianas aprendidas con escalar colinas y K2.



(b) Mostramos el segundo conjunto de clasificadores que utilizan un conjunto de datos que utilizan un conjunto de datos acumulado para la predicción de los microrrechupes. Las tendencias son similares a excepción que la tomada por el SVM con función de núcleo Pearson VII.

Figura 5.14: Resultados obtenidos durante la comprobación de los modelos generados a través de un conjunto de datos acumulado con el fin de la detección de posibles problemas de microrrechupes. Se muestran los datos en términos de precisión para cada uno de los porcentajes de compresión. Aunque se observan fluctuaciones en los comportamientos de los clasificadores, sí que se percibe una tendencia que indica que cuando mayor es la compresión, mejor es la adaptación que sufren todos y cada uno de los modelos predictivos.

el compuesto por las redes bayesianas ingenuas aumentadas con un árbol (en inglés “*Tree Augmented Naïve*”) y los bosques aleatorios. Estos clasificadores obtienen resultados muy parejos y tendencias muy similares, sin embargo, y como ya viniera pasando anteriormente, las redes bayesianas que utilizan el algoritmo de aprendizaje estructural TAN son las que destacan durante los grados de compresión más bajos. Esto, como ya fue explicado, se debe a que este ha sido el modelo utilizado para la extracción de la distribución probabilística y, al volverlo a generar, los datos están perfectamente ajustados a su forma de funcionamiento. Por otra parte, encontramos el árbol de decisión aprendido a través del algoritmo C4.5 que a pesar de contar con una tendencia similar, las fluctuaciones lo hacen separarse del resto de los clasificadores. En este grupo, aunque en el rango de máximas compresiones es el que peores resultados obtiene, en el rango de mínimas compresiones se mantiene por encima de los demás clasificadores bayesianos. Estos clasificadores bayesianos parecen no ajustarse demasiado bien al proceso de compresión. Los resultados que llegan a alcanzar en la zona central del espectro de compresiones son mucho más bajos que los que pueden lograr con el conjunto original. Sin embargo, a la hora de acercarse a los altos grados de compresión sufren una subida vertiginosa de rendimiento, llegando a igualarse al resto de los clasificadores.

En el segundo grupo de clasificadores (ver Figura 5.14b) obtenemos unos resultados muy similares. Observamos que la gran mayoría de los clasificadores desarrolla un comportamiento tal que los resultados de clasificación obtenidos sin realizar ninguna compresión sobre el conjunto de datos son mucho mejores que en el caso de comenzar con bajos porcentajes de compresión. Alrededor de una generación de evidencias equivalente al 63% de instancias sobre el conjunto original se obtiene el mínimo de su comportamiento para, desde ahí, comenzar a mejorar las tasas de acierto. De esta forma, comprobamos que estos clasificadores se ven afectados de la misma forma que los pertenecientes al grupo anterior. Habitualmente, este segundo grupo solía contar con un mayor grado de dispersión en los clasificadores, es decir, que los resultados relativos a la precisión eran muy dispares. A pesar de esto, al aplicar un proceso de aprendizaje acumulativo todos los clasificadores (a excepción de las máquinas de soporte vectorial con función de núcleo Pearson VII) se agrupan obteniendo unos resultados muy similares.

Dentro de este grupo también detectamos claramente las dos secciones de rendimiento. Cabe destacar también que, los porcentajes de compresión sobre los que comienza la mejora de rendimiento coinciden con los del primero de los grupos. Y además, con la reducción del rango, el mismo que en el primer grupo

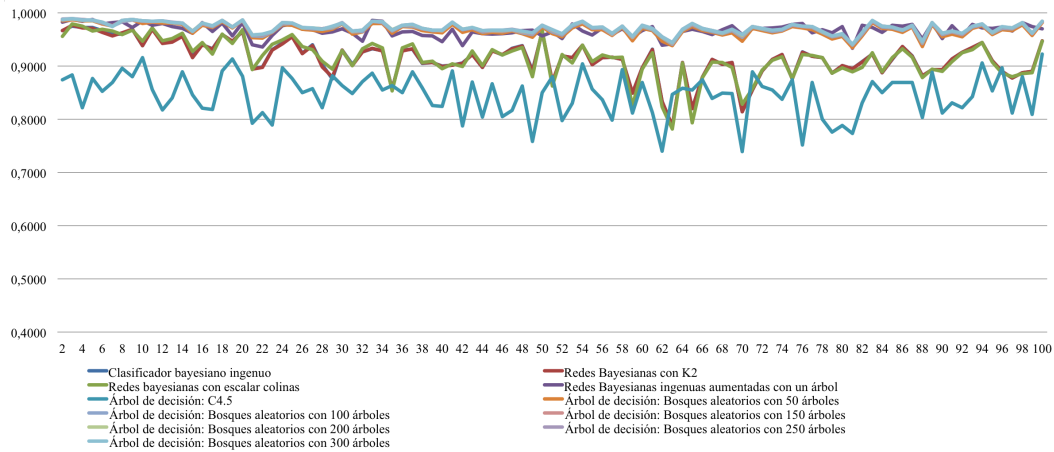
5. Adaptando los modelos al proceso cambiante

[2,10], nos volvemos a encontrar que la precisión no desciende del 90 %. De esta forma los clasificadores aproximadamente mantienen o mejoran los resultados obtenidos con el conjunto de datos original. Aspecto harto interesante, el hecho de que aplicando este método de reducción, hemos conseguido ajustar el proceso, obteniendo un porcentaje de clasificación superior al 90 % para todos los clasificadores excepto uno, lo que supone una gran mejora frente a los experimentos mostrados en el capítulo 4. Y por ende, mejorando los integrantes del meta-clasificador, mejoraremos los sistemas de meta-clasificación.

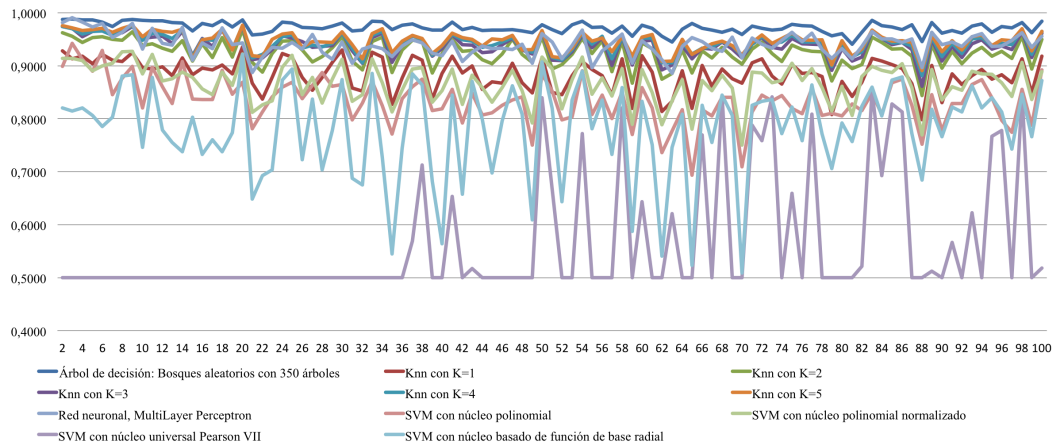
Dentro de este segundo grupo cabe destacar dos comportamientos que se desvían de la normalidad. El primero de ellos es el que desarrolla el SVM con función de núcleo RBF (del inglés “*Radial Basis Function*”). Este clasificador comparte una tendencia en su precisión muy similar a la del árbol de decisión creado con C4.5. En otras palabras, durante bajos niveles de compresión se mantiene por la zona baja de la tendencia normal y presenta grandes fluctuaciones que le hacen bajar hasta unos aciertos del 75 %. Sin embargo, al acercarse a la zona de rendimiento óptimo, este clasificador mejora significativamente para, finalmente, volver a unirse con el resto de los clasificadores obteniendo resultados superiores al 90 %. El otro de los clasificadores que merece ser comentado es el SVM con núcleo Pearson VII. Al igual que en el resto de los experimentos, observamos un rendimiento muy bajo. A parte de que ya empieza con un bajo porcentaje de acierto cuando no aplicamos ningún tipo de compresión, se mantiene por debajo del conjunto de clasificación normal. Además, sufre de unas amplias fluctuaciones. De largo es el peor de los clasificadores en la predicción de los microrrechupes mediante la utilización del sistema de aprendizaje acumulado presentado en este capítulo. No obstante, este clasificador también consigue mejorar durante las etapas de compresión máxima, consiguiendo superar la barrera del 75 %, la cual no es capaz de sobrepasar cuando se hace uso del conjunto de datos original al completo (73,54 %).

En lo que se refiere a los errores, podemos observar que el comportamiento es muy similar al de la precisión. La Figura 5.15a ilustra los resultados obtenidos en términos de área bajo la curva ROC por los clasificadores pertenecientes al primero de los grupos. De esta forma, cabe desatacar los resultados que se han obtenido. En general, muchos de los clasificadores se mantienen durante todo el proceso de compresión muy cercanos a los valores óptimos de área. Otros se mantienen cerca de la barrera de los 0,90 para acabar acercándose al óptimo (con valores superiores al 0,95). Estos últimos clasificadores son las diferentes redes bayesianas, aquellas que utilizan el algoritmo K2 o el método

5.3 Validación empírica



(a) La figura muestra las tasas de error alcanzadas por un primer grupo de clasificadores basado en un conjunto de datos acumulado. Como en el caso de la precisión, son gestionadas mejor las tasas de error cuando la compresión ha sido muy grande. Los clasificadores que más se desvían son los árboles C4.5 y las redes bayesianas aprendidas con K2 y escalar colinas.



(b) En este segundo grupo encontramos que la variación es mucho mayor. Sin embargo, las tendencias de los clasificadores siguen siendo las mismas. Las máquinas de soporte vectorial con funciones de núcleo Pearson VII y RBF no pueden igualar al resto de los clasificadores.

Figura 5.15: Tasas de error obtenidas durante la comprobación de los modelos generados a través de un conjunto de datos acumulado con el fin de detectar posibles problemas de microrrechupes. Se muestran los datos en términos de área bajo la curva ROC para cada uno de los porcentajes de compresión. Mayoritariamente, los resultados se encuentran cerca del óptimo a excepción de algunos SVM.

5. Adaptando los modelos al proceso cambiante

escalar colinas para su aprendizaje estructural y, también, la red bayesiana ingenua.

Claramente, los comportamientos en la gestión de los errores se asemejan a los de la precisión. Volvemos a disponer de tres grupos diferenciados. En el primero de ellos englobamos a los bosques aleatorios y a la red bayesiana TAN. Estos clasificadores llevan una precisa gestión de los errores, a lo largo de todo el espectro de compresiones se encuentran prácticamente rozando los valores óptimos. Por otro lado, tenemos el resto de los clasificadores bayesianos. Éstos obtienen buenos resultados, pero el proceso de compresión afecta a su gestión de errores de la misma forma que le afectaba a su precisión. De este modo, mantienen unos valores inferiores a los que son capaces de obtener con el conjunto de datos original. Posteriormente, cuanto mayor es la compresión del conjunto de datos, mayor es la mejora. Así, una vez en la zona de compresión máxima, estos clasificadores permiten gestionar los errores tan bien como lo hacían los anteriores clasificadores. Por último, volvemos a tener el árbol de decisión C4.5. A pesar de ser capaz de obtener unos resultados de precisión no muy bajos, la gestión de errores que desarrolla no es tan buena. Por tanto, a parte de mantener un alto grado de fluctuaciones con una gran amplitud, no consigue llegar a mejorar la gestión de los errores que se tenía sobre el conjunto original de datos.

Para la gestión de errores también encontramos dos secciones separadas en las que los rendimientos son diferentes. Dejando de lado el árbol C4.5 que es el más irregular de todos, el resto de clasificadores sufre una mejora en su gestión de errores cuando los porcentajes de compresión superan el 76 %. De forma análoga a la precisión, volvemos a obtener unos resultados cercanos a los logrados por los clasificadores sin el efecto de la compresión bayesiana basada en métodos de Montecarlo. Por consiguiente, podemos determinar que los modelos resultantes de una acumulación a altos grados de compresión disponen de un envejecimiento óptimo y respetan correctamente el proceso productivo que se está adaptando.

Por otra parte, en la Figura 5.15b podemos observar el comportamiento en la gestión de errores que tienen los clasificadores del segundo grupo. En este segundo caso, volvemos a comprobar que los resultados entre los diferentes clasificadores vuelven a dispersarse. Sin embargo, esa dispersión no es tan grande como la que aparece en anteriores experimentos. También, apartándonos de la normalidad encontramos ciertas excepciones que serán comentadas a continuación.

En primer lugar, hay una serie de clasificadores, como los bosques aleato-

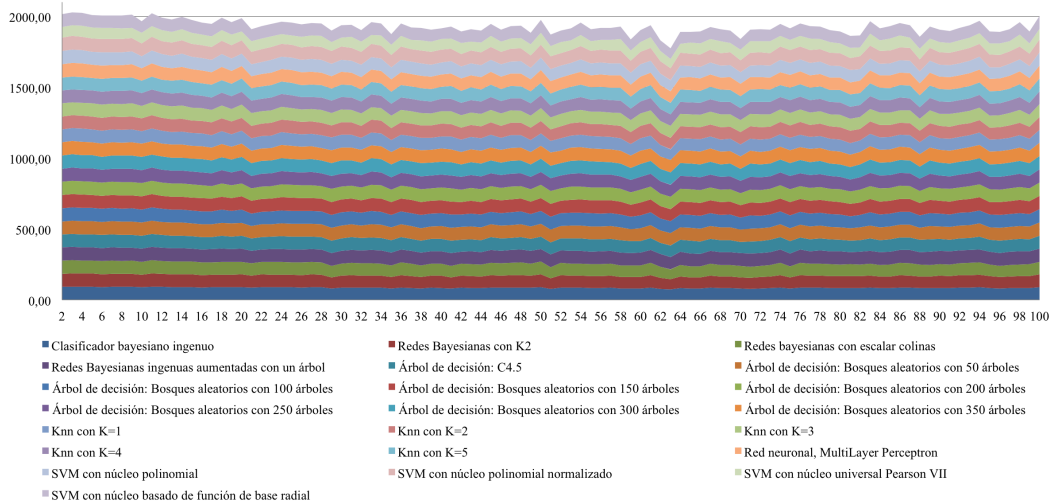
rios con 350 árboles, las redes neuronales artificiales y los KNN con valores de $K \in [2, 5]$, $K \forall \mathbb{N}$, que siguen las mismas pautas que la normalidad del primero de los grupos. Estos clasificadores sufren un pequeño descenso en el área bajo la curva ROC en la zona central del espectro de compresiones. No obstante, los resultados que obtienen a lo largo de todos los porcentajes casi siempre son superiores a las 0,90 unidades y muy cercanas al valor óptimo.

En segundo lugar, podemos agrupar a los clasificadores SVM con funciones de núcleo polinomial y polinomial normalizado y el KNN con $K = 1$. Estos clasificadores siguen aproximadamente la tendencia que llevan los ya comentados. No obstante, sus resultados son más bajos, por lo que su gestión de errores no es tan buena. Además, no son tan constantes en su rendimiento. Esto puede observarse debido a la gran cantidad de fluctuaciones con grandes valores de amplitud con las que cuenta su curva de comportamiento. Pero a pesar de ello, los resultados que siguen obteniendo se pueden equiparar a los iniciales obtenidos con el conjunto de datos al completo. Por lo que el método de compresión bayesiana basado en métodos de Montecarlo sigue siendo una buena técnica de adaptación del proceso.

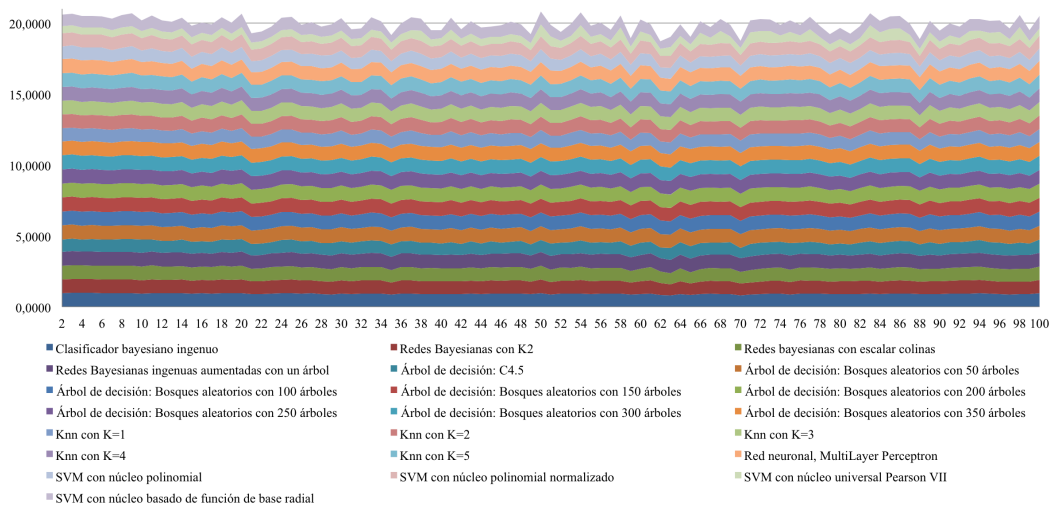
Finalmente, tenemos el grupo formado por los SVM con funciones de núcleo RBF y Pearson VII. El primero de ellos mantiene la tendencia de descenso para compresiones bajas y ascenso de compresiones altas. Pero en este caso, el resultado que se obtiene se encuentra muy cercano a los resultados originales (se aproxima sin lograr sobrepasarlo). Igualmente, este es uno de los clasificadores que mayor amplitud de fluctuación tiene, con lo que el método le ha afectado bastante en la gestión de los errores. A pesar de ello, este clasificador no es el que peor gestiona los errores, de hecho, ese es el segundo de los SVM que estamos comentando. En particular, el clasificador SVM con función de núcleo Pearson VII muestra un comportamiento bastante extraño en la gestión de errores. Básicamente, intenta reproducir lo que ya visualizamos en otros experimentos, es decir, mantenerse en una tasa de gestión de errores de 0,50 unidades. Sin embargo, sufre grandes desviaciones con gran amplitud (llega a alcanzar más de las 0,80 unidades). En los casos anteriores, también observábamos este tipo de fluctuaciones, aunque eran mucho más pequeñas. Gracias a este comportamiento irregular, a su tasa en la gestión de errores y a su precisión, consideramos que este clasificador es el peor de los dos grupos.

La dificultad de visualizar los diferentes comportamientos en las gráficas anteriores nos hace mostrar la información de una segunda forma que complementa a la primera. Es por ello, que la Figura 5.16a muestra un resumen apilado de los resultados obtenidos por los clasificadores en términos de pre-

5. Adaptando los modelos al proceso cambiante



(a) A modo de simplificación, la primera figura muestra las tendencias seguidas por todos los clasificadores en términos de precisión en la predicción de la aparición de microrrechupes analizada en este experimento. Como se puede observar, las tendencias seguidas por todos ellos son muy similares.



(b) En el segundo de los gráficos apilamos todos los resultados del área bajo la curva ROC alcanzados por los clasificadores a la hora de prever la aparición de microrrechupes.

Figura 5.16: Resultados apilados obtenidos en los experimentos orientados a la prueba de los modelos acumulados para la predicción de defectos de microrrechupe en las piezas extraídas de la fundición. La figura recoge tanto los valores alcanzados en la precisión como los alcanzados en el área bajo la curva ROC.

cisión. Ahora, queda ilustrado de forma mucho más clara cómo las tendencias seguidas por todos ellos son muy similares. Igualmente, se comprueba cómo el comportamiento que siguen cuenta con una zona (alrededor de un 40% de compresión) en la que encontramos la mínima precisión del comportamiento. También, observamos cómo en los grados de compresión máxima obtenemos los mejores resultados, superando los totales apilados que generan los clasificadores al realizar el proceso de aprendizajes y pruebas sin ningún tipo de compresión y acumulación.

Asimismo, la Figura 5.16b hace lo propio con los resultados del área bajo la curva ROC que obtuvieron los clasificadores. Como se puede observar, al apilar los resultados se distingue mejor las tendencias similares que siguen los clasificadores. Del mismo modo, y como ya ilustrase la Figura 5.16a, también somos capaces de advertir la zona de resultados más baja (aquella en la que el grado de compresión de los conjuntos de datos corresponde al 38%, muy cercano al descubierto en la precisión). También nos muestra que a grados de compresión altos se mejoran los resultados, y teniendo en cuenta los valores apilados de todos los clasificadores, se llega a mejorar el rendimiento en cuestión para manejo de errores.

A modo de conclusión, este experimento nos ha demostrado el buen rendimiento que podemos alcanzar cuando realizamos la adaptación del proceso. Es ya conocido el no determinismo de los métodos de Montecarlo, que hacen que encontremos fluctuaciones, sin embargo, las tendencias generales que se observan nos llevan a conseguir un sistema incluso mejor que el que se puede lograr utilizando todos los datos.

Otra característica que hemos probado es la no existencia de desajustes producidos por la compresión acumulativa de los datos. Llegamos a esa conclusión ya que estos resultados, tanto en términos de precisión como en gestión de tasas de error, se encuentran dentro de un rango de funcionamiento más que aceptable para el trabajo diario en una planta de producción. Por lo tanto, y a pesar de que queda un nuevo experimento que comprobará estos modelos con el conjunto de datos original, podemos asegurar la total viabilidad del proceso de compresión bayesiana basado en métodos de Montecarlo como la herramienta necesaria en el sistema de autoadaptación de los modelos MPC.

Una vez finalizado el primer experimento, y probado el buen rendimiento que se obtiene en los modelos adaptados a pesar de llevar a cabo un proceso de compresión y aprendizaje acumulativo, y observando que los resultados son equiparables a lo que realizan los modelos bajo un aprendizaje realizado con los datos completos, ahora queremos probar qué comportamiento tienen

5. Adaptando los modelos al proceso cambiante

estos modelos generados mediante compresiones acumuladas frente a los datos originales del sistema.

Así, en este experimento hemos utilizado los mismos clasificadores con las mismas configuraciones que llevamos usando a lo largo de todos los experimentos de esta tesis doctoral. La aplicación del conjunto de datos original nos permite comprobar el comportamiento real esperado por el sistema de adaptación consiguiendo desarrollar una simulación mucho más real de lo que sucedería en un planta de producción. Un punto clave para realizar las comparaciones pertinentes con el experimento con validación cruzada es la utilización del mismo conjunto de datos. Los métodos de Montecarlo no son deterministas, por lo que los conjuntos de datos generados para un porcentaje de compresión en concreto diferirían cada una de las veces, con lo que en ciertas ocasiones puede ajustarse mucho mejor que en otras. Entonces, con el fin de eliminar esa problemática vamos a emplear los mismos conjuntos de datos. Y para terminar de describir el entorno de experimentación utilizado comentaremos que en este caso se ha prescindido del proceso de validación cruzada, por lo que los conjuntos de aprendizaje serán utilizados completamente en la etapa de generación del modelo. Por otra parte, como conjunto de prueba usaremos todas y cada una de las evidencias recogidas en el conjunto de datos original.

Las mediciones serán realizadas en los mismos términos que en los experimentos anteriores, así, la primera de las medidas obtenidas para los clasificadores es el porcentaje de precisión que han alcanzado al realizar las pruebas. Esta información queda recogida en las Figuras 5.17a y 5.17b. También, se realizará una medición de la gestión de las tasas de error por cada uno de los clasificadores. Así, el área existente bajo la curva ROC ha sido seleccionada para este fin. La relación de errores queda recogida en las Figuras 5.18a y 5.18b. Con el fin de evitar problemas, debemos destacar que la numeración del eje de abscisas en todas las gráficas es el porcentaje de instancias generadas durante el proceso de compresión.

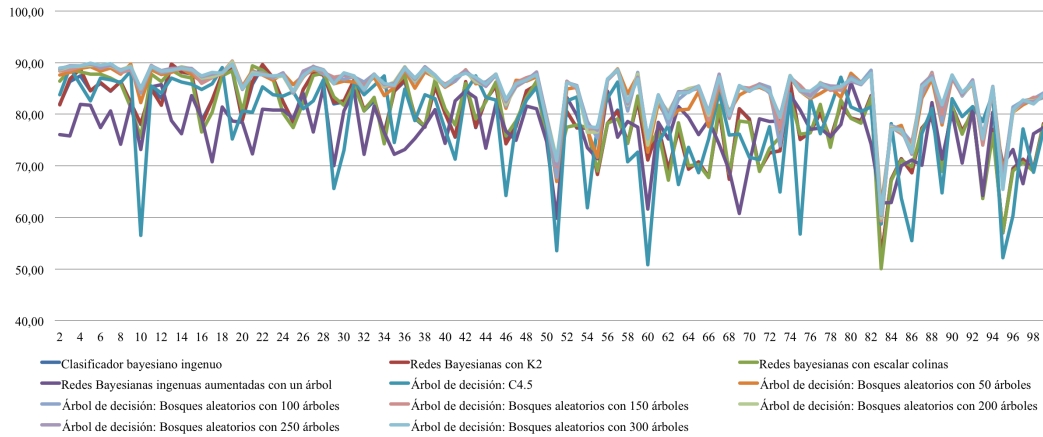
Específicamente, en la Figura 5.17a podemos observar los resultados en términos de precisión que han sido obtenidos por los clasificadores del primero de los grupos. A la hora de realizar la comprobación de los modelos con un conjunto de datos acumulado vemos cómo desaparece ese punto mínimo en la zona media del espectro de compresiones. Ahora únicamente podemos percibir una tendencia de incremento de la precisión según mayor ha sido la compresión de los conjuntos de datos. De esta forma, el envejecimiento de muestras es bien recogido por los clasificadores, ajustándose de una forma sorprendente al proceso y permitiendo hacer que todos excepto un clasificador

(la red bayesiana ingenua aumentada con un árbol) sean capaces de mejorar el rendimiento logrado cuando no ha habido ningún tipo de compresión.

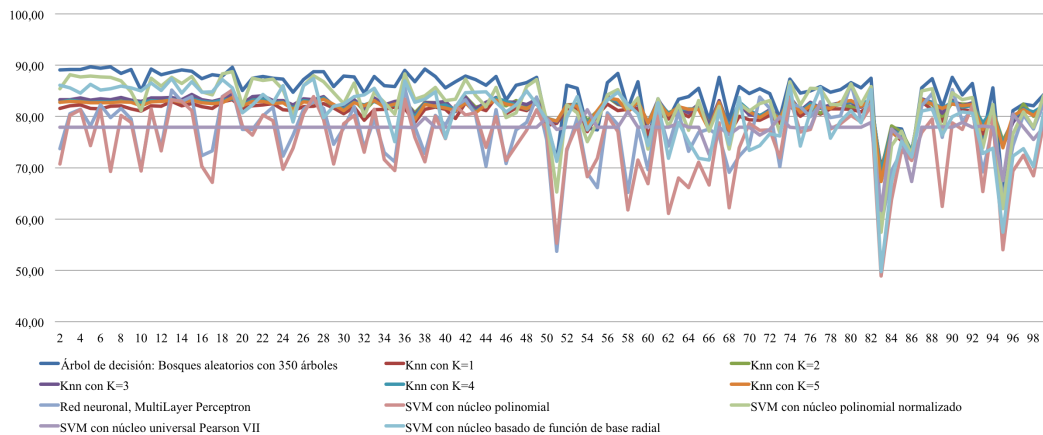
De esta forma, ahora no podemos hablar de dos secciones en las que se obtienen un buen rendimiento y un bajo rendimiento, ya que la tendencia siempre es creciente. Sin embargo, sí que podemos observar que hay una zona de compresiones en las que los resultados obtenidos son mucho más variables y las fluctuaciones tienen grandes amplitudes. Al realizar la comparación con las zonas en las que las compresiones de los conjuntos de datos son mayores y sus comportamientos más estables (mantienen fluctuaciones pero no de igual manera) podemos tomar en cuenta que el rango de compresiones que va desde el 1% al 64% es muy cambiante e inestable. En el rango contrario, el que va del 65% al 98% vemos cómo la tendencia general es la de un crecimiento asociado a unas fluctuaciones muy pequeñas. La única gran desviación es la que se encuentra en el grado de compresión del 90% (generación de únicamente el 10% de las instancias). Por consiguiente, nos encontramos que aunque los resultados sean sensiblemente más bajos, algo que ya nos sucedió en los experimentos de los modelos reducidos enfrentados al conjunto original, el proceso de adaptación mejora los resultados iniciales obtenidos al realizar las predicciones con el conjunto de datos sin procesar.

En este grupo, prácticamente todos los clasificadores se comportan de forma similar, tanto en el proceso de altas fluctuaciones como en el de mejora continua. Además, casi todos ellos consiguen una precisión en altos grados de compresión de entre el 80 y el 90%. Únicamente destacaríamos las redes bayesianas ingenuas aumentadas con un árbol ya que, siendo el modelo que genera la distribución probabilística, en el momento de realizar la prueba con el conjunto de datos original no son capaces de obtener buenos resultados. Algo similar ya nos sucedió en el experimento que se presentaba en la sección 5.3.2.1. En aquel momento, la red bayesiana TAN era la que más empeoraba en su precisión cuando aumentaban los grados de compresión. En este caso, a parte de ser el clasificador que peor precisión obtiene en esa zona, también se trata del clasificador con un comportamiento más irregular de todos ellos. Sus fluctuaciones son muy constantes en número y en ningún momento llega a superar a ninguno de sus compañeros. Del mismo modo, otro de los clasificadores que mantiene un comportamiento similar es el árbol de decisión generado con el algoritmo C4.5. Además de ser el que mayor amplitud obtiene en sus fluctuaciones, es el que en todo momento obtiene los peores resultados. Sin embargo, y en contraposición a lo que hacen las redes bayesianas TAN, este clasificador es capaz de mejorar al igual que lo hace el resto, por lo que,

5. Adaptando los modelos al proceso cambiante



(a) La figura muestra los resultados de precisión alcanzados por el primer grupo de clasificadores. Debemos recordar que los resultados son más bajos al obviar la ponderación de la validación cruzada. Además, a más compresión, en general, mejores resultados. Los clasificadores con precisión más baja son los árboles C4.5 y las “*Tree Augmented Naïve*”.



(b) El segundo grupo de clasificadores han mantenido las mismas tendencias que ya se han dado en la figura anterior. Así que, volvemos a observar que a mayor compresión, mejor es la adaptación del modelo. En este caso, los peores clasificadores han sido los SVM, superando el Pearson VII al polinomial y al polinomial normalizado.

Figura 5.17: Resultados obtenidos durante la comprobación de los modelos generados a través de un conjunto de datos acumulados frente al conjunto de datos original con el fin de prever la aparición de microrrechupes. Se muestran los datos en términos de precisión para cada uno de los porcentajes de compresión. El comportamiento de los clasificadores tiene varias fluctuaciones, aunque, cuanto mayor ha sido la compresión, mejores resultados se han alcanzado.

una vez son aplicados los rangos de compresión máxima, el árbol C4.5 vuelve a pertenecer a los clasificadores que representan la normalidad.

El resultado del resto de los clasificadores queda recogido en la Figura 5.17b. En particular, este grupo genera la misma tendencia de crecimiento en la precisión según se van comprimiendo en mayor número los datos de los conjuntos acumulados. No obstante, en este caso el comportamiento de los clasificadores es algo más heterogéneo. Así, nos encontramos con tres comportamientos diferentes. En primer lugar, contamos con los clasificadores que sí que han sido capaces de realizar una mejora total de los resultados de precisión gracias al sistema de compresión bayesiana basado en métodos de Montecarlo. Los clasificadores que se comportan así son el último de los bosques aleatorios (el que se construye con 350 árboles) y dos de las máquinas de soporte vectorial, las que han sido creadas con funciones de núcleo polinomial normalizado y RBF.

En un segundo lugar, observamos otro grupo de clasificadores que no han sido capaces de mejorar los resultados iniciales, pero que tras el bajón inicial de rendimiento que presentan todos los clasificadores, han podido recuperarse hasta el punto de desarrollar el mismo rendimiento que al comienzo de la aplicación del método de compresión. Los clasificadores que se encuentran dentro de este grupo son todos los KNN y el SVM con función de núcleo Pearson VII. De hecho, este último, a excepción del gran pico que encontramos en los niveles de compresión cercanos al 13-14 %, tiene un comportamiento lineal a lo largo de todo el espectro de compresiones posibles, en otras palabras, constantemente obtiene los mismo resultados.

Y por último, está el grupo de los clasificadores que reducen levemente su precisión en comparación con los resultados obtenidos en primer lugar (grados de compresión muy bajos). Estos son los clasificadores SVM con funciones de núcleo polinomial y las redes neuronales artificiales.

A pesar de estos hechos, las tendencias generales de todos ellos son similares. Así, durante la sección que engloba los grados de compresión comprendidos entre el 1 y el 50 % se ven las mayores fluctuaciones y los mayores descensos en el rendimiento de todos los clasificadores. Sin embargo, en el siguiente grupo de compresión el comportamiento se estabiliza. No obstante, cabe destacar que aquellos clasificadores que peores resultados han obtenido son los que más fluctuaciones sufren a lo largo de todo el espectro de compresiones, y hasta ellos reducen sus amplitudes al aplicar altos grados de compresión. Finalmente, a excepción de dos clasificadores, la precisión se mantiene entre el 80 % y el 90 %, por lo que contamos con un conjunto de 19 clasificadores que son

5. Adaptando los modelos al proceso cambiante

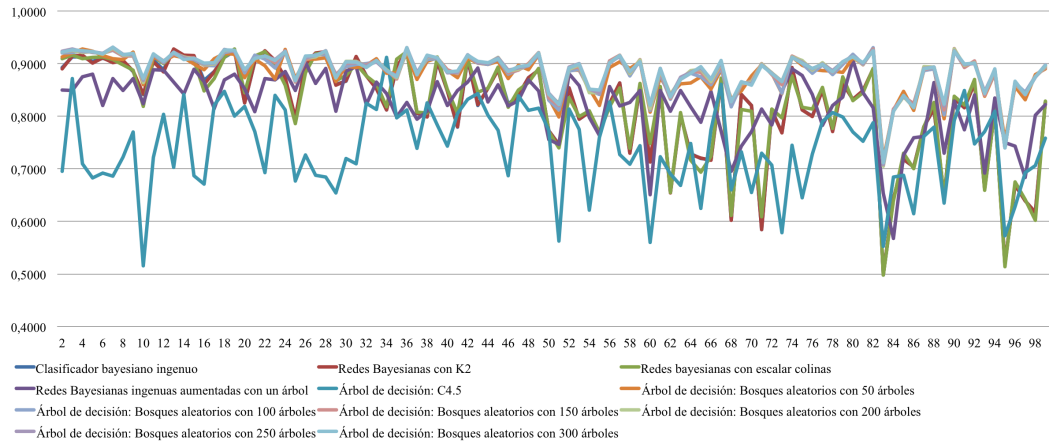
capaces de obtener buenos resultados en la predicción de los defectos de microrrechupe a pesar de haber llevado a cabo un proceso iterativo de adaptación del sistema.

En cuanto a la gestión de errores llevada a cabo por los clasificadores, la Figura 5.18a muestra los valores del área bajo la curva ROC alcanzados en el experimento. Concretamente, observamos grandes similitudes con los resultados alcanzados en términos de precisión. De esta forma, la primera característica que llama la atención es que el grado de fluctuaciones presentes a lo largo del ciclo de compresiones es igual de elevado que en la precisión. Además, los puntos de bajo rendimiento en la medición de los errores coinciden perfectamente con los puntos de bajo rendimiento para la precisión, por lo que los conjuntos de datos acumulados afectan de igual manera a ambas mediciones. No obstante, los resultados alcanzados, desde el punto de vista de las tendencias que siguen los clasificadores, son realmente interesantes. Así, cuanto menor es el grado de compresión en los conjuntos de datos acumulados, los resultados sufren de muchas más fluctuaciones, con más amplitud y peores resultados. Los resultados aquí conseguidos son más bajos que al hacer la prueba de los modelos en sí mismos, sin embargo, dentro de la zona de compresiones máximas, 9 de los 10 clasificadores obtienen valores superiores al 0,90, siendo 1 el valor óptimo para el área bajo la curva ROC.

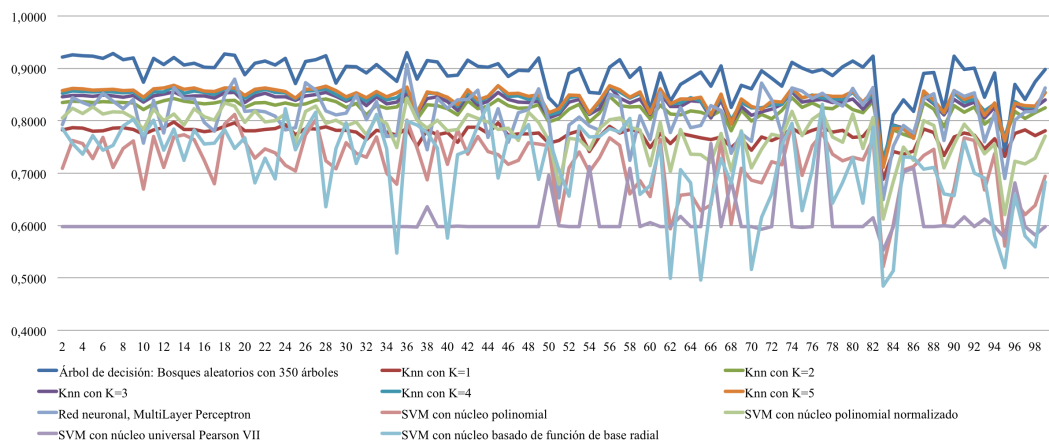
La zona de confort para este experimento estaba definida entre los porcentajes de compresión del 64 y el 98 %. Aproximadamente, para la gestión de errores encontramos que esa área es similar. Incluso, si no trabajásemos con todos ellos a la vez, sino que para cada uno de los clasificadores determináramos el grado de compresión que debe seguir, podríamos indicar que clasificadores como los bosques aleatorios comienzan a tener un comportamiento muy equilibrado a partir de niveles de compresión del 51 %, las redes bayesianas a partir del 70 % y el árbol de decisión C4.5 no consigue realizar una buena gestión de los resultados.

Particularmente, dentro de este grupo nos gustaría destacar el comportamiento de algunos de los clasificadores. Por una parte, el conjunto de los clasificadores bayesianos obtiene muy malos resultados a lo largo de las primeras 50 unidades de compresión. Sus fluctuaciones llegan a colocarlas como los peores clasificadores. Concretamente, las redes bayesianas que utilizaron el algoritmo de escalar colinas llegan a descender de la barrera de las 0,5 unidades (este punto se encuentra en el momento de reducir los conjuntos de datos generando el 83 % de las instancias). Sin embargo, según se aumenta el porcentaje de compresión, estas redes se empiezan a igualar con la normalidad.

5.3 Validación empírica



(a) Esta figura ilustra los resultados en términos de tasas de error que han sido alcanzados por los clasificadores que realizaron su aprendizaje con un conjunto de datos acumulado. Entre los 50 y los 99 se vislumbra la zona más irregular. No obstante, a mayor compresión, mejor resultado. Los árboles C4.5 vuelven a ser los peores.



(b) Las tendencias visualizadas en esta figura indican que el comportamiento es coherente al mantenerse, en general, para todos los clasificadores. Este grupo también cuenta con la misma zona de perturbaciones y cumple que a más porcentaje de compresión, se obtiene una adaptación más estable del proceso. Los SVM con Pearson VII vuelven a ser los peores.

Figura 5.18: Tasas de error obtenidas al enfrentar los modelos aprendidos con el conjunto acumulado a los datos originales en la predicción de microrrechupes. Se muestran los datos en términos de área bajo la curva ROC. A pesar de que observamos que la gestión de los errores no es tan buena, la mayoría de los clasificadores mantienen o mejoran su tasa de errores cuando la compresión es alta.

5. Adaptando los modelos al proceso cambiante

A pesar de ello, encontramos una que no es capaz de realizarlo, esa es la red bayesiana ingenua aumentada con un árbol. Como ya sucediera anteriormente, aunque este clasificador ha sido la base para realizar los métodos de reducción, no gestiona los errores tan bien como el resto de los clasificadores. A pesar de que no se puede comparar con el resto de las redes bayesianas o los bosques aleatorios, los resultados obtenidos con esta red siguen siendo similares a los que se obtendrían sin realizar ningún tipo de adaptación. Por lo que el método de compresión bayesiana basado en métodos de Montecarlo permite realizar una adaptación adecuada para el modelo.

Por otra parte, el árbol de decisión generado utilizando el método de aprendizaje C4.5, no obtiene la misma tendencia que el resto de los compañeros del grupo. Como se puede observar, este clasificador no para de fluctuar con una gran amplitud, consiguiendo porcentajes de compresión con grandes resultados y otros realmente malos. Los puntos de descenso máximo del rendimiento coinciden con los de los demás clasificadores, sin embargo, éste incluye muchos más puntos de descenso que el resto de los clasificadores. Observando detenidamente la curva que dibuja a lo largo del espectro de compresión, no podemos determinar que haya alguna zona en la que logre mejores resultados en el tratamiento de errores, por lo que es un clasificador que, al igual que en los experimentos anteriores, no es capaz de realizar una buena gestión de las tasas de error cuando le aplicamos los métodos de reducción.

El resto de los clasificadores quedan aglomerados en un segundo grupo y la Figura 5.18b muestra los resultados que obtuvieron. Este segundo grupo también sigue un alto grado de concordancia con lo acaecido durante la medición de la precisión de los clasificadores. A pesar de la diversidad de valores alcanzados (debido a que los propios clasificadores no son capaces de obtener mejores resultados), las tendencias que siguen son igual de lineales que en la precisión, y específicamente, mucho más lineales que las de los clasificadores del grupo anterior. También, al tener esa tendencia mucho más lineal podemos observar fácilmente que su proceso de adaptación ha sido correcto. La mayoría de los modelos son capaces de mantener su rendimiento en la gestión de errores, o incluso, mejorarlo.

Dada esa tendencia tan rectilínea, no es sencillo determinar el punto que marca la zona de confort de los clasificadores. Sin embargo, y dado que su tendencia en el rendimiento es directamente proporcional al grado de compresión, en otras palabras, los resultados son mejores cuanto mayores son los porcentajes de compresión, podemos determinar que la zona de confort utilizada para el primero de los grupos es igualmente útil para este segundo, pudiendo ser

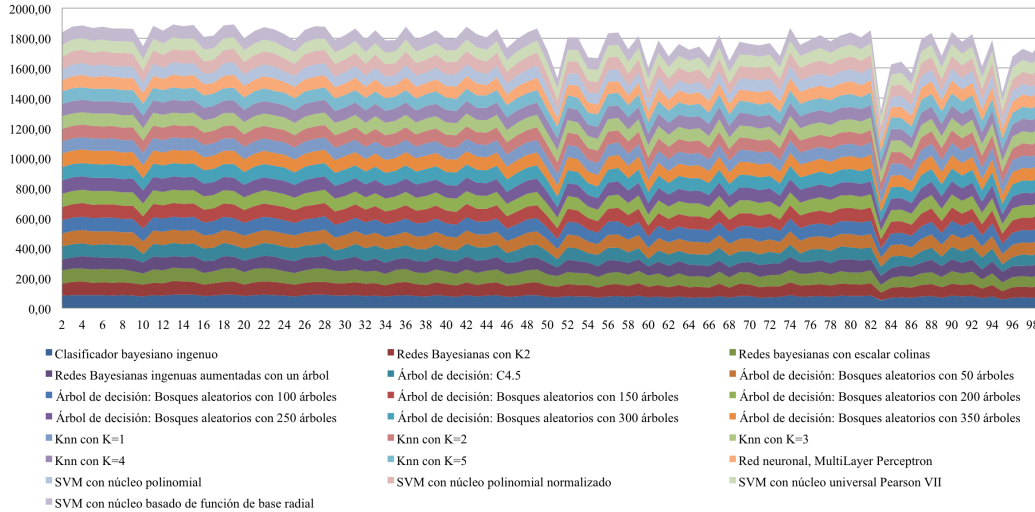
ampliada en el caso de utilizar solamente los clasificadores de este conjunto. Únicamente, encontramos un par de casos que no siguen la normalidad marcada por la mayoría. En primer lugar, las redes neuronales artificiales que su tendencia, a pesar de ser bastante lineal, no logra mejorar los resultados y se queda un poco por debajo de lo que eran capaz de hacer sin el proceso de adaptación. No obstante, su grado de desviación es tan bajo que puede considerarse como asumible. Y en segundo lugar, algunos de los clasificadores SVM obtienen muchos más picos a lo largo de todo el espectro de compresión. A pesar de ello, y a excepción del SVM generado con una función de núcleo Pearson VII, sus resultados mejoran cuando los comparamos con el punto de compresión nulo.

Dentro de este grupo advertimos que el bosque aleatorio generado con 350 árboles es el clasificador que mejores resultados obtiene. Posteriormente encontramos el grupo formado por todos los KNN, los cuales obtienen los mismos resultados con una pequeña variación de décimas, por ejemplo, el KNN con $K = 1$ mantiene la misma tendencia que el resto de los KNN pero no es capaz de igualarlos, logrando peores resultados y, por ello, aparece más abajo en la gráfica. A pesar de este hecho, debido a que su curva es como la del resto lo agrupamos con ellos. Por otra parte, agrupamos las máquinas de soporte vectorial (SVM) y las redes neuronales artificiales. Su tendencia no es tan clara y las fluctuaciones están muy presentes a lo largo de todos los grados de compresión. Finalmente, el clasificadores SVM con función de núcleo Pearson VII vuelve a repetir el comportamiento de otros experimentos. Mantiene, a lo largo de todos los porcentajes de compresión, una tendencia lineal ajustada al valor 0,50, contando con algunas excepciones en forma de altas fluctuaciones para los conjuntos de datos que utilizan bajos grados de compresión. Así, podemos llegar a las mismas conclusiones que se obtuvieron en anteriores experimentos. Es un clasificador al que el método de compresión le afecta negativamente y/o no es capaz de adaptarse al modelo que hay detrás de los datos.

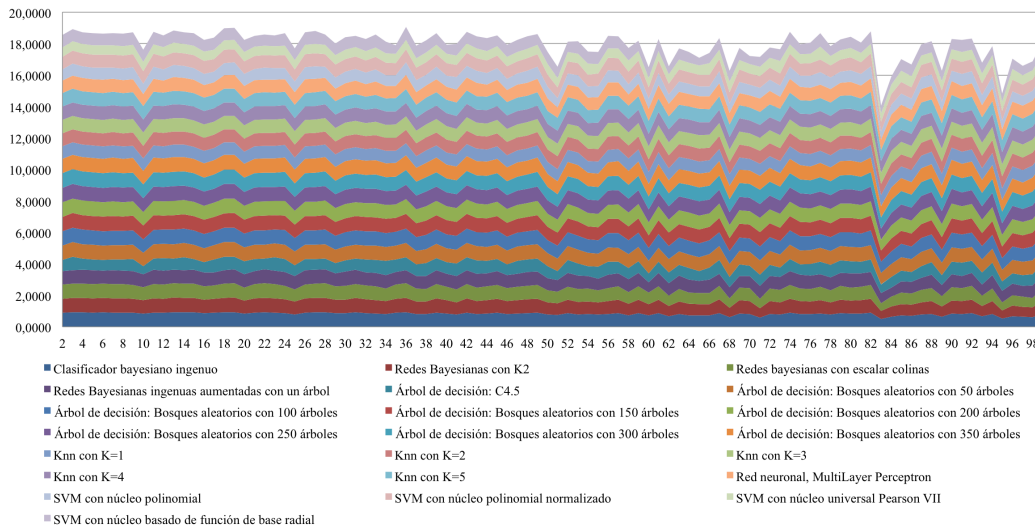
La visualización de los resultados a través de las gráficas presentadas puede ser algo complicada, por ello, y con la intención de mostrar un resumen de todo aquello comentado, las Figuras 5.19a y 5.19b muestran los resultados en términos de precisión y el área bajo la curva ROC, respectivamente, apilando cada una de las curvas de comportamiento relativas a los clasificadores.

Más concretamente, en la Figura 5.19a nos deja bien claro cómo las tendencias de todos los clasificadores son similares y que cuanto mayor es el grado de compresión, mejores resultados se han obtenido. También, localiza claramente

5. Adaptando los modelos al proceso cambiante



(a) La primera figura muestra los diferentes resultados de precisión alcanzados por los clasificadores mostrando las tendencias seguidas por todos ellos de una forma mucho más legible. Como se puede observar, la línea que trazan todos ellos es muy similar.



(b) La segunda gráfica muestra los resultados apilados relativos a las tasas de error, más concretamente, el área bajo la curva ROC que alcanzarán los clasificadores en su proceso de predicción de microrrechupes.

Figura 5.19: Resultados apilados obtenidos en los experimentos orientados a la prueba de los modelos acumulados, probándolos con el conjunto de datos original, para la predicción de microrrechupes. Se recogen tanto los valores alcanzados en la precisión como aquellos relativos al área bajo la curva ROC.

los puntos más bajos de rendimiento y cómo afecta a todos los clasificadores por igual. Y por último, nos permite visualizar la zona de confort de los clasificadores y las fluctuaciones que sufren éstos cuando los grados de compresión son más bajos.

Por otra parte, la Figura 5.19b nos permite, en primer lugar, observar cómo el comportamiento en la gestión de errores está totalmente relacionado con la precisión de los clasificadores. Así, atisbamos una serie de descensos en los rendimientos de los clasificadores localizados en las mismas posiciones que en la precisión. Del mismo modo, la gráfica muestra que los resultados logrados tras la aplicación de altos grados de compresión mejoran los totales obtenidos sin su aplicación y que la tendencia general seguida por todos los clasificadores es similar.

Como conclusión, el desarrollo de este experimento nos permite completar las pruebas del método de compresión bayesiana basado en métodos de Montecarlo para la adaptación de los modelos orientados a realizar la predicción de problemas de microrrechupes en el estado estacionario $t + 1$. Así, tras probar el método para la generación de modelos reducidos y la generación de conjuntos acumulados podemos determinar que los resultados obtenidos por el método desarrollado son muy interesantes, consiguiendo crear modelos que obtienen similares o mejores resultados que los que somos capaces de lograr con los datos originales. Para probarlo hemos realizado un testeo de los modelos en sí mismos y, posteriormente, a través de una simulación algo más real al enfrentarlos con el conjunto de datos original. Por consiguiente, estos resultados indican que el método desarrollado es suficientemente capaz de llevar a cabo los procesos de adaptación de los modelos con un envejecimiento de muestras para aportar mucha más importancia a los hechos que actualmente suceden en la planta pero sin olvidar aquello que sucedió tiempo atrás.

5.3.3.2 Propiedades mecánicas

Una vez finalizados los experimentos que prueban el modelo de compresión bayesiana basado en métodos de Montecarlo para los clasificadores que realizan las predicciones de los microrrechupes, es momento de realizar los experimentos que determinen el comportamiento de los modelos que siguen la misma técnica pero con un objetivo de predicción diferente. En este caso, y como hemos realizado anteriormente, nos disponemos a testar el comportamiento de aquellos clasificadores que se encargan de prever los valores de carga máxima, una de las propiedades mecánicas de las piezas de fundición.

5. Adaptando los modelos al proceso cambiante

Hasta este momento, y en relación con la carga máxima, hemos realizado las pruebas que nos demuestran que la reducción de los conjuntos de datos genera una serie de modelos de clasificación que (i) mantienen la distribución de probabilidad original que representa el proceso de producción y (ii) son tan generales que las predicciones de la carga máxima son muy semejantes (o mejores) a los modelos generados al utilizar el conjunto de datos original. En este experimento queremos pasar a comprobar si la generación de un conjunto de datos al que se le ha aplicado múltiples veces el proceso de adaptación pierde rendimiento frente al aprendizaje del conjunto al completo. Por lo tanto, con el fin de alcanzar este objetivo, y siguiendo la misma metodología que para los experimentos de microrrechepe, hemos aplicado el proceso de adaptación automática explicado en 5.2.2 con un “*quantum*” de tiempo t ajustado a medio año. La razón es la misma que en el experimento anterior, los datos con los que estamos trabajando han sido extraídos a lo largo de dos años del proceso de producción. Además, los cambios que sufre el proceso no son tantos como para tener que realizar un ciclo de adaptación más bajo. Por esa razón, hemos optado por la utilización de un ciclo no excesivamente pequeño ni excesivamente grande que permitirá la adaptación más adecuada al ritmo de trabajo seguido en la fundición de la que fueron adquiridos los datos originales. Por consiguiente, se han tomado cuatro conjuntos de datos diferentes, asociados a cada medio año, a los que se les han aplicado todos los porcentajes de compresión posibles para, después, ir acumulando las nuevas evidencias que llegaban a lo largo del periodo de tiempo t restante.

A continuación, pasamos a comentar los resultados obtenidos. En estos experimentos hemos seguido la misma metodología que en los anteriores, de esta forma, comprobaremos: (i) el comportamiento de los clasificadores con el fin de validar si los modelos generados cumplen con las mismas bases que un modelo aprendido con un único conjunto de datos reducido y (ii) enfrentaremos esos modelos con el conjunto de datos original para comprobar la generalización del proceso de adaptación desarrollado.

Comenzando por el primero de ellos, las Figuras 5.20a y 5.20b muestran los resultados alcanzados por los diferentes clasificadores una vez aplicado el proceso de adaptación. Aquí, al igual que en los experimentos previos, realizamos la medición de la precisión alcanzada por cada uno de estos clasificadores a lo largo del espectro de posibles porcentajes de compresión. Posteriormente, las Figuras 5.21a y 5.21b hacen referencia a la gestión de errores que han tenido los clasificadores. Para ello, efectuamos las mediciones como el área que se encuentra bajo la curva ROC. Nótese, que al igual que en anteriores ex-

perimentos, el eje de las abscisas nos muestra el porcentaje de instancias que han sido generadas en cada una de las reducciones del proceso de adaptación. El lector debe tener en cuenta que siempre que se ha realizado una reducción del conjunto de datos le hemos aplicado el mismo grado de reducción independientemente del número de reducciones que se hayan llevado a cabo.

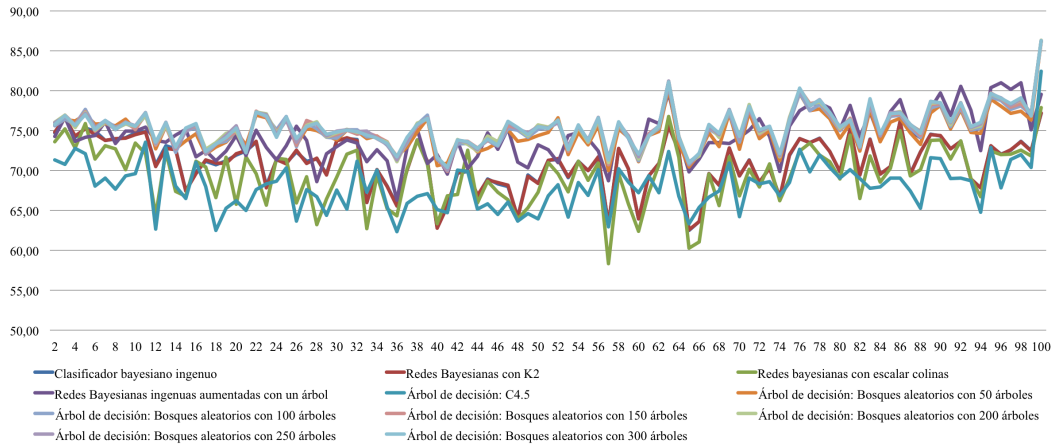
Concretamente, la Figura 5.20a nos ilustra con los resultados obtenidos en términos de precisión para el primero de los grupos de clasificación. La división en dos grupos se ha realizado con el fin de facilitar la legibilidad de las gráficas generadas. Del mismo modo, y con el propósito de permitir realizar las comparaciones entre los experimentos de la forma más sencilla posible, las agrupaciones han sido generadas utilizando los mismos clasificadores que en los experimentos previos.

De forma general, observamos que el comportamiento seguido por los clasificadores es muy cercano a lo sucedido en las pruebas de los microrrechupes (ver la sección 5.3.3.1), observando que los grados de compresión extremos son los que mejores resultados logran. Por el contrario, la zona central tiene una reducción en el rendimiento tal que el mínimo se alcanza al aplicar un porcentaje de compresión cercano al 54-58 %. Así, la tendencia de este grupo es la de reducir el rendimiento alcanzado por los clasificadores para luego comenzar a retomar mejores porcentajes cuando se comienza a realizar la aplicación de altos porcentajes de compresión. Concretamente, la precisión alcanzada dentro de las dos zonas con mejor rendimiento ronda el 70-78 %.

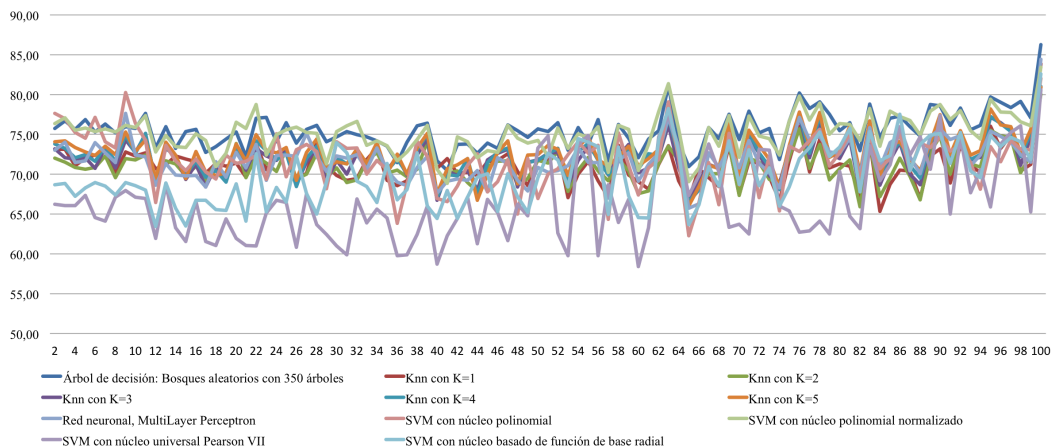
Sin embargo, este experimento sí que dispone de un aspecto que lo diferencia. Los resultados aquí alcanzados no son capaces de superar o igualar el rendimiento alcanzado por los clasificadores sin aplicar ningún método de adaptación. La reducción en la precisión ronda las 10 unidades. Un valor que no es despreciable pero que para el proceso de adaptación podría ser asumido al, en primer lugar, simplificar las etapas de aprendizaje de los modelos de predicción y, en segundo lugar, ir dotando de más importancia a los nuevos casos. Esto no es una mala noticia, ya que esa pequeña pérdida de precisión a la hora de hacer una reducción de la información estaba contemplada. Pensamos, principalmente, que lo que hace que no se alcance esa mejora de resultados es la calidad de los datos de partida, aspectos que ya han sido comentados a lo largo de los experimentos anteriores. No obstante, y a pesar de esa pequeña reducción, el envejecimiento de muestras funciona al igual que ha estado haciéndolo para los problemas de los microrrechupes. Por eso, las tendencias seguidas por los clasificadores nos resultan familiares.

En cuanto al comportamiento específico de los clasificadores, podemos in-

5. Adaptando los modelos al proceso cambiante



(a) El primer grupo de clasificadores se comporta de forma similar. Los defectos en los datos hacen que las fluctuaciones sean mucho mayores. Sin embargo, destaca que los resultados alcanzados con la mayor compresión son muy similares al los alcanzados por la menor de ellas.



(b) La segunda figura muestra que los patrones de comportamiento del primer grupo vuelven a repetirse en este segundo. Vemos una reducción entre el conjunto de datos original y el primer grado de compresión. Estos últimos igualan los resultados de compresiones mayores.

Figura 5.20: Resultados obtenidos durante la comprobación de los modelos generados a través de un conjunto de datos acumulado con el fin de la detección de posibles problemas de carga máxima. Se muestran los datos en términos de precisión para cada uno de los porcentajes de compresión. Observamos que después del decremento inicial de la precisión, la tendencia seguida por los clasificadores es la misma que en el caso de los microrrechupes.

dicar que al igual que viene pasando en experimentos anteriores, la conducta desarrollada por los bosques aleatorios es muy similar al que muestra la red bayesiana ingenua aumentada con un árbol. Esta segunda, vuelve a ser el mejor de los clasificadores mientras los porcentajes de compresión son relativamente bajos. Sin embargo, cuanto más nos acercamos a compresiones mayores, reduce su rendimiento para ser superada por los bosques aleatorios. Por otra parte, el resto de los clasificadores bayesianos (clasificador bayesiano ingenuo y los aprendidos con los métodos de escalar colinas y K2) sufren de múltiples fluctuaciones de gran amplitud (las mayores de todos los clasificadores). Esa amplitud va reduciéndose según se aumenta el grado de compresión del conjunto de datos acumulado. Por último, observamos que la pauta de actuación seguida por el árbol de decisión C4.5 vuelve a mostrar signos que ya conocemos y hemos comentado en la sección 5.3.3.1. Específicamente, este clasificador se mantiene como el peor del grupo, a excepción de puntos concretos en los que otros clasificadores obtienen peores resultados. También, aunque se da la mejora final del rendimiento del clasificador, este árbol no puede llegar a igualarles por lo que se mantiene por debajo de sus compañeros. Además, este modelo es el que mayor descenso de rendimiento obtiene al aplicar los métodos de adaptación (aproximadamente unas 12 unidades).

En lo que respecta al segundo grupo de clasificadores, la Figura 5.20b muestra los resultados que fueron capaces de lograr los modelos en el proceso de predicción de la carga máxima a través de un conjunto de datos acumulado. En este caso, comprobamos que los modelos tienen el mismo comportamiento que siguieron los clasificadores del primero de los grupos. Así, comprobamos cómo las tendencias seguidas por todo el grupo, incluso por los peores de ellos, comienzan con un proceso de descenso de la precisión hasta que finalmente alcanzan su punto mínimo (aproximadamente en la zona de los 54-58 % de compresión, i.e., una creación de instancias de un 42-48 %). Una vez superado este punto de compresión, el rendimiento vuelve a mejorar según reducimos el número de instancias generadas. Estas tendencias similares muestran que el método de la compresión bayesiana tiene el mismo efecto en todos los clasificadores, debiéndose a que está modificando los datos utilizados para el aprendizaje.

No obstante, este segundo grupo sufre del mismo problema que el primero, y es la reducción del rendimiento. No conseguimos mejorar o igualar el resultado que tienen los clasificadores al utilizar el conjunto de datos al completo. Y como ya comprobamos en el primero de los grupos, el descenso de la precisión no supera las 10 unidades. Debemos tener en cuenta que, a pesar del descenso

5. Adaptando los modelos al proceso cambiante

en el rendimiento, los datos que utilizamos no son todos los existentes y, aún así, los resultados finales de precisión rondan el 72-75 %. Interesantes valores ya que se equiparan a los resultados alcanzados por el primero de los grupos. Por consiguiente, la configuración que sea ajustada en el sistema de adaptación automático afectará de igual manera a todos los modelos de clasificación.

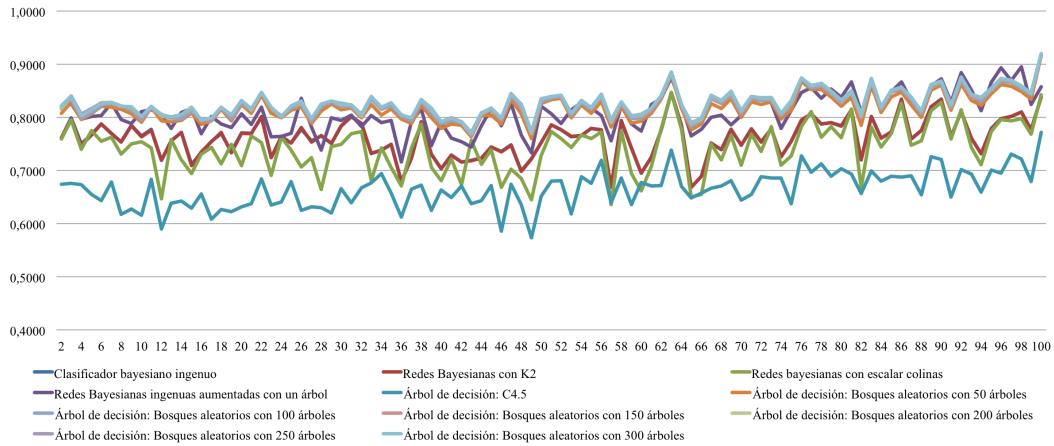
Dentro de este grupo, encontramos la misma agrupación de rendimientos que en otras ocasiones. De esta forma, encontramos los bosques aleatorios formados con 350 árboles como uno de los mejores clasificadores. En esta ocasión se ven acompañados por dos de las máquinas de soporte vectorial. Concretamente, las creadas con funciones de núcleo polinomial y polinomial normalizado; y que son capaces de superar en ciertas ocasiones los resultados de los bosques aleatorios. En un segundo grupo, y como viene siendo habitual, encontramos a todos los KNN, con los valores de K que van de 1 a 5, compartiendo los máximos y mínimos con un ínfimo desplazamiento vertical entre todos ellos. Junto a ellos encontramos las redes neuronales artificiales. A pesar de que los clasificadores KNN no tienen una etapa de entrenamiento como tal, pueden compararse a un modelo mucho más complejo como las redes neuronales artificiales. Además, siguiendo la línea que trazan tanto los KNN como las redes neuronales observamos que el comportamiento es prácticamente similar. Finalmente, nos encontramos las máquinas de soporte vectorial que fueron creadas con funciones de núcleo Pearson VII y RBF. Estos dos clasificadores, vuelven a mantener un comportamiento que no está totalmente en concordancia con lo que hacen sus compañeros, tanto máquinas de soporte vectorial como otro tipo de clasificadores. Las fluctuaciones de ambos son de una gran amplitud y hacen oscilar a los clasificadores entre las últimas y las primeras posiciones mientras los porcentajes de compresión son muy bajos. Posteriormente, cuando el grado de compresión es aumentado al máximo se acaban equilibrando (reduciendo el número y la amplitud de las fluctuaciones) pero manteniéndose como los dos últimos clasificadores de ambos grupos.

La medición de las tasas de error relacionadas con los clasificadores del primero de los grupos puede visualizarse en la Figura 5.21a. Claramente, observamos que el comportamiento que desarrollan los clasificadores es muy similar al de la precisión. La tendencia fundamental del grupo se repite. Una vez se comienzan a aplicar a los datos diferentes grados de compresión para generar los conjuntos acumulados, comienza a reducirse el rendimiento de los modelos. Posteriormente, llega un punto en el que todos los clasificadores marcan un mínimo en su comportamiento, más concretamente, cuando el porcentaje de compresión es de 48-50 %. A partir de ese punto, el rendimien-

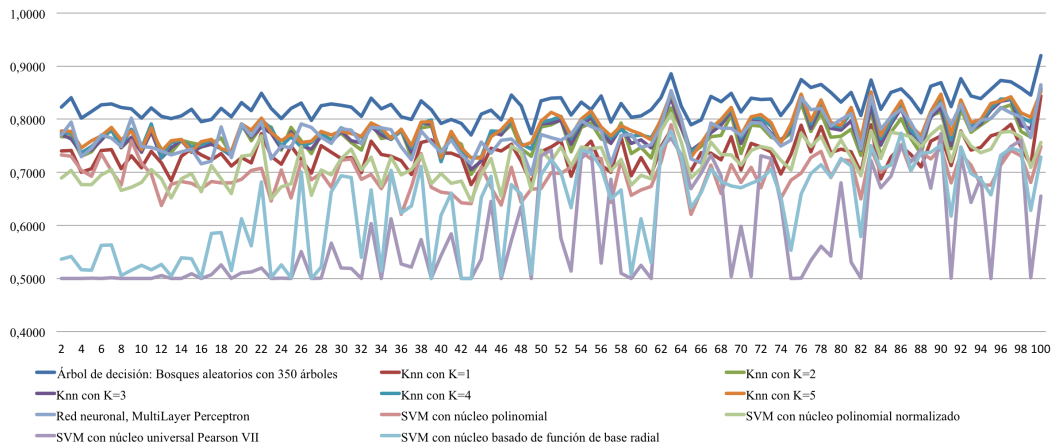
to de los clasificadores mejora hasta alcanzar un valor del área bajo la curva ROC que se encuentra entre 0,75 y 0,85 unidades. Debemos recordar que todos los experimentos relacionados con la carga máxima no han sido capaces de obtener unos grandes resultados en el área bajo la curva ROC. Por lo que éstos se asemejan bastante a los alcanzados por los clasificadores unitarios. No obstante, no llegan a igualarlos o a superarlos. Específicamente, cada uno de los clasificadores sufre un decremento en su rendimiento de aproximadamente una décima. A pesar del problema de la reducción del comportamiento, todos los datos adquiridos hacen visualizar que el envejecimiento de las muestras es efectivo, ya que según la compresión es mayor, esos datos son menos importantes respecto a los datos más nuevos, con lo que el modelo mejora al llevar a cabo la adaptación del proceso.

Al igual que en anteriores ocasiones, el comportamiento de este grupo se separa en diferentes conjuntos de clasificadores según el rendimiento que han alcanzado. En primer lugar, contamos con los bosques aleatorios y las redes bayesianas TAN (un clasificador bayesiano ingenuo aumentado con un árbol). Estos clasificadores vuelven a destacarse del resto al conseguir los mejores resultados en la gestión de los errores. Asombrosamente, en otras ocasiones, las redes bayesianas TAN solían realizar una buena gestión de los errores, pero al aplicar una gran compresión, descendía su rendimiento. No obstante, en este caso su comportamiento varía para que finalmente acabe en el mismo rango que los bosques aleatorios (entre las 0,80 y las 0,84 unidades). Este último clasificador es el que ha sido utilizado como la base de la reducción, sin embargo, los resultados en la gestión de errores no pueden llegar a superar a los bosques aleatorios. Este tipo de resultados nos hace pensar en una posible modificación del modelo utilizado para obtener la distribución probabilística de comportamiento de la planta. En un segundo grupo, las redes bayesianas que han sido creadas con el algoritmo escalar colinas y el método K2 obtienen unos resultados de área bajo la curva ROC muy similares, su curva es muy parecida. Del mismo modo, su comportamiento es también muy cercano al del los anteriores clasificadores, pero la amplitud en las fluctuaciones es mayor para el caso de los clasificadores bayesianos. Finalmente, en la última posición se encuentra el árbol de decisión que utiliza el método C4.5 para su construcción. El modelo no era capaz de obtener la misma precisión que el resto de sus compañeros, de este modo, aplica el mismo comportamiento al resultado de la gestión de errores. Aunque sus resultados sean más pobres que el de sus compañeros, debemos destacar que su línea de comportamiento en la gestión de los errores es equiparable a la de otros clasificadores. Por eso, po-

5. Adaptando los modelos al proceso cambiante



(a) La figura muestra las tasas de error alcanzadas por un primer grupo de clasificadores basado en un conjunto de datos acumulado. Casi todos los clasificadores se comportan de una forma similar ante el método de compresión y realizan una gestión de los errores aceptable. El árbol con C4.5 sigue siendo el modelo cuyo comportamiento no se ajusta al del resto de los clasificadores.



(b) Las variaciones del segundo grupo son mayores. A pesar de este hecho, el método de compresión afecta a casi todos los modelos por igual. En esta ocasión, los SVM con funciones de núcleo Pearson VII y RBF no consiguen seguir al resto de los clasificadores.

Figura 5.21: Tasas de error obtenidas durante la comprobación de los modelos generados a través de un conjunto de datos acumulado con el fin de detectar posibles problemas en la carga máxima. Se muestran los datos en términos de área bajo la curva ROC para cada uno de los porcentajes de compresión. Las tendencias en la gestión de los errores vuelven a ser similares a las de la precisión. Así, observamos ese descenso inicial y luego una similitud en valores de compresión bajos y altos.

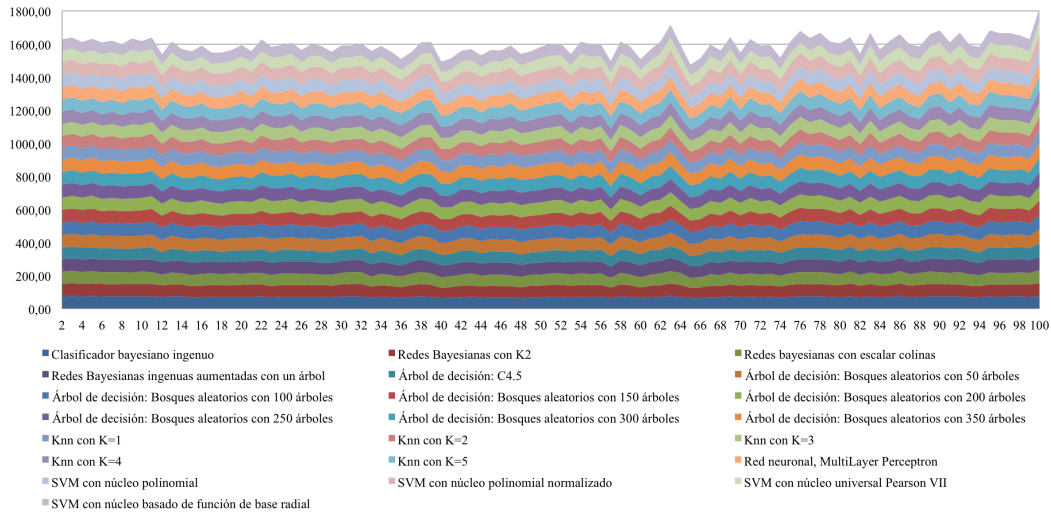
demos indicar que la adaptación a través de la compresión bayesiana basada en métodos de Montecarlo no le aporta ningún beneficio a este clasificador.

La Figura 5.21b completa los resultados relacionados con las tasas de error en la clasificación de la carga máxima. Al igual que en el primero de los grupos, estos clasificadores mantienen una tendencia similar en todos sus comportamientos. Así, es claramente detectable los puntos en los que se producen los máximos y los mínimos para todos ellos (por ejemplo, quedan identificados con un porcentaje de generación de instancias del 91 % uno de los mínimos más bajo y para el 64 % uno de los mayores máximos). Pero, como ya sucediera con el primero de los grupos, obtenemos un descenso en el resultado de su gestión de errores. No obstante, en este caso no somos capaces de identificar una tendencia de comportamiento cóncava, en la que se identifique un punto mínimo a partir del cual el comportamiento crece. Particularmente, en este segundo grupo, una vez se ha producido la reducción de rendimiento (con una compresión del 1 %), el comportamiento se mantiene relativamente estable (sin tener en cuenta las pequeñas fluctuaciones) a lo largo de todos los porcentajes de compresión.

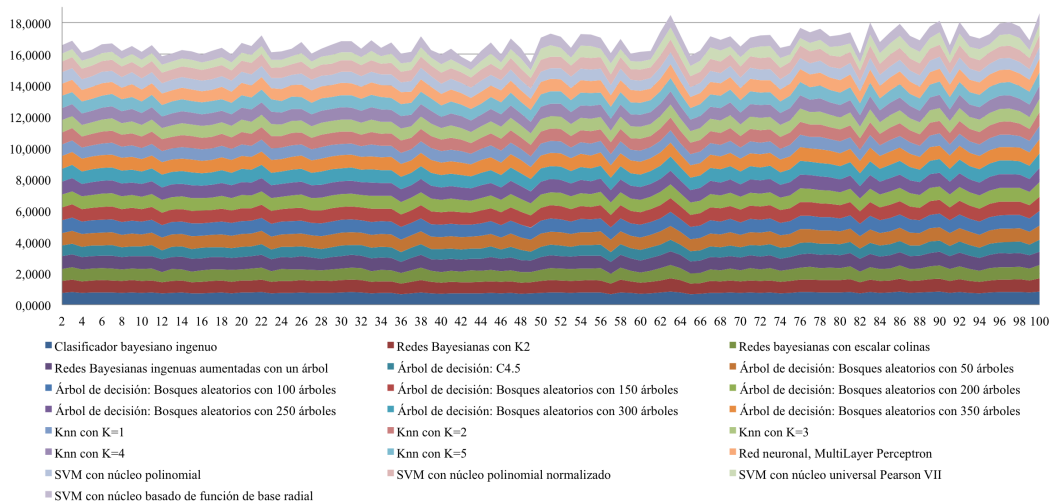
Los resultados alcanzados por estos clasificadores son algo más dispersos que los alcanzados por los clasificadores del grupo anterior. La razón es que el comportamiento que tienen estos modelos sufre de esa misma diversificación. Es decir, en este segundo grupo, los clasificadores se ajustan de una forma más diversa al proceso, con lo que no pueden obtener todos el mismo valor en la gestión de los errores. Y dada esa diversidad de resultados, el rango de mediciones en la gestión de los errores marcado por el primero de los grupos es ampliado, quedando definido como el rango de valores $[0,69, 0,84]$.

Los dos clasificadores que se separan de la normalidad, marcada tanto por este segundo grupo como por el primero, son las máquinas de soporte vectorial (SVM) creadas con funciones de núcleo RBF y Pearson VII. Estos dos clasificadores han sido los que peor comportamiento han logrado a la hora de gestionar errores en todos los experimentos realizados hasta ahora con la compresión bayesiana basada en métodos de Montecarlo. Aquí, vuelven a mantener ese comportamiento, consiguiendo ser los clasificadores más irregulares y con una amplitud mayor en sus fluctuaciones. Sorprendentemente, en el caso de la carga máxima no vemos cómo estos clasificadores llevan una tendencia lineal cercana a las 0,50 unidades. Hasta ahora, estos clasificadores para tasas muy altas de generación de instancias mantenían el comportamiento de la mayoría para, posteriormente, en tasas de generación muy bajas (compresiones muy altas) acercarse a las 0,5 unidades. En este caso mantienen este comportamiento

5. Adaptando los modelos al proceso cambiante



(a) A modo de simplificación, la primera figura muestra las tendencias seguidas por todos los clasificadores en términos de precisión en su proceso de predicción. Como se puede observar, las tendencias seguidas por todos ellos son muy similares.



(b) En el segundo de los gráficos apilamos todos los resultados del área bajo la curva ROC alcanzados por los clasificadores a la hora de prever la aparición de un valor no válido en la propiedad mecánica de carga máxima.

Figura 5.22: Resultados apilados obtenidos en los experimentos orientados a la prueba de los modelos acumulados para la predicción de defectos de carga máxima en las piezas extraídas de la fundición. La figura recoge tanto los valores alcanzados en la precisión como los alcanzados en el área bajo la curva ROC.

pero mostrando unas desviaciones de una gran amplitud. Particularmente, el SVM generado con una función de núcleo Pearson VII comienza su descenso hacia las 0,50 unidades mucho antes que el generado con la función de núcleo RBF (60 y 84 % respectivamente).

Las Figuras 5.22a y 5.22b muestran los resultados en términos de precisión y el área bajo la curva ROC respectivamente, haciendo uso de la técnica de generación de áreas apiladas. Utilizamos esta nueva forma de visualización con el fin de facilitar al legibilidad de los resultados. De esta manera, comprobamos fácilmente el lugar en el que se encuentran las mayoría de las variaciones o fluctuaciones en el comportamiento desarrollado por cada clasificador. Igualmente, nos permite observar si esas tendencias en el comportamiento son similares y cómo varían de forma similar unas con las otras.

Las pruebas de los modelos en sí mismos realizadas para determinar el valor que alcanzará la carga máxima en un estado localizado en un futuro próximo nos ha permitido comprobar que el aprendizaje basado en conjuntos de datos acumulados (en los que ha sido utilizado el método de compresión bayesiana) mantiene la representación estadística a pesar de las repeticiones del proceso. No obstante, los resultados que se obtienen no son tan buenos como en el problema de los microrrechupes. Tal y como se ha observado, sufrimos un pequeño descenso en el rendimiento. De todas formas, en el planteamiento inicial ya contábamos con ello. Dado que ese descenso es reducido, y puesto que el proceso marca que, a pesar de grandes compresiones, los modelos se adaptan aproximadamente igual el sistema, podemos indicar que el envejecimiento de las muestras está funcionando correctamente. Y el método de adaptación resulta exitoso para la carga máxima a falta de realizar la prueba que simulará su comportamiento en un entorno mucho más real.

Tras comprobar los resultados que alcanzan los diferentes clasificadores a la hora de utilizar un conjunto de datos acumulado, pasaremos a realizar la prueba que los enfrenta directamente con el conjunto de datos real. De esta forma, los modelos de clasificación serán comprobados sin utilizar la técnica de validación cruzada, por lo que los modelos serán aprendidos con los datos acumulados al completo y testados directamente con el conjunto original.

Para el desarrollo de este segundo experimento utilizaremos los mismos clasificadores y las mismas configuraciones que ya fueron empleadas en el primero de ellos. Gracias a la comprobación del conjunto original de datos frente a las generaciones acumuladas seremos capaces de visualizar el comportamiento de cada uno de los clasificadores para un entorno simulado mucho más cercano a la realidad. También, y dado que los procesos de muestreo que utilizan los

5. Adaptando los modelos al proceso cambiante

métodos de Montecarlo no son deterministas, utilizaremos los mismos conjuntos acumulados que fueron generados para el primero de los experimentos. Esto nos permite evitar las variaciones que podrían darse en esa etapa y, así, poder comparar los resultados generados en ambos experimentos.

Los resultados logrados por los clasificadores han sido medidos utilizando el porcentaje de precisión alcanzado y el área generada bajo la curva ROC, indicando cómo es la gestión de los errores, para cada uno de los clasificadores a lo largo del espectro completo de compresiones. Más concretamente, las Figuras 5.23a y 5.23b visualizarán los resultados de los dos grupos de clasificación en términos de precisión, y las Figuras 5.24a y 5.24b mostrarán los resultados de la gestión de los errores. Nótese que la numeración del eje de abscisas, al igual que en experimentos anteriores, hace referencia al porcentaje de instancias que han sido generadas para hacer cada una de las reducciones.

Específicamente, la Figura 5.23a hace referencia a los resultados alcanzados por el primero de los grupos con el objetivo de clasificar el estado final en el que se encontrará la carga máxima. Como podemos observar, los resultados al realizar las pruebas con los conjuntos de datos originales cambian significativamente. No obstante, la tendencia que sigue la predicción de la carga máxima es la misma que ya analizamos en el proceso de predicción de micro-rechupes (ver la sección 5.3.3.1). A mayores grados de compresión mejor es la adaptación del modelo y los resultados obtenidos son mejores. Incluso, la aplicación del aprendizaje con conjuntos de datos acumulados mejora el resultado de la predicción. En el caso de este grupo, todos los clasificadores mejoraron el rendimiento alcanzado cuando sus porcentajes de compresión han sido altos.

Como viene siendo habitual, todo ellos sufren de una serie de fluctuaciones. Del mismo modo que el rendimiento mejora al acercarse a valores de compresión altos, las fluctuaciones también mejoran en esta zona de compresión. En concreto, las fluctuaciones reducen su amplitud y frecuencia, aproximando progresivamente el comportamiento a una trayectoria mucho más lineal. La reducción de la variabilidad y la mejora del rendimiento se traduce en que la mayoría de los clasificadores desarrolla una precisión que se encuentra entre el 79 y el 87%. Casualmente, estos valores son equiparables a los resultados obtenidos al probar la predicción de los modelos haciendo uso del método de validación cruzada (ver sección 4.3.2.2). De esta forma, la aplicación de la compresión bayesiana utilizando métodos de Montecarlo adapta exitosamente los datos y consigue generar modelos que clasifican los estados estacionarios $t + 1$ manteniendo un rendimiento similar al que obtenemos al utilizar los datos al completo.

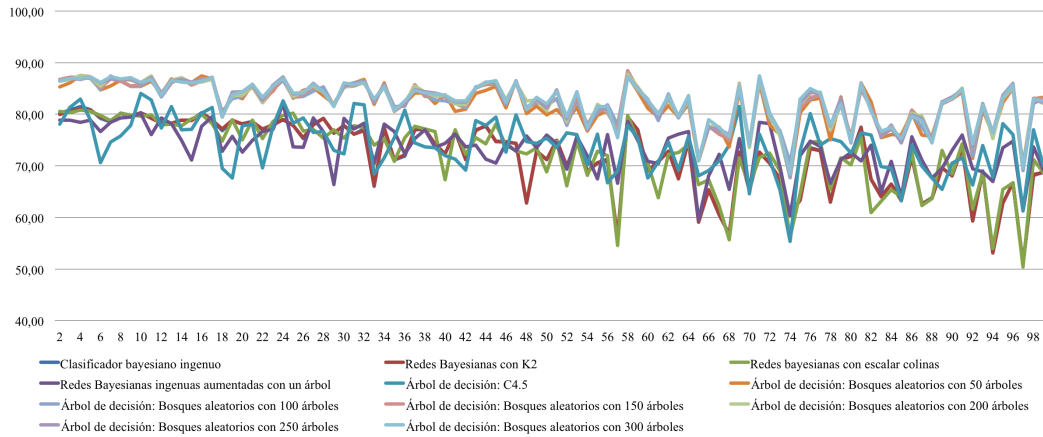
Pero dentro del comportamiento de estos clasificadores volvemos a observar que algunos de ellos se ajustan mucho mejor a las predicciones de la carga máxima. De esta forma, en primer lugar, encontramos todos los algoritmos de bosques aleatorios. Estos clasificadores son los que mejores resultados acaban alcanzando tanto para este grupo como para el segundo. Originariamente eran los que mejores resultados lograban, pero la compresión acumulada ha mejorado su precisión en unas 5 unidades.

En segundo lugar, tomamos el grupo formado por las redes bayesianas creadas con los métodos K2 y escalar colinas. Los clasificadores restantes alcanzan una precisión similar a estas dos, pero su comportamiento es más cambiante, por lo que serán agrupados formando el tercer conjunto de clasificadores. Estas dos redes bayesianas van logrando unos resultados prácticamente iguales a lo largo de todo el espectro de compresiones. Por ello, el comportamiento que siguen es equiparable. Por consiguiente, ambos métodos se adaptan de la misma manera al proceso de predicción de la carga máxima y la compresión de los conjuntos de datos.

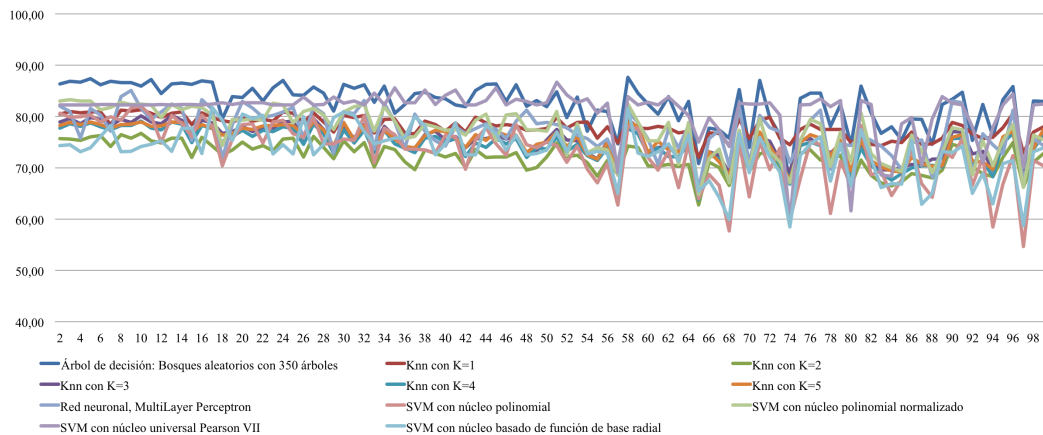
Finalmente, para el último grupo dejamos las redes bayesianas generadas con “*Tree Augmented Naïve*” y el árbol de decisión construido con C4.5. Estos dos clasificadores tienen un comportamiento mucho más irregular a lo largo de todos los valores posibles de compresión. El primero de ellos, a pesar de haber sido el método utilizado para las compresiones (con él extraemos la distribución probabilística del comportamiento de la planta) no es capaz de ajustarse tan correctamente. Así, acaba igualando el rendimiento de las redes bayesianas que utilizan K2 y escalar colinas cuando llegamos al grado de compresión máximo. El segundo de los clasificadores, el árbol de decisión, aunque logra unos resultados muy similares a las redes bayesianas, mantiene las fluctuaciones incluso durante los porcentajes de compresión máximos. Al igual que en otros experimentos, vemos que este clasificador es capaz de obtener unos resultados razonables pero que dado su grado de variabilidad no asegura que siempre funcione de la misma manera.

La Figura 5.23b muestra los resultados alcanzados por el segundo de los grupos de clasificadores. A primera vista, los resultados en términos de precisión son muy similares a los analizados para el primer grupo. Básicamente, se observa una tendencia ascendente en el nivel de precisión según aumenta el nivel de compresión. También, cabe destacar que todos los clasificadores de este grupo han sido capaces de superar o igualar los resultados obtenidos al no aplicar ningún tipo de procesamiento en los datos. No obstante, como ya fue comprobado en experimentos anteriores, este grupo obtiene unos resulta-

5. Adaptando los modelos al proceso cambiante



(a) La primera figura muestra el comportamiento en niveles de precisión obtenido por el primer grupo de clasificadores. En este caso, debido a la calidad inferior de los datos comprobamos que las fluctuaciones aumentan en número y en desplazamiento sobre todo en la zona de compresiones más pequeñas. Sin embargo, el comportamiento similar se mantiene.



(b) La segunda figura muestra cómo se han comportado el segundo grupo de clasificadores. Observamos que, como en todas las gráficas anteriores, el comportamiento es similar obteniendo mejores resultados a mayor nivel de compresión. Sin embargo, en esta ocasión no hay un clasificador que destaque por un bajo rendimiento.

Figura 5.23: Resultados obtenidos durante la comprobación de los modelos generados a través de un conjunto de datos acumulado frente al conjunto de datos original con el fin de la detección de posibles problemas de carga máxima. Se muestran los datos en términos de precisión para cada uno de los porcentajes de compresión. Para este experimento no se ha llevado a cabo la validación cruzada.

dos mucho más dispersos. Sin embargo, al realizar este experimento, un único clasificador destaca por encima de los demás mientras que el resto de ellos mantienen un comportamiento muy parejo.

Por otra parte, y tratando el tema de las fluctuaciones que sufren los resultados, este segundo grupo también las padece. A pesar de ello, sabemos que se deben al no determinismo de los métodos de Montecarlo. Pero al igual que el primero de los grupos, según se aumenta el grado de compresión estas se reducen en número y amplitud. En concreto, cuando el grado de compresión supera el 46 % (un grado de generación de instancias igual al 54 %) las tendencias comienzan a asemejarse a una trayectoria mucho más lineal. Los clasificadores, a excepción del bosque aleatorio generado con 350 árboles (el mejor de ellos y que iguala el rendimiento de los bosques aleatorios que se encuentran en el primer grupo), se engloban dentro del rango de precisión que va desde 73 al 82 %. Por consiguiente, con una pequeña ampliación del rango definido en el primer grupo, podemos indicar que los clasificadores tienen un rendimiento de entre el 73 y el 87 % en términos de precisión.

Por último, y relacionado con el comportamiento concreto de los clasificadores, dentro de este segundo grupo, todos ellos funcionan de una forma muy similar. Pero no podemos dejar de lado el comportamiento alcanzado por aquellos que habitualmente han obtenido peores resultados como son los SVM con funciones de núcleo Pearson VII y RBF. En el caso que nos ocupa, estos dos clasificadores se convierten en el tercer mejor clasificador de este grupo (por detrás de los bosques aleatorios y el SVM con núcleo polinomial normalizado) y el peor de todos pero a menos de una unidad de diferencia con el KNN de $K = 2$ respectivamente.

El primero de ellos, durante los diferentes porcentajes de precisión, mantiene el mismo comportamiento que el conjunto general de clasificadores. Pero, al llegar a una compresión del 68 % aumenta increíblemente su rendimiento manteniendo, a partir de ese momento, una tendencia totalmente lineal. Por otra parte, el segundo de ellos también mantiene la tendencia general del grupo de clasificadores. Sin embargo, al llegar al punto donde el conjunto de datos se comprime un 80 %, su rendimiento baja considerablemente (aproximadamente 5 unidades) hasta convertirlo en el peor de los clasificadores. A pesar de estos hechos, los resultados alcanzados son muy interesantes y nos hacen pensar que, a falta de los resultados sobre las tasas de error, el método funciona de manera precisa para la adaptación de diferentes modelos y diferentes conjuntos de datos.

Centrándonos ahora en los errores, la Figura 5.24a ilustra el comportamien-

5. Adaptando los modelos al proceso cambiante

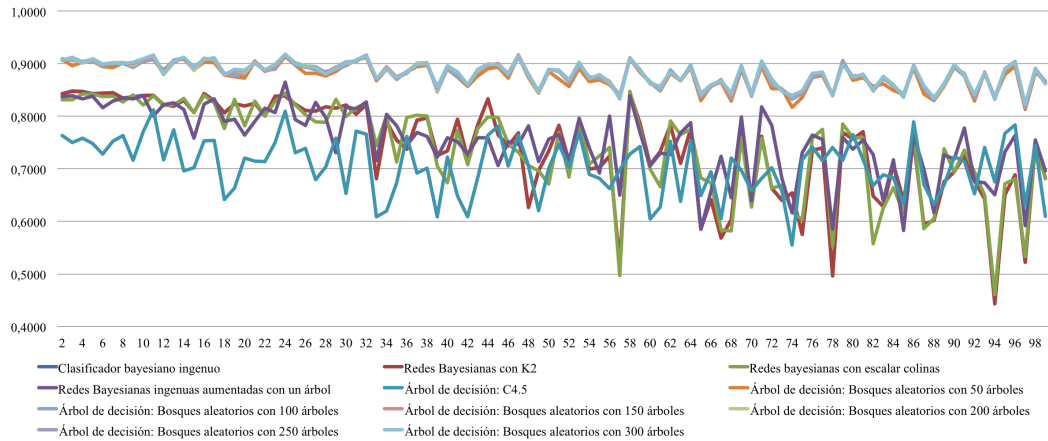
to de los clasificadores del primero de los grupos. Prácticamente, la tendencia seguida por todos ellos vuelve a ser la misma, es decir, cuanto mayor es el grado de compresión de los conjuntos de datos, mejor gestionan los errores. Además, como ya se vio para la precisión en el acierto, todos los clasificadores han sido capaces de mejorar el rendimiento gestionando sus errores. Incluso, el árbol de decisión C4.5, a pesar de ser el que peores resultados ha alcanzado en este grupo. En este caso, a pesar de su posición, se encuentra muy cercano al comportamiento de otros clasificadores (menos de una décima) y consiguiendo una mejora en su propio rendimiento de más de décima y media. Las fluctuaciones que sufren los diferentes clasificadores vuelven a verse reducidas según aumenta el porcentaje de compresión, así, una vez es superado un porcentaje de compresión del 46 % (generación del 54 % de las instancias) los resultados comienzan a ajustarse a una recta.

Esta vez, separaremos el comportamiento en dos grupos bien diferenciados. En el primero encontramos los bosques aleatorios. Estos son los mejores clasificadores en términos de error. En la zona de compresiones máximas consiguen unos valores de 0.91 unidades para el área bajo la curva ROC. Pero a pesar de ser los que mejor rendimiento logran, estos clasificadores también son los que mantienen una tendencia más clara y con fluctuaciones mucho más pequeñas. Es cierto que los máximos y los mínimos existentes en la curva de su comportamiento coinciden con los del resto de los clasificadores, pero con una amplitud mucho menor. La mejora que sufren estos clasificadores ronda la media décima.

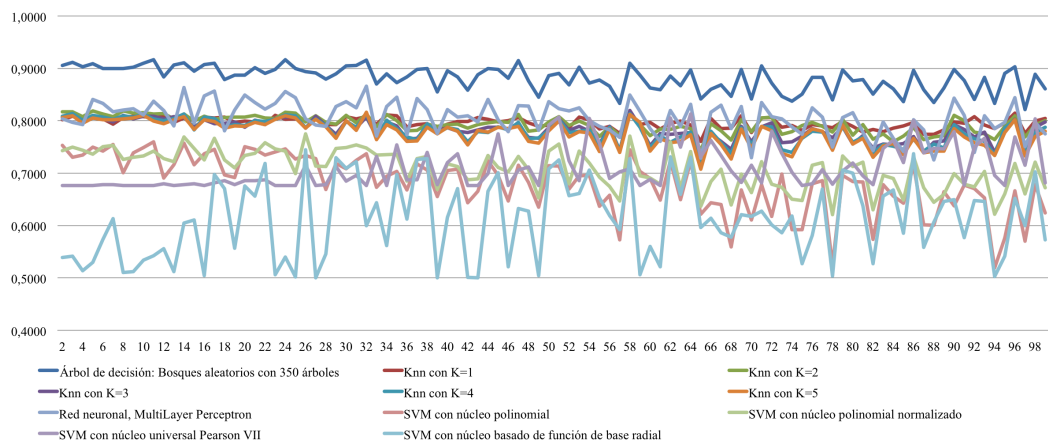
El segundo de los grupos es el formado por las redes bayesianas y el árbol de decisión C4.5. En este caso, todos los clasificadores bayesianos obtienen los mismos resultados para las tasas de error. A lo largo del espectro de compresión puede haber pequeñas desviaciones, pero en general, todas se comportan de la misma forma. En este experimento, la red bayesiana ingenua aumentada con un árbol se ajusta bastante bien al proceso. Este resultado es sorprendente ya que en experimentos anteriores, en los que se probaron los clasificadores frente al conjunto de datos original, siempre encontrábamos que no era capaz de mantener el mismo buen comportamiento que había alcanzado en la precisión; aquí, por contra, sí lo ha hecho.

Por otra parte, el árbol de decisión C4.5 sufre de un comportamiento muy variable en lo que respecta a las tasas de error. Comparándolo con el comportamiento seguido por este clasificador cuando realizamos las mediciones de los porcentajes de acierto, obtenemos dos curvas que cuentan con bastantes similitudes. Esto se debe a que durante la gran mayoría del espectro de com-

5.3 Validación empírica



(a) En la primera de las figuras todos los clasificadores siguen la misma tendencia. No obstante, los podemos unir en tres grupos, lo que mejores resultados adquieren (los bosques aleatorios), los que peores resultados (el árbol C4.5) y los que se quedan en una zona central (el resto de los clasificadores).



(b) La segunda figura muestra que la línea obtenida para la clasificación es muy similar a la de los demás clasificadores. Sin embargo, el bosque aleatorio con 350 árboles obtiene mejores resultados. Por otra parte, los SVM con funciones de núcleo Pearson VII y RBF son lo peores de todos ellos. Y en general, todos mejoran en el manejo de errores a altas compresiones.

Figura 5.24: Tasas de error obtenidas al enfrentar los modelos aprendidos con el conjunto acumulado a los datos originales y probados contra el conjunto original manteniendo el objetivo de predecir la carga máxima. Se muestran los datos en términos de área bajo la curva ROC para cada uno de los porcentajes de compresión. Por lo general, todos los clasificadores consiguen manejar de mejor forma los errores cuanto mayor ha sido la tase de compresión.

5. Adaptando los modelos al proceso cambiante

presión, su curva permanece unida a lo que hacen las redes bayesianas. Sin embargo, una vez se supera un porcentaje de generación de instancias del 44 % (i.e., aplicando una compresión del 56 %), su curva comienza a separarse de los logros alcanzados por las redes bayesianas colocándose a aproximadamente una décima de diferencia.

Los resultados en la gestión de errores para el segundo grupo de clasificadores queda recogido en la Figura 5.24b. Ahora, como viene siendo habitual, volvemos a vislumbrar la diversidad que hay en este segundo grupo. Los resultados que obtienen estos clasificadores se dispersan a lo largo de aproximadamente la mitad del espacio (concretamente se encuentran entre las 0,5 y las 0,92 unidades). No obstante, a pesar de esa variedad de valores, el comportamiento general del grupo es el mismo. Es decir, las líneas que dibujan los clasificadores son muy parecidas. Además, el rendimiento mejora cuando se van realizando compresiones mayores para la generación de los conjuntos de datos acumulados. Concretamente, de entre los once clasificadores, sólo encontramos uno que no es capaz de al menos igualar el comportamiento obtenido sin aplicar el proceso de reducción. Ese clasificador es el SVM generado con una función de núcleo RBF (del inglés “*Radial Basis Function*”).

En este grupo, a primera vista, parece que las fluctuaciones que sufren los clasificadores son mayores que las analizadas en el primero de ellos, sin embargo, se trata de un efecto visual provocado por las fluctuaciones de una de las máquinas de soporte vectorial. El resto de los clasificadores las siguen teniendo, pero con una amplitud en sus ondas mucho más pequeña. Además, como en el grupo anterior, las ondulaciones de las curvas de comportamiento son inversamente proporcionales al porcentaje de compresión, por ello, cuanto mayor es el porcentaje, menores son las fluctuaciones. El punto a partir del cual se comienzan a reducir es el de una compresión del 46 %, en otras palabras, una generación de instancias del 54 %.

En este segundo grupo encontramos tres agrupaciones de comportamientos. La primera de ellas es la relacionada con el último de los bosques aleatorios creado mediante 350 árboles. Este clasificador es el que mejor rendimiento alcanza en el grupo, pero además, imita casi a la perfección el comportamiento que siguen el resto de los bosques aleatorios en el primero de los grupos. Por ello, este tipo de clasificador es considerado como el mejor para la gestión de los errores a la hora de manejar conjuntos de datos que han sido aprendidos mediante la acumulación y compresión iterativa.

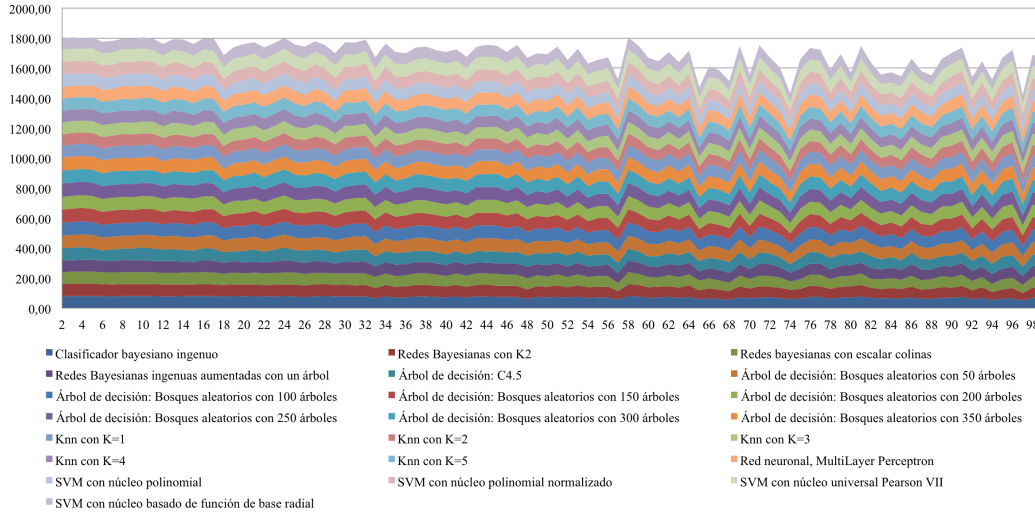
Posteriormente, como en experimentos anteriores, podemos realizar la agrupación de los diferentes KNN y la red neuronal artificial. Específicamente, la

red neuronal no se asemeja a los KNN tanto como los KNN entre ellos mismos. Pero las zonas en las que se mueve sí que son muy similares. Por eso, y debido al resultado alcanzado cuando la compresión es máxima, la introducimos dentro del grupo. Estos clasificadores obtienen unos resultados relativamente buenos para lo que se ha conseguido obtener en la predicción de la carga máxima. En este caso se encuentran en el rango que va de las 0,80 a las 0,82 unidades. Fijándonos en esos valores, estos clasificadores coinciden con las redes bayesianas del primer grupo.

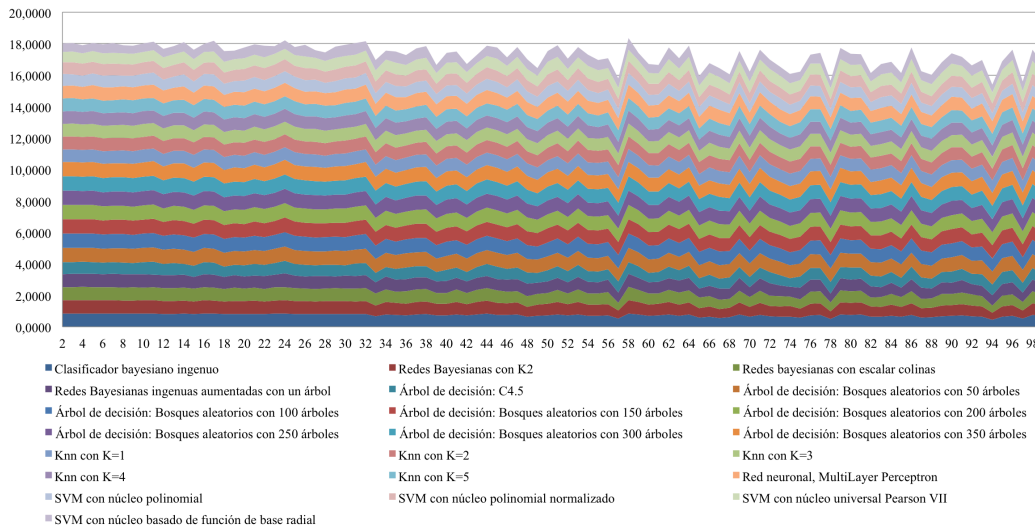
Y para finalizar, cabe señalar el comportamiento tan dispar seguido por las máquinas de soporte vectorial. En concreto, dos de ellas, las creadas con funciones de núcleo polinomial y polinomial normalizado, mantienen comportamientos similares. Ambos clasificadores podrían incluirse en el grupo anterior, sin embargo, los resultados que alcanzan son más bajos (i.e., alrededor de las 0,75 unidades). Pero los SVM creados con funciones de núcleo Pearson VII y RBF se alejan de la normalidad al adoptar cada uno su propio esquema de comportamiento. Al menos, el primero de ellos es capaz de mejorar los resultados cuando el grado de compresión es máximo. Al comparar los rendimientos logrados tanto en la precisión como en los errores observamos que en ambos aparece una sección eminentemente lineal durante los últimos porcentajes de compresión. El segundo de ellos es el peor de los clasificadores a la hora de realizar la gestión de los errores. Éste sufre las fluctuaciones de mayor amplitud. Además, éstas no llegan a reducirse a lo largo de todo el espectro de compresiones. Podemos añadir que este es el único clasificador que no mejora y que tiende a reducir su gestión de los errores alcanzando unos valores cercanos a las 0,50 unidades. A pesar de contar con picos que llegan al 0,80, éstos se encuentran rodeados por valores muy bajos que acaban restándole importancia. Particularmente, parece que los clasificadores que más afectados se ven por las operaciones de adaptación que utilizan la compresión bayesiana basada en métodos de Montecarlo son la máquinas de soporte vectorial.

Una forma mucho mejor de visualizar los resultados es mediante la utilización de áreas apiladas. Estos gráficos nos permiten detectar de una forma mucho más sencilla el lugar en el que se encuentran los mejores y los peores rendimientos generales para el conjunto de clasificadores. Además, nos ayudan a identificar cómo las tendencias que siguen éstos son muy parejas. Y finalmente, este tipo de gráficas es idóneo para que puedan visualizarse los datos cuando los gráficos de líneas pueden ser confusos debido a la cantidad de información que muestran de forma concurrente. Concretamente, la Figura 5.25a expone toda la información citada en relación a la precisión que alcan-

5. Adaptando los modelos al proceso cambiante



(a) Esta primera figura ilustra los diferentes resultados de precisión alcanzados por los clasificadores mostrando las tendencias seguidas por todos ellos de una forma mucho más legible. Como se puede observar, las tendencias son muy similares entre sí.



(b) La segunda gráfica muestra los resultados apilados relativos a las tasas de error, más concretamente, el área bajo la curva ROC que alcanzaron los clasificadores en su proceso de predicción de la carga máxima.

Figura 5.25: Resultados apilados obtenidos en los experimentos que prueban los modelos acumulados, testados frente al conjunto de datos original, para la predicción de defectos en la carga máxima. Se recogen tanto los valores alcanzados en la precisión como aquellos relativos al área bajo la curva ROC.

zaron los clasificadores a la hora de realizar las predicciones del estado de la carga máxima en el futuro. También, la Figura 5.25b hace lo propio con la gestión de las tasas de error. De esta forma, muestra los valores apilados para las diferentes áreas bajo la curva ROC que lograron todos y cada uno de los clasificadores para cada uno de los porcentajes de compresión posible.

A modo de breve conclusión, los resultados obtenidos en los experimentos sobre los conjuntos de datos acumulados completan los vistos en la sección 5.3.2.2. Entonces, una vez probado el método que genera conjuntos de datos que van acumulando la información (proceso propuesto para la adaptación) podemos determinar que los resultados obtenidos han sido satisfactorios. Claramente muestran que la utilización de la compresión bayesiana basada en métodos de Montecarlo puede ser un método válido para el envejecimiento de la muestras. A parte de ser útil, independientemente del defecto que se monitorice, mantiene la distribución probabilística inicial y su rendimiento es aproximadamente igual a la utilización del conjunto completo de datos (en ocasiones incluso lo mejora). Por consiguiente, estos resultados indican que el método desarrollado es capaz de llevar a cabo los procesos de adaptación de los modelos con un envejecimiento de muestras que aportará más importancia a los hechos que actualmente suceden en la planta pero sin dejar de lado aquello que sucedió tiempo atrás.

5.4 Discusión de los resultados

Los resultados aquí presentados muestran que el modelo de compresión para el envejecimiento de datos es viable para procesos de adaptación automática. No obstante, existen una serie de consideraciones respecto al método desarrollado.

En primer lugar, el método utilizado para realizar la compresión se basa en la adquisición de la distribución probabilística a través de una red bayesiana. En este caso, la red bayesiana es el mero instrumento que nos simplifica los cálculos. Sin embargo, tal y como se puede observar en las secciones 4.3.2.1 y 4.3.2.2, este modelo no fue el que mejor se ajustaba al proceso a la hora de realizar las predicciones de ambos defectos, microrrechupes y carga máxima. Del mismo modo, en caso de no ajustarse correctamente, en el momento de realizar el muestreo con los métodos de Montecarlo estaremos partiendo desde una base errónea, consiguiendo generar un conjunto de datos que realmente no cumple con el comportamiento real de la planta. No obstante, en el caso en el que hemos trabajado, las redes bayesianas no obtienen malos resultados

5. Adaptando los modelos al proceso cambiante

que acaben distorsionando la realidad y produciendo ruido, pero podría haber situaciones en las que sí lo causen. Por ello, sería necesario la generación de la distribución probabilística mediante algún método capaz de extraerla y que a su vez sea robusto y fácil de probar.

En segundo lugar, como se comenta en [Bin79] los métodos de Montecarlo, dada la pseudo-aleatoriedad, no son deterministas. Debido a esto, los experimentos han sido desarrollados manteniendo los ficheros. Además, las fluctuaciones que se dan dentro de los resultados son en parte gracias a este hecho. Por ello, puede que haciendo un segundo muestreo, los resultados que alcancen los clasificadores sean mucho mejores, o peores. Por ello, una solución es la búsqueda de alternativas a los métodos de Montecarlo como podrían ser la *simulación directa* [Fis70] u otras soluciones como las que se presentan en [Yap75, CM97, Gad11].

En último lugar, este método nos permite no solo trabajar en la reducción del conjunto de datos, sino en la generación de más instancias dada la distribución estadística del comportamiento. En otras palabras, este método puede generar nuevas instancias, una ampliación del conjunto de datos, pero que a pesar de crecer en cantidad siempre cumplirá la distribución extraída de los datos originales. De esta forma, se podría conseguir modificar el conjunto de datos como lo hace una de las técnicas existentes en la literatura y que tiene como fin balancear los conjuntos de datos, "*Synthetic Minority Over-sampling TEchnique*" (SMOTE) [CBHK11].

5.5 Sumario

En este capítulo hemos introducido un nuevo método que nos permite realizar el proceso de adaptación automático en un MPC. Concretamente, hemos realizado dos contribuciones fundamentales, (i) la generación del método que permite realizar el envejecimiento de las muestras para adaptar los modelos predictivos y (ii) definir y probar el proceso de adaptación continua propuesto.

Para la primera de las contribuciones hemos identificado las herramientas estadísticas existentes en la literatura que nos permitan trabajar con los conjuntos de datos de aprendizaje. En este proceso establecemos las operaciones básicas para nuestro método como son la compresión de datos basada en redes bayesianas y las técnicas de Montecarlo que permiten realizar muestreos basados en una distribución probabilística dada. Posteriormente, y basándonos en estas herramientas, hemos definido y desarrollado un método que permite

trabajar sobre los datos utilizados en las etapas de aprendizaje de los modelos con una base estadística. El objetivo que busca es reducir la importancia de la información antigua y aumentar la de la nueva, pero sin llegar a eliminar el conocimiento adquirido. Este método se conoce como *compresión bayesiana basada en métodos de Montecarlo*, y utiliza las redes bayesianas para la extracción de la distribución probabilística inherente en los datos con el objetivo de emplearla, posteriormente, para muestrear un nuevo conjunto de datos con menos instancias.

La segunda de las contribuciones está relacionada con el proceso de adaptación automático. Sencillamente, se trata en un proceso iterativo mediante el cual generaremos nuevos conjuntos de datos de aprendizaje, obtendremos nuevos modelos y los sustituiremos en caliente sobre el sistema MPC que está actuando en planta. Este tipo de adaptación conlleva que el aprendizaje de los modelos sea realizado mediante la aplicación de múltiples reducciones. Así, con el fin de demostrar la viabilidad de este nuevo modelo se han desarrollado los experimentos que permiten probar si las distribuciones estadísticas se mantienen al reducir los datos, si la generalización de los modelos reducidos es suficientemente buena y si la adaptación acumulativa no estropea el conocimiento ya adquirido.

Los resultados obtenidos por el enfoque aquí presentado han sido satisfactorios, con lo que consideramos que el método presentado puede abrir un nuevo camino en la adaptación de modelos probabilísticos generados a través de un proceso de aprendizaje automático supervisado.

«Es mejor corregir tus propios fallos que corregir los de los demás.»

Demócrito (460 a.C.-370 a.C.)

6

Generación de nuevos parámetros de producción

EL funcionamiento de un sistema MPC pasa por etapas tales como la predicción del estado estacionario $t + 1$, la actualización de los subsistemas con el fin de mantener un MPC ajustado al proceso o la determinación de que valores deben tener los diferentes parámetros de la producción para solucionar el futuro problema que habrá en la planta. Los dos primeros procesos fueron definidos y solucionados en los capítulos 4 y 5 respectivamente. Así, y con el objetivo de completar un sistema MPC de nueva generación, este capítulo se centrará en la última de ellas, también conocida como el *proceso de realimentación de la planta*.

Esta última etapa requiere desarrollar diferentes subtareas como son: (i) la determinación del estado al que se desea llegar, estado que solucionará los objetivos problemáticos de la planta; (ii) la determinación del camino de adaptación, proceso por el que se conseguirá determinar la ruta a través de la que caminará la producción permitiendo alcanzar el estado determinado en la etapa anterior; y (iii) la salida de valores a modificar, etapa que se encarga de comunicar los datos generados (etapas extraídas de la Figura 2.1 presentada en la sección 2.2.4). No obstante, para poder llevar a cabo cada una de es-

6. Generación de nuevos parámetros de producción

tas tareas debemos cumplir una serie de aspectos que determinarán el éxito del sistema y mantendrán la producción dentro de unos niveles normales de trabajo. Concretamente, esas ideas son las siguientes [CB04]:

- Todos los valores generados deben mantener las restricciones de entrada y de salida marcadas por el proceso de producción.
- Los cambios o modificaciones que vayan a realizarse sobre el proceso productivo no pueden ser bruscos.

Frente a este contexto, para el primero de los pasos encontramos el problema de que la determinación del estado al que queremos llegar hace necesario un conocimiento previo del proceso. No debemos olvidar que el método que estamos desarrollando trata de ser una generalización útil en diversos procesos productivos. De esta forma, debemos evitar la utilización de un conocimiento que no pueda ser adquirido de forma automática. Por consiguiente, la solución a este problema se basará en emplear una condición común que debe cumplirse en todos los procesos productivos y que permita la creación de ese sistema genérico. Concretamente, la condición de la que hablamos incorpora la ideas ya mencionadas, para ello, trata de generar una evidencia cercana a la fallida, cuyos valores que estén dentro de los permitidos por el proceso y de la que tenemos la certeza de que mantendrá el proceso a lo largo de la ruta óptima de producción.

En segundo lugar, debemos determinar el camino que debe seguir la producción para evitar los problemas detectados por los predictores. Concretamente, entre las diferentes opciones con las que se cuenta (para más información ir a la sección 2.2.4.6) y dado que se busca la optimización a través de la instancia que se encuentre lo más próxima posible a la situación actual no deseada, el sistema se encargará de adaptar la planta en un único movimiento. En otras palabras, el horizonte de control¹ estará igualado a 0, permitiéndonos evitar los posibles problemas que podrían darse mientras se está ajustando el proceso productivo. El resultado obtenido de la optimización será utilizado para cubrir las necesidades de realimentación de la planta.

Las soluciones tomadas por las diversas aproximaciones quedaron recogidas en la sección 2.3. A modo de resumen, existen casos que utilizan modelos matemáticos, ya sea el mismo que fue utilizado para la predicción o uno es-

¹El horizonte temporal sobre el que el sistema MPC llevará el control de la producción con el fin de acercar la situación actual a la situación deseada.

pecífico para el cálculo de la realimentación de la planta. Este tipo de soluciones sobre todo fue adoptado por los primeros métodos que se dieron a conocer [Kal60, RRTP76, GM86]. Otro tipo de métodos más avanzados, hacen uso de las redes neuronales artificiales como sistema de apoyo en la generación de la realimentación de la planta [QB00]. Por último, existe la posibilidad de la generación de los datos necesarios para la realimentación mediante la utilización de sistemas expertos basados en reglas [LD88].

Todas estas soluciones cuentan con la premisa del conocimiento previo del proceso, lo que posibilita el desarrollo del sistema de retroalimentación. Sin embargo, en nuestro caso, no podemos partir de ella, ya que fue desechada en los capítulos anteriores. Tampoco podemos marcar un punto fijo de partida sobre el que trabajar, con lo que el método a utilizar debe permitir un inicio basado en el desconocimiento total del proceso. De esta forma, la solución idónea es el uso de la programación evolutiva plasmada en un algoritmo genético [Hol75]. Este tipo de algoritmos está inspirado en la evolución genética. Sencillamente, los algoritmos hacen evolucionar una población de individuos tal y como sucede en el mundo real. Para la generación de los nuevos individuos se realizan operaciones entre las que se encuentran: (i) el cruce entre genomas (recombinación genética), (ii) aparición de ciertas mutaciones y (iii) el proceso de selección que mantiene a los más fuertes del grupo. A través de los diferentes ciclos (generaciones), este tipo de programación acaba convergiendo en probabilidad al óptimo, obteniendo una respuesta a la búsqueda que estamos realizando.

En los últimos años, la programación evolutiva ha sido utilizada con el fin de dotar de una no linealidad a los sistemas MPC [OBK⁺97, ADN01, SB03, YLW03]. Del mismo modo, los autores señalan que están utilizando algoritmos genéticos directamente diseñados para un dominio y solución en concreto [SB03]. No obstante, nosotros presentamos una aproximación que utiliza la programación evolutiva como herramienta para la optimización de los parámetros de producción con independencia del proceso de producción o sus objetivos. Además hace uso de los sistemas predictivos híbridos, lo que nos permite mantener las ventajas de ambos tipos de sistemas (lineales y no lineales) para el proceso de realimentación de la planta.

A modo de resumen, las contribuciones realizadas en este capítulo quedan recogidas en la siguiente lista:

- Mostramos cómo emplear la programación evolutiva como una herramienta que permite generar la información de realimentación de la plan-

6. Generación de nuevos parámetros de producción

ta con el objetivo de evitar los problemas del proceso productivo.

- Proveemos un algoritmo genético con cálculos concurrentes y que se mantiene actualizado de forma automática gracias a la utilización de los modelos de predicción autoadaptativos que presentamos en los capítulos anteriores.
- Describimos cómo combinar los sistemas de predicción de estados estacionarios y la medición de similitudes para obtener una función de idoneidad adecuada para nuestro algoritmo genético.
- Evaluamos de forma empírica el rendimiento del algoritmo genético desarrollado determinando los parámetros de uso óptimos y el número de generaciones necesarias para encontrar la solución.

El resto del capítulo queda organizado de la siguiente manera. La sección 6.1 muestra las teorías sobre la programación evolutiva y qué es un algoritmo genético. Además, detallará las diferentes técnicas que han sido utilizadas para el desarrollo de la herramienta capaz de realimentar la planta de producción. Posteriormente, la sección 6.2 indicará la metodología que ha sido seguida en la experimentación de la nueva herramienta desarrollada y mostrará los resultados que fueron alcanzados para ambos defectos. En la sección 6.3 discutiremos diferentes aspectos relacionados con la solución propuesta y, finalmente, la sección 6.4 recogerá un breve resumen de las aportaciones realizadas en este capítulo.

6.1 Programación evolutiva

La programación evolutiva, a través de la utilización de *algoritmos evolutivos*, trata de realizar un procedimiento de búsqueda o de optimización basándose en las ideas del mundo biológico. Concretamente, la teoría de la evolución de Darwin [Dar37] con la supervivencia del individuo mejor adaptado al medio es la base de la que parte este tipo de programación. De esta forma, generalizando las ideas de Darwin, la programación evolutiva trata de solucionar problemas que requieran de una adaptación, búsqueda u optimización. No obstante, para asegurar el éxito de este tipo de programación debemos tener en cuenta que esta teoría de la evolución no debe ser aplicada estrictamente. Las aproximaciones que mantienen estas ideas como inspiración recogen mejores frutos de la programación evolutiva. Este tipo de soluciones han obtenido grandes

resultados en dominios muy diferentes como el empresarial, la ingeniería o la ciencia [Gol94].

Los algoritmos evolutivos o algoritmos genéticos, como los llamaremos a partir de ahora, tienen como objetivo la adaptación de una población de individuos dado un entorno. Para ello, tienen en cuenta las reglas presentes en la naturaleza, aquellas que intentan promover la supervivencia de los individuos que son más útiles, mientras eliminan aquellos que no lo son. Específicamente, cada uno de esos individuos es una de las posibles soluciones que el algoritmo genético está estudiando y evolucionando.

Más formalmente, un algoritmo genético [Hol75] se trata de un proceso iterativo de evolución, el cual mantiene constantemente una población de individuos de entre unas decenas o unos cientos que, originariamente, fueron creados con valores aleatorios o bajo una heurística conocida. A lo largo de cada una de las iteraciones, llamada generación, los individuos de la población actual son evaluados, en otras palabras, se calcula el valor que indica su idoneidad para la solución del problema inicial. Para la creación de una nueva generación de individuos, estos son seleccionados con una probabilidad proporcional a su idoneidad relativa. De esta forma, aseguramos que el número de veces que un sujeto es elegido es aproximadamente proporcional a su comportamiento relativo en la población. De esta forma, los mejores individuos tienen una oportunidad mayor de reproducirse. Únicamente el uso de la selección no permite la generación de nuevos puntos que no se encuentren ya en el espacio inicial definido. Por ello, para solucionar esta problemática se hace uso de dos operaciones genéticas:

- **Cruce (en inglés “*Crossover*”).** Es el operador de combinación más importante. En esta operación se recogen dos sujetos, llamados padres, y se intercambia información de ambos con el objetivo de generar dos nuevos, llamados hijos.
- **Mutación.** Se trata de la adición de ruido que permita evitar el prematuro descubrimiento de una solución óptima local. En concreto, es la modificación de información en los individuos de forma casual, pero ajustada a un ratio prefijado que intenta simular lo que sucede en la naturaleza.

El proceso finalizará al llegar a generar un número predefinido de generaciones o cuando se haya encontrado la solución óptima que se estaba buscando.

6. Generación de nuevos parámetros de producción

En forma de resumen, el Algoritmo 3 recoge el proceso que acaba de ser descrito. El algoritmo genético utilizado en este experimento ha sido desarrollado basándose en estas ideas fundamentales y en el algoritmo general. No obstante, el dominio sobre el que trabajamos incorpora una serie de restricciones que deben ser sorteadas y que nos obligan a modificar el funcionamiento habitual.

Algoritmo 3: Algoritmo genético estándar.

```
Producir población inicial de individuos
while no se cumpla la condición de finalización do
    Evaluar la idoneidad de los individuos
    Seleccionar los mejores individuos para la reproducción
    Producir nuevos individuos
    Generar la nueva población (eliminar peores y añadir mejores)
    Mutar algunos individuos
end
```

La primera y más importante de esas restricciones es el tiempo limitado con el que cuenta el algoritmo genético para la resolución del problema de búsqueda. Nótese que los procesos productivos son muy diferentes. A pesar de ello, el algoritmo genético no dispone de un tiempo finito para lograr su objetivo. De esta forma, y tratando de mejorar el rendimiento de este tipo de programación hemos utilizado la siguiente evolución de los algoritmos genéticos. Concretamente, estamos hablando de la utilización de *Algoritmos Evolutivos Paralelos* [CP98].

En [CP98] y [Tom95] podemos comprobar que existen diversas formas de desarrollar este tipo de evolución. En nuestra investigación nos hemos centrado en la utilización del poder de computación con el que cuentan los ordenadores actuales. De esta forma, y gracias al empleo de la programación concurrente incorporada dentro del propio algoritmo genético, compartimos el procesamiento entre los diversos núcleos de nuestros procesadores. Claramente podemos observar que la mejora que buscamos es la reducción del tiempo de computación. No obstante, este tipo de desarrollo también permite beneficiarnos de un ajuste en paralelo desde el punto de vista algorítmico, es decir, el comportamiento análogo a lo que sucede en la evolución natural paralela de poblaciones espacialmente separadas [Tom95].

En nuestro caso, el tiempo necesario para el cálculo de la idoneidad de cada uno de nuestros individuos es muy elevado al compararlo con cualquiera del resto de las operaciones que desarrolla el algoritmo genético. Para solu-

cionarlos, hemos optado por la evaluación simultánea de todos los individuos utilizando los diferentes núcleos de procesamiento existente en las máquinas actuales. Habitualmente, existen más individuos que núcleos de procesamiento, por lo que los individuos son divididos entre los diferentes núcleos de procesamiento de la forma más equitativa posible. En nuestro caso, hemos optado por seleccionar un número de individuos que permita ser evaluados de forma concurrente. En concreto, 10 individuos. El ordenador sobre el que se han corrido los experimentos cuenta con 12 núcleos de procesamiento¹, con lo que utilizamos 10 para las evaluaciones de idoneidad y mantenemos 2 para otros usos. Cabe destacar, que en caso de no contar con un sistema *multicore* el algoritmo desarrollado funciona correctamente, de forma secuencial, sin recibir ninguna penalización temporal al compararlo con la solución secuencial del mismo. Así, realizando esta modificación, la secuencia de operaciones que finalmente hemos utilizado sigue el pseudocódigo mostrado en el Algoritmo 4.

Algoritmo 4: Algoritmo genético con evaluación concurrente.

```
Producir población inicial de individuos
while no se cumpla la condición de finalización do
    for Cada individuo do in parallel
        | Evaluar la idoneidad
    end
    Seleccionar los mejores individuos para la reproducción
    Producir nuevos individuos
    Generar la nueva población (eliminar peores y añadir mejores)
    Mutar algunos individuos
end
```

A continuación pasaremos a explicar cada una de las tareas que se desarrollan dentro del algoritmo genético. Concretamente, ampliaremos la información del proceso de inicialización de los genomas, la generación de nuevos individuos y la forma utilizada para realizar la valoración de éstos.

6.1.1 Inicialización

El proceso de inicialización es el que permite generar la población original que va a ser evolucionada a través del algoritmo genético y entre la

¹Intel® Core™ i7-3930 CPU a 3,20 GHz, 24GB de memoria RAM, sistema operativo Windows 7 de 64 bits

6. Generación de nuevos parámetros de producción

que se encontrará la solución final al problema. De esta forma, en nuestro caso contaremos con una población $\mathfrak{P} = \{G_1, G_2, \dots, G_9, G_{10}\}$, donde cada G_i es un genoma o individuo. A su vez, cada G puede definirse como $G = \{Gn_1, Gn_2, \dots, Gn_{n-1}, Gn_n\}$ donde Gn_i es cada uno de los genes que componen el genoma. En otras palabras, los genes (Gn) serán las variables que conforman cada una de las evidencias o soluciones, G .

La inicialización, como en todo algoritmo genético se hace a través de un proceso aleatorio. Sin embargo, un sistema MPC debe mantener en todo momento las soluciones dentro de las restricciones marcadas por el proceso de producción [ML97]. Esta fue una de las ventajas que añadieron estos sistemas y no debe ser eliminada. Para conseguir mantener la generación de instancias bajo unas restricciones, al realizar la inicialización de cada gen, se les hará llegar la información básica de sí mismos, es decir, el nombre del gen (variable del proceso de producción) y los estados que pueden alcanzar. Por consiguiente, a la hora de realizar una generación aleatoria del estado actual para ese gen no serán violadas las restricciones del proceso. Es interesante comentar que los límites son adquiridos automáticamente a través de los modelos aprendidos para la predicción. Esto significa que en cuanto sea actualizado el sistema de predicción, las restricciones para la generación de la realimentación de la planta también quedarán actualizadas.

A las restricciones de la planta hay que añadir una nueva. Dado que existen variables que no pueden ser modificadas en el proceso, estas deberán mantener el estado con el que se inició la optimización. De esta forma, durante el proceso de inicialización únicamente recibirán un único valor al configurar sus restricciones. Así, durante el proceso de generación aleatoria siempre será seleccionado el mismo valor. Un ejemplo de este tipo de variables es la referencia de la pieza. Si este gen es modificado, habremos indicado que estamos trabajando en la producción de una pieza diferente, algo totalmente falso y que puede perturbar el proceso de optimización.

Finalmente, una vez que toda la configuración de límites ha sido finalizada se procede a realizar la generación aleatoria habitual.

6.1.2 Creación de nuevas generaciones

El proceso de generación de nuevos individuos es uno de los más importantes dentro de los algoritmos genéticos. Gracias a los resultados que alcance esta tarea, el conjunto completo de individuos irá evolucionando hasta encontrar la solución final. Particularmente, y como ya comentamos al comienzo de la

sección, la creación se hace en base a dos operaciones genéticas, el cruce de los individuos y la mutación. En esta sección pasaremos a detallar el modo de funcionamiento de cada una de estas operaciones genéticas para la solución adoptada como método de realimentación de la planta.

6.1.2.1 Cruce de individuos

El cruce entre individuos es la primera de las operaciones genéticas que permite realizar la evolución de los mismos. Para ello, dados dos individuos conocidos como padres, G_1 y G_2 , seleccionados a través del método explicado en la sección 6.1.2.3, pasaremos a obtener dos nuevos individuos, conocidos como hijos, H_1 y H_2 , que formarán parte de la siguiente generación de genomas. Nótese que este proceso de cruce se llevará a cabo únicamente si se supera un ratio marcado para esta operación (en inglés “*crossover rate*”). Dado que se busca la máxima similitud con lo que supone la teoría de la evolución en el mundo real, este valor debe ser bastante alto. En el caso que nos ocupa, hemos tomado un valor del 80 %. Esto significa que la gran mayoría de las veces que se intente realizar el cruce entre dos genomas tendrá éxito.

Cuando se cumpla el ratio de cruce marcado, el proceso seleccionado para la unión de padres se trata de uno de los más simples que se conocen [Tom95]. La razón de su selección es la reducción de la complejidad computacional del algoritmo. De esta forma, dados los dos padres que van a cruzarse, será seleccionado un gen al azar y a partir de éste intercambiaremos los genes de los padres. Como resultado se obtendrán dos nuevos hijos que comparten características de ambos padres y en los que la cantidad que comparten de cada uno de ellos es generada de forma aleatoria.

6.1.2.2 Mutación de individuos

La segunda de las operaciones genéticas es la mutación. Esta segunda operación es necesaria ya que, en caso de trabajar únicamente con el cruce de genomas, siempre obtendríamos individuos de un espacio de soluciones reducido, siendo incapaces de encontrar la mejor solución al problema (obtendríamos una solución óptima local). Al igual que en el caso anterior, la operación también se producirá cuando se cumpla un ratio de mutación (en inglés “*mutation rate*”). A diferencia del ratio de cruce, éste debe ser bajo, ya que la mutación no es tan habitual como el cruce entre genomas. De esta forma, para el algoritmo desarrollado hemos optado por una probabilidad en la mutación del 5 % .

6. Generación de nuevos parámetros de producción

Para realizar la mutación, partiendo del genoma generado a través del cruce llevaremos a cabo una comprobación gen a gen para ver si se cumple el ratio de mutación marcado. En caso de hacerlo, el gen mutará, en otras palabras, obtendrá un nuevo valor de estado generado de forma aleatoria pero manteniéndose dentro de las restricciones marcadas para cada una de las variables del proceso de producción.

6.1.2.3 Selección de individuos

El proceso de selección es otra de las partes vitales del algoritmo. Esta tarea se centra en seleccionar los mejores individuos de la población para combinarlos entre ellos posibilitando encontrar la solución más óptima. Para desarrollar esta tarea hemos elegido el método de *selección proporcional a la idoneidad* (en inglés “*Fitness Proportionate Selection*”) también conocido como *selección de la ruleta* (traducción literal de la voz inglesa “*Roulette-wheel selection*”) [Bäc96].

En este método, cada uno de los valores de idoneidad tiene asociada una probabilidad de selección de cada uno de los genomas. De esta forma, si tomamos que el valor de idoneidad para los genomas es representado como I_i la probabilidad de ser seleccionado ese genoma es la que puede calcularse con la ecuación 6.1.

$$p_i = \frac{I_i}{\sum_{j=1}^N I_j} \quad (6.1)$$

donde N es el número de individuos de cada generación.

Este método recibe el nombre de la ruleta por la similitud que mantiene con la ruleta de los casinos. Usualmente, una proporción de la ruleta es asignada a cada una de las posibles selecciones basándonos en los valores de idoneidad. Esto lo podemos conseguir al realizar una división de la idoneidad de una selección entre la suma total de la idoneidad, permitiendo así, normalizar el valor a 1. Posteriormente, se realizará la elección de un valor de forma aleatoria, de la misma forma a como lo hace la ruleta de un casino cuando está girando.

Utilizando este método, obtenemos un método de selección que cuenta con una alta probabilidad de escoger a los mejores individuos, pero sin descartar la posibilidad de poder desecharlos, algo que ciertamente puede pasar en la naturaleza. Además, esta característica puede ser útil, ya que parte de la información de los genomas menos idóneos puede ser interesante para la generación

posterior de un genoma mucho mejor. Este método está en contraposición con otros métodos más sencillos como la selección por truncamiento¹.

Para llegar a seleccionar el estado es necesario llevar a cabo una búsqueda entre la representación de valores que ha sido utilizada. De esta forma y con el fin de mejorar el rendimiento del propio sistema hemos optado por una búsqueda binaria en lugar de una búsqueda secuencial.

A parte de este método, existe otra forma de ayudar al proceso de selección de ejemplares. Concretamente, ésta recibe el nombre de *elitismo*. Cuando activamos esta opción, el mejor de los individuos se va a mantener entre las generaciones a pesar de realizar el cruce de otros individuos. En nuestro caso, hemos realizado una selección elitista a la hora de crear una nueva generación para conseguir converger en la solución óptima mucho antes.

6.1.3 Valoración de individuos

El método de la valoración de los individuos es uno de los más importantes [Bäc96]. Esta tarea es la que permite valorar cómo de cerca se encuentran cada una de las soluciones o genomas de la solución final. Los valores obtenidos durante esta tarea son utilizados para el proceso de selección descrito en la sección 6.1.2.3.

La valoración de los individuos se hace en base a la función de idoneidad o “*fitness function*” en inglés. Concretamente, se trata de un tipo de función objetivo utilizada para medir, en forma del mérito conseguido, cómo de cerca se encuentra un genoma Gn_i de obtener el propósito inicial marcado. Básicamente, el éxito de un algoritmo genético está totalmente relacionado con la función de evaluación que utiliza.

Además, esta función de valoración debe contar con una ejecución rápida, dado que el algoritmo genético necesita efectuar múltiples iteraciones con múltiples llamadas a la evaluación de una población \mathfrak{P} de genomas, razón por la cual hemos optado por paralelizar el proceso de valoraciones reduciendo el tiempo de procesamiento.

Haciendo una simple clasificación de las funciones de evaluación, podemos diferenciar dos tipos [Bäc96]: (i) las funciones que no cambian a lo largo del ciclo de vida del software y (ii) funciones que son mutables. Concretamente, la solución propuesta para esta tesis doctoral utiliza una función del primer tipo. A lo largo de la vida del sistema de realimentación a planta la función

¹Método de selección que siempre elimina un porcentaje de los genomas más débiles.

6. Generación de nuevos parámetros de producción

va a ser la misma. Sin embargo, el cálculo que se realiza utiliza ciertas partes del sistema que sí van a ser actualizadas, con lo que gracias a esta unión de elementos podemos mantener una función fija que se mantiene adaptada al proceso de producción en todo momento. En concreto, nuestra función de idoneidad tratará de determinar si los nuevos parámetros de producción generados producen una pieza sin defectos y, de ser así, determinarán la proximidad de esta solución al estado actual de la planta, que cuenta con los valores que fueron detectados como problemáticos. De esta manera, nos aseguraremos de obtener un producto sin defectos y que únicamente requiere mínimos cambios en el proceso productivo.

De forma mucho más específica, el cálculo de la idoneidad será realizado únicamente cuando podamos asegurar que la solución generada supera la condición inicial. Así, en caso de que los parámetros de producción del genoma provoquen un producto defectuoso, su idoneidad será directamente la más baja posible, es decir, 0. Si el genoma supera la evaluación inicial, entonces será medida la similitud con aquel que provocó el error. Este proceso de evaluación queda recogido en el Algoritmo 5.

Algoritmo 5: Evaluación de la idoneidad.

```
Input: Genoma actual,  $Gn$   
Input: Genoma erróneo,  $Gn_e$   
Output: Valor de idoneidad,  $f \in [0, 1]$   
predicción  $\leftarrow$  PredecirEstadoEstacionario( $Gn$ )  
if ExisteError(predicción) then  
  |  $f \leftarrow 0$ ;  
end  
else  
  |  $\vec{v}_1 \leftarrow$  ObtenerVector( $Gn$ )  
  |  $\vec{v}_2 \leftarrow$  ObtenerVector( $Gn_e$ )  
  |  $f \leftarrow$  SimilitudDelCoseno( $\vec{v}_1, \vec{v}_2$ )  
end  
return  $f$ 
```

Para desarrollar el algoritmo presentado, identificamos dos tareas fundamentales:

- **Predicción de estados estacionarios $t + 1$.** La forma de determinar si la nueva solución va a ser exitosa es la evaluación a través de los modelos de predicción generados y probados en el capítulo 4. Además, estos

modelos son fuente de información para la inicialización de cada uno de los genes, marcando las restricciones de la producción. No obstante, la actualización constante de estos modelos va a permitir mantener actualizada la función de idoneidad en todo momento (para más información ver el capítulo 5).

- **Medición de similitudes mediante la similitud del coseno.** Cada una de las soluciones propuestas, los genomas, pueden ser transformados en vectores, y así, comprobar la similitud entre ellos. Concretamente, para conocer cómo de similares son utilizaremos la medición del coseno que forman los dos vectores [TSK⁺06]. El coseno de 0 es 1, en otras palabras, es el vector más próximo y, por lo tanto, es el que mayor valor de idoneidad debe obtener. Dados dos vectores A y B , la función de similitud entre ellos $S(A, B)$ se obtiene a través del coseno de los vectores reflejado en la ecuación 6.2, y que está basada en el cálculo del *producto escalar* de dos vectores.

$$S(A, B) = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (6.2)$$

6.2 Validación empírica

Esta sección recogerá los experimentos que han sido desarrollados con el fin de testar el proceso de realimentación de la planta. Concretamente, se trata del experimento que permite demostrar el funcionamiento del algoritmo genético que realiza la búsqueda de los mejores parámetros de producción que tratan de evitar defectos y un gran número de cambios en el proceso.

6.2.1 Metodología general

Para la realización de los experimentos que solucionan los futuros problemas en la planta, hemos empleado los mismos datos que ya fueron utilizados en los experimentos para la predicción de los estados estacionarios $t + 1$ o la adaptación de los modelos. Estos datos han sido extraídos de una fundición que fabrica piezas que acabarán en la industria automovilística y que, al ser piezas de alta precisión y la base de la seguridad activa de un vehículo, sufren unos

6. Generación de nuevos parámetros de producción

fuertes controles de calidad. Estos datos están clasificados según el defecto que intentan minimizar, concretamente los microrrechupes y valores bajos en la carga máxima. Para más información revise las reseñas presentadas en el capítulo 3. Con este conjunto de datos tratamos de responder una serie de preguntas relacionadas con la investigación:

¿Un proceso evolutivo es capaz de obtener o encontrar la solución que optimiza los valores de los parámetros de producción con el fin de no producir productos defectuosos?

¿Es posible desarrollar una única tarea para la modificación del proceso productivo utilizando, para ello, una función de idoneidad común a todo objetivo de optimización marcado?

Dado que un sistema MPC trabaja en intervalos temporales regulares, ¿existen problemas relacionados con el tiempo de ejecución que nos obliguen a generar un sistema subóptimo? o, por contra, ¿podría ser utilizado para cualquier tipo de proceso productivo?

¿Cuál es el número de generaciones que nos permite alcanzar el mayor número de soluciones óptimas?

¿El sistema de realimentación desarrollado aconseja la modificación de una cantidad pequeña de variables, pudiendo aplicar los cambios en un único paso, es decir, sin necesidad de controlar la producción durante una frontera temporal superior a 0?

Como se puede vislumbrar, las tres primeras preguntas hacen referencia al grado de generalización de la aproximación desarrollada. Además, se plantea la generación desde los diferentes puntos de vista posibles, el objetivo de optimización del sistema, la evaluación y la ejecución. Las últimas preguntas hacen referencia al rendimiento esperado del sistema de realimentación propuesto. De esta forma, con el fin de encontrar las respuestas a estas preguntas hemos llevado a cabo la siguiente metodología de investigación:

- **Adquisición de datos.** Este primer paso es el que nos proveerá de los datos necesarios para el desarrollo de la demostración o validación empírica del sistema de realimentación. Este proceso es compartido y sus resultados, el conjunto de datos de trabajo, son los mismos que fueron

utilizados en los procesos de predicción de estados estacionarios $t + 1$ o en la adaptación continua de los modelos. Cabe destacar que estos datos son la base de todos los experimentos, por consiguiente, a mayor calidad en el proceso de adquisición, mejores resultados serán alcanzados en las siguientes operaciones.

- **Evaluación de instancias a través de modelos predictivos.** En este momento contamos con el sistema MPC al completo, por lo que, vamos a poner a prueba el comportamiento de todo el conjunto de métodos desarrollados. De esta forma, cada una de las instancias adquiridas en la etapa anterior será evaluada para buscar futuros problemas. En caso de que se haya detectado alguno, se proseguirá con el resto de pasos de la metodología. Sin embargo, no se llevará a cabo tarea alguna cuando no descubramos defectos, pasando, así, al siguiente análisis. Cabe destacar que, para realizar el estudio de cada una de las evidencias, emplearemos los meta-clasificadores que mejor se ajustaron a cada uno de los defectos. Concretamente, para evitar los microrrechupes se empleará una combinación de clasificadores a través del método *stacking*, mientras que para la carga máxima emplearemos un meta-clasificador basado en *grading*. Ambos usan como apoyo los clasificadores bayesianos ingenuos aumentados con un árbol (para más información sobre la selección de meta-clasificadores o su configuración diríjase al capítulo 4).
- **Búsqueda de la solución óptima.** Una vez se ha determinado que la evidencia sobre la que estamos trabajando podría causar problemas de calidad en las piezas generadas, es el momento de poner en funcionamiento el sistema de búsqueda de soluciones basado en programación evolutiva. Como ya hemos explicado, en nuestra aproximación contamos con un algoritmo genético con evaluación de idoneidad concurrente. Para su funcionamiento ha sido marcado un ratio de cruce del 80 %, un ratio de mutación del 5 % y se utiliza una población de 10 individuos. Además, mantenemos siempre el mejor de los individuos a través de la técnica del elitismo y la selección de los genomas que vamos a cruzar es realizada a través del método de la ruleta. Debemos recordar que la función de idoneidad utilizada es la misma para ambos defectos y se basa en realizar una predicción inicial que identifica las soluciones que evitan los defectos en las piezas y, posteriormente, una medición de similitud entre evidencias. La única diferencia entre los dos defectos es que cada uno de los ellos utilizará un modelo predictivo diferente, consiguiendo un algoritmo genético estático pero con comportamiento dinámico.

6. Generación de nuevos parámetros de producción

- **Evaluación de resultados.** Una vez ha finalizada la ejecución del proceso de búsqueda de los nuevos parámetros de producción, procederemos a realizar la evaluación de los resultados. En este caso nos interesa fijarnos en los parámetros que han sido generados y, por ello, basaremos este proceso en las siguientes métricas:
 - *Ajuste del sistema.* En primer lugar, comprobaremos cómo se ha ajustado el sistema completo y si ha llegado a solucionar los problemas detectados en la planta. De esta forma, partiendo de la base de que conocemos la cantidad de evidencias defectuosas, sabremos cómo de idóneo es el sistema global que presentamos en esta tesis doctoral.
 - *Desplazamiento.* En segundo lugar, nos interesa realizar una serie de mediciones que determinen el desplazamiento que se ha producido entre la evidencia problemática y los nuevos parámetros generados. De esta forma, realizaremos una medición del número de parámetros que ha sido necesario modificar para solucionar futuros problemas en el proceso.
 - *Proceso.* En tercer lugar, queremos analizar el proceso de búsqueda de información mediante programación evolutiva, por ello, realizaremos una serie de mediciones que monitorizarán el número de generaciones necesarias para obtener los valores óptimos de cada solución. A través del análisis del proceso podremos seleccionar las generaciones necesarias obteniendo un proceso altamente óptimo y con una buena velocidad de ejecución.
 - *Tiempo de ejecución.* Finalmente, y muy relacionado con el anterior punto, pretendemos medir el tiempo necesario para encontrar la solución, aspecto harto importante ya que determinará el tiempo de ciclo necesario para el sistema MPC desarrollado.

6.2.2 Realimentación de la planta a través de programación evolutiva paralela

Aplicando la metodología explicada en la sección 6.2.1 y el algoritmo genético con la configuración ya comentada en la sección 6.1 hemos obtenido los siguientes resultados. Nótese que, con el propósito de simplificar la lectura y aumentar la legibilidad de los mismos, vamos a proceder a la división de los resultados según el defecto en el que se centran.

6.2.2.1 Microrrechupes

Como ya ha sido comentado, las soluciones basadas en la programación evolutiva han sido la base de varias investigaciones para conseguir dotar de no linealidad a los sistemas MPC [OBK⁺97, ADN01, SB03, YLW03]. Sin embargo, todas ellas han estado centradas para la optimización de objetivos únicos y fueron diseñadas para un proceso específico. A lo largo de esta sección mostramos los resultados obtenidos para combatir la aparición de microrrechupes a través nuestra solución genérica independiente de proceso y que hace uso de las características híbridas con las que hemos dotado a nuestro sistema MPC.

Concretamente, los resultados que serán comentados hacen referencia a diferentes aspectos del comportamiento del sistema de realimentación de la planta. En primer lugar, dado que para el desarrollo de los experimentos hemos utilizado el sistema MPC completo, mostraremos cómo ha actuado en relación a la solución de los problemas conocidos de antemano. Posteriormente, nos centramos en aspectos tales como el número de variaciones necesarias para corregir el proceso o el número de generaciones necesarias para la solución del problema de búsqueda planteado, así como el tiempo utilizado por el sistema de realimentación para encontrar la solución óptima.

Tabla 6.1: Resumen de los resultados obtenidos en lo referente a las instancias tratadas y corregidas para los defectos de microrrechupes.

Correcciones	
Instancias totales	951
Instancias con defecto	261
Instancias corregidas	261

De esta forma, la Tabla 6.1 muestra el comportamiento del sistema MPC completo. Como podemos observar, de todas las evidencias con las que contaba el conjunto de datos de trabajo, 951, 261 eran defectuosas. A través de la utilización del sistema MPC, contando con un meta-clasificador que realiza la combinación de resultados a través del método *stacking con clasificadores "Tree Augmented Naïve"* para el segundo nivel de pila, han sido detectadas y corregidas el 100% de las instancias defectuosas.

En cuanto a la segunda de las características a medir, hemos trabajado con el número de modificaciones llevadas a cabo por el sistema para cada una de las 261 instancias corregidas. La Tabla 6.2 muestra los resultados alcanzados por el algoritmo genético genérico a la hora de optimizar las evidencias defectuosas. Concretamente, podemos observar que la menor de las modifi-

6. Generación de nuevos parámetros de producción

Tabla 6.2: Resumen de los resultados obtenidos en lo referente a las variables del proceso modificadas con el fin de solucionar y evitar los problemas de microrrechupes.

Modificación de variables	
Número mínimo de modificaciones	1
Número máximo de modificaciones	19
Número medio de modificaciones	2,67
Desviación estándar	1,94
Número de modificaciones marcado por la moda	2
Número de modificaciones marcado por la mediana	2
Número de modificaciones de 1 variable	37
Número de modificaciones de 2 variable	113
Número de modificaciones de 3 variable	87
Número de modificaciones de 4 variable	9
Número de modificaciones de 5 variable	0
Número de modificaciones de más de 5 variable	15

caciones ha sido de una única variable. Resultado harto lógico, ya que la no modificación de ninguna de las variables volvería a obtener una pieza defectuosa, tal y como fue detectado por el sistema MPC en su etapa de predicción del estado estacionario. Hay ciertas ocasiones en las que los patrones identificados en el funcionamiento de la planta no nos permiten afinar tanto el proceso de optimización de los parámetros de producción, así nos lo hace ver el valor máximo de variables modificadas. En este caso, existe una evidencia en la que han tenido que ser rectificadas 19 variables (más de la mitad de las variables con las que trabajamos). A pesar de este valor tan alto, el resto de las mediciones realizadas nos indican que la mayoría de los casos se centran en el rango que va desde una única variable a la modificación de cuatro de ellas. Únicamente, 15 optimizaciones (un 5,75 %) quedaron por encima de este espacio. Además, como se puede observar en la Tabla 6.2 a través de diferentes medidas estadísticas (media, moda y mediana), el valor de modificación más representativo es el de 2 variables. Al calcular el valor medio, podemos notar que la desviación estándar es de un 1,94 variables, un valor muy pequeño y que nos indica que no hay una alta fluctuación en los valores de modificación.

En tercer lugar, hablaremos sobre las generaciones que han sido necesarias para obtener los óptimos en cada uno de los casos. Cabe destacar que, dado el diseño de la función de idoneidad, no vamos a obtener nunca el valor óptimo, ya que eso significaría obtener un vector totalmente similar al defectuoso, pero sin serlo. De esta forma, la finalización de la ejecución del algoritmo genético

Tabla 6.3: Resumen de los resultados obtenidos en lo referente a las generaciones necesarias para obtener la solución óptima que elimina los problemas de microrrechupes.

Generaciones necesarias para el óptimo	
Número mínimo de generaciones	115
Número máximo de generaciones	1900
Número medio de generaciones	587,35
Desviación estándar	433,62
Número de generaciones marcado por la mediana	449
Óptimos logrados con 500 generaciones	147
Óptimos logrados con 1000 generaciones	219
Óptimos logrados con 1500 generaciones	245

estará marcada por un número n de generaciones. De esta forma, y dejando correr el algoritmo durante 2000 generaciones, hemos tomado las mediciones que se muestran en la Tabla 6.3. Al igual que para la medición de modificaciones hemos tomado los valores máximos y mínimos necesarios para encontrar el óptimo. En este caso, vemos que el rango es muy amplio, desde las 115 a las 1900 generaciones. Algo que era de presuponer, ya que la programación evolutiva, dadas las mismas operaciones de generación de individuos, no obtiene el mismo comportamiento en cada una de sus ejecuciones [Bäc96]. Así, valores como la media o la mediana se quedan por debajo de las 590 generaciones, pero la desviación estándar nos marca que existe una variación muy grande, de unas 433,52 unidades.

También, hemos realizado unas mediciones que nos indican cuantos óptimos seríamos capaces de alcanzar con un menor número de generaciones. No debemos olvidar que debido a diversos problemas, por ejemplo el tiempo del ciclo de procesamiento, los sistemas MPC pueden ser subóptimos [ML91, MR93], lo que nos permitiría trabajar con un número de generaciones reducido. Concretamente, utilizando únicamente 500 generaciones el sistema es capaz de alcanzar el 56,32% de los óptimos. En segundo lugar, y con un valor de 1000 generaciones el sistema alcanza una precisión mucho mayor, un 83,91%. Finalmente, con 1500 generaciones alcanzamos un 93,87%. Nótese que a pesar de no encontrar el óptimo, todas las soluciones generadas habrían sido comprobadas por el sistema de predicción, confirmando que los productos finales se encontrarán libres de fallos.

En último lugar, hemos evaluado el tiempo necesario para la ejecución del proceso de optimización de la planta utilizando las 2000 generaciones. Estas

6. Generación de nuevos parámetros de producción

Tabla 6.4: Resumen de los resultados obtenidos en lo referente a los tiempos de ejecución para obtener 2000 generaciones de individuos que han tratado de eliminar los defectos de microrrechupe.

Tiempos de ejecución	
Tiempo mínimo	00:02:40
Tiempo máximo	00:04:25
Tiempo medio	00:03:21
Tiempo medio por generación	0,1005 seg.
Tiempo medio para 500 generaciones	00:00:50
Tiempo medio para 1000 generaciones	00:01:40
Tiempo medio para 1500 generaciones	00:02:31

últimas mediciones cuentan con una gran importancia porque, dado el tiempo necesario para el ciclo, puede que debamos reducir el número de generaciones y obtener así un sistema subóptimo. En la Tabla 6.3 podemos observar los resultados obtenidos. El tiempo mínimo necesitado por el sistema ha sido de 02:40 minutos, mientras que el máximo ha sido de 04:25 minutos. Haciendo la media de tiempos calculados en la resolución de los 261 defectos obtenemos que los parámetros óptimos que solucionan los problemas de microrrechupes pueden descubrirse en aproximadamente 03:21 minutos. Extendiendo los cálculos anteriores, podemos determinar que se ha empleado un tiempo medio de 0,1 segundos en cada una de las iteraciones, es decir, en la creación de nuevas generaciones de individuos.

De esta forma, y tomando el tiempo medio de la generación, podemos determinar que, aproximadamente, el tiempo necesario para la optimización de los parámetros de producción al utilizar un número diferente de generaciones. Así, para lograr alcanzar el 56,32% de los óptimos (500 generaciones), el sistema ha tenido que ser ejecutado durante 50 segundos, mientras que para alcanzar un número mayor de óptimos, un 83,91%, el sistema a efectuado la evolución de 1000 generaciones a lo largo de los 01:40 minutos que estuvo ejecutándose. Finalmente, la itearación de 1500 generaciones ha permitido obtener una cantidad de óptimos mucho mayor, el 93,87%, necesitando un tiempo de 02:31 minutos.

Como breve conclusión, con los resultados aquí recogidos podemos asegurar que las herramientas basadas en programación evolutiva son de gran utilidad para la resolución de los problemas que buscan los parámetros óptimos de producción. Además, la solución presentada utiliza los sistemas predictivos híbridos para el cálculo de la idoneidad de esos parámetros. Por último, los

resultados alcanzados son muy interesantes, ya que muestran la posibilidad de solucionar los problemas dados en la planta con mínimos desplazamientos y bajo un tiempo de procesamiento reducido.

6.2.2.2 Propiedades mecánicas

Al igual que en la solución de problemas de microrrechupes, vamos a comprobar el funcionamiento de nuestra aproximación para la optimización de los parámetros de producción de la planta al solucionar problemas en las propiedades mecánicas, concretamente, en la carga máxima. El único cambio respecto al experimento realizado para los microrrechupes ha sido el modelo específico para la predicción de propiedades mecánicas. De esta forma, presentaremos cómo este desarrollo mejora a los anteriores al mantener un sistema de generación evolutivo común con una función de idoneidad única pero adaptada gracias al modelo híbrido utilizado.

Asimismo, para este segundo experimento volveremos a medir los mismos elementos que marcan el comportamiento del sistema. De hecho, esas características son las siguientes: (i) el modo en el que ha actuado el sistema MPC completo a la hora de detectar y corregir los problemas relacionados con las propiedades mecánicas, (ii) el desplazamiento producido por el sistema de realimentación de la planta, en otras palabras, las modificaciones que han sido necesarias para adaptar los parámetros de producción y obtener de nuevo un producto sin defectos, (iii) las iteraciones o generaciones que han sido necesarias para obtener el resultado óptimo y (iv) el tiempo de procesamiento a la hora de buscar los nuevos parámetros de la planta.

Tabla 6.5: Resumen de los resultados obtenidos en lo referente a las instancias tratadas y corregidas para los defectos en las propiedades mecánicas.

Correcciones	
Instancias totales	889
Instancias con defecto	244
Instancias corregidas	232

En primer lugar, y teniendo en cuenta el comportamiento del sistema MPC que utiliza la combinación de clasificadores basada en *grading con clasificadores de primer nivel creados con "Tree Augmented Naïve"*, la Tabla 6.5 nos muestra que nuestro sistema ha procesado 889 evidencias diferentes. De la cantidad total recogida, únicamente 244 de ellas cuentan con problemas en la

6. Generación de nuevos parámetros de producción

carga máxima. Entonces, una vez puesto en funcionamiento el Modelo Predictivo de Control se han detectado y corregido 232 de ellas. En este caso el sistema no ha sido capaz de solucionar completamente los problemas de la planta, pero debemos recordar que, como observamos en el capítulo 4, los modelos para la predicción de problemas de carga máxima cuentan con una menor precisión debido a los problemas de ruido en los datos. No obstante, el sistema no ha sido capaz de procesar únicamente el 5 % de las evidencias. A pesar de este resultado, el sistema MPC sigue manteniendo un control sorprendentemente alto sobre lo que está aconteciendo en la planta.

Tabla 6.6: Resumen de los resultados obtenidos en lo referente a las variables del proceso modificadas con el fin de solucionar y evitar los problemas en las propiedades mecánicas.

Modificación de variables	
Número mínimo de modificaciones	1
Número máximo de modificaciones	12
Número medio de modificaciones	2,68
Desviación estándar	2,16
Número de modificaciones marcado por la moda	2
Número de modificaciones marcado por la mediana	2
Número de modificaciones de 1 variable	44
Número de modificaciones de 2 variable	126
Número de modificaciones de 3 variable	32
Número de modificaciones de 4 variable	5
Número de modificaciones de 5 variable	3
Número de modificaciones de más de 5 variable	22

En segundo lugar, hemos vuelto a medir el número de modificaciones realizados para volver a encaminar la producción a su ruta óptima. Estas medidas marcarán gran parte del éxito o fracaso del sistema de realimentación a planta porque, tal y como quedó recogido en la sección 2.2.4, un sistema MPC debe tratar de obtener un desplazamiento mínimo entre lo que está sucediendo en la planta y la solución óptima a la que se desea llegar. De este modo, tras la medición de las 232 correcciones realizadas, los resultados han quedado recogidos en la Tabla 6.6.

Claramente, esta tabla nos muestra que a pesar de no haber sido capaz de solucionar todos los casos con problemas, el sistema de realimentación ha sido capaz de utilizar un desplazamiento menor al solucionar problemas de carga máxima que de microrrechupes. Como valores extremos vemos que las modificaciones van desde de una única variable a 12 de ellas. A pesar de

que el cambio máximo vuelve a acercarse al 50 % de las variables que estamos manejando, al igual que en el caso anterior, la mayoría de los cambios se centran en las 2 variables. Específicamente, fuera del rango de 5 modificaciones únicamente se encuentran el 9,28 % de las correcciones. De nuevo, la moda y la mediana marcan el cambio de 2 variables como el más común, mientras que la media indica un valor cercano a 3 modificaciones con una desviación estándar de 2,16 unidades. Con estos resultados comprobamos que el comportamiento para este defecto es prácticamente similar al anterior, con lo que la función de idoneidad se encarga de marcar como óptimas aquellas soluciones que se encuentran más cerca del punto de partida (evidencia defectuosa) pero que no producen ningún tipo de problema.

Tabla 6.7: Resumen de los resultados obtenidos en lo referente a las generaciones necesarias para obtener la solución óptima que elimina los problemas en las propiedades mecánicas.

Generaciones necesarias para el óptimo	
Número mínimo de generaciones	21
Número máximo de generaciones	1993
Número medio de generaciones	337,40
Desviación estándar	385,67
Número de generaciones marcado por la mediana	168
Óptimos logrados con 500 generaciones	181
Óptimos logrados con 1000 generaciones	219
Óptimos logrados con 1500 generaciones	224

Como ya hicimos para los defectos de microrrechupe, mediremos el número de generaciones necesario para encontrar el óptimo en cada una de las operaciones de búsqueda. Este aspecto es también de gran importancia ya que se encuentra totalmente relacionado con el tiempo necesario para optimizar el proceso. De este modo, comprobando los resultados tras 2000 generaciones, observamos en la Tabla 6.7 que el número mínimo de generaciones necesarias es de 21. Por otra parte, el número máximo fue de 1993.

Al igual que en el caso centrado en solucionar problemas de microrrechupes, mantenemos un amplio rango de valores que bien queda registrado a través del cálculo del valor medio (337,40 generaciones) y la desviación estándar que sufre (385,67 generaciones). De todas formas, y pensando en las posibilidades de trabajar de forma subóptima, hemos realizado la cuenta del número de óptimos alcanzados con 500, 1000 y 1500 generaciones. Para cada uno de estos valores obtenemos el 76,37 % (181 óptimos), 94,41 % (219 óptimos) y

6. Generación de nuevos parámetros de producción

94,51 % (224 óptimos) de las soluciones óptimas respectivamente. Nótese, que a pesar de reducir el número de generaciones y no conseguir corregir el 100 % de los errores, todas las soluciones han sido evaluadas con el sistema para la predicción de estados estacionarios y se ha llegado a determinar que los productos finales cumplen con los requisitos de calidad establecidos.

Tabla 6.8: Resumen de los resultados obtenidos en lo referente a los tiempos de ejecución para obtener 2000 generaciones de individuos que han tratado de eliminar los defectos en las propiedades mecánicas.

Tiempos de ejecución	
Tiempo mínimo	00:04:00
Tiempo máximo	00:04:59
Tiempo medio	00:04:29
Tiempo medio por generación	0,1345 seg.
Tiempo medio para 500 generaciones	00:01:07
Tiempo medio para 1000 generaciones	00:02:15
Tiempo medio para 1500 generaciones	00:03:22

Finalmente, otra de las características del comportamiento del sistema que hemos medido, y que cuenta con una gran importancia, es el tiempo de procesamiento utilizado para encontrar la solución adecuada. En el caso de la búsqueda de los parámetros que resuelven los problemas de carga máxima, el tiempo ha sido ligeramente superior al de los microrrechupes. Así, la Tabla 6.7 nos ilustra con los diferentes registros de tiempo. Concretamente, el rango de tiempos se mueve entre los 04:00 minutos y los 04:59 minutos, dando un tiempo medio de unos 04:29 minutos. El cuello de botella del sistema es la predicción de los defectos, por lo que podemos deducir que el tiempo necesario para combinar los resultados a través de *stacking* es menor que con *grading*.

Volviendo a calcular el tiempo medio necesario para realizar cada generación podemos aproximar la duración del proceso de optimización para los diferentes valores de generaciones marcados anteriormente. El tiempo de una única generación es de 0,14 segundos. Por lo tanto, el tiempo medio para la evolución de 500 generaciones será de 01:07 minutos, para 1000 generaciones de 02:15 minutos y para las 1500 generaciones de 03:22 minutos.

A pesar de ese incremento de tiempo al compararlo con el experimento anterior, podemos determinar que esa variación no es estadísticamente representativa, por lo que al igual que en el experimento para la solución de los microrrechupes, el tiempo necesario para la optimización y realimentación no es problemático para la gestión de un proceso como el de la fundición.

En resumen, observando los resultados obtenidos al optimizar los parámetros tanto para evitar problemas de microrrechupe como en las propiedades mecánicas, podemos afirmar que la programación evolutiva es una gran herramienta para lograr resolver los problemas de la planta, siempre que la función de idoneidad esté bien escogida.

De esta forma, hemos logrado resolver diferentes tipos de problemas con el mismo algoritmo, en un tiempo bastante reducido a través de una función de idoneidad fija pero con comportamiento variable gracias al sistema de evaluación de estados estacionarios $t + 1$. Así, todas las partes del sistema se benefician del desarrollo de modelos híbridos que son automáticamente adaptados, tal y como fue explicado en el capítulo 5.

6.3 Discusión de los resultados

El presente capítulo se ha centrado en la descripción del método adoptado para la optimización de los parámetros de producción cuando se ha detectado un posible problema en los productos finales. Para llevar a cabo el proceso hemos optado por la utilización de la programación evolutiva, concretamente, por un algoritmo genético con evaluaciones de idoneidad paralelas.

Este tipo de soluciones, dada la naturaleza de la propia programación evolutiva, aporta la siguiente serie de ventajas al proceso de optimización [Tom95]: (i) la incorporación de la robustez al proceso de optimización, (ii) permite solucionar problemas sin tener conocimientos específicos del dominio, (iii) trabajar con múltiples soluciones de forma simultánea y (iv) en los casos de maximización o minimización de una función, como en el caso que tratamos, esta tarea se ve menos afectada por los máximos o mínimos locales.

Además, gracias a la unión del sistema de realimentación con el sistema de predicciones híbrido, mantenemos un sistema con características híbridas y totalmente independiente del proceso, mejorando las soluciones anteriores [OBK⁺97, ADN01, SB03, YLW03]. A pesar de estos progresos, la programación evolutiva también cuenta con una serie de apreciaciones que merecen ser comentadas.

En primer lugar, la primera de las limitaciones que presenta este tipo de soluciones es que no puede asegurar que la solución óptima será alcanzada. Esta situación se debe a que la finalización de nuestro algoritmo genético sucede cuando hemos creado un número n de generaciones. De esta forma, y como se puede comprobar en los experimentos descritos en la sección 6.2.2, puede

6. Generación de nuevos parámetros de producción

darse un funcionamiento subóptimo del sistema ,en mayor o menor grado, dependiendo de la cantidad de generaciones que sean seleccionadas. La forma de solucionarlo es mediante el empleo de otras estrategias de búsqueda que sí aseguren alcanzar el valor óptimo, por ejemplo, la búsqueda en profundidad o a lo ancho [Pea84, RN03].

En segundo lugar, al igual que en algoritmos de búsqueda como el escalar colinas, puede darse la situación de que se escogen los individuos que se creen que ayudarán a converger cuanto antes en la solución idónea [Bäc96]. Sin embargo, puede que no sean los adecuados. En parte hemos solucionado este problema al utilizar el *elitismo* y mantener siempre el mejor individuo de cada generación. Pero también puede solucionarse a través del *pangenetismo*, es decir, la selección de un único padre y la generación de los hijos a través de unos cruces que únicamente involucran a este único padre. Ambas soluciones hacen que los algoritmos genéticos aumenten la seguridad de que convergerán en la solución óptima.

En último lugar, uno de los aspectos más importantes es el tiempo. Puede que el tiempo logrado con este sistema no sea suficiente para el proceso productivo sobre el que se está trabajando. Del mismo modo, puede que el proceso no soporte la utilización de sistemas subóptimos, con lo que es necesario trabajar en una nueva solución. No obstante, y dado que el cuello de botella ha sido identificado en los procesos de predicción, reduciendo el número de predicciones necesarias, reduciremos el tiempo necesario para obtener la solución.

De esta forma, proponemos, de nuevo, la utilización de los métodos de búsqueda habituales [Pea84] que tomarán como punto de partida la evidencia en la que se detectaron los errores, de esta forma, no será necesario realizar tantas predicciones como las necesarias para la creación de 2000 generaciones. En otras palabras, evitamos tener que predecir unas 10.000 soluciones diferentes y puede que en sólo 23 predicciones (generación de las 23 diferentes soluciones con un cambio de una variable a través de un método de búsqueda en anchura), aproximadamente 1 segundo, hayamos encontrado la solución idónea.

Este nuevo método no está exento de otros problemas, por ejemplo, el desconocimiento del tiempo necesario para optimizar los problemas de la planta. Además, ese tiempo es dependiente del número de variables del proceso que necesiten ser modificadas, en otras palabras, el número de ramas del árbol de búsqueda que iremos recorriendo (i.e., predicciones que haremos) será lo que marque el tiempo de ejecución de la optimización.

6.4 Sumario

A lo largo de este capítulo se ha presentado la herramienta que completa el sistema MPC que presentamos en esta investigación. Concretamente, se ha realizado la descripción del sistema de realimentación de la planta encargado de optimizar los parámetros de producción y comunicar los valores que deben utilizarse en las variables independientes sobre las que podemos trabajar. Así, hemos contribuido al avance del estado de la técnica al presentar un sistema de optimización de parámetros independiente del proceso y del objetivo de predicción fijado haciendo uso de la programación evolutiva paralela, es decir, simulando la teoría de la evolución de Darwin [Dar37], y utilizando la potencia actual de los sistemas de computación concurrente.

De esta forma, el sistema de realimentación cuenta con un algoritmo genético que realiza la búsqueda de los parámetros óptimos de la planta mediante la utilización de las operaciones genéticas de cruce y de mutación. La primera de ellas permite que, una vez seleccionados dos individuos de forma proporcional a la idoneidad, generemos dos nuevos individuos que comparten características de ambos padres. La segunda de las operaciones permite ampliar el espacio de soluciones a través de la modificación de los genes que componen cada solución. Ambas operaciones cumplen las restricciones que existen en la naturaleza al trabajar bajo unos ratios de cruce del 80 % y de mutación del 5 %.

La gran mejora de este sistema viene dada por el cálculo de la idoneidad. Gracias a él seremos capaces de determinar cuál de los genomas, soluciones para nosotros, es más cercano al resultado deseado y será tenido en cuenta para evolucionarlo hacia la solución óptima. La función desarrollada nos permite trabajar en cualquier dominio y con cualquier objetivo sin necesitar ningún cambio en ella. Esto es conseguido gracias a la integración del sistema de predicción de estados estacionarios en el sistema de optimización de la producción. De esta forma, cada una de las soluciones generadas será pasada por el sistema de predicción para determinar que no va a tener errores. Posteriormente, en caso de pasar este corte, el sistema presentado realiza un análisis de semejanza basado en la similitud del coseno. Esta medición de la similitud nos permite determinar cómo de cerca se encuentra la nueva solución generada de la evidencia con problemas asegurando, así, que la modificación que habrá que realizar en la planta será la mínima posible.

Este sistema nos ha permitido obtener unos resultados muy interesantes para dos tipos de defectos diferentes. De esta forma, demostramos que el sistema es independiente del objetivo. Por otra parte, la función de idoneidad

6. Generación de nuevos parámetros de producción

logra los propósitos para los que fue diseñada, ya que, en ambos defectos, la mayoría de las correcciones únicamente necesitan de la modificación de dos parámetros de producción. Igualmente, dado que el para la optimización utilizaremos el sistema de predicción, éste es capaz de mantenerse actualizado en todo momento a través del proceso automático explicado en el capítulo 5. Todas estas características hacen que sea un sistema idóneo para la optimización del proceso en el MPC autoadaptativo e híbrido que presentamos en esta tesis doctoral.

«Conclusión es el lugar donde
llegaste cansado de pensar.»

Anónimo

7

Conclusiones

UNA vez definidos y descritos los procesos de investigación, desarrollo y evaluación incluidos en esta tesis doctoral, es momento de volver la vista atrás y evaluar las contribuciones realizadas. Para ello, resumiremos brevemente las contribuciones de esta investigación doctoral y realizaremos la comparación pertinente con la hipótesis y los objetivos marcados en el inicio del proceso.

De esta forma, el capítulo se encuentra organizado como sigue. La sección 7.1 recopila las diferentes validaciones que se han desarrollado en esta investigación. La sección 7.2 resume brevemente los resultados alcanzados y los evalúa dentro del dominio de investigación seleccionado: la informática industrial para la optimización de procesos. Posteriormente, la sección 7.3 nos ilustra con las diferentes áreas de aplicación en las que la investigación realizada será de gran utilidad. Las limitaciones de la solución propuesta serán recogidas en la sección 7.4. Por otra parte, en la sección 7.5 recogemos las líneas futuras de investigación que quedan abiertas tras la finalización de esta tesis doctoral. En último lugar, la sección 7.6 concluye esta disertación y da la visión personal del autor sobre las fábricas del futuro.

7.1 Síntesis de la validación del sistema

A lo largo de este trabajo de investigación se han desarrollado múltiples validaciones empíricas con el objetivo de comprobar cada uno de los componentes básicos que nos han permitido crear un Modelo Predictivo de Control (en inglés “*Model Predictive Control*”, MPC) autoadaptativo y que hace uso de modelos híbridos para la predicción de los estados estacionarios.

En primer lugar, hemos validado si un proceso productivo tiene la capacidad de generar una representación de conocimiento útil para la optimización de los procesos. Además, hemos ratificado que ese proceso de generación de conocimiento puede realizarse a través de métodos de aprendizaje automático, reduciendo el esfuerzo de generar esa nube de conocimiento. Con el fin de llevar a cabo esta comprobación, y las posteriores, decidimos seleccionar como caso de uso la fundición de hierro grafito esferoidal que produce piezas de alta precisión para la industria automovilística. Igualmente, hemos tenido en cuenta más de un defecto (i.e., microrrechupes y propiedades mecánicas) para determinar el funcionamiento multiobjetivo de las aproximaciones presentadas.

Así, fuimos capaces de comprobar la gran capacidad de los modelos de aprendizaje automático. Prácticamente todos los modelos probados extrajeron y representaron la información del proceso productivo de forma precisa, consiguiendo ratios superiores al 94% para determinar que sucederá en un estado futuro en relación a la aparición de microrrechupes y cercanos al 87% para determinar el estado de la carga máxima (propiedad mecánica clasificable dentro de la fuerza). En ambos casos alcanzamos los mejores resultados al aplicar un clasificador basado en la agrupación de clasificadores débiles, concretamente, los bosques aleatorios. A pesar de obtener los mejores resultados con el mismo clasificador, las configuraciones realizadas para cada uno de los defectos no fueron las mismas. Este hecho imposibilita determinar cuál es el clasificador óptimo para la predicción de cualquier tipo de defecto.

No obstante, a lo largo de la historia de los sistemas MPC se han desarrollado modelos que estén cercanos a la naturaleza del proceso que tratan de controlar. En otras palabras, los actuales MPC cuentan con modelos lineales o no lineales. De esta forma, con el objetivo de solucionar el problema de la selección de los clasificadores, y a su vez, permitirnos generar un modelo que cuente con los beneficios de ambos tipos de modelos, hemos comprobado la capacidad de representar el proceso productivo a través de una colmena de clasificadores heterogéneos combinados a través de diversos métodos (i.e.,

7.1 Síntesis de la validación del sistema

stacking, grading, por voto y multiesquema). Los resultados obtenidos fueron muy relevantes. No sólo fuimos capaces de conseguir mantener la precisión en la predicción con el mejor de los meta-clasificadores, sino que la mayoría de los modelos combinados mantuvieron una precisión muy próxima entre sí. E incluso, en el caso de los microrrechupes, pudimos mejorar la precisión del mejor modelo de clasificación. Unido a esto, conseguimos que las tasas de error disminuyeran, por lo que logramos demostrar que un sistema de meta-clasificación era la solución más conveniente para los propósitos de esta tesis doctoral.

Cabe destacar que la aportación más importante de este tipo de aproximación es que el método de generación de modelos aquí presentado no necesita ni de personal especializado en el proceso, ni de conocimiento previo sobre cómo se desempeñan las tareas de producción o sus restricciones. Este tipo de aprendizaje extraerá de forma automática las tendencias y patrones reconocibles del proceso de producción. Igualmente, mejoramos los modelos al disponer de un sistema híbrido, y por consiguiente, no utilizar modelos que, por su naturaleza, no son capaces de adaptarse al proceso productivo. En definitiva, contaremos con los valiosos resultados que todos ellos pueden aportar a la hora de realizar una clasificación.

En segundo lugar, apareció una contrariedad derivada de la selección de los métodos de predicción del estado estacionario $t + 1$. Asimismo, esta cuestión se trata de uno de los mayores problemas con los que cuentan los sistemas MPC y que les viene persiguiendo desde mucho tiempo atrás. Concretamente, estamos hablando de la actualización de los modelos para conseguir mantenerlos ajustados al cambiante proceso de producción. Para ello, tanto los datos antiguos como los nuevos cuentan con una gran importancia. Sin embargo, la nueva información que fue añadida recientemente, a pesar de ser lo que está sucediendo actualmente en la planta, no es tan representativa como nos gustaría que fuera. De esta manera, desarrollamos un proceso de actualización de los modelos predictivos evitando tener que mantener en todo momento la totalidad de datos históricos del proceso productivo y que agiliza la adaptación, dotando de una mayor importancia a los datos más nuevos.

En particular, presentamos un método que permite realizar la reducción de los datos históricos para que en cantidad sean menos representativos. Este tipo de técnicas son conocidas como *factores de envejecimiento* o *factores de desvanecimiento*. Nuestro método trata de extraer la distribución probabilística inherente en los datos históricos del proceso a través de una red bayesiana para, posteriormente, pasar a generar un nuevo conjunto de datos reducido

7. Conclusiones

(mediante la utilización de los métodos de Montecarlo) que sea capaz de crear un modelo prácticamente igual de preciso y con la misma gestión de errores que los modelos aprendidos con el conjunto de datos completo. Así, una vez unida la nueva información, estamos modificando la distribución probabilística anterior pero aportando una mayor importancia a los nuevos datos (son más representativos debido a la reducción de la cantidad de datos históricos). Este proceso se repetirá a lo largo del tiempo n veces, generando los conjuntos de datos que utilizaremos para el aprendizaje de los modelos. Nótese que dada la selección inicial de modelos, necesitamos realizar un aprendizaje completo ya que nuestros modelos no son automáticamente actualizables.

Los experimentos realizados para la validación de este método de actualización de modelos demuestran que el sistema de compresión bayesiana basada en métodos de Montecarlo es capaz de reducir ampliamente los datos manteniendo su distribución. Tal es su comportamiento que, los modelos generados al comprimir el conjunto de datos original un 98 % únicamente reducen su precisión en aproximadamente 20 unidades. Sin llegar a tales niveles de compresión y manteniéndonos en torno al 40 ó 50 % de compresión, los modelos aprendidos para ambos defectos mantienen prácticamente su precisión y su gestión de errores. Seguidamente, sabiendo que el funcionamiento de los modelos comprimidos es el esperado, desarrollamos los experimentos para determinar si un aprendizaje acumulativo, que aplica este tipo de técnica como factor de envejecimiento, es válido y mantiene un sistema de predicción de estados estacionarios actualizado. En este último caso, y tras simular el aprendizaje acumulativo de dos años en la planta de producción, alcanzamos unos resultados que indican que este método es altamente factible como sistema de actualización de modelos. De hecho, en los extremos, ante altos niveles de compresión o bajos niveles de compresión es donde mejores resultados consiguen los modelos. Pero lo que los diferencia es que al utilizar los factores de compresión máxima, no solo igualamos los resultados alcanzados con el conjunto de datos original, sino que en el caso de los microrrechupes los conseguimos mejorar. La explicación a este fenómeno es que en la generación de los nuevos datos reducidos se han eliminado las ocurrencias anecdóticas que hacían perder precisión a los modelos, en otras palabras, el ruido.

En ultimo lugar, y como última contribución de esta tesis doctoral, se ha finalizado el sistema MPC al incorporar un medio de adaptación de los parámetros de producción para la realimentación de la planta. En este caso, la optimización únicamente debe darse cuando los sistemas de predicción de estados estacionarios han detectado problemas en la planta. Además, la

7.2 Resumen de los resultados alcanzados

solución dada no debe implicar la modificación de excesivas variables. Para solventarlo, hemos mejorado el estado de la técnica al desarrollar un algoritmo genético independiente del proceso y del objetivo, ya que la función de idoneidad desarrollada hará uso de los sistemas de predicción, quedando el algoritmo adaptado completamente al proceso productivo en el que se está trabajando. Del mismo modo, al realizar esa unión, contamos con las ventajas de la utilización de los modelos híbridos y mantendremos el sistema de optimización actualizado en todo momento al aplicar el método anteriormente comentado. Por otra parte, y dado que debemos generar los parámetros que eviten modificar en demasía el proceso productivo, utilizamos la técnica de la similitud del coseno para determinar cómo de próximos se encuentran los vectores solución del que fue predicho como defectuoso.

En lo que respecta a los resultados de este último sistema, cumplimos con las premisas iniciales que nos marcamos. Específicamente, el sistema ha sido capaz de corregir el 100 % de los casos que presentaron defectos de microrrechupes y aproximadamente el 95 % de los casos con faltas en las propiedades mecánicas. También, a pesar de encontrar ciertos casos que necesitan de una modificación de un gran número de variables (alrededor de la mitad de las variables independientes que controlamos) la tendencia mayoritaria es la de modificar entre 2 y 3 variables del proceso. Y el aspecto más problemático, el tiempo, nos muestra que es necesario alrededor de 3 minutos para optimizar el proceso completo de producción. En caso de no ser suficiente, una modificación del sistema de optimización nos puede llevar a alcanzar conclusiones subóptimas en aproximadamente 1 minuto asegurando que aproximadamente el 50 % de las soluciones dadas sí que alcanzarán el óptimo.

A modo de resumen, las aportaciones de esta tesis doctoral han permitido mejorar los sistemas MPC al eliminar ciertas limitaciones como: (i) la realimentación ineficiente, (ii) las ineficiencias en la actualización del modelo, (iii) la necesidad de personal experto en ambos dominios, en el proceso productivo y en el desarrollo matemático de sistemas MPC, (iv) la utilización de modelos que trabajan con una única función objetivo y (v) la dificultad de generación y gestión de sistemas MPC no lineales.

7.2 Resumen de los resultados alcanzados

En esta sección resumiremos todas las contribuciones aportadas por nuestro trabajo de investigación. Así, partiendo de la definición teórica de un sistema

7. Conclusiones

MPC [Ber95] y su definición matemática [BM99b], hemos trabajado en la eliminación de algunos de los problemas con los que cuentan estos sistemas (descritos en la sección 2.5), consiguiendo, así, mejorar el estado de la técnica con las siguientes contribuciones fundamentales:

1. **Un método para la predicción de estados estacionarios $t + 1$ basado en técnicas de aprendizaje supervisado.** Hemos presentado una forma automatizada de extraer la información de una planta de producción a través de la identificación estadística de patrones para, posteriormente, generar una serie de modelos capaces de prever que es lo que sucederá con la producción en un instante de tiempo futuro. Para poder desarrollarlo hemos empleado diferentes técnicas de aprendizaje automático supervisado, como las redes bayesianas, los árboles de decisión y las máquinas de soporte vectorial, entre otros. Gracias a este método, disponemos de un sistema de predicción que cuenta con una precisión cercana al 95 % para los defectos de microrrechupe y de un 86 % para los de carga máxima.
2. **El primer método para la generación de sistemas MPC híbridos basado en la combinación de clasificadores heterogéneos.** Evolucionando las ideas iniciales presentadas en la primera contribución, hemos construido el primer sistema predictivo de control que cuenta con un modelo híbrido. En otras palabras, combina diferentes clasificadores, tanto lineales como no lineales, obteniendo beneficios tales como evitar la selección de un único clasificador y asegurar que los resultados son una gran aproximación de la solución que podría alcanzar el mejor de los clasificadores. Además, el método presentado sigue siendo útil para la generación de sistemas MPC solamente lineales o no lineales. En ese caso, únicamente deberemos combinar clasificadores de tipos lineal o no lineales. Los resultados logrados son similares a los anteriores, una precisión cercana al 95 % para los microrrechupes y de un 86 % para la carga máxima, pero, en este caso, sufrimos una reducción de las tasas de error.
3. **Un método estadístico para la reducción/compresión de conjuntos de datos manteniendo la misma distribución probabilística original.** Hemos desarrollado un nuevo método que permite realizar una reducción de un conjunto de datos con la posibilidad de reducir costes computacionales, esos costes derivados por el espacio necesario para el almacenamiento de información o para ser utilizado como método

de envejecimiento de muestras en la actualización del modelo. Además, el conjunto de datos resultante es totalmente representativo porque es capaz de obtener modelos semejantes a los que logramos cuando empleamos el conjunto de datos original. Este método combina técnicas estadísticas como son las redes bayesianas y los métodos de muestreo de Montecarlo. Los resultados alcanzados son muy interesantes, ya que aplicando los mayores grados de compresión, es decir, un 98 %, la precisión de los modelos generados con ese conjunto de datos, para ambos defectos, únicamente es reducida en 20 unidades. No obstante, los porcentajes de compresión de la zona media, entre el 40 % y el 60 % de compresión, muestran un comportamiento similar al alcanzado por el conjunto de datos original.

4. **Un procedimiento para la adecuación de los modelos predictivos a través de la aplicación de factores de envejecimiento y un aprendizaje acumulativo.** Como complemento de la anterior aportación, proponemos un procedimiento para permitir la actualización de los sistemas de predicción de estados estacionarios que se basa en un proceso de aprendizaje iterativo, en el cual aplicamos la compresión bayesiana basada en los métodos de Montecarlo para desvanecer las muestras. Además, se presenta el funcionamiento del sistema concurrente que realiza estas tareas y permite modificar los modelos del sistema MPC. Los resultados alcanzados son sorprendentes, ya que el método de actualización nos permite mantener los mismos resultados (algo ligeramente superiores para la predicción de microrrechupes) cuando aplicamos altos grados de compresión (entre el 80 y el 98 %) en el envejecimiento de los datos.
5. **Una mejora de los sistemas de optimización de parámetros basados en la programación evolutiva.** Hemos evolucionado las soluciones desarrolladas por otros autores al preparar un algoritmo válido para cualquier proceso o cualquier objetivo. Nuestro algoritmo se adapta fácilmente al proceso y se actualiza de forma automática ya que se basa en partes del sistema MPC que sí que cambian a lo largo del ciclo de vida del sistema. Además, dados los problemas temporales con los que cuenta la programación evolutiva, hemos hecho uso de la teoría de la programación evolutiva paralela con el propósito de conseguir un sistema que en un tiempo insignificante es capaz de optimizar los parámetros de la producción produciendo los mínimos cambios posibles (el algoritmo cambia únicamente dos variables en la mayoría de las optimizaciones).

7. Conclusiones

Gracias a las contribuciones que acabamos de presentar hemos sido capaces de cumplir con todos los objetivos específicos que definimos en el capítulo 1:

- Desarrollar y evaluar un modelo general de predicción de defectos y características utilizando técnicas de aprendizaje automático.
- Desarrollar y evaluar un modelo basado en técnicas de toma de decisiones y análisis de sensibilidades que permita ofrecer la información necesaria para que el proceso de producción sea modificado para mantener la producción dentro de las restricciones del proceso.
- Desarrollar y evaluar una metodología que permita mantener el sistema totalmente adaptado en todo momento ante los posibles cambios que puedan darse en el proceso de producción.

Del mismo modo, estas contribuciones han posibilitado cumplir los objetivos operacionales iniciales que nos planteamos:

- Diseñar e implementar un marco de trabajo para la detección de defectos y características en las etapas iniciales del proceso productivo.
- Diseñar e implementar un sistema basado en un meta-clasificador que permita predecir la etapa de la planta en un periodo de tiempo $t + 1$ basado en la decisión tomada a través de múltiples predictores, maximizando la precisión y minimizando el error medido en tasas de falsos positivos y falsos negativos.
- Diseñar e implementar un método de representación del conocimiento que sirva como base al sistema cuyo objetivo sea indicar los parámetros que deben ser introducidos en el proceso de producción para evitar las anomalías.
- Diseñar e implementar un sistema que permita dar indicaciones de cómo configurar el proceso que se está llevando en la planta de producción.
- Diseñar e implementar un método que permita mantener el sistema actualizado frente a los cambios que se producen en la producción.

Dado que hemos sido capaces de lograr todos y cada uno de los objetivos planteados al comienzo del proceso de investigación de la tesis doctoral, consideramos como superado el objetivo principal marcado tras la hipótesis de

investigación, y que no es otro que: “*Con el fin de cubrir las necesidades que ya han identificado las empresas en cuanto a calidad, reducción de costes o eco-fabricación, así como la capacidad de obtener una mejora competitiva, se pretende diseñar, desarrollar, evaluar e implantar un sistema que sea capaz de readaptar la producción para cumplir con las restricciones definidas en el proceso.*”. Finalmente, al conseguir desarrollar el Modelo Predictivo de Control autoadaptativo e híbrido presentado y gracias a los logros comentados, creemos demostrada la hipótesis inicial de partida:

Es posible modelar el flujo de negocio de un proceso productivo como una nube de conocimiento a través de la creación de un modelo predictivo de control híbrido (lineal y no lineal) basado en técnicas de aprendizaje automático actuales, manteniéndolo actualizado y llevando el control de la producción para la optimización de alguno de sus parámetros críticos.

7.3 Aplicaciones de la investigación

Como puede entenderse, las contribuciones realizadas en esta tesis doctoral están claramente enfocadas a un dominio industrial que busca la fabricación inteligente como la herramienta clave en el proceso de fabricación.

En este contexto, la conjunción de todas las contribuciones permite que procesos tan antiguos como la fundición puedan llevar un control mucho más estricto de lo que se lleva hoy en día. Además, otros procesos productivos que ya contaban con este tipo de sistemas los verán mejorados, dado que la eliminación de importantes restricciones ha sido el objetivo primordial de la investigación aquí presentada. De esta forma, el método de actualización, una de las contribuciones más importantes de esta tesis doctoral, permitirá que los sistemas MPC estén ajustados al proceso en todo momento, evitando los ya comentados problemas de adaptación.

Igualmente, la modificación de los modelos utilizados para este tipo de sistemas aporta dos ventajas importantes a los procesos productivos. La primera de ellas es que no es necesaria la perturbación del proceso productivo actual para la generación del modelo, es decir, no hay que detener la producción con el fin de generar un conjunto de datos inicial, ni hay que contar con los

7. Conclusiones

expertos en el dominio para que definan las etapas, estados de la planta y situaciones críticas, entre otros. Habitualmente el proceso productivo está diseñado para que dado un producto, las operaciones llevadas a cabo sean lo más parecidas entre sí. Gracias a esta premisa de partida, los patrones de producción pueden ser aprendidos de forma automática de los datos que ya tiene la planta de producción. La segunda de las ventajas, y derivada de la anterior, es que se simplifica el proceso de generación, con lo que podremos tener una planta optimizada en un tiempo mucho menor y con un coste reducido.

A pesar de que la investigación aquí detallada ha estado dirigida únicamente a la detección de defectos en los productos manufacturados, los sistemas MPC pueden tomar cualquier otro objetivo. Hoy en día, áreas de investigación relacionadas con la energía o con el medio ambiente pueden beneficiarse de este desarrollo. Nuestro sistema MPC es capaz de trabajar con objetivos como la cantidad de energía utilizada o desaprovechada, la cantidad de desperdicios generados, o el impacto medioambiental del proceso. Por consiguiente, y dado que la solución presentada es multiobjetivo, podremos mantener una producción ecológica, de bajo coste y con 0 defectos.

No obstante, los procesos productivos no son los únicos que pueden servirse de nuestras contribuciones. Por ejemplo, la generación de espacios inteligentes (traducción literal de la voz inglesa “*smart spaces*”), sobre todo dedicados al cuidado de niños o personas mayores, podrían tomar la solución MPC con el objetivo de realizar un aprendizaje automático de patrones de comportamiento de individuos (en lugar de realizar al aprendizaje del comportamiento de un proceso productivo), adelantarse a futuros casos de peligro a través de los sistemas de predicción, determinar anomalías en los comportamientos (desviaciones en las rutas óptimas de los comportamientos) y tomar las decisiones oportunas para solucionarlas. A este nuevo dominio podemos agregar otros como el control inteligente del tráfico. Gracias a los sistemas MPC se podrían prevenir futuros embotellamientos y tomar las decisiones adecuadas que desviasen el tráfico o que modificasen las frecuencias de apertura y cierre de los semáforos.

En último lugar, y de forma específica, cada una de las aportaciones pueden ser empleadas en cualquier tipo de soluciones en las que el aprendizaje automático sea su base fundamental. De esta forma, la demostración de que un sistema de meta-clasificación puede igualar o incluso superar al mejor de los clasificadores seleccionados al hablar en términos de precisión, e incluso, reducir las tasas de errores con las que trabajan los clasificadores, puede alentar a mejorar los sistemas de clasificación basados en un único clasificador. Asi-

mismo, el modelo de compresión bayesiana basado en métodos de Montecarlo está enfocado a la optimización de clasificadores creados a través de procesos de aprendizaje automático. Por consiguiente, el sistema de adaptación presentado podrá utilizarse en todos los desarrollos que requieran la actualización de sus modelos. Finalmente, esta técnica también puede ayudar a la generación de conjuntos de datos más pequeños y que mantendrán las distribuciones probabilísticas del conjunto de datos inicial.

7.4 Limitaciones del modelo

A pesar del avance que supone para el estado de la técnica, el sistema MPC híbrido y autoadaptativo presentado cuenta con una serie de limitaciones.

1. la posibilidad de aparición de ruido a lo largo del proceso,
2. no existe una herramienta que nos permita validar los datos de entrada,
3. sacrificamos variables del proceso para poder llevar a cabo sistemas más manejables o
4. las medidas de la incertidumbre no han sido introducidas en los modelos predictivos.

A pesar de éstas, ya conocidas, los métodos utilizados en esta tesis doctoral cuentan con otro tipo de limitaciones que deberán unirse a las anteriores y que tendremos que tener en cuenta a la hora de evolucionar este tipo de sistemas. Así, en primer lugar, podemos identificar una limitación relacionada con los modelos predictivos. Actualmente, contamos con una única dimensión temporal. En otras palabras, se realiza una predicción en un estado final de la producción. De esta forma, el sistema es mucho más complicado y debe contar con muchas más variables a controlar. Por ello, la división de un sistema MPC en mini sistemas MPC apilados y con comunicación entre ellos (Modelo Predictivo de Control basado en agentes) podría ser la forma de conseguir superar esta limitación.

En segundo lugar, el tipo de aprendizaje seleccionado cuenta con sus propias limitaciones que quedarán incorporadas a las anteriormente mencionadas. Por una parte, el aprendizaje supervisado supone un gran esfuerzo ya que necesitamos disponer de todas las evidencias etiquetadas, es decir, debemos indicar el resultado que han obtenido. No obstante, esto queda solventado, en

7. Conclusiones

parte, porque en muchos de los procesos productivos ya cuentan con un registro que almacena toda la información del proceso junto con su resultado final. Igualmente, la aplicación de otro tipo de aprendizaje, como el aprendizaje semi-supervisado, puede reducirla.

En tercer lugar, el sistema comienza a trabajar directamente con los datos que le son proporcionados. Pero a lo largo de las investigaciones realizadas hemos encontrado problemas relacionados con ruido incorporado en el proceso de adquisición de datos. De esta forma, llegamos a la conclusión que el sistema cuenta con la limitación de no disponer una etapa de adaptación de datos y reducción de ruido que, probablemente, sea capaz de mejorar el proceso de predicción de los estados estacionarios.

En cuarto lugar, el proceso de adaptación desarrollado obtiene buenos resultados pero cuenta con dos limitaciones heredadas de los métodos utilizados. Por una parte, la selección de la red bayesiana puede no ser la más adecuada. A pesar de que en la literatura se comenta que este tipo de clasificadores son los mejores para representar la distribución [HS86], los experimentos nos han mostrado que otros modelos son capaces de adaptarse mejor al proceso productivo, por lo que sus distribuciones estadísticas estarán más ajustadas a lo que sucede realmente. En segundo lugar, el método de muestreo basado en los métodos de Montecarlo es no determinista, por eso no podemos saber con seguridad el resultado exacto que obtendrá el método de compresión, únicamente tenemos conocimiento de la tendencia que seguirá a lo largo de los diferentes niveles de compresiones. Por ello, la eliminación de alguna, o ambas limitaciones, hará que el proceso de actualización sea mucho mejor y fiable.

Finalmente, no podemos olvidarnos del tiempo de ejecución. El sistema desarrollado cuenta con una duración del proceso de optimización bastante buena. No obstante, pueden existir procesos de producción muy críticos y que necesiten un ciclo de ejecución de análisis muy corto. En ese caso, el sistema de realimentación debe ser mejorado en lo que a tiempo se refiere. A pesar de este hecho, debemos destacar que el sistema actual no tendría ningún problema de funcionamiento, ya que el tiempo de procesamiento puede ser reducido a cambio de trabajar con un sistema subóptimo.

7.5 Trabajo futuro

Hemos identificado cuatro líneas de investigación para el trabajo futuro en torno a esta investigación doctoral que pasarán a ser descritas a continuación.

7.5.1 Mejora del proceso de captura de datos

Actualmente, el proceso de adquisición de datos se realiza de forma manual, lo que provoca la aparición de irregularidades en los datos que representan el proceso. En este caso, la investigación puede dirigirse de dos formas.

En primer lugar, y probablemente la más óptima de las dos, se trataría de trabajar en el desarrollo de un sistema SCADA (de la voz inglesa “*Supervisory Control And Data Acquisition systems*”) [DS99] genérico para permitir la captura de datos automatizada. La mayor problemática de este tipo de desarrollos es la heterogeneidad de los procesos productivos. De esta forma, es muy complicado un sistema que permita conectar cualquier tipo de maquinaria o sensorica al sistema. Por ello, el desarrollo, en lugar de estar orientado en la conexión de las máquinas directamente al sistema SCADA, podría estar enfocado en realizar una captura simple de información visual, es decir, a través de la utilización de técnicas de visión artificial para la identificación de los parámetros del equipo de producción. En la actualidad, este tipo de técnicas ya ayudan a identificar signos vitales en biomedicina [Mil94], son utilizadas en la agricultura [CCK02] o para la inspección de productos [Nov92, TVTV99]. No obstante, los datos recogidos seguirán contando con el ruido producido por los sensores, lo que obligará a utilizar algún tipo de técnica que lo reduzca, por ejemplo, los filtros de Kalman [WB95].

En segundo lugar, la opción es el desarrollo de un proceso inicial que permita trabajar con los datos. De esta forma, podríamos determinar cuáles son correctos y cuáles no lo son, corrigiendo o desechando estos últimos. En otras palabras, estamos introduciendo la etapa de preprocesamiento de los datos con el objetivo de mejorar los modelos a desarrollar. Así, unas de las técnicas que permitirían reducir el ruido es la descomposición en valores singulares (del inglés “*Singular Value Decomposition*”) o el análisis de la semántica latente (traducción de la voz inglesa “*Latent Semantic Analysis*”). Este tipo de técnicas ha contado con una gran aceptación en los procesos de recuperación de documentos [SFS08] con grandes resultados en la reducción de ruido y ambigüedades. A pesar de que pueda parecer extraño, la descomposición en las 3 matrices de trabajo se realizaría de forma análoga a los textos: una primera matriz que contiene las piezas (los documentos), una segunda que contiene las variables (los términos) y la tercera que relaciona las dos anteriores.

En ambos casos se espera que los sistemas predictivos sean mucho más precisos consiguiendo manejar el proceso productivo de una forma mucho más óptima.

7.5.2 Optimización temporal del proceso de realimentación de la planta

La solución propuesta para el proceso de optimización de parámetros y realimentación de la planta ha obtenido unos resultados realmente interesantes, ya que en un corto periodo de tiempo ha sido capaz de realizar una optimización casi perfecta en al menos el 95 % de los casos. En otras palabras, ha sido capaz de obtener los consejos de modificación de la producción con un número reducido de cambios en el proceso (en la mayoría de los casos se realizan cambios en dos variables de las veinticuatro que fueron incorporadas).

A pesar de ello, identificamos el tiempo como una de sus mayores limitaciones. Un sistema MPC marca una restricción temporal para el procesamiento (análisis, predicción y optimización) que se basa en el intervalo de actuación del propio sistema. De esta forma, los tiempos alcanzados en nuestra experimentación pueden no ser suficiente para otro tipo de procesos productivos que sean muy diferentes al de la fundición. Por consiguiente, y con el objetivo de obtener un sistema lo más genérico posible, este es uno de los puntos sobre los que primero pensamos trabajar.

La investigación para la mejora del tiempo de procesamiento puede estar dirigida hacia la utilización de otro tipo de métodos de la optimización. Como ya fue comentado en el capítulo 6, el cuello de botella es el gran número de predicciones que deben llevarse a cabo, por lo que, un método que reduzca ese número obtendrá la solución óptima de una forma mucho más eficiente. Así, creemos que la utilización de los algoritmos de búsqueda habituales [Pea84] podrían acelerarlo.

El cambio fundamental entre el algoritmo genético y los algoritmos de búsqueda es que vamos a partir desde un punto en concreto. Ese punto es la propia evidencia en la que fue detectado el error. Así, recortamos el tiempo de procesamiento para evolucionar las evidencias, empezando directamente con las comprobaciones de aquellas posibles soluciones identificadas en el espacio circundante a la evidencia errónea. Esta aproximación puede lograr la solución óptima en la primera de las pruebas que se realice, con lo que alcanzaríamos una interesante reducción que nos haría huir de la posibilidad de trabajar con un sistema MPC subóptimo.

Completando lo anterior, la nueva línea de investigación propuesta eliminaría otra de las limitaciones impuesta por la programación evolutiva. Concretamente, utilizando ciertos tipos de búsquedas podríamos asegurar que siempre va a ser encontrada la solución óptima global [RN03].

7.5.3 Evolución a sistemas distribuidos, Modelos Predictivos de Control basados en agentes

El sistema aquí presentado se trata de una solución completa para los procesos productivos. La aproximación inicial de los MPC hace que recuperen los datos referentes a todas las etapas del proceso para realizar las predicciones y la optimización de la producción. No obstante, esta forma de solucionar el problema, la seleccionada por nosotros, en caso de que el proceso sea muy grande, nos puede llevar a tener que controlar unos modelos predictivos de incommensurable tamaño (i.e., con miles de variables que necesitan de una optimización muy rápida). Así, como evolución al sistema MPC presentado, surge una línea de investigación para lograr un Modelo Predictivo de Control Distribuido haciendo uso de agentes híbridos y autoadaptativos.

La nueva aproximación nos permitirá distribuir pequeños sistemas MPC con las características del sistema presentado, pero significativamente más pequeños y simples dado que la cantidad de proceso que controlan es mucho menor. Estos sistemas MPC contarán con un añadido, específicamente, con una capa que les permite comunicarse entre ellos y enviar alertas para la optimización de parámetros de agentes que se encuentren gestionando cualquier otra parte de la cadena de producción. La comunidad científica ha estado trabajando sobre este tipo de aproximaciones [CJKT02, NDSH08, Sca09, DOC10], sin embargo, siguen manteniendo las mismas limitaciones que ya conocíamos y que fueron explicadas en la sección 2.5. Por lo que, la unión de la aproximación que hemos desarrollado en esta tesis doctoral y las actuales hacia el desarrollo distribuido de software volverá a reducir las limitaciones con las que cuentan los sistemas actuales.

Del mismo modo, dada la simplificación de cada uno de los modelos, esta aproximación permitirá la gestión de sistemas muchos más grandes, o incluso, sistemas que se encuentren divididos entre diferentes localizaciones.

7.5.4 Perfeccionar las infraestructuras de trabajo

Las estructuras de ejecución de los sistemas MPC son computadores (más simples o más complejos) que se encuentran emplazados en cada una de las plantas de producción. Esto conlleva que las empresas tengan que realizar un gran desembolso para la adquisición de las infraestructuras necesarias para equipar todas y cada una de sus plantas. Igualmente, estos sistemas deben estar funcionando en todo momento con una buena refrigeración, a la par que

7. Conclusiones

estar bien cuidados por expertos. Pero la realidad no es así, el duro entorno en el que se desarrolla el proceso de manufacturación, sobre todo en el campo de la fundición, no permite contar con las condiciones más adecuadas para los sistemas informáticos.

A la luz de estos hechos, creemos que la descentralización de la computación puede ser una solución. Así, las plantas serán meros clientes de un servicio ofrecido. Con lo que, en primer lugar, ya no será necesario contar con un sistema en cada una de ellas. Gracias al concepto conocido como *la nube* (trasladar el procesamiento a servidores dedicados conectados a través de Inet), podemos disponer de sistemas bien cuidados y dedicados exclusivamente a la predicción y control de plantas. Del mismo modo, serán capaces de dar servicio a múltiples clientes. Siguiendo esta estrategia, los costes de la incorporación de estas soluciones de control y optimización de la producción serán muy inferiores. Sin embargo, para lograr este objetivo hay que diseñar cómo utilizar los sistemas de predicción y clasificación en entornos de “*cloud computing*“. Actualmente existen aproximaciones como Apache Mahout¹ que pueden hacer factible esta solución. No obstante, no todos los modelos utilizados están desarrollados, y mucho menos, la combinación de clasificadores o la adaptación, con lo que sería necesario su desarrollo para proporcionarles la capacidad de ser ejecutados en la nube.

Por otro lado, y dado que hay cientos de operaciones repetitivas y que pueden ser ejecutadas de forma simultánea, surge una segunda línea de investigación basada en la computación GPGPU (del inglés “*General-Purpose computation on Graphical Processing Units*”). Las últimas generaciones de tarjetas gráficas permiten la ejecución de código fuente específico pero que consigue realizar unas grandes mejoras en aplicaciones con un alto grado de paralelismo [OLG⁺07]. Por ello, la concurrencia ya conseguida en la aproximación que presentamos puede ser mejorada. Además, esta nueva línea de investigación es fácilmente combinable con la primera de ellas mediante la creación de granjas de computadoras provistos con las últimas generaciones de tarjetas gráficas. De esta forma, conseguiríamos unas máquinas con la potencia de un supercomputador pero con un precio reducido. Sin embargo, y como acabamos de comentar, el desarrollo del software para funcionar en este tipo de unidades gráficas es específico, y actualmente, no disponemos de las mismas herramientas que en la programación habitual, por lo que, de nuevo, se necesitaría la migración de nuestro sistema a estas infraestructuras.

¹ <http://mahout.apache.org/>

7.6 Consideraciones finales

Los procesos productivos estarán con nosotros por muchos años. Ellos son la base de la actual sociedad, y sin ellos no habríamos llegado a evolucionar como lo hemos hecho. Éstos están tan imbricados en nuestra vida que si nos faltaran no sabríamos cómo desempeñar nuestro día a día.

Pero estos procesos de manufacturación adolecen de los mismos problemas que nuestra sociedad. Su sufrimiento es el mundo globalizado en el que vivimos y la gran competencia que está surgiendo ante los mercados diversificados con los que contamos. Así, se está obligando a que los productores sean capaces de responder lo más rápido y fiablemente posible. Durante los próximos años, y sobre todo, teniendo en cuenta la situación de crisis en la que nos encontramos, las empresas de producción que quieran sobrevivir deben ser capaces de generar sus productos con la máxima calidad al menor coste.

Así, como solución de estos problemas, nos encontramos dando los primeros pasos para la optimización de la producción. Pero no solo en lo relacionado con la idea del “0 defectos”, la eco-conciencia con la que cuenta ahora la sociedad nos hace intentar mejorar otros aspectos de la producción hasta conseguir un proceso totalmente eficaz y eficiente. También, los investigadores van generando sistemas cada vez más inteligentes, los cuales seguirán su evolución hasta llegar, incluso, a superar al ser humano. Esto, unido junto a otras técnicas de control y automatización de procesos nos hará llegar hasta un punto en el que. . .

«La fábrica del futuro sólo tendrá dos empleados, un hombre y un perro. El hombre estará ahí para alimentar al perro. El perro estará ahí para evitar que el hombre entre en contacto con el equipo.»

Warren G. Bennis
(1925-presente)

‘Conclusion is the place you reach tired of thinking.’

Anonymous

8

Conclusions

ONCE we have defined and described the research, development and evaluation processes of this dissertation, it is time to look back and evaluate the contributions and the achievements. In this way, we present a brief summary of our main contributions and we compare our accomplishments with the hypothesis and objectives defined at the beginning of the research.

The chapter is organised as follows. Section 8.1 summarises the results obtained through the different validation processes of this dissertation. Section 8.2 briefly summarises the achieved successes and evaluates them for the selected research field: industrial informatics for the optimisation of the manufacturing processes. Later, section 8.3 shows the potential applications of our research. The limitations of the proposed approach are showed in section 8.4. Moreover, section 8.5 outlines the avenues of the future work that arises from the performed research. Finally, section 8.6 concludes this dissertation and summarises our personal point of view about the future factories.

8.1 Synthesis of the system’s validation

In this research work, we have validated our approach throughout several evaluations in order to check each of the basic components that allow us to build a Model Predictive Control (MPC) system, which is self-adaptive and

8. Conclusions

employs a hybrid model to predict steady states in a future temporal frontier $t + 1$.

Firstly, we validated whether a production process is capable of generating a useful knowledge representation for process optimisation or not. In addition, we verified that the process of knowledge extraction can be performed using machine-learning methods, reducing the effort to model the manufacturing process as an expert knowledge cloud. To perform this evaluation, and the subsequent ones, we selected the foundry process specialised on graphite spheroidal iron as our use case. This foundry produces precision castings for the automotive industry. We also took into account more than one defect (i.e., microshrinkages and mechanical properties) to determine the performance of our multi-objective approach.

The majority of the tested machine-learning models were able to extract the information from the manufacturing process in a very precise manner. In fact, we achieved accuracy rates above 95 % in determining the appearance of microshrinkages (i.e., tiny porosities inside the casting), and close to 86 % foreseeing the value of the ultimate tensile strength (i.e., a mechanical property that is the needed strength to break a casting when we pull its ends). In both cases, we achieved the best results when we applied a model based on grouping weak classifiers, specifically, random forests. Although the same classifier accomplished the best results for both defects, the configuration for each one was not the same. This renders impossible to determine which is the optimal classifier for the prediction of every possible defect, suggesting each type of objective needs different approaches to tackle it.

Moreover, the MPC models are developed trying to be as close as they can to the manufacturing process, in other words, the MPC models, along the years, are usually built with linear or non-linear models depending on the process they are going to control. In this manner, in order to solve the problem of the selection of classifiers and the generation a model that includes the benefits of every type of machine-learning algorithms, we tested the ability of representing the production process using a set of heterogeneous classifiers combined through several methods (i.e., *stacking*, *grading*, *by vote* and *MultiSchema*). The results were sound, indeed, the best meta-classifier maintained the same accuracy that we achieved in the previous work. Furthermore, the performance of the majority of the combination methods was similar. Regarding microshrinkages, the meta-classification approach even improved the execution of the best single classifier. In addition, the error rates were decreased, proving that a meta-classification system is the most suitable

8.1 Synthesis of the system's validation

one for the purposes of this research work.

Note that the main contribution of this new approach for generating the models of the MPC systems is that it is not required neither human expert knowledge staff, nor the prior notions on how to perform de manufacturing tasks, nor the constraints of the process. Actually, this type of learning automatically extracts trends and recognizable patterns without any human interaction. Similarly, we improved the previous models creating a hybrid strategy for the classification steps.

Secondly, a drawback appeared due to the selected classifiers for the prediction of the steady states. This problem is one of the biggest issues MPC systems have faced along the years. More accurately, it is the updating method to keep models adjusted to the process despite its evolution. To update the model, both old and new data have great importance, however, the new information added recently, in spite of representing the current state of what is currently happening in the plant, is not as representative as the old one when we generate the classifiers. Thus, we designed a new method for renewing the knowledge of the predictive models avoiding the negative impact of the historical data in the new and up to date information. In this way, our new approach speeds up the adaptation and gives more importance to the newest data over the old one.

Particularly, we presented a method able to reduce historical data in order to make it less representative. This kind of technique is known as *fading factors*. Our method tries to extract the probability distribution inherent in the historical data of the process through a Bayesian network to, subsequently, generate a newly reduced dataset using Monte Carlo methods, which is able to build a model almost as accurate and with the same error rates as the model learned with the entire dataset. Later, we append the newest data to this reduced dataset, modifying the aforementioned probability distribution and giving more importance to the newest one. The process will be repeated generating the datasets that will be employed to learn the new predictive models. Note that given the initial selection of models, we need to make a complete learning because our models are not upgradeable.

The experiments generated to validate the updating of the models using the Monte Carlo-based Bayesian compression method show that reduce the volume of data while maintaining its distribution. Actually, we are able to create models, which using a 98% of compression rate, are only 20 units less accurate than the models learned with the original dataset. Notwithstanding, without reaching such levels of compression, for instance about a 40% or 50%

8. Conclusions

of compression, the learnt models for both defects obtain practically the same accuracy and have similar error rates when compared to the uncompressed predictor. Once we knew the performance of the reduced models was what we expected, we conducted new experiments to determine if a cumulative learning, which employs the proposed method for fading the old evidences, is valid for maintaining the predictive system up-to-date. Therefore, after the simulation of a two-year cumulative learning for the foundry process, the achieved results illustrate that our new method is highly useful to keep the models adjusted to the manufacturing process. In fact, when we compressed the historical data using the highest compression percentages (i.e., 98 % of compression), the models reached their best performance. In addition, the models were able to keep up with the behaviour of the original models, indeed, in the microshrinkages case, the cumulative models overcame the previous models. The explanation for this phenomenon is that when reducing the dataset with the compression methods, we removed the anecdotic occurrences, solving the noise problems.

Finally, as a final contribution of this dissertation, we completed the MPC system by adding a method for optimising the parameters of the manufacturing process. In this case, the optimisation should only occur when forecasting systems encounter a problem in a near future. Furthermore, as other constraint, the proposed solution should not involve excessive modifications in the process. To achieve this goal, we improved the state-of-the-art developing a genetic algorithm fully independent of the objective or the process. The algorithm is completely adapted to the production process due to the fitness function already employed in the aforementioned predictive models. Likewise, and thanks to this union, our optimisation system will be updated when the prediction models are. Besides, we used the cosine similarity distance to determine how far the proposed solution is to the behaviour predicted as faulty casting in order to assure that we are going to select the solution that changes the less number of independent variables in the manufacturing process.

Regarding the results of the experiments, on the one hand, we fulfilled the initial constraints. On the other hand, the system corrected the 100 % of the castings flawed with microshrinkages and about a 95 % of evidences with a low ultimate tensile strength. Also, in spite of the fact that we found some cases requiring the modification of a huge amount of variables (i.e., about half of the independent variables that are monitored), the main trend is a change of only 2 or 3 variables. Nevertheless, the most problematic issue is the time. Our system needs, more or less, 3 minutes to carry out the optimisation process. This time is not problematic for our use case, however, there may be

other manufacturing process requiring a lower time lapse for the optimisation process. In this case, we could work with a suboptimal MPC system that employs 1 minute to obtain the final solution and is able to answer with an optimal solution to approximately the 50 % of the performed analyses.

Summarising, the contributions of this thesis have improved MPC systems by overcoming common limitations such as: (i) inefficient feedback, (ii) inefficiencies updating the model, (iii) the need of skilled people in both domains, the production process and the mathematical development of MPC systems, (iv) the use of models that work with a single objective function and (v) the difficulty of generating and managing non-linear MPC systems.

8.2 Summary of the achieved results

In this section we sum up all the contributions of our research. Based on the theoretical definition of an MPC system [Ber95] and its mathematical definition [BM99b], we have worked to tackle some of the problems these systems have (for more details see the section 2.5), thus, we have improved the state of the art with the following key contributions:

1. **A method for predicting the steady state $t+1$ based on supervised machine-learning techniques.** We have presented an automated way to extract information from a production process, employing the statistical identification of patterns, to generate a series of models that are able to predict what will happen in the near future. We have employed several supervised machine-learning algorithms such as Bayesian networks, decision trees and support vector machines, among others. We have developed a prediction system with an accuracy of 95 % for the prediction of microshrinkages and an accuracy of 86 % to foresee the ultimate tensile strength.
2. **The first method for creating hybrid MPC systems based on the combination of heterogeneous classifiers.** Evolving the aforementioned ideas, we have built the first predictive control system composed by a hybrid model. In other words, it combines different classifiers, both linear and non-linear, obtaining benefits such as avoiding the selection of a single classifier or ensuring that results are a close approximation of the solution that could be achieved with the best single classifier. Furthermore, this method is still useful for creating linear or

8. Conclusions

non-linear MPC systems (in this case, we only have to combine classifiers of only one type). The achieved results are similar to the previous ones, an accuracy close to a 95 % for microshrinkages and an 86 % for the ultimate tensile strength. However, when using this approach where consistently reduced error rates.

3. **A statistical method for the reduction / compression of datasets while maintaining the original probability distribution.** We provide a new method that compresses a dataset with the objective of reducing computational costs, the space required for storage and fading the old instances or samples to keep the model up to date. In addition, the resulting dataset is fully representative because our method is able to obtain models similar to those contain the original data. This method combines statistical techniques such as Bayesian networks and Monte Carlo sampling methods. The accuracy of the generated models when applying the highest compression factors (for instance, 98 %), for both defects, is only reduced in 20 units. Notwithstanding, percentages of compression close to the 50 % show a similar behaviour when compared to the one achieved by the original dataset.
4. **A process to adjust predictive models by applying fading factors and cumulative learning.** As a complement of our previous contribution, we propose a method to update the prediction systems based on an iterative learning process, in which we apply the Monte Carlo-based Bayesian compression for fading the old evidences. In addition, we present a concurrent system that performs these tasks (i.e., the compression and the iterative learning) and modifies the models employed by the MPC system. The update method allows us to maintain the same performance (just slightly better for the prediction of microshrinkages) when we apply high levels of compression (i.e., between 80 % and 98 %) in order to fade the historical data.
5. **An improvement of parameter optimisation system based on evolutionary programming.** We have evolved the solutions developed by other authors creating an algorithm valid for any process or optimisation goal. Although our algorithm is fixed (the source code does not need to be changed), it will adapt itself to the process and automatically update because it employs other parts of the MPC system to perform its optimisation tasks. Moreover, given the known problem of the evolutionary programming is related to the needed time to find the optimal

8.2 Summary of the achieved results

solution, we have developed a genetic algorithm that implements the theory of parallel evolutionary programming and is able to optimise the independent variables producing as few changes as possible (i.e, the algorithm only modifies two variables in most optimisations) in about 3 minutes.

Thanks to the contributions presented above, we were able to fulfill all the specific objectives defined in chapter 1:

- Developing and testing a general model for foreseeing defects and features using machine learning techniques.
- Developing and evaluating a model based on decision-making techniques and sensitivity analysis that allows to provide the needed information to optimise the manufacturing process and keeps it under the known constraints.
- Developing and testing a methodology to maintain the system fully adapted to the several changes that may occur in the manufacturing process.

Similarly, these contributions allowed us to meet the initial operational objectives that we set:

- Designing and implementing a framework for the detection of defects and features in the initial stages of the manufacturing process.
- Designing and implementing a meta-classification system for predicting the step $t+1$ of the plant based on the decision made by multiple predictors, maximizing the accuracy and minimizing the error rates measured as false positive and false negatives rates.
- Designing and implementing a method for knowledge representation, which is employed in the system, to calculate the parameters that must be introduced into the production process to avoid anomalies.
- Designing and implementing a system to offer recommendations on how to configure the process that is being carried out on the plant.
- Designing and implementing a method to keep the system updated in view of the changes that occur in the process.

8. Conclusions

Since we were able to achieve every goal we defined at the beginning of this research, we believe we have successfully fulfilled the main objective behind the working hypothesis: ‘*With the aim of meeting the needs that have already been identified by the companies in terms of quality, cost reduction and eco-making, as well as the ability to obtain a competitive improvement, we are going to design, develop, evaluate and implement a system that is able to readjust the parameters of the production process obeying the constraints defined in the process.*’. Finally, due to we have developed a hybrid and self-adaptive Model Predictive Control system, and thanks to the accomplished achievements, we think we have demonstrated the initial hypothesis:

It is possible to model the business flow of a manufacturing process as a knowledge cloud through the creation of a hybrid (linear and non-linear) model predictive control based on current machine learning techniques, keeping it updated and carrying the manufacturing control to optimize some of its critical parameters.

8.3 Research applications

Since the aim behind this research is to provide an intelligent manufacturing process that could make a difference between companies, the contributions included in this thesis are focused on the industry

In light of this context, all these contributions allow ancient processes, such as the foundry, to keep a much tighter control of what is going on than when using the commonly outdated methods. Moreover, other manufacturing processes that have already employed this kind of systems will be improved because we removed some limitations of the MPC systems as the principal ambition of this research. Therefore, the new updating approach allows MPC systems to be adjusted in real time, avoiding the aforementioned adaptation problems.

Similarly, the type of predictors selected in our new approach provides two important advantages to the manufacturing process. On the one hand, it is not necessary to disturb the current process to generate the final model, i.e., we do not have to stop the production in order to generate an initial dataset

or use expert human knowledge to define the stages of the plant or the critical situations. In this way, given a product, the manufacturing process is usually designed to carry out always the same tasks with, approximately, the same settings. Accordingly, and taking into account this idea, we can extract some patterns from the data that already exists in the plant. On the other hand, and derived from the previous one, we simplify the building process of MPC systems, thus, we can optimise the manufacturing process in a shorter time and with a lower cost.

Although our research has been focused on detecting defects in manufactured products, MPC systems can be applied for any other optimisation purpose. Nowadays, some research areas related to energy or the environment can benefit from this development. Our MPC system is able to work with goals such as the amount of energy used or wasted, the amount of waste generated, or the environmental impact of the process. Consequently, the presented solution is a multi-objective MPC system than can actually carry out a low cost environmental friendly production with 0 defects.

Manufacturing processes are not the only ones which can benefit from our contributions. For example, the MPC solution could be a great control system for smart spaces focused on taking care of children and elderly people. Our system can learn the behaviour patterns of each person, anticipate future dangerous situations through the prediction systems, identify anomalies, deviations from the optimal path of each behaviour; and make the decisions to avoid those situations. In addition, our approach could be useful to provide an intelligent traffic control system. MPC systems can predict future bottlenecks and make the appropriate decisions to divert the traffic or change the traffic lights.

Finally, each contribution could be employed in any type of solutions based on machine learning methods. In this manner, we proved that a meta-classification system matches or even overcomes the best single classifier in terms of accuracy. Also, the combination of classifiers reduces the error rates. Hence, these conclusions can encourage developers to improve the classification systems based on a single classifier. Furthermore, the Monte Carlo-based Bayesian compression method was developed to enhance the updating process of classifiers created through automatic learning processes. Accordingly, the presented adaptive system can be used in all developments that require a way to keep up-to-date its models. In fact, this technique can even help to re-sample larger datasets maintaining the probability distributions.

8.4 Limitations of the model

Although our hybrid and self-adaptive MPC system advances to the state of the art, it presents some limitations.

1. The possibility of appearance of noise during the process,
2. there is no tool that allows us to validate the input data,
3. the predictive systems are incomplete because we remove some independent variables to simplify the final systems or
4. the uncertainty has not been introduced in the predictive models yet.

Despite these already known limitations, the methods employed in our work add other drawbacks that must be joined to the previous ones increasing the list that we should resolve with our future work. First, we identify a limitation related to the predictive models. Our current MPC only deals with a single temporal dimension; in other words, the prediction is performed for the results of the whole manufacturing process. Thus, the predictors are much more complicated and are composed by much more variables to control. Therefore, the division in mini MPC systems with a new layer for communicating between them (Agent-based Model Predictive Control systems) might be the way that will lead us to eliminate this limitation.

Second, the type of learning adds its own limitations to the aforementioned drawbacks. Supervised learning requires the labelling of all the evidences, i.e., we must indicate the results achieved after the manufacturing process for each dependent variable, which means a huge effort. However, many plants have solved this drawback because they already maintain a registry that stores information from the process. Another solution for those foundries not having the possibility to offer that registry would be to reduce the labelling efforts by selecting another type of learning such as semi-supervised learning.

Third, our MPC system directly works with the provided data and, while conducting our experiments, we have found some noise problems generated because the data acquisition is performed in a manually manner. Therefore, we conclude that the inexistence of a previous step for adapting the data and reducing the noise is another limitation. Probably, we would improve the results achieved by the predictive systems only with the development of this new task.

Fourth, the adaptation process performs well but has two limitations inherited from the methods we used. On the one hand, Although the literature says that a Bayesian network is the best classifier to represent the probability distribution [HS86], it may not be the most appropriate way of representing the process. Actually, our experiments show that other models are able to adapt to the process better than the Bayesian networks. On the other hand, the Monte Carlo sampling method is not deterministic. In this way, we cannot know the result that we will get using our compression method; we can only assure the trend that goes through the different compression levels. It seems clear that the removal of one or both constraints would make the updating process better and more reliable.

Finally, we cannot forget the execution time. The optimisation process carried out on our developed system offers quite good time response. However, there may be very critical manufacturing processes that require a very short execution cycle. In that case, the feedback system should be improved to speed its tasks. Despite this fact, we must emphasize that the current system does not have any problem of running under that situation, since the processing time can be reduced turning it into a suboptimal MPC system.

8.5 Avenues for further work

We have identified four research lines for future work related to the MPC system presented in this dissertation. These areas will be described below.

8.5.1 Improving the data gathering process

Currently, the data gathering process is performed manually, which causes the appearance of irregularities in the data is used to represent the process. In this case, the research may be driven in two main ways.

On the one hand, and probably the best of both solutions, we could design and develop a generic SCADA system (Supervisory Control And Data Acquisition systems) [DS99] that would be able to automatically extract the data from the process. The biggest problem of this kind of developments is the heterogeneity of manufacturing processes. Therefore, it is nearly impossible to build a system capable of connecting every kind of equipment or sensor. In this way, instead of focusing our efforts towards connecting the hardware, we can collect the information of the process in a visual way, in other words,

8. Conclusions

we would be able to identify the settings of the machines through the use of artificial vision techniques. Nowadays, these techniques are helpful for identifying vital signs in biomedicine [Mil94], in agriculture [CCK02] or for product inspections [Nov92, TVTV99]. Notwithstanding, the collected data will still have noise problems, in this case, produced by the sensors. This fact forces us to use some techniques to reduce or remove that noise, for instance the Kalman filters [WB95].

On the other hand, we can include an initial step to handle the input data. In this manner, we could determine which instances are wrong or not, correcting or discarding them. Moreover, we can analyse the variables in order to find the importance factor of each one. Actually, we are introducing the data pre-processing stage with the aim of improving the predictive models. Some techniques, such as *Singular value Decomposition* or *Latent Semantic Analysis* have been used in document retrieval with successful results in reducing the noise and ambiguities [SFS08]. Although it may seem impossible the application of these methods to our domain, the decomposition in matrices would be conducted as follows: a first matrix containing castings (documents), a second containing the variables (terms) and the third that relates the previous ones.

In both cases, we expect the predictive systems to be much more accurate and our MPC system, able to manage the manufacturing processes in a much more optimal manner.

8.5.2 Time optimisation in the feedback process

The proposed solution to optimise the independent variables and give a feedback to the plant accomplished really interesting results. Currently, our genetic algorithm was able to perfectly tune at least the 95 % of cases in a short period of time. In other words, it was able to give modification advices to improve the manufacturing process changing a reduced number of variables (the majority of the advises only asked to modify 2 out of 24 variables).

Nevertheless, we identified the optimisation time as the most important limitation of our MPC system. In fact, a MPC system defines a temporal constraint for the processing (analysis, prediction and optimisation) that is fixed by the time interval between executions. Thus, the time needed for our genetic algorithm (about 3 minutes, but achieving always the best possible solution for each evidence) may not be enough for other manufacturing processes. Therefore, as said before this is one of the first limitations to solve if we want to get a more generic MPC system.

The main direction to solve this drawback is the employment of other optimisation methods. As we mentioned in chapter 6, the bottleneck is the large amount of predictions that we have to carry out. In this way, the method able to reduce this number will be a much more efficient solution. Thus, we believe that we could accelerate the optimisation process changing the evolutionary programming by the typical searching algorithms [Pea84]. The difference between both methods is that in the second one we could start from a particular point, the faulty evidence. Therefore, we would reduce the time to arrive at the relevant area of the solutions space, by directly starting to check solutions that are close to the faulty one. This approach could allow us to find the optimal solution in the first iterations because it is highly probably we can obtain the optimal solution after the change of a few independent variables, avoiding the possibility of having to work with a suboptimal MPC system.

Moreover, this new approach would remove another limitation inherent to the evolutionary programming. More accurately, we can assure we are going to find always the global optimal solution when we employ certain types of search algorithms.

8.5.3 Evolving into distributed systems, agent-based Model Predictive Control systems

The usual approach of MPC systems is to collect data from every step of the process in order to predict and optimise the final result of the entire manufacturing process. Nevertheless, this manner of working, which was selected by us, makes the predictive models handle a huge amount of variables (i.e., thousands of variables that need to be tuned very fast). Hence, the presented MPC approach could evolve to become a Distributed Model Predictive Control system based on hybrid and auto-adaptive agents.

The new approach will allow us to distribute small MPC systems with the presented functionality but significantly smaller. Those systems would be simpler since the amount of variables to be controlled is much lower. Specifically, this kind of MPC systems would have a new layer to allow the communication between them. Thanks to this new feature, agents managing any other stage of the process would be informed of the alerts that happen along the manufacturing process. Nowadays, the scientific community is working on this kind of approaches [CJKT02, NDSH08, Sca09, DOC10], however, their MPC systems still maintain the same limitations we already explained and showed in section 2.5. Therefore, the binding between the approaches introduced along this

8. Conclusions

dissertation and the current tendencies of distributed systems would improve the performance and reduce the limitations of MPC systems.

In the same way, due to the simplification of each model, the new Agent-based Model Predictive Control system would allow the management of larger systems, even if they are divided in several locations.

8.5.4 Improving the working infrastructure

MPC systems usually run in computers (more simplex or complex), which are placed in each manufacturing plant. This fact implies that companies must make an investment in acquiring new infrastructure to equip each and every one of its plants. Moreover, these systems have several working requirements (such as good cooling, being taken care by experts and so on) that are not satisfied because the hard conditions of the manufacturing process do not provide the most suitable environment for the computer systems, especially in foundries.

Against this background, we believe that the solution could be to decentralise the computing. In this way, plants would only need to demand its service to control the manufacturing process. Hence, it would not be needed to deploy one MPC system in each plant. They could be provided with powerful and fully dedicated systems, well maintained and in an appropriate environment, to predict and control the manufacturing process thanks to *cloud computing* techniques. Similarly, these computers could be able to manage the process of more than one plant. Therefore, we would reduce the costs of deploying these control solutions. Notwithstanding, we need to design the prediction and the optimisation systems to run in a cloud environment in order to be capable of developing it. Currently, there are approaches like Apache Mahout¹ that can make feasible this approach, although all the tested methods are not available yet.

Moreover, there are hundreds of repetitive tasks that can be performed simultaneously. In this way, the proposed methods could be improved using the GPGPU computing (*General-Purpose computation on Graphical Processing Units*). The latest generations of graphical units allow us to execute a specific source code to get big time improvements applying a high degree of parallelism [OLG⁺07]. Hence, the achieved concurrency could be augmented using these devices. In addition, GPGPU computing can be easily combined with cloud

¹<http://mahout.apache.org/>

computing, creating computing farms equipped with the latest graphical units; in other words, we would have a low cost supercomputer. Nevertheless, the software that can run in graphical units has a specific implementation, thus, it would be mandatory to migrate the majority of the algorithms to this new programming paradigm, which does not currently rely on the same tools.

8.6 Final considerations

Manufacturing processes will be with us for many years. They are the axes of our modern society and we would not have the hundreds of products we use every day without them. These are so included in our lives that if they go missing, we would not know how to live in our day to day.

Nevertheless, the globalised world in which we are living and the emerging rivalry caused by the diversified markets are forcing companies to be able to product as quickly and reliably as possible. In the upcoming years, taking into account the current crisis in which we are involved, the manufacturing companies that want to survive should be capable of generating their products with the highest quality at the lowest cost.

Hence, as a solution to these problems, we are walking the first steps to optimising the manufacturing. Actually, although we worked with the idea of '0 defects', the eco-conscience of the society encourage us to improve other aspects of the manufacturing process to achieve a fully effective and efficient process. Moreover, researchers are generating increasingly intelligent systems, which will continue their evolution until reach, or even overcome, the human beings. Everything together, and new techniques for controlling and automating processes, will take us the point where...

'The factory of the future will have only two employees, a man and a dog. The man will be there to feed the dog. The dog will be there to keep the man from touching the equipment.'

Warren G. Bennis
(1925-presente)

Bibliografía

- [ABQ⁺99] F. Allgöwer, T.A. Badgwell, JS Qin, J.B. Rawlings, y S.J. Wright. Nonlinear Predictive Control and Moving Horizon Estimation—An introductory overview. *Advances in Control, Highlights of ECC*, 99:391–449, 1999.
- [ADN01] H. Al-Duwaish y W. Naeem. Nonlinear model predictive control of hammerstein and wiener models using genetic algorithms. En *Control Applications, 2001.(CCA'01). Proceedings of the 2001 IEEE International Conference on*, páginas 465–469. IEEE, 2001.
- [AG97] Y. Amit y D. Geman. Shape quantization and recognition with randomized trees. *Neural computation*, 9(7):1545–1588, 1997.
- [AJGG⁺01] R.A. Abou-Jeyab, Y.P. Gupta, J.R. Gervais, P.A. Branchi, y S.S. Woo. Constrained multivariable control of a distillation column using a simplified model predictive control algorithm. *Journal of Process Control*, 11(5):509–517, 2001.
- [Aka73] H. Akaike. Information theory and an extension of the maximum likelihood principle. En *Second international symposium on information theory*, tomo 1, páginas 267–281. Springer Verlag, 1973.
- [AT00] R.M. Ansari y M.O. Tadó. Non-linear model-based process control: applications in petroleum refining. *Measurement Science and Technology*, 11:1830, 2000.
- [AW99] S. Amari y S. Wu. Improving Support Vector Machine Classifiers by Modifying Kernel Functions. *Neural Networks*, 12(6):783–789, 1999.

BIBLIOGRAFÍA

- [BA02] K.P. Burnham y D.R. Anderson. *Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach*. Springer Verlag, 2002.
- [Bäc96] T. Bäck. *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms*. Oxford University Press, USA, 1996.
- [Bae09] SA Baeurle. Multiscale modeling of polymer materials using field-theoretic methodologies: a survey about recent developments. *Journal of mathematical chemistry*, 46(2):363–426, 2009.
- [Bay63] T. Bayes. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, 53:370–418, 1763.
- [BBHK10] M. Berthold, C. Borgelt, F. Höppner, y F. Klawonn. *Guide to Intelligent Data Analysis: How to Intelligently Make Sense of Real Data*. Springer, 2010.
- [BCP00] J.G. Bekker, I.K. Craig, y P.C. Pistorius. Model predictive control of an electric arc furnace off-gas process. *Control Engineering Practice*, 8(4):445–455, 2000.
- [Ber95] R. Berber. *Methods of Model based Process Control*. 293. Springer, 1995.
- [Bha99] H.K.D.H. Bhadeshia. Neural networks in materials science. *ISIJ International*, 39:966–979, 1999.
- [Bin79] K. Binder. *Monte-Carlo Methods*. Wiley Online Library, 1979.
- [Bis95] C.M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995. ISBN 0198538642.
- [Bis06] C.M. Bishop. *Pattern recognition and machine learning*, tomo 4. Springer New York, 2006.
- [BM99a] A. Bemporad y M. Moran. Control of systems integrating logic, dynamics, and constraints. *Automatica-Oxford-*, 35:407–427, 1999.

- [BM99b] A. Bemporad y M. Morari. Robust model predictive control: A survey. *Robustness in identification and control*, páginas 207–226, 1999.
- [BM00] A. Bemporad y M. Morari. Predictive control of constrained hybrid systems. *Nonlinear model predictive control*, páginas 71–98, 2000.
- [Bor00] A.C. Bordons. Control predictivo: metodología, tecnología y nuevas perspectivas, departamento de ingeniería sistemas y automática. *Informe Técnico, Universidad de Sevilla. Aguadulce, Almería*, 2000.
- [Bra47] J.L. Bray. *Non-ferrous production metallurgy*. J. Wiley, 1947.
- [Bre96] L. Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.
- [Bre01] L. Breiman. Random Forests. *Machine Learning*, 45(1):5–32, 2001.
- [BSoM95] P.R. Beeley, R.F. Smart, y Institute of Metals. *Investment casting*. Institute of Materials, 1995.
- [Bun96] W. Buntine. A guide to the literature on learning probabilistic networks from data. *Knowledge and Data Engineering, IEEE Transactions on*, 8(2):195–210, 1996.
- [BWG91] R.P. Bithmead, V. Wertz, y M. Gerers. *Adaptive Optimal Control: The Thinking Man's GPC*. Prentice Hall Professional Technical Reference, 1991.
- [CA98] H. Chen y F. Allgöwer. A Quasi-Infinite Horizon Nonlinear Model Predictive Control Scheme with Guaranteed Stability. *Automatica*, 34(10):1205–1217, 1998.
- [CB97] E.F. Camacho y C.A. Bordons. *Model Predictive Control in the Process Industry*. Springer-Verlag New York, Inc., 1997.
- [CB04] E.F. Camacho y C. Bordons. *Model predictive control*. Springer Verlag, 2004.

BIBLIOGRAFÍA

- [CBHK11] N.V. Chawla, K.W. Bowyer, L.O. Hall, y W.P. Kegelmeyer. Smote: Synthetic minority over-sampling technique. *Arxiv preprint arXiv:1106.1813*, 2011.
- [CCK02] Y.R. Chen, K. Chao, y M.S. Kim. Machine vision technology for agricultural applications. *Computers and Electronics in Agriculture*, 36(2):173–191, 2002.
- [CD09] P.F. Christoffersen y F.X. Diebold. Optimal prediction under asymmetric loss. *Econometric Theory*, 13(06):808–817, 2009.
- [CGH96] E. Castillo, J.M. Gutiérrez, y A.S. Hadi. *Expert Systems and Probabilistic Network Models*. Springer, New York, NY, USA, erste edición, December 1996. ISBN 0387948589.
- [CH91] G.F. Cooper y E. Herskovits. A bayesian method for constructing bayesian belief networks from databases. En *Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence*, páginas 86–94. Morgan Kaufmann Publishers Inc., 1991.
- [CH92] G.F. Cooper y E. Herskovits. A bayesian method for the induction of probabilistic networks from data. *Machine learning*, 9(4):309–347, 1992.
- [CHL06] J.R. Cano, F. Herrera, y M. Lozano. On the Combination of Evolutionary Algorithms and Stratified Strategies for Training Set Selection in Data Mining. *Applied Soft Computing Journal*, 6(3):323–332, 2006.
- [CJ06] I. Czarnowski y P. Jedrzejowicz. Instance Reduction Approach to Machine Learning and Multi-Database Mining. En *Proceedings of the Scientific Session organized during XXI Fall Meeting of the Polish Information Processing Society, Informatica, ANNALES Universitatis Mariae Curie-Skłodowska, Lublin*, páginas 60–71. 2006.
- [CJKT02] E. Camponogara, D. Jia, B.H. Krogh, y S. Talukdar. Distributed model predictive control. *Control Systems Magazine, IEEE*, 22(1):44–52, 2002.
- [CKY08] R. Caruana, N. Karampatziakis, y A. Yessenalina. An empirical evaluation of supervised learning in high dimensions. En

- Proceedings of the 25th international conference on Machine learning*, páginas 96–103. ACM, 2008.
- [CL68] C. Chow y C. Liu. Approximating discrete probability distributions with dependence trees. *Information Theory, IEEE Transactions on*, 14(3):462–467, 1968.
- [Cla96] DW Clarke. Adaptive predictive control. *Annual Reviews in Control*, 20:83–94, 1996.
- [Cle81] W. Clegg. Faster data collection without loss of precision. An extension of the learnt profile method. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 37(1):22–28, 1981.
- [CM97] P. Chowdhuri y S. Mehairjan. Alternative to monte carlo method for the estimation of lightning incidence to overhead lines. En *Generation, Transmission and Distribution, IEE Proceedings-*, tomo 144, páginas 129–131. IET, 1997.
- [CMH83] C. Cutler, A. Morshedi, y J. Haydel. An Industrial Perspective on Advanced Control. En *AICHE annual meeting*. 1983.
- [CMT87] D.W. Clarke, C. Mohtadi, y P.S. Tuffs. Generalized predictive control system, part i. the basic algorithm. *Automatica*, 23(2):137–148, 1987. ISSN 0005-1098.
- [Cor94] DMC Corp. Technology overview, julio 1994. Product literature from DMC Corp.
- [CP98] E. Cantú-Paz. A survey of parallel genetic algorithms. *Calculateurs paralleles, reseaux et systems repartis*, 10(2):141–171, 1998.
- [CR80] C.R. Cutler y B L. Ramaker. Dynamic Matrix Control - A Computer Control Algorithm. En *Proceedings de la joint automatic control conference*, páginas 13–15. 1980.
- [Dar37] C. Darwin. *The Origin of Species*. Collier, 1937.
- [Dav92] JH Davenport. Primality testing revisited. En *Papers from the international symposium on Symbolic and algebraic computation*, páginas 123–129. ACM, 1992.

BIBLIOGRAFÍA

- [Day87] W.H.E. Day. Consensus methods as tools for data analysis. En *1. Conference of the International Federation of Classification Societies*, páginas 317–324. 1987.
- [def03a] Defra (2003a) Countryside Stewardship and Environmentally Sensitive Areas Schemes Annual Report 2002–03. Informe técnico, Report No. PB 9023. London SW1P 3JR, 2003.
- [def03b] Defra (2003b). The Countryside Stewardship Scheme 2004. Information and how to apply. Informe técnico, Report No. PB 8723. DEFRA, London, 2003.
- [Die00a] T. Dietterich. Ensemble methods in machine learning. *Multiple classifier systems*, páginas 1–15, 2000.
- [Die00b] T.G. Dietterich. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine learning*, 40(2):139–157, 2000.
- [DKA01] J. Durbin, S.J. Koopman, y AC Atkinson. *Time Series Analysis by State Space Methods*, tomo 15. Oxford University Press Oxford, 2001.
- [DM99] S. Davies y A. Moore. Bayesian networks for lossless dataset compression. páginas 387–391, 1999.
- [DMM⁺93] R. Dollar, L.L. Melton, A.M. Morshedi, D.T. Glasgow, y K.W. Repsher. Consider adaptive multivariable predictive controllers. *Hydrocarbon Processing*, 10:109–113, 1993.
- [DO96] S.L. De Oliveira. *Model Predictive Control for Constrained Nonlinear Systems*. vdf Hochschulverlag AG, 1996.
- [DOC10] L.B. De Oliveira y E. Camponogara. Multi-agent model predictive control of signaling split in urban traffic networks. *Transportation Research Part C: Emerging Technologies*, 18(1):120–139, 2010.
- [DS99] A. Daneels y W. Salter. What is SCADA? En *Proceedings de la International Conference on Accelerator and Large Experimental Physics Control System, Trieste, Italy*, páginas 339–343. 1999.

- [FGG97] N. Friedman, D. Geiger, y M. Goldszmidt. Bayesian network classifiers. *Machine Learning*, 29(2):131–163, 1997.
- [FH52] E. Fix y J.L. Hodges. Discriminatory analysis-nonparametric discrimination: Small sample performance. *Technical Report Project 21-49-004, Report Number 11*, 1952.
- [Fis70] D.h Fischer. The direct simulation method - an alternative to the monte carlo method. En *Proceedings of the fourth annual conference on Applications of simulation*, páginas 134–145. Winter Simulation Conference, 1970.
- [FLM98] B.A. Foss, B. Lohmann, y W. Marquardt. A field study of the industrial modeling process. *Journal of Process Control*, 8(5):325–338, 1998.
- [FM90] J.B. Froisy y T. Matsko. IDCOM-M Application to the Shell Fundamental Control Problem. En *AIChE annual meeting*. 1990.
- [FR97] R. Findeisen y J.B. Rawlings. Suboptimal Infinite Horizon Nonlinear model Predictive Control for Discrete Time Systems. Informe técnico, Technical Report, 1997.
- [Fre97] BJ Frey. *Bayesian Networks for Pattern Classification, Data Compression and Channel Coding*. Tesis Doctoral, Toronto Canada: Department of Electrical and Computer Engineering, University of Toronto, 1997.
- [Fro94] J.B. Froisy. Model Predictive Control: Past, Present and Future. *ISA Transactions*, 33(3):235–243, 1994.
- [fTM08a] American Society for Testing y Materials. ASTM B489 - e1 Standard Practice for Bend Test for Ductility of Electrodeposited and Autocatalytically Deposited Metal Coatings on Metals, 2008.
- [fTM08b] American Society for Testing y Materials. ASTM D1062 - Standard Test Method for Cleavage Strength of Metal-to-Metal Adhesive Bonds, 2008.

BIBLIOGRAFÍA

- [FW95] S. Floyd y M.K. Warmuth. Sample compression, learnability, and the vapnik-chervonenkis dimension. *Machine Learning*, 21(3):269–304, 1995.
- [Gad11] M.H. Gadallah. An alternative to monte carlo simulation method. *International Journal of Experimental Design and Process Optimisation*, 2(2):93–101, 2011.
- [Gar95] S.R. Garner. Weka: The Waikato Environment for Knowledge Analysis. En *Proceedings de la New Zealand Computer Science Research Students Conference*, páginas 57–64. 1995.
- [GCFC06] R. Gonzaga-Cinco y J. Fernández-Carrasquilla. Mechanical properties dependency on chemical composition of spheroidal graphite cast iron. *Revista de Metalurgia*, 42:91–102, 2006.
- [GFH88] P. Grosdidier, B. Froisy, y M. Hammann. The IDCOM-M Controller. En *Model Based Process Control: Proceedings de la IFAC Workshop, Atlanta, Georgia, USA, 13-14 Junio 1988*, páginas 31–36. Pergamon Pr, 1988.
- [GGS01] G.C. Goodwin, S.F. Graebe, y M.E. Salgado. *Control System Design*, tomo 240. Prentice Hall New Jersey, 2001.
- [GM82] C.E. Garcia y M. Morari. Internal Model Control. A Unifying Review and some new Results. *Industrial & Engineering Chemistry Process Design and Development*, 21(2):308–323, 1982.
- [GM86] C.E. Garcia y AM Morshedi. Quadratic Programming Solution of Dynamic Matrix Control (QDMC). *Chemical Engineering Communications*, 46(1):73–87, 1986.
- [GN95] H. Genecili y M. Nikolaou. Design of Robust Constrained Model-predictive Controllers with Volterra Series. *AIChE Journal*, 41(9):2098–2107, 1995.
- [GNDC05] H. Gupta, V. Navda, S.R. Das, y V. Chowdhary. Efficient gathering of correlated data in sensor networks. En *Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing*, páginas 402–413. ACM, 2005.

-
- [Gol79] L.M. Golden. The effect of surface roughness on the transmission of microwave radiation through a planetary surface. *Icarus*, 38(3):451–455, 1979.
- [Gol94] D.E. Goldberg. Genetic and evolutionary algorithms come of age. *Communications of the ACM*, 37(3):113–119, 1994.
- [GPM89] C.E. Garcia, D.M. Prett, y M. Morari. Model Predictive Control: Theory and Practice—A Survey. *Automatica*, 25(3):335–348, 1989.
- [GSDD05] G. Goodwin, M.M. Seron, y J.A. De Don'á. *Constrained control and estimation: an optimisation approach*. Springer Verlag, 2005.
- [Hen98] M.A. Henson. Nonlinear model predictive control: Current status and future directions. *Computers and Chemical Engineering*, 23(2):187–202, 1998.
- [Ho98] T.K. Ho. The random subspace method for constructing decision forests. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(8):832–844, 1998.
- [Hol75] J.H. Holland. *Adaptation in Natural and Artificial Systems*. 53. University of Michigan press, 1975.
- [HS86] G.E. Hinton y T.J. Sejnowski. Learning and relearning in boltzmann machines. *MIT Press, Cambridge, Mass.*, 1:282–317, 1986.
- [HT98] T. Hastie y R. Tibshirani. Classification by pairwise coupling. *The annals of statistics*, 26(2):451–471, 1998.
- [HW07] Q.P. He y J. Wang. Fault detection using the k-nearest neighbor rule for semiconductor manufacturing processes. *Semiconductor Manufacturing, IEEE Transactions on*, 20(4):345–354, 2007.
- [Inc93] Setpoint Inc. SMC-IDCOM: A State-of-the-Art Multivariable Predictive Controller, oct 1993. Product literature from Setpoint, Inc.

BIBLIOGRAFÍA

- [Inc95] Honeywell Inc. RMPCT concepts reference, oct 1995. Product literature from Honeywell Inc.
- [IPDNCT04] L. Int Panis, L. De Nocker, E. Cornelis, y R. Torfs. An uncertainty analysis of air pollution externalities from road transport in belgium in 2010. *Science of the total environment*, 334:287–298, 2004.
- [IPRDNT02] L. Int Panis, A. Rabl, L. De Nocker, y R. Torfs. Diesel or petrol? an environmental comparison hampered by uncertainty. En *Proceedings of 11th International Symposium, Transport and Air Pollution, Held Graz University of Technology, Austria, 19-21 July 2002*, tomo 1. 2002.
- [JL95] G.H. John y P. Langley. Estimating continuous distributions in bayesian classifiers. En *Proceedings of the eleventh conference on uncertainty in artificial intelligence*, páginas 338–345. Morgan Kaufmann Publishers Inc., 1995.
- [Kai80] T. Kailath. *Linear Systems*, tomo 1. Prentice-Hall Englewood Cliffs, NJ, 1980.
- [Kal60] R.E. Kalman. Contributions to the Theory of Optimal Control. *Bol. Soc. Mat. Mexicana*, 5(2):102–119, 1960.
- [KB61] R.E. Kalman y R.S. Bucy. New results in linear filtering and prediction theory. *Journal of Basic Engineering*, 83(3):95–108, 1961.
- [KBD01] L.I. Kuncheva, J.C. Bezdek, y R.P.W. Duin. Decision templates for multiple classifier fusion: an experimental comparison. *Pattern Recognition*, 34(2):299–314, 2001.
- [KBH00a] D.E. Kassmann, T.A. Badgwell, y R.B. Hawkins. Robust steady-state target calculation for model predictive control. *AIChE Journal*, 46(5):1007–1024, 2000.
- [KBH00b] D.E. Kassmann, T.A. Badgwell, y R.B. Hawkins. Robust steady-state target calculation for model predictive control. *AIChE Journal*, 46(5):1007–1024, 2000.

- [KD08] K. Kryszczuk y A. Drygajlo. Credence estimation and error prediction in biometric identity verification. *Signal Processing*, 88(4):916–925, 2008.
- [KG88] S.S. Keerthi y E.G. Gilbert. Optimal Infinite-horizon Feedback Laws for a General Class of Constrained Discrete-Time Systems: Stability and Moving-Horizon approximations. *Journal of Optimization Theory and Applications*, 57(2):265–293, 1988.
- [KHDM98] J. Kittler, M. Hatef, R.P.W. Duin, y J. Matas. On combining classifiers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(3):226–239, 1998.
- [Kle70] D. Kleinman. An Easy Way to Stabilize a Linear Constant System. *IEEE Transactions on Automatic Control*, 15(6):692–692, 1970.
- [KM08] U.B. Kjærulff y A.L. Madsen. *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis*. Information Science and Statistics. Springer, 2008.
- [KSBM01] S.S. Keerthi, S.K. Shevade, C. Bhattacharyya, y K.R.K. Murthy. Improvements to platt’s smo algorithm for svm classifier design. *Neural Computation*, 13(3):637–649, 2001.
- [KSH00] T. Kailath, A.H. Sayed, y B. Hassibi. Linear estimation. *Upper Saddle River, NJ*, 2000.
- [Kun04] Ludmila I. Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, 2004. ISBN 0471210781.
- [KW09] M.H. Kalos y P.A. Whitlock. *Monte Carlo Methods*. Wiley-VCH, 2009.
- [KWL98] M.T. Koterba, F.D. Wilde, y W.W. Lapham. Ground-water data-collection protocols and procedures for the National Water-Quality Assessment Program: Collection and documentation of water-quality samples and related data. *U.S. Geological Survey, Earth Science Information Center, Open-File Reports Section, BOX 25286, MS 517, Denver, CO 80225(USA).*[nd]., 98(4107), 1998.

BIBLIOGRAFÍA

- [Kwo94] W.H. Kwon. Advances in Predictive Control: Theory and Application. En *Asian Control Conference, Tokyo*. 1994.
- [KZP07] SB Kotsiantis, ID Zaharakis, y PE Pintelas. Supervised machine learning: A review of classification techniques. *Frontiers in Artificial Intelligence and Applications*, 160:3, 2007.
- [Lar90] W E Larimore. Canonical variate analysis in identification, filtering, and adaptive control. *29th IEEE Conference on Decision and Control*, páginas 596–604, 1990.
- [LC97] J.H. Lee y B. Cooley. Recent advances in model predictive control and other related areas. En *AIChE Symposium Series*, tomo 93, páginas 201–216. New York, NY: American Institute of Chemical Engineers, 1971-c2002., 1997.
- [LD88] W.J. Leech y T.N. Dean. Automated rule based process control method with feedback and apparatus therefor, 1988. US Patent 4,754,410.
- [Lew98] D. Lewis. Naive (bayes) at forty: The independence assumption in information retrieval. *Machine Learning: ECML-98*, páginas 4–15, 1998.
- [LHWB06] M. Lazar, W. Heemels, S. Weiland, y A. Bemporad. Stabilizing model predictive control of hybrid systems. *Automatic Control, IEEE Transactions on*, 51(11):1813–1818, 2006.
- [Lju99] L. Ljung. *System Identification: Theory for the User (2nd Edition)*. Prentice Hall, 2 edición, jan 1999.
- [LM92] C.W. Lung y N.H. March. *Mechanical Properties of Metals: Atomistic and Fractal Continuum Approaches*. World Scientific Pub Co Inc, Julio 1992.
- [LM01] H. Liu y H. Motoda. *Instance Selection and Construction for Data Mining*. Kluwer Academic Pub, 2001.
- [LM08] H. Liu y H. Motoda. *Computational Methods of Feature Selection*. Chapman & Hall/CRC, 2008.
- [LMG94] J.H. Lee, M. Morari, y C.E. Garcia. State-space interpretation of model predictive control. *Automatica*, 30(4):707–717, apr 1994.

- [LSOdM98] A. Lazaro, I. Serrano, J.P. Oria, y C. de Miguel. Ultrasonic sensing classification of foundry pieces applying neural networks. En *5th International Workshop on Advanced Motion Control, 1998. AMC'98 - Coimbra., 1998*, páginas 653–658. IEEE, 1998.
- [LSS06] P. Larrañaga, J. Sertucha, y R. Suárez. Análisis del proceso de solidificación en fundiciones gráficas esféricas. *Revista de Metalurgia*, 42:244–255, 2006.
- [Mac02] J.M. Maciejowski. *Predictive control: with constraints*. Pearson education, 2002.
- [May97] D.Q. Mayne. Nonlinear Model Predictive Control: An Assessment. En *Chemical Process Control-V: Assessment and New Directions for Research: Proceedings of the Fifth International Conference on Chemical Process Control, Tahoe City, California, January 7-12, 1996*, páginas 217–231. American Institute of Chemical Engineers, 1997.
- [May00] D.Q. Mayne. Nonlinear Model Predictive Control: Challenges and Opportunities. *Nonlinear model predictive control*, páginas 23–44, 2000.
- [MB88] P. Marquis y J.P. Broustail. SMOC, a Bridge Between State Space and Model Predictive Controllers: Application to the Automation of a Hydrotreating Unit. En *Model-Based Process Control of the 1988 IFAC Workshop, Georgia, USA*, páginas 37–45. 1988.
- [MBP⁺02] G.D. Martin, E. Boe, S. Piche, J.D. Keeler, D. Timmer, M. Gerules, y J.P. Havener. Method and apparatus for modeling dynamic and steady-state processes for prediction, control and optimization, noviembre 22 2002. US Patent App. 10/302,923.
- [MDM⁺82] H.T. MacGillivray, R.J. Dodd, B.V. McNally, J.F. Lightfoot, H.G. Corwin, y S.R. Heathcote. Monte-carlo simulations of galaxy systems. *Astrophysics and Space Science*, 81(1):231–250, 1982.
- [MHER95] E.S. Meadows, M.A. Henson, J.W. Eaton, y J.B. Rawlings. Receding Horizon Control and Discontinuous State Feedback

BIBLIOGRAFÍA

- Stabilization. *International Journal of Control*, 62(5):1217–1229, 1995.
- [Mil94] B. Miller. Vital signs of identity [biometrics]. *Spectrum, IEEE*, 31(2):22–30, 1994.
- [MJ98] G. Martin y D. Johnston. Continuous Model-based Optimization. En *Hydrocarbon processing's process optimization conference*. 1998.
- [ML91] M. Morari y J.H. Lee. Model Predictive Control: The Good, the Bad, and the Ugly. páginas 419–444. 1991.
- [ML97] M. Morari y J H. Lee. Model predictive control: Past, present and future. *Computers and Chemical Engineering*, 23:667–682, 1997.
- [MLP08] M. Maragoudakis, E. Loukis, y P.P. Pantelides. Random forests identification of gas turbine faults. En *ICSENG '08: Proceedings de 2008 en la 19ª International Conference on Systems Engineering*, páginas 127–132. IEEE Computer Society, Washington, DC, USA, 2008. ISBN 978-0-7695-3331-5.
- [MM90] D.Q. Mayne y H. Michalska. Receding Horizon Control of Nonlinear Systems. *IEEE Transactions on Automatic Control*, 35(7):814–824, 1990.
- [MM93] H. Michalska y D.Q. Mayne. Robust Receding Horizon Control of Constrained Nonlinear systems. *IEEE Transactions on Automatic Control*, 38(11):1623–1633, 1993.
- [Mos95] E. Mosca. *Optimal, Predictive, and Adaptive Control*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1995. ISBN 0-13-847609-8.
- [MR93] K.R. Muske y J.B. Rawlings. Model Predictive Control with Linear Models. *AIChE Journal*, 39(2):262–287, 1993.
- [MRRS00] D.Q. Mayne, J.B. Rawlings, C.V. Rao, y P.O.M. Scokaert. Constrained model predictive control: Stability and optimality. *Automatica*, 36(6):789–814, 2000.

- [MS93] M. Milik y J. Skolnick. Insertion of peptide chains into lipid membranes: An off-lattice monte carlo dynamics model. *Proteins: Structure, Function, and Bioinformatics*, 15(1):10–25, 1993.
- [MSTC94] D. Michie, D.J. Spiegelhalter, C.C. Taylor, y J. Campbell, editores. *Machine Learning, Neural and Statistical Classification*. Ellis Horwood, Upper Saddle River, NJ, USA, 1994. ISBN 0-13-106360-X.
- [MZ89] M. Morari y E. Zafiriou. *Robust Process Control*. Morari, 1989.
- [NDSH08] RR Negenborn, B. De Schutter, y J. Hellendoorn. Multi-agent model predictive control for transportation networks: Serial versus parallel schemes. *Engineering Applications of Artificial Intelligence*, 21(3):353–366, 2008.
- [NMS00] G. Nicolao, L. Magni, y R. Scattolini. Stability and robustness of nonlinear receding horizon control. *Nonlinear model predictive control*, páginas 3–22, 2000.
- [Nov92] A.R. Novini. Machine vision inspection system and method for transparent containers, 1992. US Patent 5,095,204.
- [OBK⁺97] C. Onnen, R. Babuska, U. Kaymak, JM Sousa, HB Verbruggen, y R. Isermann. Genetic algorithms for optimization in predictive control. *Control Engineering Practice*, 5(10):1363–1372, 1997.
- [OGLC09] P. Ojeda, M.E. Garcia, A. Londoño, y N.Y. Chen. Monte carlo simulations of proteins in cages: Influence of confinement on the stability of intermediate states. *Biophysical journal*, 96(3):1076–1082, 2009.
- [OLG⁺07] J.D. Owens, D. Luebke, N. Govindaraju, M. Harris, J. Krüger, A.E. Lefohn, y T.J. Purcell. A survey of general-purpose computation on graphics hardware. En *Computer graphics forum*, tomo 26, páginas 80–113. Wiley Online Library, 2007.
- [OW84] T.J. Ostrand y E.J. Weyuker. Collecting and categorizing software error data in an industrial environment. *Journal of Systems and Software*, 4(4):289–300, 1984.

BIBLIOGRAFÍA

- [PBZ08] Y.K. Peña, P.G. Bringas, y A. Zabala. Advanced fault prediction in high-precision foundry production. En *Industrial Informatics, 2008. INDIN 2008. 6th IEEE International Conference on*, páginas 1672–1677. IEEE, 2008.
- [Pea84] J. Pearl. *Heuristics: Intelligent Search Strategies for Computer Problem Solving*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1984.
- [PF97] F. Provost y T. Fawcett. Analysis and Visualization of Classifier Performance: Comparison under Imprecise Class and Cost Distributions. En *Proceedings of the third international conference on knowledge discovery and data mining*, páginas 43–48. Amer Assn for Artificial, 1997.
- [PG80] D.M. Prett y R.D. Gillette. Optimization and Constrained Multivariable Control of a Catalytic Cracking Unit. En *Proceedings de la joint automatic control conference*, tomo 1. 1980.
- [PG88] D.M. Prett y C.E. García. *Fundamental Process Control*. Butterworths Stoneham, MA, 1988.
- [PK03] M. Perzyk y A. Kočański. Detection of causes of casting defects assisted by artificial neural networks. tomo 217, páginas 1279–1284. Sage Publications, 2003.
- [Pla98] J.C. Platt. 12 fast training of support vector machines using sequential minimal optimization. 1998.
- [PLD05] H. Peng, F. Long, y C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(8):1226–1238, 2005.
- [PMM⁺07] D. Pham, Z. Muhamad, M. Mahmuddin, A. Ghanbarzadeh, E. Koç, y S. Otri. Using the bees algorithm to optimise a support vector machine for wood defect classification. En *Memorias del Innovative Production Machines and Systems Virtual Conference*. 2007.
- [PR00] J. Pearl y S. Russell. Bayesian networks. Informe Técnico Tech. Rep. R-216, Computer Science Department, University of California, Los Angeles, 2000.

- [PSRJG00] S. Piche, B. Sayyar-Rodsari, D. Johnson, y M. Gerules. Non-linear model predictive control using neural networks. *Control Systems Magazine, IEEE*, 20(3):53–62, 2000.
- [PZH08] T. Peng, W. Zuo, y F. He. SVM based Adaptive Learning Method for Text Classification from Positive and Unlabeled Documents. *Knowledge and Information Systems*, 16(3):281–301, 2008.
- [QB97] S.J. Qin y T.A. Badgwell. An overview of industrial model predictive control technology. En *AIChE Symposium Series*, tomo 93, páginas 232–256. New York, NY: American Institute of Chemical Engineers, 1971-c2002., 1997.
- [QB00] S.J. Qin y T.A. Badgwell. An Overview of Nonlinear Model Predictive Control Applications. *Nonlinear Model Predictive Control*, 26:369–392, 2000.
- [QB03] S.J. Qin y T.A. Badgwell. A survey of industrial model predictive control technology. *Control engineering practice*, 11(7):733–764, 2003.
- [Qui86] JR Quinlan. Induction of Decision Trees. *Machine learning*, 1(1):81–106, 1986.
- [Qui93] J.R Quinlan. *C4.5 Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.
- [Ram93] W.F. Ramirez. *Process Control and Identification*. Academic Press San Diego, 1993.
- [Raw00] J.B. Rawlings. Tutorial overview of model predictive control. *Control Systems, IEEE*, 20(3):38–52, 2000.
- [Rip87] B.D. Ripley. *Stochastic Simulation*, tomo 183. Wiley Online Library, 1987.
- [RL79] J. Rissanen y G.G. Langdon. Arithmetic coding. *IBM Journal of Research and Development*, 23(2):149–162, 1979.
- [RMM94] J.B. Rawlings, E.S. Meadows, y K.R. Muske. Nonlinear Model Predictive Control: A Tutorial and Survey. En *Advanced control of chemical processes (ADCHEM'94): IFAC symposium*,

BIBLIOGRAFÍA

- Kyoto, Japan, 25-27 May 1994*, páginas 203–214. Pergamon Pr, 1994.
- [RN03] S.J. Russell y Norvig. *Artificial Intelligence: A Modern Approach (Second Edition)*. Prentice Hall, 2003.
- [RR99] C.V. Rao y J.B. Rawlings. Steady states and constraints in model predictive control. *AIChE Journal*, 45(6):1266–1278, 1999.
- [RRTP76] J. Richalet, A. Rault, JL Testud, y J. Papon. Algorithmic Control of Industrial Processes. En *Identification and system parameter estimation: proceedings del 4º IFAC Symposium, Tbilisi, USSR, 21-27 September 1976*, páginas 1119–1167. North-Holland Pub. Co., 1976.
- [RRTP78] J. Richalet, A. Rault, J.L. Testud, y J. Papon. Model Predictive Heuristic Control: Applications to Industrial Processes. *Automatica*, 14(5):413–428, 1978.
- [RWR98] C.V. Rao, S.J. Wright, y J.B. Rawlings. Application of interior-point methods to model predictive control. *Journal of optimization theory and applications*, 99(3):723–757, 1998.
- [Saw03] S.S. Sawilowsky. You think you’ve got trivials. *Journal of Modern Applied Statistical Methods*, 2(1):218–225, 2003.
- [SB03] H. Sarimveis y G. Bafas. Fuzzy model predictive control of non-linear processes using genetic algorithms. *Fuzzy sets and systems*, 139(1):59–80, 2003.
- [SBGZ98] G.B. Sentoni, L.T. Biegler, J.B. Guiver, y H. Zhao. State-space nonlinear process modeling: Identification and universality. *AI-ChE journal*, 44(10):2229–2239, 1998.
- [Sca09] R. Scattolini. Architectures for distributed and hierarchical model predictive control - a review. *Journal of Process Control*, 19(5):723–731, 2009.
- [Sch78] G. Schwarz. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

- [SE00] Chan K. Saltelli, A. and y Scott E.M. *Sensitivity Analysis: Gauging the Worth of Scientific Models*. Springer, New York, NY, USA, wiley edición, August 2000. ISBN 9780471998921.
- [Seg04] M.R. Segal. Machine learning benchmarks and random forest regression. 2004.
- [SF97] O. Slupphaug y B.A. Foss. Model predictive control for a class of hybrid systems. En *In Proc. European Control Conf.* 1997.
- [SF98] O. Slupphaug y B.A. Foss. Bilinear Matrix Inequalities and Robust Stability of Nonlinear Multi-model MPC. En *Proceedings of the American Control Conference*, tomo 3, páginas 1689–1694. 1998.
- [SF01] A. Seewald y J. Fürnkranz. An evaluation of grading classifiers. *Advances in Intelligent Data Analysis*, páginas 115–124, 2001.
- [SFS08] S. Sakellaridi, H. Fang, y Y. Saad. Graph-based multilevel dimensionality reduction with applications to eigenfaces and latent semantic indexing. En *ICMLA '08: Proceedings of the 2008 Seventh International Conference on Machine Learning and Applications*, páginas 194–200. IEEE Computer Society, Washington, DC, USA, 2008.
- [SGS00] P. Spirtes, C.N. Glymour, y R. Scheines. *Causation, Prediction, and Search*, tomo 81. The MIT Press, 2000.
- [Shi94] F.G. Shinskey. *Feedback Controllers for the Process Industries*. McGraw-Hill Professional, 1994.
- [Sin89] A.K. Sinha. Ferrous physical metallurgy. 1989.
- [SKM09] Y. Singh, A. Kaur, y R. Malhotra. Comparative Analysis of Regression and Machine Learning Methods for Predicting Fault Proneness Models. *International Journal of Computer Applications in Technology*, 35(2):183–193, 2009.
- [SLS06] J. Sertucha, A. Loizaga, y R. Suárez. Improvement opportunities for simulation tools. En *Proceedings del 16º European Conference and Exhibition on Digital Simulation for Virtual Engineering*. 2006. Charla invitada.

BIBLIOGRAFÍA

- [Slu98] O. Slupphaug. *On Robust Constrained Nonlinear Control and Hybrid Control: BMI- and MPC-based State-feedback Schemes*. Department of Engineering Cybernetics, Norwegian University of Science and Technology, 1998.
- [SM98] P.O.M. Scokaert y D.Q. Mayne. Min-max feedback model predictive control for constrained linear systems. *IEEE Transactions on Automatic Control*, 43(8):1136–1142, 1998.
- [Soe92] R. Soeterboek. *Predictive Control: A Unified Approach*. Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1992.
- [SR98] P.O.M. Scokaert y J.B. Rawlings. Constrained Linear Quadratic Regulation. *IEEE Transactions on Automatic Control*, 43(8):1163–1169, 1998.
- [SS05] Kalpakjian S. y S. Schmid. *Manufacturing, Engineering and Technology*. Pentice Hall, 2005. ISBN 0131489658.
- [SSLG07] J. Sertucha, R. Suárez, J. Legazpi, y P. Gacetabeitia. Influence of moulding conditions and mould characteristics on the contraction defects appearance in ductile iron castings. *Revista de Metalurgia*, 43:188–195, 2007.
- [SV10] D. Silver y J. Veness. Monte-carlo planning in large pomdps. *Advances in Neural Information Processing Systems (NIPS)*, 46, 2010.
- [THM02] C.G. Thomas, R.A. Harshman, y R.S. Menon. Noise Reduction in BOLD-based fMRI using Component Analysis. *Neuroimage*, 17(3):1521–1537, 2002.
- [Tom95] M. Tomassini. A survey of genetic algorithms. *Annual Reviews of Computational Physics*, 3(2):87–118, 1995.
- [Tor03] K. Torkkola. Feature Extraction by Non Parametric Mutual Information Maximization. *The Journal of Machine Learning Research*, 3:1415–1438, 2003.
- [TSK⁺06] P.N. Tan, M. Steinbach, V. Kumar, et al. *Introduction to Data Mining*. Pearson Addison Wesley Boston, 2006.

- [TVBDW⁺03] U. Thissen, R. Van Brakel, AP De Weijer, WJ Melssen, y LMC Buydens. Using support vector machines for time series prediction. *Chemometrics and intelligent laboratory systems*, 69(1-2):35–49, 2003.
- [TVT⁺99] P. Tantaswadi, J. Vilainatre, N. Tamaree, y P. Viraiivan. Machine vision for automated visual inspection of cotton quality in textile industries using color isodiscrimination contour. *Computers & industrial engineering*, 37(1-2):347–350, 1999.
- [ÜMB06] B. Üstün, W.J. Melssen, y L.M.C. Buydens. Facilitating the application of support vector regression by using a universal pearson vii function based kernel. *Chemometrics and Intelligent Laboratory Systems*, 81(1):29–40, 2006.
- [Van83] W. Vandaele. *Applied Time Series and Box-Jenkins Models*. Academic Press, 1983.
- [Vap00] V.N. Vapnik. *The Nature of Statistical Learning Theory*. Springer, 2000.
- [VGN95] P. Vuthandam, H. Genceli, y M. Nikolaou. Performance Bounds for Robust Quadratic Dynamic Matrix Control with end condition. *AIChE Journal*, 41(9):2083–2097, 1995.
- [VMB02] M. Vasudevan, M. Muruganath, y A.K. Bhaduri. Application of bayesian neural network for modelling and prediction of ferrite number in austenitic stainless steel welds. *ser. Mathematical Modelling of Weld Phenomena - VI. London: Institute of Materials*, páginas 1079–1100, 2002.
- [VOF10] J.D. Velásquez, Y. Olaya, y C.J. Franco. Time Series Prediction Using Support Vector Machines. *Ingeniare. Revista chilena de ingeniería*, 18:64–75, 2010.
- [Vor01] J. Voracek. Prediction of mechanical properties of cast irons. *Applied Soft Computing*, 1(2):119–125, 2001.
- [Wah91] B. Wahlberg. System identification using laguerre models. *Automatic Control, IEEE Transactions on*, 36(5):551–562, 1991.

BIBLIOGRAFÍA

- [WB95] G. Welch y G. Bishop. An Introduction to the Kalman Filter. *University of North Carolina at Chapel Hill, Chapel Hill, NC*, 7(1), 1995.
- [WMG06] L. Wang, Z. Mu, y H. Guo. Application of support vector machine in the prediction of mechanical property of steel materials. *Journal of University of Science and Technology Beijing, Mineral, Metallurgy, Material*, 13(6):512–515, 2006.
- [WNC87] I.H. Witten, R.M. Neal, y J.G. Cleary. Arithmetic coding for data compression. *Communications of the ACM*, 30(6):520–540, 1987.
- [Wol89] D.H. Wolpert. A benchmark for how well neural nets generalize. *Biological Cybernetics*, 61(4):303–313, 1989.
- [Wol92] D.H. Wolpert. Stacked generalization. *Neural networks*, 5(2):241–259, 1992.
- [Wri97] S. Wright. Applying new optimization algorithms to model predictive control. En *Fifth International Conference on Chemical Process Control – CPC V*, páginas 147–155. CACHE Publications, 1997.
- [WY07] A. Widodo y B.S. Yang. Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing*, 21(6):2560–2574, 2007.
- [YAO93] XIN YAO. Evolutionary artificial neural networks. *International Journal of Neural Systems*, 4(3):203–222, 1993.
- [Yap75] Lakshman S. Yapa. Analytical alternatives to the monte carlo simulation of spatial diffusion. *Annals of the Association of American Geographers*, 65(2):163–176, 1975.
- [YLW03] J. Yang, M. Liu, y C. Wu. Genetic algorithm based nonlinear model predictive control method. *Control and Decision*, 18(2):141–144, 2003.
- [YP93] T.H. Yang y E. Polak. Moving Horizon Control of Nonlinear Systems with Input Saturation, Disturbances and Plant Uncertainty. *International Journal of Control*, 58(4):875–903, 1993.

- [YT91] C. Yousfi y R. Tournier. Steady State Optimization Inside Model Predictive Control. En *American Control Conference, 1991*, páginas 1866–1870. IEEE, 1991.
- [YZZ07] J. Yang, Y. Zhang, y Y. Zhu. Intelligent fault diagnosis of rolling element bearing based on SVMs and fractal dimension. *Mechanical Systems and Signal Processing*, 21(5):2012–2024, 2007.
- [Zaf90] E. Zafiriou. Robust Model Predictive Control of Processes with Hard Constraints. *Computers & Chemical Engineering*, 14(4-5):359–371, 1990.
- [ZGS98] H. Zhao, J. Guiver, y G. Sentoni. An identification approach to nonlinear state space model for industrial multivariable model predictive control. En *American Control Conference, 1998. Proceedings of the 1998*, tomo 2, páginas 796–800. IEEE, 1998.
- [ZXD08] P. Zhang, Z. Xu, y F. Du. Optimizing casting parameters of ingot based on neural network and genetic algorithm. En *Fourth International Conference on Natural Computation, 2008. ICNC'08. Washington, DC, Los Alamitos USA*, tomo 4, páginas 545–548. IEEE, 2008.
- [ZZ06] J. Zhang y M. Zulkernine. A hybrid network intrusion detection technique using random forests. En *ARES '06: Proceedings de la First International Conference on Availability, Reliability and Security*, páginas 262–269. IEEE Computer Society, Washington, DC, USA, 2006. ISBN 0-7695-2567-9.

Índice alfabético

A

- adaptación continua, 11, 15, 19, 284
 - compresión bayesiana*, 11, 154, 155, 161, 168, 170, 173, 182, 187, 189, 197, 205, 212, 213, 215, 219, 222, 225, 233, 236, 243, 245, 280, 283, 287
 - compresión bayesiana basada en métodos de Montecarlo*, 154, 155, 161, 168, 170, 173, 182, 187, 189, 197, 205, 212, 213, 215, 219, 222, 225, 233, 236, 243, 245, 280, 283, 287
 - envejecimiento de muestras*, 170
 - envejecimiento de muestras*, 154, 157, 161, 162, 205, 207, 216, 227, 231, 235, 245
 - factores de desvanecimiento*, 11, 154, 279
 - factores de envejecimiento*, 279, 280, 283
 - métodos de Montecarlo*, 169
 - métodos de Montecarlo*, 154–158, 160–162, 167, 168, 170, 173, 180, 182, 187, 189, 197, 205, 212, 213, 215, 216, 219, 222, 225, 233, 236, 239, 243, 245, 280, 283, 287, 288
- adaptación de los modelos, 155, 161, 167, 168, 188, 205, 225, 245
- adquisición de datos, 10, 124, 149, 170, 262, 288, 289
 - manual*, 10, 149, 289
 - SCADA*, 10, 149, 289
- afinación, 34
- ajuste del sistema, 264
- algoritmo de las abejas, 9, 76
- algoritmo genético, 9, 76
- análisis de componentes principales, 156
- análisis de la semántica latente, 149, 289
- análisis de sensibilidades, 11, 14, 160, 284
- análisis multivariable, 96
- aprendizaje acumulativo, 184, 205, 210, 280, 283
- aprendizaje automático, 7, 8, 10, 13, 14, 21, 28, 75, 85, 94, 95, 102, 104, 124, 132, 148, 156, 172, 182, 278, 282, 284–287
 - clasificadores estadísticos*, 8
 - semi-supervisado*, 288
 - supervisado*, 94–96, 102, 104, 106, 111, 114, 124, 125, 132, 155,

ÍNDICE ALFABÉTICO

- 167, 168, 182, 282, 287
 - aprendizaje del modelo, 125, 171
 - aprendizaje estructural, 97, 125, 134
 - clasificador bayesiano ingenuo*, 98–100, 125, 127, 130, 171, 192, 212, 229
 - clasificador bayesiano ingenuo aumentado con un árbol*, 97, 101, 125, 127, 130, 139, 141, 143, 146, 164, 171, 175, 180, 191, 192, 194, 199, 200, 202, 209, 212, 217, 222, 229, 231, 237, 240, 263, 265, 269
 - escalar colinas*, 97, 98, 125, 164, 171, 192, 212, 220, 229, 231, 237
 - K2*, 97–99, 164, 171, 192, 210, 229, 231, 237
 - NPC*, 134
 - PC*, 134
 - aprendizaje paramétrico, 97
 - árboles de decisión, 9, 109, 113, 126–128, 130, 134, 136, 137, 171, 191, 194, 199, 202, 210, 212, 217, 222, 240, 282
 - bosques aleatorios*, 111, 112, 126, 128, 130, 132, 134, 137, 171, 176, 185, 189, 192, 195, 200, 209, 212, 219, 222, 223, 229–231, 237, 239, 240, 242, 278
 - C4.5*, 111, 112, 126–128, 143, 171, 176, 182, 184, 185, 191, 194, 199, 202, 209, 210, 212, 217, 220, 222, 229, 231, 240
 - áreas de aplicación, 277
 - arena, 86
 - arquitectura, 16
 - arquitectura global
 - controlador*, 18
 - evaluador*, 18
 - generador de nuevos modelos*, 18
 - lector de evidencias*, 16
 - predictores*, 18
 - preprocesador*, 16
 - sistema experto*, 18
 - autoadaptativos, 8, 161, 167, 278, 285, 287
- ## B
- búsqueda de la solución óptima, 263
 - búsquedas
 - algoritmos de búsqueda*, 160
 - binaria*, 259
 - secuencial*, 259
 - bosques aleatorios, 111, 112, 126, 128, 130, 132, 134, 137, 171, 176, 185, 189, 192, 195, 200, 209, 212, 219, 222, 223, 229–231, 237, 239, 240, 242, 278
- ## C
- C4.5*, 111, 112, 126, 128, 143, 171, 176, 182, 184, 185, 191, 194, 199, 202, 209, 210, 212, 217, 220, 222, 229, 231, 240
 - calidad de la mezcla, 86
 - características, 80, 84, 85, 88, 124
 - propiedades mecánicas*, 86, 88–90, 124, 132, 134, 188, 191, 225, 269, 273, 278, 281
 - carga máxima, 89, 90, 124, 134, 136, 143, 146, 148, 169, 188, 191, 200, 202, 225, 229, 231, 233, 235–237, 243, 245, 262, 263, 269, 270, 272, 278, 282
 - clasificación colectiva, 94
 - clasificador bayesiano ingenuo, 98–100, 125, 127, 130, 171, 192, 212,

- 229
- clasificador bayesiano ingenuo aumentado con un árbol, 97, 101, 125, 127, 130, 139, 141, 143, 146, 164, 171, 175, 180, 191, 192, 194, 199, 200, 202, 209, 212, 217, 222, 229, 231, 237, 240, 263, 265, 269
- clasificadores, 95, 119, 124, 127, 128, 130, 137, 167, 173, 176, 178, 180, 197, 200, 202, 207, 212, 213, 216, 219, 223, 225–227, 230, 231, 235, 237, 240, 242, 243
- árboles de decisión, 9, 109, 113, 126–128, 130, 134, 136, 137, 171, 191, 194, 199, 202, 210, 212, 217, 222, 240, 282
- k* vecinos más próximos, 9, 76, 102, 104, 125, 127, 130, 134, 136, 144, 171, 195, 203, 213, 219, 223, 230, 239, 242
- máquinas de soporte vectorial, 9, 76, 106, 109, 126, 128, 130, 136, 171, 175, 176, 182, 185, 192, 195, 200, 203, 209, 210, 213, 219, 223, 230, 233, 239, 242, 243, 282
- meta-clasificador, 11, 15, 119–121, 132, 134, 136, 141–144, 146, 148, 210, 265, 279, 284
- redes bayesianas, 9, 75, 96–99, 101, 125, 128, 130, 132, 154–157, 161–163, 166, 171, 173, 191, 210, 220, 222, 231, 237, 240, 242, 243, 245, 279, 283, 288
- redes neuronales artificiales, 8, 39, 40, 43, 44, 75, 94, 104, 125, 134, 136, 171, 203, 213, 219, 223, 230, 242, 251
- clasificadores estadísticos, 8, 75
- cloud computing, 292
- colada, 84
- combinación de clasificadores, 18, 95, 124–126, 137, 139, 144, 269, 282
- combinación por voto, 115, 126, 139, 141, 143, 146, 279
- grading, 115, 119–122, 127, 139, 143, 144, 146, 263, 269, 272, 279
- multiesquema, 115, 122, 123, 127, 139, 141, 143, 146, 279
- stacking, 115, 121, 122, 127, 139, 141, 143, 263, 265, 272, 279
- combinación por voto, 115, 126
- la mayoría decide, 115, 116, 126, 139, 142, 143, 146
- mayoría simple, 115
- pluralidad, 115, 116
- regla de decisión bayesiana, 116, 117
- regla de la media, 117, 126, 141
- regla de la probabilidad mínima, 119, 127, 141, 144
- regla de la probabilidad máxima, 118, 127, 141
- regla de la suma, 117, 118
- regla del producto, 116, 117, 119, 126, 141, 144
- unanimidad, 115, 116
- combinar resultados, 18
- composición, 86
- compresión bayesiana, 11, 154, 155, 161, 168, 170, 173, 182, 187, 189, 197, 205, 212, 213, 215, 219, 222, 225, 233, 236, 243, 245, 280, 283, 287

ÍNDICE ALFABÉTICO

compresión bayesiana basada en métodos de Montecarlo, 154, 155, 161, 168, 170, 173, 182, 187, 189, 197, 205, 212, 213, 215, 219, 222, 225, 233, 236, 243, 245, 280, 283, 287

concepto general, 28

conciencia colectiva, 19

condiciones de ruta necesarias, 134

conjunto de particiones, 121

contracción secundaria, 87

control, 80

control de procesos, 5

controlador, 18, 35, 44, 63, 64, 71

objetivos principales, 44

controles de la producción, 2

criterio de información bayesiana, 156

criterio de los mínimos cuadrados, 42

crossover rate, 257

D

defectos, 80, 85, 87, 126, 134, 188, 261, 263, 264, 278, 281, 286

carga máxima, 89, 90, 124, 134, 136, 143, 146, 148, 169, 188, 191, 200, 202, 225, 229, 231, 233, 235–237, 243, 245, 262, 263, 269, 270, 272, 278, 282

microrrechupes, 84, 86–88, 124, 128, 132, 134, 136, 137, 143, 144, 146, 148, 169, 172, 178, 182, 188, 189, 205, 206, 210, 225–227, 235, 236, 245, 262, 263, 265, 269, 271–273, 278, 280–283

descomposición de Cholesky, 43

descomposición en valores singulares, 149, 289

desplazamiento, 264, 269, 270

desviación de la normalidad, 19

dimensión, 86

diseño, 34

distancia euclidiana, 103

distribución estadística, 97, 154, 166, 173, 246, 288

distribución probabilística, 97, 116, 157–159, 161, 163, 166, 168, 175, 187, 191, 205, 209, 237, 245, 246, 279, 280, 282

E

elitismo, 259, 263, 274

embudo, 51

enfriamiento, 84, 85

envejecimiento de muestras, 154, 157, 161, 162, 170, 205, 207, 216, 227, 231, 235, 245

equilibrio interno, 44

escalar colinas, 97, 98, 125, 164, 171, 192, 212, 220, 229, 231, 237, 274

espacio de estados, 32

etapas de la fundición

control final, 84

desarrollo de patrones, 84

enfriamiento, 84

finalización, 84

fusión y colada, 84

moldes de arena y creación de mazotas, 84

evaluación de instancias, 263

evaluación de resultados, 264

evaluación simultánea, 255

evaluador, 18

F

fábrica del futuro, 277, 293

- factores de desvanecimiento, 11, 154, 279
- factores de envejecimiento, 279, 280, 283
- filtro de Kalman, 46, 54, 59, 65, 66, 71, 289
- formulación, 32
- función de idoneidad, 252, 259, 261–263, 266, 269, 273, 281
mutable, 259
no cambia, 259, 273
- función de Lyapunov, 60
- función de núcleo, 108, 126
Pearson VII, 109, 126, 130, 132, 136, 171, 175, 176, 182, 184, 187, 192, 195, 200, 203, 209, 210, 213, 219, 223, 230, 233, 239, 243
polinomial, 109, 126, 130, 171, 213, 219, 230, 243
polinomial normalizado, 109, 126, 128, 136, 171, 187, 203, 213, 219, 230, 239, 243
radial basis function, 109, 126, 171, 178, 182, 187, 195, 203, 210, 213, 219, 230, 233, 239, 243
- función objetivo, 29, 49, 50, 58, 60, 67, 98
- fundición, 75, 76, 79–81, 83, 85, 86, 123, 124, 128, 134, 136, 137, 144, 148, 153, 188, 225, 261, 272, 278, 285, 290, 292
características, 80
defectos, 80
etapas, 84
variables del proceso, 80
- fusión, 84, 85
- G**
- generación de nuevos individuos, 255, 256
- generador de nuevos modelos, 18
- geometría, 86
- GPGPU, 292
- grading, 115, 119–122, 127, 139, 143, 144, 146, 263, 269, 272, 279
- H**
- hipótesis, 13, 14, 16, 19
- horizonte de predicción, 52
- I**
- identificación de la tecnología, 41
- identificación de modelos lineales, 41
- identificación de modelos no lineales, 41
- identificación del modelo, 42
- incertidumbre, 31, 64, 72, 287
- infraestructuras, 292
- ingeniería de control industrial, 28
- inicialización, 255
- inspecciones
a posteriori, 80
destructivas, 87, 89
- inteligencia colectiva, 114
- K**
- k vecinos más próximos, 9, 76, 102, 104, 125, 127, 130, 134, 136, 144, 171, 195, 203, 213, 219, 223, 230, 239, 242
- K2, 97–99, 164, 171, 192, 210, 229, 231, 237
- L**
- líneas futuras, 277, 288
- la mayoría decide, 115, 116, 126, 139, 142, 143, 146

ÍNDICE ALFABÉTICO

- lector de evidencias, 16
ley de los MPC, 34
limitaciones, 7, 13, 70, 273, 277, 281, 287
- M**
- mínimos cuadrados parciales, 44
máquinas de soporte vectorial, 9, 76, 106, 109, 126, 128, 130, 136, 171, 175, 176, 182, 185, 192, 195, 200, 203, 209, 210, 213, 219, 223, 230, 233, 239, 242, 243, 282
función de núcleo, 108, 126
método de la máxima verosimilitud, 100
método Gauss-Newton, 43
métodos de control, 29
métodos de Montecarlo, 154–158, 160–162, 167–170, 173, 180, 182, 187, 189, 197, 205, 212, 213, 215, 216, 219, 222, 225, 233, 236, 239, 243, 245, 280, 283, 287, 288
propósitos, 160
métricas, 21
múltiples funciones objetivo, 8, 11
maldición de la dimensionalidad, 101, 150
manual, 10
margen funcional, 106
mayoría simple, 115
mazarotas, 84
meta-clasificación, 94, 137, 143, 144, 210, 279, 286
meta-clasificador, 11, 119–121, 132, 134, 136, 141–144, 146, 148, 210, 265, 279
metales
ferrosos, 81–83
no ferrosos, 81
metalurgia
fabricación, 81
ferroaleaciones, 81
fundición, 81
granel de metales ferrosos, 81
hierro forjado, 81
lingotes, 81
mecanizado, 81
metales no ferrosos a granel, 81
metodología, 19, 123, 124, 128, 137, 155, 169, 172, 206, 226, 252, 261, 262, 264
microrrechupes, 84, 86–88, 124, 128, 132, 134, 136, 137, 143, 144, 146, 148, 169, 172, 178, 182, 188, 189, 205, 206, 210, 225–227, 235, 236, 245, 262, 263, 265, 269, 271–273, 278, 280–283
modelado, 41
modelo autorregresivo con entradas exógenas, 37
modelo de Bob-Jenkins, 37, 43
modelo de Laguerre, 43
modelo de predicción, 29
modelo de respuesta de paso finito, 38
modelo predictivo de control, 2, 3, 5, 8–10, 13, 15, 19, 27, 28, 34, 35, 53, 56, 57, 59, 64, 68, 70, 71, 73, 93, 94, 97, 113, 132, 137, 142, 143, 148–150, 153, 167, 168, 205, 215, 249, 251, 256, 262–267, 269, 270, 278–283, 285–287, 290, 291
académicos, 28
afinación, 34

- aplicaciones industriales*, 68
autoadaptativos, 8, 161, 167, 278, 285, 287
basado en agentes, 287, 291
basado en principios fundamentales, 35, 36, 41, 93, 153
comercial, 35
concepto general, 28
diseño, 34
empírico, 35, 36, 38, 93, 153
formulación, 32
híbrido, 13, 94, 115, 265, 278, 282, 285, 287
historia, 53
incertidumbre, 31
industriales, 28
limitaciones, 7, 70
lineal, 6, 13, 30, 35, 36, 41, 42, 68, 93, 113, 278, 282, 285
múltiples funciones objetivo, 8, 11
no lineal, 6, 8, 13, 30, 35, 36, 38, 42, 60, 68, 70, 93, 113, 132, 278, 281, 282, 285
principios fundamentales, 35
robusto, 8, 31
siguiente generación, 7, 10, 73
trayectoria ideal, 2
modelos subóptimos, 50, 71
moldeado, 86
moldeo, 85
moldes de arena, 84, 85
MPC comerciales
DMC, 54, 56, 70
DMCPlus, 54, 59
HIECON, 54, 58
IDCOM, 54–56, 70
IDCOM-M, 54, 57
DMC, 56
LQG, 54
QDMC, 54, 57
RMPCT, 43, 54, 59
SMCA, 54, 58
SMOC, 54, 58, 71
multiesquema, 115, 122, 123, 127, 139, 141, 143, 146, 279
multilayer perceptron, 104, 125, 171
mutation rate, 257
- ## N
- números pseudo-aleatorios*, 158, 159, 162, 163, 167
características, 158
niveles de riesgo, 88
NMPC industrial
modelo de entradas/salidas, 66
modelo de estado-espacio, 65
modelo de principios fundamentales, 67
optimización de los estados estacionarios y dinámicos, 67
realimentación de la salida, 66
NMPC teórico
contractivo, 62
de horizonte cuasi-infinito, 63
de modo dual, 63
estabilidad, 60
horizontes finitos, 61
restricción de igualdad de estado terminal a cero, 62
robusto, 64
NPC, 134
nube de conocimiento, 13, 148, 278, 285
- ## O
- objetivos*
específicos, 13, 14, 284
operacionales, 13, 15, 16, 284

ÍNDICE ALFABÉTICO

principales, 13, 14, 284
objetivos de control, 34
operaciones genéticas
 cruce, 253, 257
 mutación, 253, 257
optimización de los resultados, 11

P
pangenetismo, 274
parametrización de entradas, 52
 en base a una función, 53
 movimiento único, 52
 movimientos múltiples, 52
particionamiento recursivo, 110
patrones, 84
PC, 134
PCA, análisis de componentes principales, 44
pluralidad, 115, 116
porcentaje de precisión, 21
potencial de nucleación, 86
predicción de características, 14, 15, 86, 284
predicción de defectos, 14, 15, 284
predicción de situaciones futuras, 10
predicción multivariable, 14
predictores, 18
preprocesador, 16
principios fundamentales, 35
problema de horizonte finito, 33
problema de optimización de bucle abierto, 33
proceso, 264
proceso productivo, 15, 42, 93, 94, 124, 128, 137, 144, 149, 154, 187, 250, 256, 260, 262, 274, 281, 285, 286
procesos productivos, 1, 3, 12, 79, 86
 fundición, 79–81

visión actual, 3
producto escalar, 261
programación concurrente, 254
programación evolutiva, 251, 252, 263–265, 273, 283, 290
 algoritmo de las abejas, 9, 76
 algoritmo genético, 9, 76, 251–255, 259, 261, 263–266, 273, 274, 281, 290
 paralela, 254, 264, 283
propiedades mecánicas, 86, 88–90
 ductilidad, 89
 dureza, 88
 elasticidad, 89
 fragilidad, 89
 fuerza, 88
 maleabilidad, 89
 plasticidad, 89
 resistencia, 89
 tenacidad, 89
protocolo de pruebas, 41
pruebas del modelo, 127, 172
punto de referencia, 51
puntos de coincidencia, 52

R
ratificación de defectos, 84
rayos X, 87
realimentación, 11, 249–252, 256, 261, 262, 264, 265, 269, 270, 281, 290
redes bayesianas, 9, 75, 96–99, 101, 125, 128, 130, 132, 154–157, 161–163, 166, 171, 173, 191, 210, 220, 222, 231, 237, 240, 242, 243, 245, 279, 283, 288
 aprendizaje estructural, 97, 125
 aprendizaje paramétrico, 97
redes neuronales artificiales, 8, 39, 40,

43, 75, 94, 104, 125, 134, 136,
171, 203, 213, 219, 223, 230,
242, 251
multilayer perceptron, 104, 125,
171
retropropagación, 105, 125, 171
reducción de datos, 150
regla de decisión bayesiana, 116, 117
regla de la media, 117, 126, 141
regla de la probabilidad mínima, 119,
127, 141, 144
regla de la probabilidad máxima, 118,
127, 141
regla de la suma, 117, 118
regla del producto, 116, 117, 119, 126,
141, 144
relaciones causa-consecuencia, 96
relevancia, 12
representación polinomial, 43
restricción, 250, 256
restricciones de estabilidad, 62
retos, 10, 13
adaptación continua, 11
adquisición de datos, 10
optimización de los resultados, 11
predicción de situaciones futuras,
10
realimentación del proceso, 11
retropropagación, 105, 125, 171
RMPCT, 43
robusto, 8, 31, 72, 101, 111, 130, 148,
273
ruido, 10, 42, 137, 149, 158, 173, 175,
180, 253, 270, 280, 287–289
ruta óptima, 250, 270

S

SCADA, 10, 14, 149, 289
selección, 258

selección de características, 150
selección de instancias, 150
selección de la ruleta, 258, 263
selección proporcional a la idoneidad,
258
siguiente generación, 7, 10, 73
similitud del coseno, 261, 281
sistema actualizado frente a los cam-
bios, 16, 284
sistema experto, 18
sistemas de control, 31
situaciones críticas, 42
SMOTE, 246
stacked generalization, 122
stacking, 115, 121, 122, 127, 139, 141,
143, 263, 265, 272, 279

T

técnicas de simulación, 2, 80
tasas de error
área bajo la curva ROC, 22, 172,
173, 184, 187, 188, 192, 195,
197, 200, 203, 205, 207, 210,
213, 215, 216, 220, 223, 226,
231, 235, 236, 240, 245
MAE, error absoluto medio, 21,
127, 128, 130, 136, 137, 141–
144, 146
*RMSE, raíz del error cuadrado
medio*, 22, 127, 128, 130, 136,
137, 142, 143, 146
tecnologías de control, 44
*determinación del camino de op-
timización*, 49, 249, 250
determinar el estado óptimo, 47,
249, 250
estimación del estado, 46
formulación de restricciones, 50

ÍNDICE ALFABÉTICO

horizonte de salida y parametrización de entradas, 52
trayectorias de entrada y salida, 51, 249

teoría de la evolución de Darwin, 252
teorema de Bayes, 96
tiempo de ejecución, 264, 267, 269, 272, 274, 288, 290
toma de decisiones, 14, 284
transformada de Laplace, 43
tratamiento térmico, 84, 86
trayectoria de referencia, 29, 51, 55
trayectoria ideal, 2, 27

U

ultrasonidos, 87
unanimidad, 115, 116

V

validación cruzada, 119, 124, 171, 197, 235, 236
validación empírica, 123, 261, 262, 278
valoración, 255, 259
variables del proceso, 80, 85
relacionadas con el metal, 85, 86
relacionadas con el molde, 85, 86
vectores de soporte, 107
velocidad de enfriamiento, 86
vertido, 86
visión artificial, 289

Z

zona, 51