

TESIS DOCTORAL

Desarrollo de técnicas de minería
de datos en procesos industriales

Modelización en líneas
de producción de acero

Ana González Marcos



UNIVERSIDAD DE LA RIOJA

TESIS DOCTORAL

Desarrollo de técnicas de minería
de datos en procesos industriales

Modelización en líneas
de producción de acero

Ana González Marcos

Universidad de La Rioja
Servicio de Publicaciones
2007

Esta tesis doctoral, dirigida por los doctores D. Joaquín B. Ordieres Mere y D. Eliseo P. Vergara González, fue leída el 31 de julio de 2006, y obtuvo la calificación de Sobresaliente Cum Laude Unanimidad

© Ana González Marcos

Edita: Universidad de La Rioja
Servicio de Publicaciones

ISBN 978-84-690-5501-4

UNIVERSIDAD DE LA RIOJA

DEPARTAMENTO DE INGENIERÍA MECÁNICA

DESARROLLO DE TÉCNICAS DE MINERÍA DE
DATOS EN PROCESOS INDUSTRIALES:
MODELIZACIÓN EN LÍNEAS DE PRODUCCIÓN
DE ACERO

Una disertación dirigida por el

PROF. DR. JOAQUÍN B. ORDIERES MERÉ

y por el

PROF. DR. ELISEO P. VERGARA GONZÁLEZ

y desarrollada por

ANA GONZÁLEZ MARCOS

a fin de optar al grado de

DOCTORA POR LA UNIVERSIDAD DE LA RIOJA

Tesis Doctoral defendida en la ciudad de Logroño

en Julio de 2006

Resumen

La minería de datos puede definirse como el proceso de extraer conocimiento útil y comprensible, previamente desconocido, a partir de grandes volúmenes de datos. En el ámbito industrial, una de las aplicaciones más interesantes del proceso de minería de datos es el modelado de sistemas.

El rápido crecimiento en la capacidad para almacenar datos que están experimentando los procesos industriales actuales, así como el desarrollo de los procesadores, proporciona nuevas posibilidades para analizar su comportamiento. Teniendo en cuenta, además, que en la mayoría de los procesos industriales, las relaciones entre variables no son lineales y la dificultad derivada de obtener modelos explícitos que definan su comportamiento, se comprende la importancia de los modelos basados en datos frente a otros modelos analíticos basados en ecuaciones explícitas.

Hoy en día, una de las herramientas más empleadas en la industria en el modelado de sistemas, por su eficiencia y simplicidad, son las redes neuronales, eje central sobre el que se desarrolla esta tesis. En ella, se propone el uso de estas técnicas, junto con otras procedentes del campo de la minería de datos, para el modelado de un proceso industrial real: una línea de acero galvanizado. En concreto, se proponen mejoras en los sistemas de control actuales mediante el desarrollo, a partir de los datos procedentes del proceso de fabricación, de modelos para predecir *on-line* las propiedades mecánicas de las bobinas de acero galvanizado, por un lado, y, de un modelo de velocidad de la banda dentro del horno de proceso, por otro.

Desafortunadamente, por las condiciones existentes en las industrias (interferencias electromagnéticas, picos de corriente en el arranque de motores, el factor humano, etc.), los datos erróneos son muy probables entre los datos almacenados. Para minimizar la influencia perjudicial del ruido presente en los datos en el entrenamiento de redes neuronales, se ha desarrollado un nuevo algoritmo de aprendizaje robusto, basado en el estimador τ no lineal y que emplea el algoritmo de aprendizaje de retropropagación del error, el cual supone una innovación en las denominadas redes neuronales robustas.

Agradecimientos

En primer lugar me gustaría dar las gracias a mis directores de tesis, Joaquín B. Ordieres Meré y Eliseo P. Vergara González. A Joaquín por haberme *contagiado* el entusiasmo por la investigación, por su guía y sabios consejos y por su constante apoyo y aliento durante la elaboración de esta tesis. A Eliseo por su disposición y optimismo, por su inestimable ayuda y por su ánimo y buenos consejos.

Y, como no, a mis compañeros del Área de Proyectos de Ingeniería, Alpha, Javi, Fernando y Manolo (quién, además, es un gran compañero de *batallas* en la Universidad de León), con quienes siempre se aprende algo nuevo y con quienes siempre se puede contar; a Antonio Ciampi, por todo lo que de él aprendí dentro y fuera de la universidad, durante mis estancias en la Universidad de McGill; a todo el personal de Aceralia por su disponibilidad, accesibilidad y eficacia durante todo el proceso de elaboración de esta tesis; y a mis actuales compañeros en la Universidad de León por hacerme más fácil el estar lejos de *casa*.

Por último, pero no en último lugar, quiero agradecerle a mi familia su apoyo incondicional, su comprensión, su cariño ... Ellos son quienes me *soportan* día a día, animándome como nadie.

Esta tesis no habría sido posible sin vuestra ayuda y saber hacer.

A todos,

GRACIAS

Índice general

1. Introducción	1
1.1. Antecedentes	1
1.2. Motivación y objetivos	5
1.3. Contribuciones	6
1.3.1. Publicaciones	6
1.3.2. Unidad temática de la tesis	7
1.4. Estructura del documento	7
2. Redes neuronales en el modelado de procesos industriales	9
2.1. Introducción	9
2.2. Algoritmo de retropropagación del error	11
2.2.1. Variantes del algoritmo de retropropagación	13
2.3. Algoritmo de retropropagación TAO-robust	16
2.3.1. Estimador τ no lineal	16
2.3.2. Implementación del estimador τ no lineal en el algoritmo de retropropagación del error	20
3. Modelado de una línea de producción de acero galvanizado	23
3.1. Introducción	23
3.2. Factores que influyen en las propiedades mecánicas del acero	24
3.2.1. Composición química	25
3.2.2. Tratamiento térmico	29
3.2.3. Proceso productivo	31
3.3. Modelado de las propiedades mecánicas de las bobinas	36

3.4. Modelado de velocidad de la banda en el horno de recocido . . .	39
4. Resultados y discusión	43
4.1. Aplicación de técnicas de minería de datos en el modelado de una línea de acero galvanizado	43
4.2. Desarrollo de un nuevo algoritmo de entrenamiento robusto . . .	45
5. Conclusiones	49
5.1. Conclusiones	49
5.2. Líneas de trabajo futuras	51
Bibliografía	53
I Publicaciones	57

Índice de figuras

1.1. Fases del modelo de referencia CRISP-DM (Fuente: Chapman et al., 2000).	2
2.1. Esquema de un perceptrón multicapa con una capa oculta.	10
3.1. Ejemplo de perfil de temperaturas en un ciclo de recocido.	30
3.2. Proceso de fabricación del acero.	32
3.3. Esquema general de una línea de acero galvanizado.	34
3.4. Modelo neuronal para la predicción de las propiedades mecánicas.	39
3.5. Modelo neuronal para la velocidad de la banda en el horno de recocido.	42
4.1. Evolución del límite elástico en función de la velocidad en las muestras recogidas para la <i>clase1</i>	45
5.1. Ejemplo de mapa del límite elástico estimado para una bobina.	50

Notación

α	velocidad de aprendizaje o tamaño de la actualización de los pesos
$f(\cdot)$	función de activación de las neuronas de la capa oculta
$f_s(\cdot)$	función de activación de las neuronas de la capa de salida
$h(v)$	función real
μ	número de patrón empleado
n	número total de patrones
ψ	función de influencia
\vec{r}	vector de residuos
S	factor de escala
t	número de época
t_k	valores objetivo de salida de la red
θ'_k	umbrales de las neuronas de la capa de salida
θ_j	umbrales de las neuronas de la capa oculta
τ_n	estimador de escala τ
W	matriz de pesos (w) y bias (θ)
w'_{kj}	pesos de las neuronas de la capa de salida
w_{ji}	pesos de las neuronas de la capa oculta x_i entradas de la red (con $i = 1, \dots, p$)
y_k	salidas de la red (con $k = 1, \dots, q$)
z_j	salidas de la capa oculta de la red (con $j = 1, \dots, l$),

Capítulo 1

Introducción

1.1. Antecedentes

Las plantas industriales, más allá de subsistir, buscan ser líderes en un ambiente competitivo, dinámico, diversificado e innovador, siendo precisamente la innovación tecnológica el instrumento con el que cuentan para lograr este objetivo. Una característica común de los procesos industriales actuales es el constante y rápido crecimiento de su capacidad para almacenar datos, es decir, día a día se dispone de mayores volúmenes de históricos que contienen información acerca de dichos procesos productivos. Es cierto que la información reduce la incertidumbre y, por tanto, permite tomar mejores decisiones. Sin embargo, al aumentar la cantidad de datos almacenados, la capacidad para asimilarlos disminuye, por lo que se hace necesario el uso de herramientas que permitan extraer *conocimiento útil* a partir de grandes conjuntos de datos. Es aquí donde tiene cabida la minería de datos (*Data Mining*), un área de investigación que pretende dar respuesta a esa necesidad de procesar y analizar 'masas' de datos con el fin último de encontrar y aprovechar el conocimiento útil contenido en ellos.

La minería de datos es, esencialmente, un proceso conducido por un problema: la respuesta a una pregunta o la solución a un problema se busca analizando los datos disponibles. El análisis de los datos forma el núcleo de la minería de datos, pero el proceso completo abarca también temas tales como la definición del problema empresarial y el desarrollo de la solución para resolverlo.

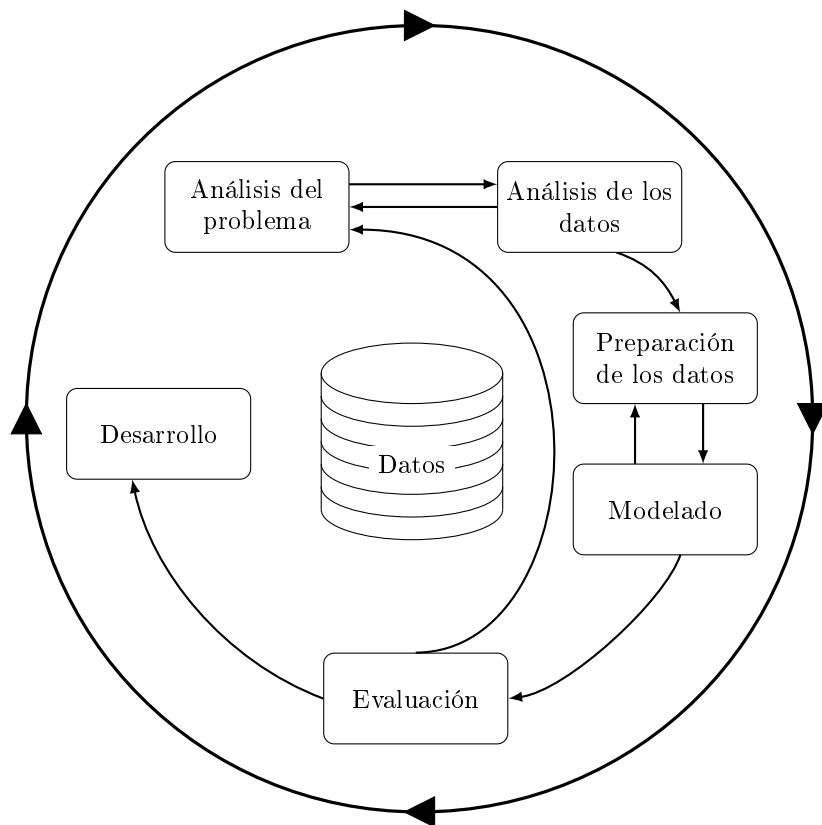


Figura 1.1: Fases del modelo de referencia CRISP-DM (Fuente: Chapman et al., 2000).

Para facilitar el proceso de minería de datos, ha habido diferentes intentos por formalizarlo dividiéndolo en un determinado número de fases secuenciales (Brachman and Anand, 1996; Chapman et al., 2000; Fayyad et al., 1996; Pyle, 1999). Aunque los nombres y contenido de estas fases difieren ligeramente, la misma idea general está presente en todos los modelos: primero el analista se familiariza con el problema y los datos; posteriormente, se preparan los datos y se construyen y evalúan los modelos para, finalmente, consolidar el nuevo conocimiento adquirido y utilizarlo para resolver el problema planteado.

Uno de los modelos del proceso de minería de datos más avanzado es el CRISP-DM (*CRoss Industry Standard Process model for Data Mining*) (Chapman et al., 2000), ilustrado en la figura 1.1, desarrollado y respaldado por un

consorcio de algunas de las mayores compañías en la industria de la minería de datos¹, el cual abarca todo el ciclo de vida de un proyecto de minería de datos. En la metodología CRISP-DM, el proceso de minería de datos se divide en seis fases: análisis del problema, análisis de los datos, preparación de los datos, modelado, evaluación y desarrollo:

- *Análisis del problema.* En primer lugar, el analista debe familiarizarse con el dominio del problema. Este conocimiento es importante en todas las fases de la minería de datos, ya que es imposible tomar decisiones sin tener un conocimiento adecuado de lo que es interesante, sorprendente o relevante con respecto al problema que se está intentando resolver. Sin el conocimiento necesario, el analista estará buscando a ciegas.
- *Análisis de los datos.* La comprensión de los datos incluye entender su origen, naturaleza y fiabilidad, así como familiarizarse con el contenido de los mismos a través de un análisis exploratorio. La adecuada preparación de los datos, selección de las herramientas de modelado y evaluación sólo es posible si el analista tiene una buena idea global, un modelo mental, de los datos.
- *Preparación de los datos.* El objetivo fundamental de esta fase es facilitar la labor de construir modelos precisos y fiables, corrigiendo errores y extrayendo nuevas características. La preparación de los datos es un tema difícil y disperso y, es tan dependiente de su aplicación, que sólo pueden darse algunas líneas generales (Pyle, 1999).
- *Modelado.* Aquí es donde se buscan soluciones al problema. Mientras que las fases previas básicamente preparaban los datos para el modelado y las posteriores se preocupan por su desarrollo en la práctica, la solución se especifica en esta fase. Esta solución puede ser un modelo predictivo o más descriptivo, como una segmentación o clasificación de los datos en un conjunto de grupos, el análisis de ciertas propiedades como dependencias entre variables, o un estimador de la función de densidad de probabilidad de los datos. Modelar es, quizá, el tema más ampliamente discutido en la literatura (veáse, por ejemplo, Bishop, 1995; Cherkassky and Mulier, 1998; Cios et al., 1998; Ripley, 1996).

¹Destacando SPSS, Inc. <http://spss.com>

- *Evaluación.* Antes de desarrollar la solución, es necesario evaluarla desde el punto de vista del problema de negocio original, para determinar si la solución encontrada es lo suficientemente buena como para ser desarrollada. Cabe señalar que, junto con la solución, el proceso de minería de datos genera nuevas percepciones, ideas y modelos secundarios, los cuales también son importantes con respecto al problema a resolver.
- *Desarrollo.* Finalmente, la solución se utiliza para resolver el problema original.

En la práctica, el proceso de minería de datos es altamente dinámico e iterativo, ya que cualquier fase puede suscitar preguntas o ideas que necesitan ser investigadas o implementadas en una fase previa, y dependiente del problema planteado, de la disponibilidad de las fuentes de datos, del conocimiento de las herramientas necesarias, de la metodología desarrollada y de los requerimientos y recursos de la empresa (Martínez de Pisón Ascacibar, 2003).

Una de las aplicaciones más interesantes del proceso de minería de datos en el ámbito industrial es el modelado de sistemas². El hecho de que en la mayoría de los procesos industriales, las relaciones entre variables sean no lineales, y la dificultad derivada de obtener modelos explícitos que definan su comportamiento, hace que se piense en recurrir a modelos basados en datos frente a otros modelos analíticos basados en ecuaciones explícitas. Hoy en día, el modelado no lineal cuenta con importantes técnicas que han alcanzado gran aplicabilidad gracias al creciente desarrollo de los procesadores. Entre estas técnicas se pueden mencionar los algoritmos genéticos, la lógica difusa, etc., si bien, una de las herramientas más empleadas en la industria, por su eficiencia y simplicidad, son las redes neuronales, eje central sobre el que se desarrolla esta tesis.

Desafortunadamente, por las condiciones existentes en las industrias (interferencias electromagnéticas, picos de corriente en el arranque de motores, el factor humano, etc.), los datos erróneos, definidos en la literatura como espurios (*outliers*), son muy probables entre los datos almacenados. Así, por ejemplo, Hampel et al. (1986) afirman que los datos rutinarios contienen entre el 1 y el 10 % de errores graves y que hasta en los registros de mejor calidad no puede garantizarse la ausencia de errores. La presencia de datos atípicos en el conjun-

²La relevancia de esta actividad radica en la posibilidad que ofrece un buen modelo de diseñar leyes de control, de predecir parámetros del proceso productivo de especial interés, de detectar fallos en algún componente del proceso, etc.

to de datos causa un peor ajuste, en ocasiones lejano del óptimo, del modelo obtenido, por lo que es muy importante tratarlos de algún modo para que no perjudiquen los modelos buscados. Una de las estrategias en este sentido, consiste en hacer uso de técnicas de diagnóstico de espurios con las que se detectan y eliminan los datos erróneos antes de proceder a la construcción del modelo (Castejón Limas et al., 2004; Rousseeuw and Leroy, 1987). Pero también, en un contexto más directo, evitando este tipo de preprocesado de datos, existe la posibilidad de emplear técnicas de modelado que sean robustas a espurios (Ahmed and Farag, 2002; Chen and Jain, 1994; Chuang et al., 2000; Huang et al., 1998; Liano, 1996; McDowall and Ham, 1997).

1.2. Motivación y objetivos

El análisis *inteligente* de los datos, abre un mundo de posibilidades completamente nuevo para muchas industrias ya que, aunque los ingenieros tienen cada vez un mayor acceso a los datos de proceso, raramente usan herramientas de estadística avanzada para resolver problemas de control del mismo. La causa más probable de esta reticencia, está en que el desarrollo de las herramientas estadísticas tuvo lugar en una época de riguroso modelado matemático, cálculo manual y pequeños conjuntos de datos. De esta forma, se crearon herramientas sofisticadas que requerían destreza teórica y cuyos requisitos, como el preprocesado de los datos, no eran entendidos por los ingenieros. Tanto si los datos no eran los apropiados para los algoritmos, como si éstos se parametrizaban mal, la fidelidad de los resultados obtenidos era muy baja, lo que pudo conducir a un rechazo del análisis estadístico en general.

Con esta tesis, en la que se recogen los frutos del trabajo realizado en el marco de diversos proyectos de investigación, se pretende, por un lado, mostrar la utilidad de las técnicas y herramientas de análisis de datos, en general, y de las redes neuronales, en particular, para la extracción de *conocimiento útil* a partir de los históricos de un proceso industrial; y, por otro, desarrollar un nuevo método de entrenamiento de redes neuronales que permita *aprender* de los datos de un modo más robusto y eficiente.

1.3. Contribuciones

1.3.1. Publicaciones

En Ordieres Meré et al. (2004), se presenta la aplicación del proceso de minería de datos en el modelado de un proceso industrial real. En concreto, el problema que se plantea es la predicción de las propiedades mecánicas de las bobinas de acero galvanizado, como una mejora de los sistemas de control de una línea de acero galvanizado. Puesto que, en la actualidad, las características del producto que se pretende fabricar no se pueden medir directamente, sino que deben llevarse a cabo ensayos en un laboratorio tras el proceso de fabricación, no es posible aplicar una estrategia de control clásica, siendo el control en lazo abierto la estrategia a adoptar. Sin embargo, con un estimador *on-line* de dichas propiedades mecánicas a partir de los datos procedentes del proceso de fabricación, sería posible introducir mejoras en los sistemas de control actuales. Los resultados obtenidos en el desarrollo de dichos modelos predictivos se presentan en esta publicación.

En Pernía Espinoza et al. (2005a), se persigue el mismo objetivo que en la publicación anterior: tratar de mejorar los sistemas de control de una línea de acero galvanizado. En este caso, se desarrolla un modelo de velocidad de la banda de acero dentro de un horno de recocido, a partir de los datos de proceso, para asegurar, mediante la regulación de velocidad, que la temperatura real de la banda a la salida de la zona de calentamiento del horno se aproxime a la deseada. De esta forma, sería posible mejorar el control del tratamiento térmico al que se someten las bandas de acero antes de su inmersión en el pote de zinc, ya que áquel es un proceso clave para la obtener las propiedades de la banda deseadas y una buena adherencia del recubrimiento. Además, en esta publicación, se muestra la aplicación en el modelado de procesos industriales, y buen comportamiento frente a espurios, de un algoritmo de aprendizaje robusto.

Por último, en Pernía Espinoza et al. (2005b), se presenta un nuevo algoritmo de aprendizaje robusto, basado en el estimador τ no lineal y denominado algoritmo de aprendizaje por retropropagación *TAO-robust*, desarrollado para minimizar la influencia perjudicial del ruido presente en los datos en el entrenamiento de redes neuronales. Esta aproximación usa un estimador de escala variable y dependiente de la función de Huber de los errores obtenidos en cada época, con lo que se soluciona una de las principales deficiencias de otros métodos robustos: la selección del factor de escala que determina los límites para

reducir los efectos de los espurios. Además, esta aproximación también posee las importantes propiedades de robustez con un punto de ruptura elevado y una alta eficiencia con errores normalmente distribuidos.

1.3.2. Unidad temática de la tesis

Todos los trabajos presentados en esta tesis se enmarcan dentro del ámbito de la *minería de datos*:

- Las dos primeras publicaciones citadas en el apartado anterior, están enfocadas en la aplicación industrial de la minería de datos. En concreto, se proponen dos modelos, basados en redes neuronales, con el fin de introducir mejoras en una línea de producción de acero galvanizado.
- La última publicación de la sección anterior, se centra en el desarrollo de un nuevo algoritmo cuya finalidad consiste en obtener modelos neuronales que se ajusten mejor a la información contenida en los datos, al utilizar en su entrenamiento un criterio de error robusto a espurios.

1.4. Estructura del documento

En este capítulo se han presentado los antecedentes, la motivación y contribuciones aportadas con esta tesis, así como su unidad temática. El resto del documento está organizado de la siguiente manera. En el Capítulo 2, se introduce el algoritmo de aprendizaje de retropropagación del error (utilizado a lo largo de toda la tesis), se analiza el criterio de error basado en el estimador τ no lineal y se muestra el modo de implementar éste en áquel, obteniéndose, de esta manera, el nuevo algoritmo de aprendizaje robusto, denominado algoritmo de aprendizaje por retropropagación *TAO-robust*. En el Capítulo 3 se introduce la problemática de los sistemas de control actuales en las líneas de fabricación de acero galvanizado y se presentan los trabajos realizados en este sentido: la predicción de las propiedades mecánicas de las bobinas de acero galvanizado y el modelo de velocidad del horno de recocido de la línea. Tras exponer la metodología utilizada en cada una de las publicaciones de la presente tesis, en el Capítulo 4, se discuten los resultados obtenidos. Esta tesis finaliza con las conclusiones del Capítulo 5.

Capítulo 2

Redes neuronales en el modelado de procesos industriales

2.1. Introducción

Para modelar procesos complejos, como es el caso de la mayoría de los procesos industriales, una alternativa eficiente y confiable a los modelos basados en ecuaciones explícitas, la constituyen las redes neuronales.

Una de las arquitecturas de redes más adecuada para la determinación de modelos de procesos, empleando datos de entrada y salida del mismo, es la llamada red perceptrón multicapa (*MultiLayer Perceptron*, MLP) (figura 2.1), por ser considerada un aproximador universal de funciones (Funahashi, 1989; Hornik et al., 1989). Más concretamente, una red MLP conteniendo al menos una capa oculta con suficientes unidades no lineales, puede aprender cualquier tipo de función o relación continua entre un grupo de variables de entrada y salida.

El algoritmo de entrenamiento por excelencia del perceptrón multicapa, es el algoritmo de retropropagación del error (*BackPropagation*, BP), que no es más que la generalización de la regla del Error Cuadrático Medio (*Least Mean Square*, LMS) al caso de perceptrones multicapa. Aunque en el entrenamiento de redes neuronales supervisadas se use, por tradición y sencillo computo, la media del error cuadrático (*Mean Squared Error*, MSE) como medida de error¹,

¹Algunas de las arquitecturas de redes y algoritmos de aprendizaje que emplean el MSE

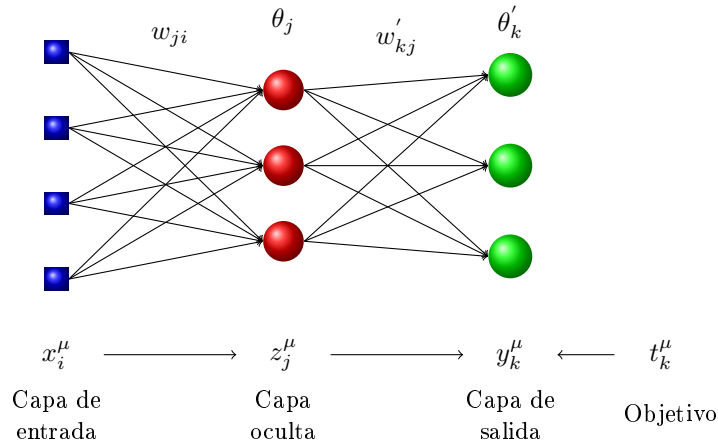


Figura 2.1: Esquema de un perceptrón multicapa con una capa oculta.

ésta no es adecuada cuando los patrones de entrenamiento están contaminados, ya que los modelos obtenidos pueden ser muy diferentes de los esperados (Chuang et al., 2000; Liano, 1996; Walczak, 1996).

En la literatura se encuentran diferentes tipos de redes neuronales robustas que pretenden mejorar el aprendizaje cuando los patrones de entrenamiento contienen datos erróneos (Ahmed and Farag, 2002; Chen and Jain, 1994; Chuang et al., 2000; Huang et al., 1998; Liano, 1996; McDowall and Ham, 1997). Estas aproximaciones poseen dos propiedades importantes: robustez con un elevado punto de ruptura y una alta eficiencia con errores normalmente distribuidos (Rousseeuw and Leroy, 1987). Sin embargo, en su uso, existen algunas deficiencias. Una de ellas se encuentra en la selección del factor de escala que determina los límites para reducir los efectos de los espurios: si es demasiado grande, los espurios pueden no ser detectados; por el contrario, si es demasiado pequeño, una gran cantidad de datos, incluyendo los *buenos*, serían discriminados. La selección de este factor no es clara, siendo diversas las propuestas que se encuentran en la literatura. Por ejemplo, (Connor et al., 1994; Hawkins, 1980; Huber, 2003; Rousseeuw and Leroy, 1987) proponen seleccionar el estimador de escala basándose en un estadístico robusto, como la mediana del error o la desviación mediana absoluta (*Median Absolute Deviation*, MAD). Ambos están

son: ADALINE (ADaptive LInear NEuron), perceptrón, retropropagación (Rumelhart et al., 1986), quickpropagation (Fahlman, 1988), RAN (*Resource Allocating Network*) (Platt, 1991), RBF (*Radial Basis Function*) (Moody and Darken, 1989), y la SuperSAB (Tollenaere, 1990).

basados en la medida del error, pero, en la primera fase de entrenamiento, esta medida no refleja el error real ya que se desconocen los datos que son espurios y aquellos que no lo son. En este caso, ambas medidas del error pueden ser, en realidad, no significativas. Otro modo de seleccionar el factor de escala (Chen and Jain, 1994) se basa en el conocimiento previo del porcentaje aproximado de espurios en el conjunto de datos, lo cual también es difícil de saber (Chuang et al., 2000).

Con la idea de solucionar este problema, en la publicación (Pernía Espinoza et al., 2005b), se presenta un nuevo algoritmo de aprendizaje robusto, denominado algoritmo de aprendizaje por retropropagación *TAO-robust*, en el cual se usa un estimador de escala variable y dependiente de la función de Huber de los errores obtenidos en cada época. Además, esta aproximación también posee las importantes propiedades de robustez con un punto de ruptura elevado y una alta eficiencia con errores normalmente distribuidos.

El resto de este capítulo está organizado de la siguiente manera: en primer lugar se introduce el algoritmo de retropropagación del error para, posteriormente, analizar el criterio de error basado en el estimador τ no lineal y el modo de implementarlo en áquel, con el propósito de obtener un nuevo algoritmo de aprendizaje robusto.

2.2. Algoritmo de retropropagación del error

Básicamente, el algoritmo de aprendizaje por retropropagación del error consta de dos pasadas a través de las diferentes capas de la red: una hacia adelante y otra hacia atrás. En la primera, se presenta un patrón de entrada (x_i^μ) a la primera capa de la red y se propaga hacia adelante a través de los pesos (W), capa por capa, hasta que se alcanzan las neuronas de la capa de salida. Durante esta fase, los pesos de la red permanecen fijos. Sin embargo, durante la pasada hacia atrás, estos pesos se ajustan con una regla de corrección del error. Concretamente, la señal de salida de la red (y_k^μ) se compara con el patrón de salida deseado (t_k^μ) y se calcula una señal de error (según un determinado criterio E), que se propaga hacia atrás, en dirección contraria a las conexiones de la red (de ahí el nombre de retropropagación del error), haciendo que los pesos se ajusten de tal forma que la salida de la red se aproxime a la respuesta deseada, es decir, que se minimice la señal de error.

De forma similar al algoritmo LMS, el algoritmo BP lleva a cabo la minimización del error mediante el *descenso del gradiente*, donde los pesos de la capa de salida y de la capa oculta se modifican aplicando una corrección Δw , proporcional a la derivada parcial de la señal del error respecto a los pesos:

$$w'_{kj} = w'_{kj} + \Delta w'_{kj} = w'_{kj} - \alpha \frac{\partial E(W)}{\partial w'_{kj}} \quad (2.1)$$

$$w_{ji} = w_{ji} + \Delta w_{ji} = w_{ji} - \alpha \frac{\partial E(W)}{\partial w_{ji}} \quad (2.2)$$

Para un determinado patrón de entrada, la salida de la neurona j -ésima de la capa oculta se calcula según la ecuación 2.3, donde $f(\cdot)$ es la función de activación de las neuronas en dicha capa, y la respuesta de la red para la neurona k -ésima de la capa de salida se calcula según la ecuación 2.4, donde $f_s(\cdot)$ es la función de activación de las neuronas en la capa de salida².

$$z_j^\mu = f \left(\sum_i w_{ji} x_i^\mu - \theta_i \right) = f(\gamma_j^\mu), \quad j = 1, 2, \dots, l \quad (2.3)$$

$$y_k^\mu = f_s \left(\sum_j w'_{kj} z_j^\mu - \theta_k^\mu \right) = f_s(v_k^\mu), \quad k = 1, 2, \dots, q \quad (2.4)$$

De esta forma, utilizando la regla de la cadena, las actualizaciones de los pesos de la capa de salida y oculta, respectivamente, se llevan a cabo según las siguientes expresiones:

$$\Delta w'_{kj} = \alpha \delta_k z_j^\mu \quad \text{con } \delta_k = - \frac{\partial E}{\partial y_k^\mu} \frac{\partial f_s(v_k^\mu)}{\partial v_k^\mu} \quad (2.5)$$

$$\Delta w_{ji} = \alpha \delta_j x_i^\mu \quad \text{con } \delta_j = \sum_k \left(\delta_k w'_{kj} \right) \frac{\partial f(\gamma_j^\mu)}{\partial \gamma_j^\mu} \quad (2.6)$$

En resumen, el algoritmo de retropropagación completo consta de las siguientes etapas:

1. Inicializar los pesos a valores pequeños aleatorios.

²Las funciones de activación de las distintas capas de la red pueden ser de tipo lineal ($F(x) = x$), sigmoideal ($F(x) = \frac{1}{1+e^{-x}}$), tangh ($F(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$), etc.

2. Escoger un patrón de entrada, x^μ , y presentarlo a la capa de entrada.
3. Propagar la activación hacia delante a través de los pesos hasta que la activación alcance las neuronas de la capa de salida.
4. Calcular los valores de δ para las capas de salida (δ_k) usando los valores objetivo para el patrón de entrada seleccionado.
5. Calcular los valores de δ para la capa oculta (δ_j).
6. Actualizar los pesos de la capa de salida y de la capa oculta, de acuerdo con las ecuaciones 2.5 y 2.6, respectivamente.
7. Repetir del paso 2 al 6 para todos los patrones de entrada.

Tomando como medida del error la media del error cuadrático (MSE)³, la actualización de los pesos queda de la siguiente manera:

$$\Delta w'_{kj} = \alpha (t_k^\mu - y_k^\mu) \frac{\partial f_s(v_k^\mu)}{\partial v_k^\mu} z_j^\mu \quad (2.7)$$

$$\Delta w_{ji} = \alpha (t_k^\mu - y_k^\mu) \frac{\partial f_s(v_k^\mu)}{\partial v_k^\mu} w'_{kj} \frac{\partial f(\gamma_j^\mu)}{\partial \gamma_j^\mu} x_i^\mu \quad (2.8)$$

2.2.1. Variantes del algoritmo de retropropagación

El algoritmo de retropropagación comentado anteriormente, es el primer algoritmo eficaz para el aprendizaje de redes MLP, aunque presenta algunos inconvenientes como su lentitud de convergencia, la posibilidad de incurrir en el denominado sobreajuste estropeando la capacidad de generalización de la red, su sensibilidad a espurios, etc. Para resolver algunos de estos inconvenientes continuamente se plantean correcciones o variantes. Buena parte de estas modificaciones tratan de resolver el problema de su lenta convergencia, mientras que otras se centran en conseguir una mejor generalización evitando el sobreajuste, por un lado, y tratando los espurios, por otro.

³El error total será $E = \frac{1}{2} \sum_k (t_k - y_k)^2$

En esta sección, se tratarán brevemente algunas de las variantes propuestas, indicando aquellas que han sido utilizadas en los trabajos desarrollados y presentados en esta tesis.

Aceleración del aprendizaje

El algoritmo BP básico no es conocido por su rapidez de convergencia y, modificar su velocidad de aprendizaje (α) para acelerar el proceso, no es una buena idea, ya que el efecto será el de introducir inestabilidad en la regla de aprendizaje causando oscilaciones violentas en los pesos aprendidos. Así, la primera variante propuesta para acelerar este método consiste en incluir en el algoritmo un término (β), denominado momento, al cambio de los pesos (según la ecuación 2.9), de tal manera que se incremente la actualización de los pesos si el nuevo cambio coincide con la dirección de los cambios anteriores y se reduzca en caso contrario (Rumelhart et al., 1986).

$$\Delta_t W = -\alpha \frac{\partial E_t}{\partial W} + \beta \Delta_{t-1} W \quad (2.9)$$

Otros métodos que permiten mejorar la velocidad de aprendizaje consisten en utilizar funciones de activación antisimétricas (como la tangente hiperbólica), preprocesar las entradas (normalizándolas y haciéndolas independientes entre sí (LeCun, 1993)), seleccionar adecuadamente los pesos iniciales de la red, asignar un ritmo de aprendizaje diferente a cada peso, ritmos adaptativos, etc. (Haykin, 1999; LeCun et al., 1998).

Ya se ha visto que el algoritmo básico del BP ajusta los pesos en la dirección de máxima pendiente de la superficie del error, dirección en la que la función de coste decrece más rápidamente. A pesar de ello, no siempre se garantiza que éste sea el camino más rápido hacia el mínimo. Para acelerar la convergencia existen los denominados métodos de segundo orden, que se basan en realizar el descenso utilizando también la información proporcionada por el ritmo de cambio de la pendiente, $H = \partial^2 E / \partial W^2$. Dentro de este grupo se encuentran los métodos de Newton y cuasi-Newton, los cuales suponen un gran esfuerzo computacional al tener que calcular, almacenar, e invertir en el método de Newton, la matriz Hessiana, H . Además, el método de Newton puede divergir violentamente si el algoritmo no está cerca de un mínimo, y la inversión de H puede presentar problemas de estabilidad numérica en algunos casos.

Con el propósito de evitar estos problemas y requerimientos computacionales se han desarrollado otros métodos, también de segundo orden, que se pueden

clasificar en un punto intermedio entre el método de descenso del gradiente y el de Newton: los algoritmos basados en la técnica del gradiente conjugado. Estos algoritmos pretenden conseguir una buena velocidad de convergencia con relación al número de operaciones que efectúan utilizando para ello un número de operaciones y memoria reducidos. Evitan el cálculo de H , utilizando una combinación lineal del gradiente actual y las direcciones de búsqueda previas (ecuación 2.10).

$$W_{t+1} = W_t + \alpha_t d_t \quad \text{con } d_t = -\frac{\partial E_t}{\partial W} + \beta_t d_{t-1} \quad \text{y } d_0 = -\frac{\partial E_0}{\partial W} \quad (2.10)$$

donde β_t es una función escalar que asegura que d_t y d_{t-1} son conjugados.

En concreto, el algoritmo de gradiente conjugado empleado en (Pernía Espinoza et al., 2005a) se denomina *actualización Fletcher-Reeves* (Fletcher, 1987), donde β_t se determina según la relación de la norma cuadrada del gradiente actual con la norma cuadrada del gradiente previo (ecuación 2.11).

$$\beta_t = \frac{\left\| \frac{\partial E_t}{\partial W} \right\|^2}{\left\| \frac{\partial E_{t-1}}{\partial W} \right\|^2} \quad (2.11)$$

Capacidad de generalización

Aunque es deseable alcanzar el mejor resultado posible durante el entrenamiento, uno de los aspectos fundamentales de una red neuronal es su capacidad de generalizar, es decir, de ofrecer buenos resultados con patrones de test que no han sido empleados en su entrenamiento. La generalización es importante no sólo para que la red funcione correctamente ante conjuntos de datos que no han sido vistos con anterioridad, sino también porque es posible que haya datos que contengan ruido, que estén distorsionados o sean incompletos.

La capacidad de generalización de una red está determinada por tres factores: el número de patrones de entrenamiento, la arquitectura de la red y la complejidad física del problema (Haykin, 1999). Estas tres circunstancias están muy relacionadas ya que, cuanto más complejo sea el problema a modelar, más grande deberá ser la red, es decir, se deberán ajustar más parámetros y, por tanto, se necesitarán más muestras para entrenarla. No teniendo un control sobre la complejidad del problema y, disponiendo habitualmente de un número limitado de patrones, en muchas ocasiones se corre el peligro de sobreajustar

la red al conjunto de datos de entrenamiento y obtener malos resultados con datos de test, si el número de parámetros de la red es muy grande. Además, hay que tener en cuenta el inconveniente de la “maldición de la dimensionalidad” (*curse of dimensionality*) (Bellman, 1961), que plantea que el número de datos necesarios para especificar una función, en general, crece exponencialmente con la dimensión del espacio de entrada.

Para evitar el sobreentrenamiento existen diferentes formas de actuar: la parada temprana, limitar el tamaño de la red, entrenar con ruido, etc.

En el método de parada temprana (utilizado tanto en Ordieres Meré et al. (2004) como en Pernía Espinoza et al. (2005a)), a la vez que se entrena la red, los resultados se validan con otro conjunto de datos, denominado de validación, debiendo detener el aprendizaje en el punto óptimo de mínimo error de validación. De esta forma, se pretende evitar que la red se adapte demasiado a las particularidades de los patrones empleados en el entrenamiento, aprendiendo incluso el ruido presente en ellos.

Algunas de las técnicas empleadas para limitar el tamaño de la red consisten en reducir el número de entradas (realizando un análisis de componentes principales, empleando ratios, etc.), para disminuir el número de parámetros ajustables; hacer que varias neuronas compartan sus pesos (*weight sharing*); podar la red (*pruning*), es decir, eliminar explícitamente ciertas conexiones; o asociar a cada peso una tendencia de disminución a cero para que desaparezcan aquellos que no son actualizados periódicamente (*weight decay*). Esta última técnica, el decaimiento de pesos, ha sido la utilizada en Ordieres Meré et al. (2004) para evitar el sobreajuste de la red, al no disponer de un número suficiente de patrones de entrenamiento en comparación con el número de variables de entrada.

Por último, el entrenamiento con ruido (*jitter*), consiste, como su propio nombre indica, en añadir una pequeña cantidad de ruido blanco a cada uno de los patrones utilizados. Así, cada vez que se presente el vector entrada-salida deseada al MLP, se añade un pequeño ruido aleatorio.

2.3. Algoritmo de retropropagación TAO-robust

2.3.1. Estimador τ no lineal

El estimador τ fue introducido por Yohai and Zamar (1988) como una nueva clase de estimador robusto para modelos de regresión lineales y, posteriormente,

generalizado por Tabatabai and Argyros (1993), al caso de regresión no lineal. En ambos casos, el punto de ruptura es de 0.5 y, ambos, poseen una alta eficiencia con los errores normalmente distribuidos.

La definición que se expone a continuación sigue de cerca la expuesta en (Tabatabai and Argyros, 1993).

Considérese el modelo de regresión no lineal $y_i = \eta(\vec{x}_i, \theta) + \varepsilon_i$, donde θ es el vector de los p parámetros desconocidos y ε_i es el error o residuo que se considera, según el modelado clásico, de media cero y varianza constante y conocida. Se asumirá, además, que la función $\eta(\vec{x}_i, \theta)$ es continuamente diferenciable dos veces en algún conjunto convexo C .

Sea $\vec{r}(\theta) = [r_1(\theta), \dots, r_n(\theta)]$ el conjunto de residuos definidos como $r_i(\theta) = y_i - \eta(\vec{x}_i, \theta)$, $i = 1, \dots, n$.

Supuesto 1. Sea h una función real que satisface las siguientes propiedades:

- (i) $h(0) = 0$
- (ii) $h(-v) = h(v)$
- (iii) Si $0 \leq v \leq \gamma$ entonces $h(v) \leq h(\gamma)$
- (iv) h es continuo.
- (v) Sea $a = \sup h(v)$; entonces $0 < a < \infty$
- (vi) Si $h(v) < \sup h(v)$ y $0 \leq v < \gamma$, entonces $h(v) < h(\gamma)$.

Huber (1981) definió el estimador M (el estimador de escala de un conjunto $\vec{v} = (v_1, \dots, v_n)$), denotado por $S(v)$, como

$$\sum_{i=1}^n h\left(\frac{v_i}{S(v)}\right) = nE_{\phi}[h(v)] \quad (2.12)$$

donde E_{ϕ} es la esperanza en la distribución normal estándar ϕ .

Rousseeuw and Yohai (1984) definieron al estimador S como el que minimiza la función de escala de un estimador M de los residuos. Es decir, es el estimador que minimiza $S(\vec{r}(\theta))$. En su trabajo plantean que este estimador pudiera tener un valor de punto de ruptura alto, pero al coste de tener una baja eficiencia en condiciones de normalidad. Para resolver este problema Yohai and Zamar (1988) definieron una clase de estimador de escala más amplio, el estimador τ , el cual alcanza simultáneamente un punto de ruptura alto y

una alta eficiencia ante condiciones de normalidad. Posteriormente, Tabatabai and Argyros (1993) propusieron una generalización de este estimador τ al caso de regresión no lineal, preservando todas las propiedades del primero, y cuya definición se presenta a continuación.

Sean h_1 y h_2 dos funciones que cumplen las propiedades (i)-(vi) del Supuesto 1, y sea S el estimado-M de escala basado en h_1 . Entonces dado un conjunto $\vec{v} = (v_1, \dots, v_n)$, el estimador de escala τ_n se define como

$$\tau_n(\vec{v}) = S(\vec{v}) \left[\frac{1}{n} \sum_{i=1}^n h_2 \left(\frac{v_i}{S(\vec{v})} \right) \right]^{1/2} \quad (2.13)$$

La estimación τ de los parámetros de regresión θ se define entonces como el vector resultante $\hat{\theta}$ que cumpla con

$$\tau_n(\vec{r}(\hat{\theta})) = \text{Min}_{\theta \in C} \tau_n(\vec{r}(\theta)) \quad (2.14)$$

$$\tau_n(\vec{r}(\theta)) = S(\vec{r}(\theta)) \left[\frac{1}{n} \sum_{i=1}^n h_2 \left(\frac{r_i(\theta)}{S(\vec{r}(\theta))} \right) \right]^{1/2} \quad (2.15)$$

y $S(\vec{r}(\theta))$ satisface

$$\sum_{i=1}^n h_1 \left(\frac{r_i}{S(\vec{r}(\theta))} \right) = nE_\phi(h_1) \quad (2.16)$$

El mínimo de esta función se consigue derivando 2.15 respecto a θ , e igualando luego a cero el resultado

$$\begin{aligned} \frac{\partial \tau_n^2(\vec{r}(\theta))}{\partial \theta} &= \frac{1}{n} \sum_{i=1}^n \left\{ 2S(\vec{r}(\theta)) \frac{\partial S(\vec{r}(\theta))}{\partial \theta} h_2 \left(\frac{r_i(\theta)}{S(\vec{r}(\theta))} \right) + \right. \\ &+ \left. \psi_2 \left(\frac{r_i(\theta)}{S(\vec{r}(\theta))} \right) \left(-S(\vec{r}(\theta)) \frac{\partial \eta(\vec{x}_i, \theta)}{\partial \theta} - r_i(\theta) \frac{\partial S(\vec{r}(\theta))}{\partial \theta} \right) \right\} = 0 \end{aligned} \quad (2.17)$$

Por otro lado, derivando 2.16 se obtiene

$$\frac{\partial S(\vec{r}(\theta))}{\partial \theta} = - \frac{\sum_{i=1}^n \psi_1 \left(\frac{r_i(\theta)}{S(\vec{r}(\theta))} \right) \frac{\partial \eta(\vec{x}_i, \theta)}{\partial \theta}}{\sum_{i=1}^n \psi_1 \left(\frac{r_i(\theta)}{S(\vec{r}(\theta))} \right) \frac{r_i(\theta)}{S(\vec{r}(\theta))}} \quad (2.18)$$

donde

$$\psi_j(v) = \frac{d}{dv} h_j(v), \quad j = 1, 2$$

Sustituyendo 2.18 en 2.17 se obtiene la ecuación de la estimación τ

$$\sum_{i=1}^n \left[w(\theta) \psi_1 \left(\frac{r_i(\theta)}{S(\bar{r}(\theta))} \right) + \psi_2 \left(\frac{r_i(\theta)}{S(\bar{r}(\theta))} \right) \right] \frac{\partial \eta(\bar{x}_i, \theta)}{\partial \theta} = \vec{0} \quad (2.19)$$

donde

$$w(\theta) = \frac{\sum_{i=1}^n \left[2h_1 \left(\frac{r_i(\theta)}{S(\bar{r}(\theta))} \right) - \psi_2 \left(\frac{r_i(\theta)}{S(\bar{r}(\theta))} \right) \frac{r_i(\theta)}{S(\bar{r}(\theta))} \right]}{\sum_{i=1}^n \psi_1 \left(\frac{r_i(\theta)}{S(\bar{r}(\theta))} \right) \frac{r_i(\theta)}{S(\bar{r}(\theta))}} \quad (2.20)$$

Supuesto 2. Se asumirá que h_2 satisface

$$2h_2(v) - \psi_2(v)v \geq 0 \quad (2.21)$$

Si esto se cumple, $w(\theta) \geq 0$ y se podría pensar en el estimador τ como en un estimado M con la función ψ de influencia adaptativa: $\psi_n(v) = w(\theta)\psi_1(v) + \psi_2(v)$, la cual es el promedio ponderado de ψ_1 y ψ_2 . Seleccionando adecuadamente $\psi_1(v)$ y $\psi_2(v)$, el estimador τ resultante tendrá un punto de ruptura elevado y una alta eficiencia con errores normalmente distribuidos.

Si se consideran la siguiente familia de funciones ψ

$$\psi(v) = \begin{cases} v \left(1 - \frac{v^2}{c^2}\right)^2 & \text{si } |v| < c \\ 0 & \text{si } |v| \geq c \end{cases} \quad (2.22)$$

con su correspondiente familia de funciones $h(v)$

$$h(v) = \begin{cases} \frac{v^2}{2} \left(1 - \frac{v^2}{c^2} + \frac{v^4}{3c^4}\right) & \text{si } |v| \leq c \\ \frac{c^2}{6} & \text{si } |v| > c \end{cases} \quad (2.23)$$

y se selecciona $c = c_1$, entonces $h(v) = h_1(v)$ y $\psi(v) = \psi_1(v)$. Del mismo modo, si $c = c_2$, entonces $h(v) = h_2(v)$ y $\psi(v) = \psi_2(v)$.

Seleccionando $c_1 = 1,56$ y $c_2 = 6,08$, el estimador τ tiene un punto de ruptura igual al 50 %, una eficiencia del 95 % bajo la hipótesis de normalidad en los residuos, y se garantiza que su función de influencia ψ es acotada (Tabatabai and Argyros, 1993).

2.3.2. Implementación del estimador τ no lineal en el algoritmo de retropropagación del error

En esta sección se presenta la manera de combinar las propiedades de robustez del estimador τ , basado en la minimización de un criterio (error τ) que hace que la estimación de parámetros sea insensible a espúreos, con el algoritmo BP, para desarrollar una red robusta a espúreos.

En este caso, la función de coste a minimizar es

$$E_{TAO} = \sum_{\mu=1}^n \tau^2(r_\mu; S(t)) \quad (2.24)$$

donde $r_\mu = \bar{y}^\mu - \hat{y}^\mu$ es el error del μ -ésimo patrón de entrenamiento, t es el número de época, $S(t)$ es el estimador de escala que depende de la iteración t , y τ^2 es el criterio de error TAO.

Según se indica en las ecuaciones 2.5 y 2.6, para implementar el error τ en el algoritmo de retropropagación, lo único que se debe cambiar es el cálculo de la derivada parcial de la métrica del error (que en este caso es $\tau_n^2(\vec{r})$) con respecto a la salida de la red (y_k^μ), ya que el resto no depende del criterio de error empleado y, por tanto, se puede dejar intacto

$$\frac{\partial E}{\partial y_k^\mu} = \frac{\partial \tau_n^2(\vec{r})}{\partial y_k^\mu} = - \sum_{\mu=1}^n \frac{1}{n} \left\{ 2S_t \frac{\partial S_t}{\partial r} h_2(v) + \psi_2(v) \left(S_t - \frac{\partial S_t}{\partial r} r_\mu \right) \right\} \quad (2.25)$$

donde

$$\frac{\partial S_t}{\partial r} = \frac{\sum_{\mu=1}^n \psi_1(v)}{\sum_{\mu=1}^n [\psi_1(v) v^\mu]} \quad \text{con } v^\mu = \frac{r_\mu}{S_t} \quad (2.26)$$

y $h(v)$ y $\psi(v)$ se toman según las ecuaciones 2.23 y 2.22.

$S_t(\vec{r})$ se puede obtener para cada iteración mediante la siguiente ecuación (Tabatabai and Argyros, 1993):

$$S_{t+1}(\vec{r}) = \left\{ \sum_{\mu=1}^n \left[\frac{S_t^2 h_1\left(\frac{r_\mu}{S_t}\right)}{n E_\phi \left[h_1\left(\frac{r_\mu}{S_t}\right) \right]} \right] \right\}^{1/2} \quad (2.27)$$

Capítulo 3

Modelado de una línea de producción de acero galvanizado

3.1. Introducción

La creciente demanda que están experimentando los aceros recubiertos mediante electrocincado y galvanizado, por sus propiedades anticorrosión, en sectores como la automoción, la fabricación de electrodomésticos y la construcción, junto con el aumento de las exigencias de los clientes, hace que el sector siderúrgico, a pesar de ser una actividad tradicional y madura, realice importantes esfuerzos en el campo de las nuevas tecnologías de fabricación y de mejora de la calidad de sus productos.

En la actualidad, las características del producto que se pretende fabricar, esto es, las propiedades mecánicas de las bobinas de acero galvanizado, no se pueden medir directamente, sino que deben llevarse a cabo ensayos en un laboratorio, empleando métodos destructivos, tras el proceso de galvanizado¹. En este caso, el problema que se plantea, es la necesidad de adoptar un control en lazo abierto, al no ser posible aplicar una estrategia de control clásica. Por ello, en Ordieres Meré et al. (2004) se pretende, a través de la aplicación de técnicas de minería de datos, crear un modelo que permita predecir *on-line* dichas propiedades mecánicas, a partir de los datos procedentes del proceso de

¹Estos ensayos mecánicos normalmente incluyen el límite elástico (E), la resistencia a la rotura (R) y el alargamiento (A).

fabricación. De esta forma, sería posible una mejora en los sistemas de control actuales, la cual repercutiría en la calidad del producto final, siendo éste uno de los objetivos básicos de todo proceso industrial.

Un proceso clave para obtener las propiedades de la banda deseadas y una buena adherencia del recubrimiento de zinc, en el galvanizado por inmersión en caliente, es el tratamiento térmico al que se someten las bandas de acero antes de su inmersión en el pote de zinc. Actualmente, la temperatura de la banda en el horno de proceso se consigue cambiando únicamente la temperatura de consigna del horno. Sin embargo, con un modelo de velocidad para la banda de acero dentro del horno, a partir de los datos de proceso, como el propuesto en Pernía Espinoza et al. (2005a), se podrían conseguir temperaturas de salida más adecuadas, mediante la regulación de la velocidad.

En la introducción de esta tesis se comentó, que la primera etapa de todo proceso de minería de datos consiste en obtener un conocimiento adecuado del problema que se pretende resolver. Por eso, en la sección 3.2, se exponen los factores que influyen en las propiedades mecánicas de las bobinas de acero galvanizado y se presenta la línea de fabricación de acero galvanizado en la que se ha trabajado para obtener los modelos de propiedades mecánicas y de velocidad en el horno de recocido. Con este conocimiento previo, en las siguientes secciones, 3.3 y 3.4, se presenta la metodología empleada en la consecución de dichos modelos.

3.2. Factores que influyen en las propiedades mecánicas del acero

La *composición química* del acero, el *tratamiento térmico* al que se somete, así como el *proceso de fabricación* utilizado, definen sus propiedades mecánicas. Aunque el elemento básico que constituye el acero es el hierro, la adición de pequeñas cantidades de otros elementos, puede tener un efecto marcado en el tipo y propiedades del acero. Además, hay que tener en cuenta que, cuando se aplican tratamientos térmicos al material, incluido el enfriamiento a una velocidad determinada desde un pico de temperatura concreto, estos elementos producen respuestas diferentes. No hay que olvidar tampoco, que el proceso de producción utiliza combinaciones de tratamientos térmicos y trabajos mecánicos, las cuales son de crítica importancia para comprender el posterior comportamien-

3.2 Factores que influyen en las propiedades mecánicas del acero 25

to del acero y lo que puede y no puede hacerse de forma satisfactoria con el material tras el proceso básico de producción.

3.2.1. Composición química

La clave para entender los efectos de la composición química y del tratamiento térmico en la metalurgia y propiedades de los aceros, es reconocer que éstas dependen de los siguientes factores (Burdekin, 1992):

1. microestructura
2. tamaño del grano
3. inclusiones no metálicas
4. precipitados entre granos o en sus contornos
5. presencia de gases absorbidos o disueltos

El acero es una aleación compuesta básicamente por hierro al que se añaden pequeñas cantidades de carbono (hasta un máximo del 2%) y otros elementos que le proporcionan determinadas propiedades mecánicas. A medida que aumenta la cantidad de carbono añadida, se incrementa la resistencia del material, disminuyendo su ductilidad y haciéndolo más sensible al tratamiento térmico. El acero así obtenido, aunque barato y útil en algunas aplicaciones (por ejemplo, armaduras para hormigón, cables, etc.), presenta problemas cuando el nivel de carbono es medio-alto y se utiliza en posteriores procesos de fabricación, particularmente cuando es necesario realizar soldaduras. Sin embargo, es posible obtener un acero más versátil manteniendo el contenido de carbono a un nivel relativamente bajo, si se añaden pequeñas cantidades de otros elementos ya que, combinados con tratamientos térmicos adecuados, pueden producir aceros resistentes con buena ductilidad, tenacidad y soldabilidad; aceros con mayor resistencia a la corrosión; aceros con mayor resistencia a altas temperaturas; etc.

Aditivos

Como ya se ha comentado, la adición de pequeñas cantidades de carbono al hierro incrementa su resistencia y sensibilidad al tratamiento térmico (o *templabilidad*²). Sin embargo, existen otros elementos que también afectan a la resistencia y templabilidad, aunque en mucha menor medida que aquel. El *manganeso*, el *romo*, el *molibdeno*, el *níquel* y el *cobre* actúan en la microestructura del acero y permiten obtener la resistencia requerida, manteniendo un nivel de carbono bajo, si el tratamiento térmico/condiciones de fabricación son los adecuados:

- El *manganeso* es uno de los elementos fundamentales e indispensables, apareciendo prácticamente en todas las aleaciones de acero, debido, principalmente, a que se añade como elemento de adición para neutralizar la perniciosa influencia del azufre y del oxígeno, que siempre suelen contener los aceros cuando se encuentran en estado líquido en los hornos durante los procesos de fabricación. El manganeso actúa también como desoxidante, aumenta la capacidad de endurecimiento y evita, en parte, que en la solidificación del acero se desprendan gases que den lugar a porosidades perjudiciales en el material.
- El *romo* aumenta la profundidad del endurecimiento y mejora la templabilidad. Asimismo, aumenta la resistencia a altas temperaturas y evita la corrosión.

Es un elemento principal de aleación en aceros inoxidables y, debido a su capacidad de formar carburos, se utiliza en revestimientos o recubrimientos duros de gran resistencia al desgaste, como émbolos, ejes, etc.

- Una de las ventajas más grandes que reporta el empleo de *níquel*, es evitar el crecimiento del grano en los tratamientos térmicos, lo que sirve para producir en ellos gran tenacidad. El níquel, además, hace descender los puntos críticos y, por ello, los tratamientos pueden hacerse a temperaturas ligeramente más bajas que la que corresponde a los aceros ordinarios. Experimentalmente se observa que con los aceros aleados con níquel se

²“Templabilidad” es un término utilizado para describir la aptitud de una aleación para endurecerse como consecuencia de un tratamiento térmico, es decir, es una medida de la profundidad a la cual una aleación específica puede endurecerse (Callister, 1996).

3.2 Factores que influyen en las propiedades mecánicas del acero 27

obtiene para una misma dureza, un límite de elasticidad ligeramente más elevado y mayores alargamientos y resistencias que con los aceros al carbono o de baja aleación.

El níquel se utiliza mucho en los aceros inoxidable, para aumentar la resistencia a la corrosión.

- El *molibdeno* también es un elemento habitual, ya que aumenta mucho la profundidad de endurecimiento del acero, así como su resistencia al impacto. Es el elemento más efectivo para mejorar la resistencia del acero a bajas temperaturas, reduciendo, además, la pérdida de resistencia por templado: añadiendo sólo pequeñas cantidades de molibdeno a los aceros cromo-níqueles, se disminuye o elimina casi totalmente el problema conocido como *fragilidad por revenido*³.
- El *cobre* aumenta la resistencia a la corrosión de aceros al carbono.

El afinamiento de la estructura del grano de los aceros conduce, al mismo tiempo, a un incremento en el límite elástico y a una mejora en la tenacidad y ductilidad, lo que es importante para obtener aceros con mejores propiedades. Aunque el tratamiento térmico y, en particular, la velocidad de enfriamiento, son factores clave para obtener el afinamiento del grano, la presencia de uno o más elementos que lo promueven, añadiendo la nucleación de nuevos granos durante el enfriamiento, también es extremadamente beneficiosa. Tales elementos son el *niobio*, el *vanadio* y el *aluminio*, los cuales pueden ser añadidos en pequeñas cantidades, hasta un 0,05 % en peso:

- El *niobio* hace posible la mejora de las características mecánicas mediante el afino del grano y la precipitación.
- El *vanadio* facilita la formación de grano pequeño y reduce la pérdida de resistencia durante el templado, aumentando, por tanto, la capacidad de endurecimiento. Es un elemento desoxidante muy fuerte y tiene una

³La *fragilidad por revenido*, o *fragilidad Krupp*, se produce al migrar los elementos a los contornos del grano si el acero alcanza una temperatura entre 500 y 600 °C. A temperatura normal, estos aceros tienen muy poca tenacidad y el fallo se produce por fractura intergranular.

gran tendencia a formar carburos, por lo que basta con añadir pequeñas cantidades. Una característica de los aceros con vanadio es su gran resistencia al ablandamiento por revenido. Se emplea principalmente para la fabricación de aceros de herramientas.

- El *aluminio* se emplea principalmente como desoxidante en la elaboración de acero, aunque también reduce el crecimiento del grano al formar óxidos y nitruros.

Existen, además, otros elementos cuya adición puede resultar beneficiosa en las propiedades de los aceros:

- El *silicio*, al igual que el manganeso, aparece en todos los aceros porque se añade intencionadamente durante el proceso de fabricación. Se emplea como elemento desoxidante complementario del manganeso, con objeto de evitar que aparezcan en el acero poros y otros defectos internos. Mejora ligeramente la templabilidad y la resistencia de los aceros y, en ciertos casos, mejora también su resistencia a la oxidación.
- El *boro* logra aumentar la capacidad de endurecimiento cuando el acero está totalmente desoxidado. Una pequeña cantidad de Boro (0.001 %) tiene un efecto marcado en el endurecimiento del acero, ya que también se combina con el carbono para formar los carburos que dan al acero características de revestimiento duro.
- El *titanio* se suele añadir en pequeñas cantidades a algunos aceros muy especiales, para desoxidar y afinar el grano. Este elemento tiene gran tendencia a formar carburos y a combinarse con el nitrógeno.

En los aceros inoxidables cromo-níquel, actúa como estabilizador de los carburos y evita la corrosión intercrystalina.

- El *cobalto* es un elemento poco habitual en los aceros, ya que disminuye la capacidad de endurecimiento. Sin embargo, se puede usar en aplicaciones donde se requiere un revestimiento duro para servicio a alta temperatura, ya que produce una gran cantidad de solución sólida endurecedora, cuando es disuelto en ferrita o austenita.

3.2 Factores que influyen en las propiedades mecánicas del acero 29

- El *wolframio* (tungsteno), es un elemento muy utilizado para la fabricación de aceros de herramientas, empleándose en especial en los aceros rápidos, aceros para herramientas de corte y aceros para trabajos en caliente. Sirve para mantener la dureza de los aceros a elevada temperatura y evitan que se desafilen o ablanden las herramientas, aunque lleguen a calentarse a 500 ó 600 °C.
- El *plomo* es un ejemplo de elemento casi insoluble en hierro. Se añade plomo a muchos tipos de acero para mejorar en gran manera su maquinabilidad.

Inclusiones no metálicas

La presencia de inclusiones no metálicas debe ser cuidadosamente controlada en aplicaciones particulares. Tales inclusiones se presentan como un residuo del mineral, de la chatarra o del combustible en el proceso de fabricación del acero, por lo que se deben seguir ciertos pasos para reducirlas al nivel requerido. Las impurezas más comunes son el *azufre* y el *fósforo*: elevadas cantidades de estos elementos reducen la resistencia, la ductilidad y pueden producir problemas de grietas en uniones soldadas. Otros elementos que pueden aparecer como impurezas y pueden tener serios efectos negativos en el acero son el *antimonio* y el *arsénico*, ya que pueden promover la *fragilidad por revenido*. Es particularmente importante asegurar que este grupo de elementos sea eliminado de los aceros de bajo contenido en carbono.

Los aceros con un alto contenido en gases disueltos, en particular *oxígeno* y *nitrógeno*, pueden comportarse de manera frágil, por lo que su nivel debe ser controlado. Incluyendo pequeñas cantidades de elementos que tengan una especial afinidad con ellos, se consiguen combinaciones que escapan del acero líquido a altas temperaturas o permanecen como inclusiones no metálicas: el *silicio* y el *aluminio* se usan como desoxidantes, aunque el aluminio también ayuda a controlar el nivel de nitrógeno libre ya que éste debe mantenerse a niveles bajos en aquellos casos en los que el fenómeno de *fragilidad por envejecimiento* es importante.

3.2.2. Tratamiento térmico

Durante la fabricación del acero, la composición química deseada se obtiene

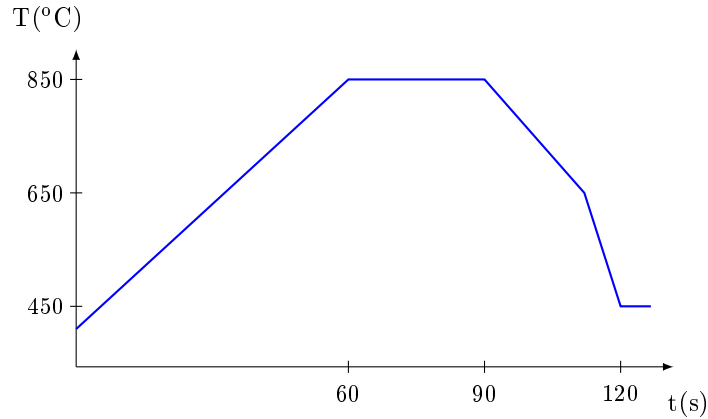


Figura 3.1: Ejemplo de perfil de temperaturas en un ciclo de recocido.

mientras está en estado líquido a altas temperaturas. A medida que el acero se enfría, se solidifica y se producen cambios sustanciales en su estructura. Sin embargo, la microestructura y propiedades de un acero pueden modificarse con tratamientos térmicos cuidadosamente elegidos tras el proceso de producción del producto base:

- *Recocido.* Con este nombre se conocen varios tratamientos donde el enfriamiento del acero, desde una temperatura entre 800 y 950 °C, es lento y se realiza esencialmente en un horno. Su objetivo principal es “ablandar” el acero para facilitar su mecanizado posterior. También es utilizado para regenerar el grano o eliminar las tensiones internas. En un ciclo de recocido se reconocen las siguientes fases (figura 3.1):

1. Subida de la temperatura hasta la temperatura máxima del perfil de recocido, para asegurar la recristalización del acero.
2. Mantenimiento a alta temperatura para permitir el crecimiento del grano.
3. Enfriamiento lento para controlar la textura del metal y obtener unas propiedades mecánicas y físicas adecuadas.
4. Enfriamiento rápido para preparar el acero al tratamiento de envejecimiento, congelando una cantidad máxima de carbono en sobre-

3.2 Factores que influyen en las propiedades mecánicas del acero 31

saturación.

5. Envejecimiento o igualación, garantizando una precipitación del carbono, para minimizar los fenómenos de envejecimiento ulterior del acero.

- *Normalizado*. El acero se enfría, también desde una temperatura entre 800 y 950 °C, al aire. Con este tratamiento se consigue un tamaño del grano más fino y, por tanto, un mayor límite elástico y una mejor tenacidad, para una composición dada.
- *Templado*. Es un proceso de calentamiento seguido de un enfriamiento, generalmente rápido con una velocidad mínima llamada “crítica”. El fin que se pretende conseguir con el ciclo del temple es aumentar la dureza y resistencia mecánica. El factor que caracteriza a la fase de enfriamiento es la velocidad del mismo. La velocidad crítica de los aceros al carbono es muy elevada, mientras que los elementos de aleación disminuyen en general la velocidad crítica de temple y en algunos tipos de alta aleación es posible realizar el temple al aire. A estos aceros se les denomina ‘autotemplantes’. Los factores que influyen en la práctica del temple son el tamaño de la pieza, la composición química del acero, el tamaño del grano y el medio de enfriamiento.
- *Revenido*. Es un tratamiento complementario del temple, que generalmente sigue a éste. Al conjunto de los dos tratamientos también se le denomina “bonificado”. El tratamiento de revenido consiste en calentar al acero después de normalizado o templado, seguido de un enfriamiento controlado que puede ser rápido cuando se pretenden resultados altos en tenacidad, o lento, para reducir al máximo las tensiones térmicas que pueden generar deformaciones. Cuando se pretenden los dos objetivos, se recurre al doble revenido, el primero con enfriamiento rápido y el segundo con enfriamiento lento hasta -300°C.

3.2.3. Proceso productivo

El acero se obtiene a partir de dos materias primas fundamentales: mineral de hierro y chatarra férrea. El proceso de fabricación de acero fundido a

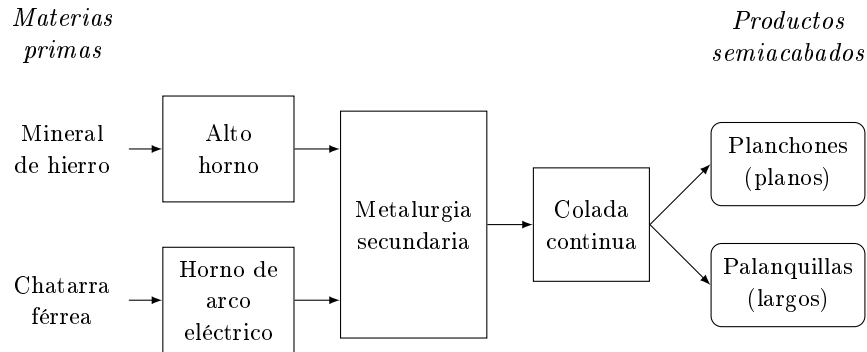


Figura 3.2: *Proceso de fabricación del acero.*

base de mineral de hierro utiliza un alto horno, mientras que el proceso con chatarra férrea recurre a un horno de arco eléctrico. Ambos procesos son complementados con Metalurgia Secundaria en hornos de cuchara, donde se refina el acero y se ajusta su composición química para producir calidades específicas. Posteriormente, el acero líquido pasa a las máquinas de colada continua para su solidificación, obteniéndose lo que se conoce como productos semiacabados: desbastes o planchones, si presentan un corte transversal rectangular, para la fabricación de productos planos (bandas, bobinas, láminas, etc.); o *blooms* o palanquillas, si tienen un corte transversal cuadrado, que se destinan para la fabricación de productos largos (barras, alambrón, etc.) (figura 3.2).

Una vez ajustada la composición química del acero en las estaciones de metalurgia secundaria y antes de ser vertido de la cuchara a la artesa de la colada continua, se toman muestras del mismo para ser analizadas, de tal forma que los resultados de este análisis representan la composición química de toda la colada.

La transformación de los planchones y palanquillas en productos acabados se realiza mediante la “laminación en caliente”. En este proceso, los productos semiacabados se calientan primero en hornos, donde son llevados a temperaturas de laminación. Este tratamiento metalúrgico permite, por medio de la oxidación generada, remover pequeños defectos superficiales y ablandar el acero para facilitar los procesos de trefilado y conformado. A continuación, el formato se transforma mecánicamente hasta la forma y tamaño desados al pa-

3.2 Factores que influyen en las propiedades mecánicas del acero 33

sar a través de los cilindros de la unidad de laminación, que reducen la sección transversal del acero mediante la presión ejercida por éstos.

Las chapas finas laminadas en caliente también pueden ser sometidas a deformación a temperatura ambiente (“laminación en frío”) para reducir el espesor y obtener bobinas laminadas en frío. Estas últimas pueden ser entregadas al mercado o continuar su procesamiento en los hornos de recocido y en los trenes de laminación de temple: el recocido es un tratamiento térmico que ablanda y recupera las características mecánicas del material deformado en frío; mientras que la laminación del temple permite eliminar las líneas de fluencia, corregir la forma y dar la terminación superficial requerida. A su vez, estas bobinas pueden continuar procesos posteriores de recubrimiento electroquímico.

Fabricación del acero galvanizado

Parte de los trabajos publicados en esta tesis, se centran en una línea de galvanizado continuo por inmersión en particular, por lo que, en esta sección, se ha preferido describir esta línea de fabricación de acero galvanizado⁴. A pesar de ello, lo aquí descrito, puede aplicarse, con ligeras modificaciones, a la mayor parte de las líneas de galvanizado continuo por inmersión instaladas en todo el mundo, por ser todas ellas muy similares.

La línea puede considerarse dividida en tres secciones (figura 3.3):

- *Sección de entrada* que comprende las unidades de debobinado, enderezado, corte y soldadura.
- *Sección de proceso* que comprende la unidad de tensión, horno de proceso, unidad de revestimiento, control de revestimiento, equipo de enfriamiento, unidad de tratamiento químico y unidades de aplanado.
- *Sección de salida* que comprende el equipo de bobinado y de salida de bobinas, tijera, aplanadora, inspección, apiladora y salida de chapas de rechazo, apilado y salida de chapas de primera calidad.

⁴En Vergara González (1999) puede encontrarse una descripción más detallada de esta línea.

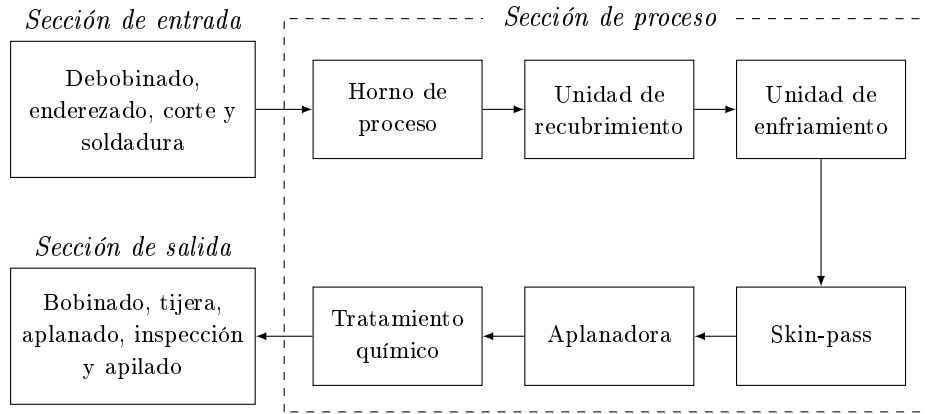


Figura 3.3: Esquema general de una línea de acero galvanizado.

La primera operación que tiene lugar en la línea, consiste en la formación de una banda continua a partir de las bobinas de acero procedentes de laminación, que ya poseen el espesor requerido. Para transformar las bobinas en una banda continua, se debobinan, despuntando la cabeza y la cola, y se sueldan a solape.

Las impurezas arrastradas de la laminación se eliminan mediante el calentamiento en atmósfera no oxidante. Una vez limpia la banda, ésta se somete a un ciclo de recocido (véase apartado 3.2.2), para mejorar las características mecánicas de la banda. El horno de proceso en el que se realizan estas operaciones, se puede dividir en cinco zonas:

- *Zona de Precalentamiento y Limpieza.* La misión de esta sección es limpiar la banda, que procede directamente del Tren Tándem, y calentarla hasta una temperatura de 450-800°C, según espesores y ciclos térmicos. La limpieza se realiza por volatilización de los aceites de laminación, arrastre mecánico de partículas de suciedad y reducción del óxido superficial, al pasar la banda entre dos filas de mecheros cuya llama, reductora y abierta, calienta las zonas hasta una temperatura máxima de 1260°C. Conseguir una perfecta limpieza de banda es fundamental para lograr una buena adherencia. Si esta sección no consigue limpiar la banda, o si, por el contrario, se oxida por una mala composición de la llama (llama oxidante), será imposible lograr un recubrimiento de calidad.

3.2 Factores que influyen en las propiedades mecánicas del acero 35

- *Zona de Calentamiento.* La banda limpia, y a una temperatura de 450-800°C, es calentada en esta sección hasta una temperatura superior a 780°C. El calentamiento se realiza por radiación, protegiéndose la banda con atmósfera de HN contra la oxidación, la cual podría producirse como consecuencia de la entrada de aire por alguna fuga o por rotura de un tubo radiante. A su paso por esta sección, la banda alcanza su máxima temperatura en todo el proceso (aproximadamente 800°C), siendo la temperatura máxima de la zona 980°C.
- *Zona de Enfriamiento Controlado.* La banda, que a su paso por la sección anterior alcanzó su máxima temperatura, durante el recorrido por esta zona se enfriará de modo controlado, para conseguir unas características mecánicas adecuadas. El enfriamiento se realiza mediante un intercambio de calor banda-atmósfera, refrigerándose esta última por medio de tubos enfriados por aire. En esta sección, una oxidación de la banda sería irreversible por estar el hidrógeno de la atmósfera a una temperatura excesivamente baja. La temperatura de banda a la salida de esta sección varía entre 600-800°C según el ciclo térmico.
- *Zona de enfriamiento Rápido (Jet Cooling).* En esta sección se enfría la banda hasta una temperatura más adecuada para realizar el recubrimiento, esto es, ligeramente superior a la del baño. Es imprescindible mantener esta sección totalmente exenta de fugas debido a que la presencia de oxígeno oxidaría la banda de forma irreversible, impidiendo la adherencia. La temperatura de la banda, a la entrada del baño, debe ser lo más estable posible, ya que las variaciones de temperatura favorecen la disolución del acero en el zinc.
- *Zona del Turn Down.* Desvía la banda hacia el pote. El conducto de bajada cierra estanco con el nivel del metal fundido en el pote.

A continuación, la banda se sumerge en un pote con zinc fundido quedando revestida de este metal. El control del revestimiento se realiza proyectando aire a alta presión sobre ambos lados de la banda recubierta. Tras el revestimiento, la banda se enfría mediante corriente de aire forzado a través de seis conductos, montados uno vertical y cinco horizontales, y mediante la proyección de agua pulverizada, sobre ambas caras, en el tanque de enfriamiento.

Si durante el almacenamiento o transporte se produjeran condiciones suaves de oxidación se formarían sobre la superficie de la banda manchas blancas correspondientes a la presencia de óxido. Este fenómeno no supone una disminución de las características anticorrosivas del recubrimiento, pero sí deslucen el aspecto de la banda. Para prevenir la aparición de esta *oxidación blanca*, se somete a la banda a un tratamiento superficial de ácido crómico en la unidad de tratamiento químico.

En dos de los rodillos deflectores que conducen la banda por debajo del pote, están localizados los equipos de la galga para medir el espesor del revestimiento. Cuando el espesor de la banda sea inferior a 2mm el aplanado se realizará por medio de la C.S.L. (*Continuous Stretch Leveling*) o aplanadora. Cuatro juegos de rodillos de tensión producen la tensión suficiente para obtener alargamientos de hasta el 2%. A la salida de esta unidad existe un equipo de rodillos para la medida de la tensión de la banda, indicando la uniformidad de la medida la planitud de la banda. En algunas ocasiones este aplanado puede venir precedido de un *skin-pass*, que dote al material de las características mecánicas, y rugosidad superficial adecuadas.

Después de este proceso, se obtiene el producto acabado bien en forma de bobinas o de chapas cortadas. Las chapas, de longitud predeterminada, se obtienen mediante una tijera y se imprimen con la marca de fabricación mediante un marcador operado mecánicamente, instalado en la mesa de inspección.

3.3. Modelado de las propiedades mecánicas de las bobinas

En la sección anterior, se ha visto cómo las propiedades mecánicas de las bobinas de acero galvanizado pueden verse afectadas a lo largo de todo el proceso de fabricación, desde la obtención del acero (que determina la composición química de la colada) hasta, prácticamente, el momento en el que se convierte en producto acabado (bien en forma de bobinas o de chapas galvanizadas). Eso quiere decir, que podría resultar factible desarrollar un modelo que permitiera predecir dichas propiedades mecánicas, a partir de los datos de proceso, e introducir mejoras en los sistemas de control existentes en las líneas de fabricación de acero galvanizado. A continuación, se expone la metodología seguida en esta línea de investigación, cuyos resultados han sido publicados en Ordieres Meré

et al. (2004).

Una vez analizado el problema y determinados los objetivos que se pretendían alcanzar, una fase crucial en el proceso de minería de datos iniciado, fue la selección de las variables que intervendrían en la generación del modelo. Tras consultar con los expertos, se decidió incluir en el estudio quince variables de composición química de la colada y tres variables de proceso procedentes de la línea de galvanizado, para predecir el límite elástico, la resistencia a la rotura y el alargamiento de las bobinas (ver cuadro 3.1).

Las fases de análisis y preparación del conjunto de datos, como paso previo al modelado, fueron llevadas a cabo, principalmente, empleando técnicas de visualización (histogramas, diagramas de dispersión de variables por pares, etc.) y técnicas de proyección (proyección Sammon (Sammon, 1969) y Análisis de Componentes Principales, PCA (Dunteman, 1989)). Estas herramientas fueron útiles para observar la estructura de los datos (asimetrías, heterogeneidad, etc.), las relaciones existentes entre las variables, la ocurrencia de datos erróneos, así como la existencia de diferentes clases de comportamiento.

Tras eliminar los espurios y dividir el conjunto de datos en las clases previamente detectadas, se procedió al entrenamiento de un conjunto de redes neuronales en aquellas clases con un número “suficiente” de patrones, teniendo en cuenta que la base de datos se dividió en tres subconjuntos (entrenamiento, validación y test) para mejorar la capacidad de generalización de las redes entrenadas. Con los patrones de entrenamiento (63,3 % de las muestras) se realizó el ajuste de los pesos de las neuronas, mientras que los de validación (31,6 % de los datos disponibles) fueron utilizados para comprobar, cada cierto número de ciclos de entrenamiento (treinta de un total de cien mil), que el error obtenido con los patrones de entrenamiento era coherente con el error que vería la red neuronal frente a nuevos datos que no hubieran participado en el ajuste. En cada caso entrenado, la red final se correspondía con aquella cuyo error de validación era mínimo. Por último, los patrones de test (5 % restante), que no habían sido vistos nunca por la red entrenada, sirvieron para comprobar la capacidad de generalización de la misma.

Además de la técnica de parada temprana empleada, para que los resultados del entrenamiento no dependieran de una determinada selección, que al ser aleatoria pudiera no ser idónea en el contenido de su información, se generaron tres conjuntos de ficheros con los que se llevó a cabo el entrenamiento, validación y test, seleccionando finalmente la red que mejor se adaptó a la información

Variable	Símbolo
<i>Variables de entrada</i>	
Temperatura media de la banda a la salida de la zona de calentamiento (°C)	tmp.avg
Velocidad media de la banda dentro del horno (m min^{-1})	vel.avg
Alargamiento en el skin-pass (%)	skin.A
Contenido en Carbono (% en peso)	C
Contenido en Manganeso (% en peso)	Mn
Contenido en Silicio (% en peso)	Si
Contenido en Azufre (% en peso)	S
Contenido en Fósforo (% en peso)	P
Contenido en Aluminio (% en peso)	Al
Contenido en Cobre (% en peso)	Cu
Contenido en Níquel (% en peso)	Ni
Contenido en Cromo (% en peso)	Cr
Contenido en Niobio (% en peso)	Nb
Contenido en Vanadio (% en peso)	V
Contenido en Titanio (% en peso)	Ti
Contenido en Boro (% en peso)	B
Contenido en Nitrógeno (% en peso)	N
Carbono equivalente (% en peso)	C_{eq}
<i>Variables de salida</i>	
Límite Elástico (MPa)	YIELD
Resistencia a la rotura (MPa)	TENSILE
Alargamiento (%)	ELONG

Cuadro 3.1: Variables utilizadas en el modelado de las propiedades mecánicas.

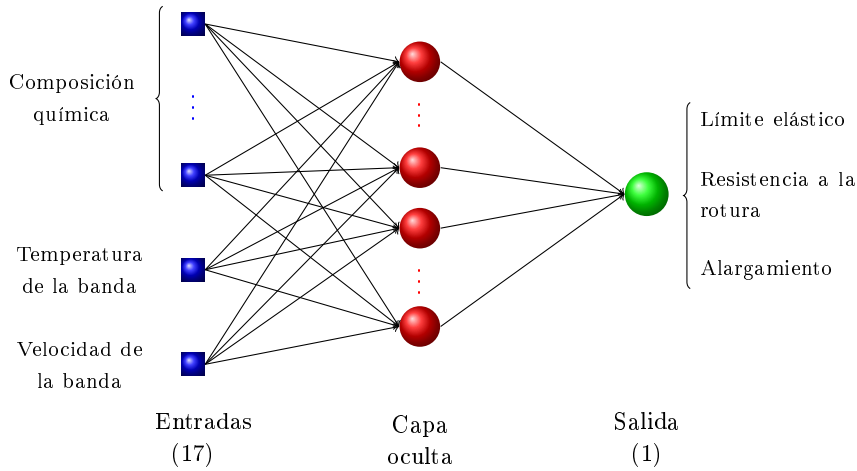


Figura 3.4: Modelo neuronal para la predicción de las propiedades mecánicas.

contenida en los datos.

En cuanto a la arquitectura de la red, se utilizó un perceptrón multicapa con diecisiete entradas, un número variable de neuronas en la capa oculta y una salida (figura 3.4). El algoritmo de aprendizaje seleccionado fue el de retropropagación del error con decaimiento de pesos, para evitar el sobreajuste de la red, al no disponer de un número suficiente de patrones de entrenamiento en comparación con el número de variables de entrada.

3.4. Modelado de velocidad de la banda en el horno de recocido

En el apartado 3.2.2, se ha tratado la influencia que los tratamientos térmicos tienen en las propiedades finales del acero. En la línea de acero galvanizado por inmersión objeto de estudio, las bobinas fabricadas son sometidas a un tratamiento de recocido antes de proceder a su revestimiento. Este proceso es clave no sólo para la obtener las propiedades de la banda deseadas, sino también para conseguir una buena adherencia del recubrimiento. Actualmente, la temperatura de la banda de acero se consigue modificando las consignas de temperatura del horno. Ya en Martínez de Pisón Ascacibar (2003), se propone un control más eficiente de la temperatura, para asegurar que la temperatura

Variable	Símbolo
<i>Variables de entrada</i>	
Espesor de la banda a la entrada del horno (mm)	THICKN
Anchura de la banda a la entrada del horno (mm)	WIDTH
Consigna de temperatura de la banda a la salida de la zona de calentamiento (°C)	TMPP2SP
Temperatura real de la banda a la entrada de la zona de calentamiento (°C)	TMPP1
Consigna de temperatura de zona 1 del horno dentro de la zona de calentamiento (°C)	TMPH1SP
Consigna de temperatura de zona 3 del horno dentro de la zona de calentamiento (°C)	TMPH3SP
Consigna de temperatura de zona 5 del horno dentro de la zona de calentamiento (°C)	TMPH5SP
<i>Variable de salida</i>	
Velocidad media de la banda dentro del horno (m min^{-1})	VEL

Cuadro 3.2: Variables utilizadas en el modelado de la velocidad de la banda en el horno de recocido.

real de la banda a la salida de la zona de calentamiento del horno se aproxime a la deseada, si, además, se regula la velocidad de la banda dentro del horno. En este sentido, y como una mejora más en el control del ciclo de recocido, se ha trabajado en el desarrollo de un modelo robusto de la velocidad de la banda dentro del horno de recocido, publicándose los resultados obtenidos en Pernía Espinoza et al. (2005a).

La base de datos con la que se desarrolló el modelo, fue tomada de Martínez de Pisón Ascacíbar (2003), donde se especifica cuál ha sido el proceso de selección de las variables más relevantes (cuadro 3.2), partiendo de un total de más de 6000. Por eso, una vez delimitados los objetivos a alcanzar, hubo que decidir el tipo de modelo robusto más conveniente. De nuevo, por su probada eficiencia y simplicidad, se prefirió el modelado mediante redes neuronales.

El conjunto de datos utilizado contenía un pequeño porcentaje de muestras, un 3 %, con valores de velocidad fuera del rango normal de funcionamiento de la línea. A pesar de que en un principio se pensó en que correspondían a espurios causados por errores de adquisición, y, por tanto, podían ser eliminados, los expertos del proceso explicaron que la bajada brusca de velocidad de la banda

era debida a fenómenos transitorios que pueden darse en la línea, tales como la soldadura de una bobina o la incorporación de una bobina atípica (con dimensiones inusuales). Al visualizar la evolución temporal de la velocidad y temperatura de salida de la banda, se vió que estas transiciones de velocidad duraban cortos espacios de tiempo y que sus efectos en la temperatura de salida de la banda no eran significativos, es decir, la calidad del producto no se veía afectada por esos cambios.

Ante esta situación, en la que no es extraño que disminuya la velocidad de forma brusca, era conveniente que el modelo neuronal aprendiera la relación entre la temperatura y la velocidad en estos casos. Sin embargo, para que su presencia en el entrenamiento no afectara negativamente al modelo, resultaba adecuado el uso de algoritmos robustos. De los diferentes tipos de redes neuronales robustas que pueden encontrarse en la literatura (Ahmed and Farag, 2002; Chen and Jain, 1994; Chuang et al., 2000; Huang et al., 1998; Liano, 1996; McDowall and Ham, 1997), se seleccionó el método propuesto por Liano (1996), denominado LMLS (*Least Mean Log Squares*), por sus importantes propiedades (robustez con un elevado punto de ruptura y una alta eficiencia con errores normalmente distribuidos) y por su facilidad de implementación en cualquier programa de entrenamiento que emplee la retropropagación. Además, para verificar su comportamiento en el modelado de procesos industriales, que hasta el momento no había tenido lugar, se comparó con el algoritmo de aprendizaje no robusto, basado en el MSE.

En el entrenamiento se emplearon redes perceptrón multicapa con siete entradas, quince neuronas en la capa oculta y una salida (figura 3.5). Para el aprendizaje, el algoritmo de retropropagación con el método de optimización del gradiente conjugado de Fletcher-Reeves, fue utilizado. En cuanto al criterio de error, como ya se ha comentado, se empleó, por un lado, el error medio del logaritmo de los cuadrados (criterio robusto) y, por otro, el error cuadrático medio (criterio no robusto).

Por último, para que el modelo de velocidad de la banda dentro del horno produjera consignas de control adecuadas, únicamente se usaron patrones cuya diferencia entre la temperatura real y la deseada fuera menor de 20°C (valor que los expertos consideraron dentro de los límites de tolerancia).

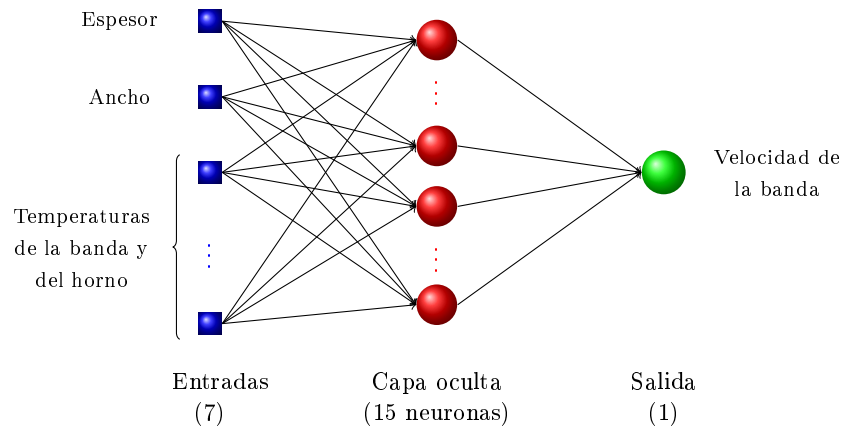


Figura 3.5: Modelo neuronal para la velocidad de la banda en el horno de recocido.

Capítulo 4

Resultados y discusión

En este capítulo se presentan los resultados obtenidos en la elaboración de esta tesis, los cuales pueden dividirse, principalmente, en dos bloques. El primero, incluiría la aplicación de las técnicas de minería de datos en el modelado de una línea de producción de acero galvanizado por inmersión en caliente; mientras que en el segundo, tendría cabida el desarrollo de un nuevo método de entrenamiento neuronal robusto.

4.1. Aplicación de técnicas de minería de datos en el modelado de una línea de acero galvanizado

Durante las primeras fases del proceso de minería de datos aplicadas al modelo de las propiedades mecánicas de las bobinas, esto es, en el análisis y preparación del conjunto de datos, se observó que la variable que medía el alargamiento en el skin-pass, permanecía constante en las muestras disponibles. Puesto que era una variable que no aportaba ningún tipo de información, y teniendo presente la “maldición de la dimensionalidad”, se obvió en el resto del proceso de minería de datos, con lo que el número total de variables de entrada pasó a ser diecisiete.

Además, se localizaron once bobinas que destacaban del resto por sus altos

contenidos en Manganeso, Silicio, Cromo, Carbono equivalente, así como por poseer una mayor resistencia a la rotura. Al ser un número tan poco significativo (representaban un 0,63 % del total) y no pudiendo incluirse en ninguna otra de las clases detectadas, se eliminaron del análisis.

Por último, y tal como se acaba de avanzar, en los datos analizados se encontraron tres clases de comportamiento, lo que quiere decir, que resultaba conveniente desarrollar un modelo para cada una de ellas.

Tras eliminar los datos atípicos y dividir el conjunto de datos en las tres clases detectadas, se observó que uno de los grupos en los que se acababa de dividir la base de datos contenía un número de patrones insuficiente para el entrenamiento adecuado de una red neuronal con diecisiete entradas. Por ello, únicamente se modelaron los dos grupos restantes (los denominados *clase1* y *clase2*). De todas las redes entrenadas para cada caso, las mejores ofrecieron resultados bastante buenos, siendo los errores medios relativos de los patrones de test, es decir, de los nunca vistos por la red durante el entrenamiento, no superiores al 4,5 %: los modelos con un mejor ajuste correspondían a la resistencia a la rotura, con un error medio relativo inferior al 2 %; en el polo opuesto, el modelo para el alargamiento de la *clase2*, poseía un error medio relativo en torno al 4,3 %; y, en un punto “intermedio”, las estimaciones del límite elástico con los patrones de test, no fueron tan buenas como para la resistencia a la rotura, pero en algunos casos mejoraban las obtenidas para el alargamiento.

A la vista de las buenas estimaciones ofrecidas por los modelos creados, éstos fueron utilizados para cuantificar el impacto que los parámetros de proceso (que son los que se pueden controlar para conseguir modificar las características del producto final) tenían sobre estas propiedades mecánicas. El objeto de este análisis no era otro que el de obtener un mayor conocimiento del efecto que el proceso de recocido tiene sobre las propiedades de las bobinas sometidas al proceso de galvanizado. Para llevarlo a cabo, se seleccionaron varios conjuntos de patrones de test; se generaron nuevas muestras, incrementando y disminuyendo las variables del horno de recocido (temperatura primero y velocidad de la banda después), y se aplicaron estos nuevos conjuntos de datos a sus correspondientes modelos neuronales para comprobar los cambios producidos o, dicho de otro modo, para comprobar la sensibilidad de estas propiedades mecánicas a las variaciones de temperatura y velocidad en el horno de recocido.

Únicamente resultó sorprendente la tendencia del límite elástico al variar la velocidad en la *clase1*, ya que por la propia razón de ser del ciclo de recocido,

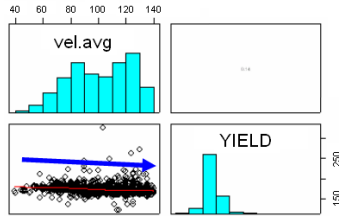


Figura 4.1: Evolución del límite elástico en función de la velocidad en las muestras recogidas para la clase1.

al aumentar la temperatura (o al reducir la velocidad y pasar más tiempo en el horno) el acero se reblandece. El motivo de que esto ocurriera, estaba en la propia distribución de los patrones utilizados para entrenar la red. En efecto, la red aprendió correctamente la evolución de esta propiedad mecánica en función de la velocidad (figura 4.1). Si el número de muestras disponibles en el rango de velocidades bajas fuera mayor, este hecho no habría tenido lugar.

En cuanto al modelo de velocidad de la banda en el horno de recocido, tal y como se podía intuir en un principio, una vez entrenadas las redes, se comprobó el mejor comportamiento del modelo robusto frente al basado en el MSE: con la red robusta, se obtuvo un 4,43 % de error medio relativo con los patrones de test, frente al 5,66 % de la red no robusta. Además, al representar las velocidades estimadas con ambos modelos y compararlas con la real, la obtenida con el modelo robusto permanecía muy próxima a ella, haciendo que el error entre las temperaturas de consigna y real fueran mínimos, mientras que los valores de la red no robusta estaban claramente desviados (por su tendencia a reducir el error cuadrático medio global que producen todos los datos, incluidos los datos contaminados).

4.2. Desarrollo de un nuevo algoritmo de entrenamiento robusto

Para probar el comportamiento de la nueva red neuronal robusta propuesta (de aquí en adelante *TAOnet*), se diseñaron diferentes experimentos y se compararon los resultados obtenidos con los de la red robusta propuesta por Liano (1996) (denominada *LMLSnet*) y los de una red no robusta entrenada con el

algoritmo basado en el *MSE*.

Tales experimentos consistían en aproximar, con una red de dos capas y diez neuronas en la capa oculta, una función ($y = |x|^{2/3}$) a la que se añadieron diferentes tipos de ruido: ruido gaussiano; ruido gaussiano con adición de espurios aleatorios de gran valor en cantidades del 10 y 49 %; y ruido gaussiano en el 51 % de los patrones y ruido de fondo uniformemente distribuido en la región de interés en el 49 % restante.

En el primer caso, con datos de buena calidad y un ruido de fondo pequeño, las respuestas de las tres tipologías de red entrenadas, se aproximaron bastante al modelo esperado en estas condiciones (ausencia de espurios). Para poder evaluar objetivamente el comportamiento de cada red, se empleó un único criterio de error: se calculó la raíz del error cuadrático medio (*Root Mean Square*, RMS) de cada respuesta con respecto al modelo ideal. En este primer experimento, la red que presentó un mejor comportamiento fue la *LMLSnet* con un error final del 0,60 %, seguida de la *TAOnet*, con un error del 0,75 %, y de la *MSE* con un error final del 1,64 %. En cuanto a su convergencia, la más rápida fue la *MSE*, seguida de la *LMLSnet* y, por último, la *TAOnet*.

El siguiente experimento realizado, en el cual se añadía el mismo ruido gaussiano que en el caso anterior, además de un 10 % de espurios aleatorios de gran valor, se observó la bondad de las redes robustas al ser sus respuestas muy próximas a la ideal esperada. En contraposición, la red no robusta *MSE* se vio severamente afectada por los espurios, ofreciendo un modelo erróneo. Los errores finales de entrenamiento fueron muy parecidos para las redes robustas: un 0,63 % para la *LMLSnet* y un 0,71 % para la *TAOnet*; mientras que la red *MSE* presentó un error del 240 %. En cuanto a las velocidades de convergencia, siendo muy similares entre las redes robustas, la *LMLSnet* fue más rápida.

Cuando se estudió el comportamiento de las redes robustas en el límite de su punto de ruptura (50 %), es decir, cuando se añadió, además del ruido gaussiano, un 49 % de espurios aleatorios de gran valor, con las redes robustas se obtuvieron, nuevamente, modelos muy próximos al ideal, mientras que la red no robusta, muy sensible a los datos erróneos, estimó un modelo completamente equivocado. En consecuencia, los errores finales de entrenamiento de las redes robustas fueron muy inferiores al obtenido con la red *MSE*, obteniéndose un error ligeramente más pequeño para la *TAOnet*, un 0,76 %, que para la *LMLSnet*, un 0,78 %. De nuevo, la *LMLSnet* convergió más rápidamente que la *TAOnet*.

4.2 Desarrollo de un nuevo algoritmo de entrenamiento robusto 47

En el último experimento, en el cual el 51 % de los patrones presentaban ruido gaussiano y el 49 % restante ruido de fondo uniformemente distribuido, se comprobó como, a pesar de la existencia del ruido uniformemente distribuido en las cercanías de los datos “buenos”, la *TAOnet* pudo estimar el modelo que correspondía a esos datos. Las otras redes, *LMLSnet* y *MSE*, presentaron considerables desviaciones del modelo ideal. Aunque la *LMLSnet* convergió, de nuevo, más rápidamente que la *TAOnet*, la única red que ofreció un comportamiento adecuado, fue esta última, con un error del 0,98 % frente al 5,2 % y 70 % de la *LMLSnet* y *MSE*, respectivamente.

En resumen, los diferentes experimentos analizados, muestran las principales propiedades del algoritmo de aprendizaje *TAO-robust* propuesto: robustez frente a espurios y alto punto de ruptura. En la mayoría de los casos, su comportamiento fue muy similar al de la red robusta propuesta por Liano (1996), la cual convergía más rápidamente en todos los entrenamientos realizados. Sin embargo, cuando los patrones contenían ruido de fondo no estructurado, el comportamiento de la red neuronal robusta *TAOnet* fue superior a las otras, siendo, por tanto, recomendable su uso en estas condiciones de ruido.

Capítulo 5

Conclusiones

5.1. Conclusiones

En este trabajo se ha mostrado la utilidad de las redes neuronales, junto con otras técnicas y herramientas procedentes de la minería de datos, en el modelado de una línea de producción de acero galvanizado por inmersión, a partir de los datos del mismo. Los resultados obtenidos han servido para resaltar las principales ventajas que aportan las redes neuronales en el modelado de procesos industriales: versatilidad para adaptarse a diversas funciones, efectividad para modelar procesos no lineales, aplicación intuitiva y facilidad de implementación. Pero, también, han servido para poner de manifiesto la necesidad de tratar los espurios, los cuales, en mayor o menor medida, siempre se encuentran presentes en los datos recogidos y almacenados, y, en este sentido, resaltar la eficacia de las redes neuronales robustas sobre las no robustas.

En primer lugar, se modelaron las propiedades mecánicas de las bobinas de acero galvanizado. Aunque cabría esperar mejores modelos con un mayor número de patrones de entrenamiento, los buenos resultados de predicción obtenidos demuestran la posibilidad de estimar *on-line* las características de las bobinas fabricadas. Las ventajas de estos modelos son evidentes, si se tiene en cuenta que la medición de estas propiedades supone un coste económico y temporal: se podrían tomar decisiones adecuadas en tiempo real de una forma más económica.

Otra ventaja derivada de estos modelos proviene de su sencillez de imple-

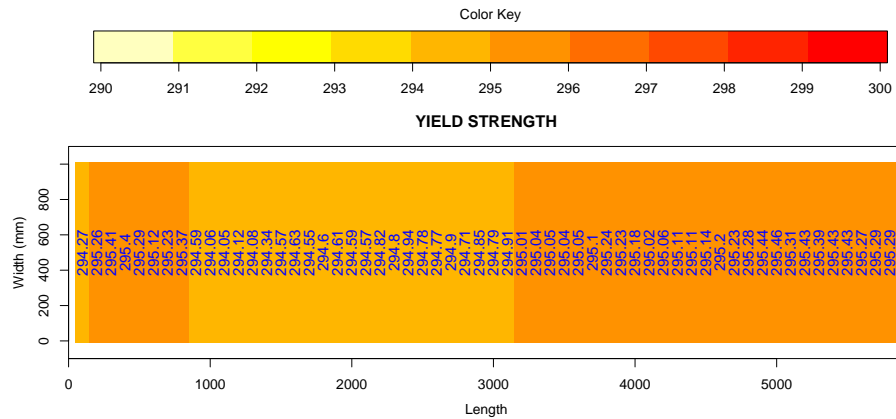


Figura 5.1: Ejemplo de mapa del límite elástico estimado para una bobina.

mentación en la línea de producción, ya que el *software* utilizado para el entrenamiento de las redes, SNNS (*Stuttgart Neural Network Simulator*)¹, posee una herramienta que genera el código C de una red neuronal, el cual puede ser fácilmente integrable en el sistema informático existente en la empresa.

Por último, teniendo en cuenta que las variables de proceso, en la línea objeto de estudio, se miden cada 100 metros, con los modelos propuestos se puede construir un mapa de características de cada bobina (como la de la figura 5.1). Una de las ventajas que aporta este mapa, es que permite conocer qué metros de una bobina cumplen con los requisitos del cliente y cuales no, con el ahorro económico que supone puesto que en lugar de desechar toda la bobina, se cortará sólo aquella zona que no cumpla con los requerimientos.

El segundo modelo creado, el de la velocidad de la banda en el horno de recocido, también ofreció buenos resultados, con lo que su implementación en la línea de fabricación, serviría para controlar de un modo más eficiente la temperatura de la bobina a la salida del horno, consiguiendo, de esta manera, temperaturas de salida más adecuadas que las obtenidas al cambiar únicamente la temperatura de consigna del horno.

Además, este modelo podría ser utilizado para establecer estrategias de funcionamiento de la línea, para establecer el orden en el que procesar las bobinas, previsualizar las condiciones de velocidad de la línea en condiciones transitorias, etc.

¹<http://www-ra.informatik.uni-tuebingen.de/SNNS/>

Aunque las mejoras individuales propuestas pueden considerarse “pequeñas” en el conjunto de la línea, el global puede resultar en una ventaja competitiva para la industria que las adopte.

Por otro lado, se han mostrado las importantes propiedades que posee el nuevo algoritmo robusto de entrenamiento desarrollado: el algoritmo de aprendizaje por retropropagación *TAO-robust*. Asimismo, se ha demostrado su superioridad frente a otros algoritmos robustos (como el LMLS), en el caso de presentar patrones de entrenamiento con ruido de fondo.

La combinación del estimador τ no lineal con el algoritmo de entrenamiento de retropropagación del error, supone una innovación en las redes neuronales robustas, al solucionar el problema, encontrado en otras redes propuestas, de seleccionar el factor de escala que determina los límites para reducir los efectos de los espurios.

5.2. Líneas de trabajo futuras

Es importante recordar que ningún modelo o algoritmo puede o debe ser usado de modo exclusivo, y que no existe el “mejor” modelo o algoritmo para un problema dado, ya que la propia naturaleza de los datos afectará a la elección de los modelos. En este sentido, resulta recomendable utilizar otras herramientas y tecnologías, como las máquinas de vectores soporte (*Support Vector Machines*, SVM), con objeto de obtener los mejores modelos posibles.

También resulta interesante aprovechar las nuevas percepciones, ideas, etc., que se generan a lo largo del proceso de minería de datos. En concreto, los modelos de predicción de las propiedades mecánicas de las bobinas, a partir de los datos de proceso, pueden utilizarse para obtener un clasificador de la calidad de las bobinas fabricadas, en función del mapa de características de cada bobina y de los requisitos impuestos por el cliente.

Del mismo modo, los buenos resultados obtenidos en la predicción de las propiedades mecánicas, hacen pensar en la búsqueda de soluciones para otros problemas relacionados. Tal es el caso, por ejemplo, de la creación de un “cerrojo artificial”, en el que ya se está trabajando², con el fin de hallar una solución

²Las primeras aproximaciones se basan en un modelo neuronal, que predice el alargamiento de las bobinas en el skin-pass en función de algunas variables del proceso de fabricación y de su composición química. De esta forma, si la diferencia entre el alargamiento que estima el modelo y el medido realmente es significativa, se hace necesario sacar la bobina de la línea

a un problema que se presenta esporádicamente, pero cuyas consecuencias pueden llegar a ser graves: el etiquetado incorrecto del grado de acero de una bobina. Bajo estas circunstancias, una bobina es tratada como si tuviera una composición química que en realidad no posee y, dado que, generalmente, estas bobinas de acero galvanizado sufren posteriores transformaciones antes de convertirse en producto final, si, por error, se envía a un cliente una bobina de un material más duro de lo esperado, se pueden ocasionar daños importantes en la maquinaria del mismo, tales como rotura de prensas, etc.

Hasta ahora, únicamente se ha tratado la problemática de una línea de producción de acero galvanizado, pero resultaría muy interesante, también, verificar la utilidad de las técnicas de minería de datos en el modelado de otros procesos industriales.

En cuanto a las redes neuronales, futuros trabajos podrían ir dirigidos a desarrollar un sistema que permita cuantificar el grado de robustez de las mismas.

Bibliografía

- Ahmed, M. and A. Farag: 2002, 'A neural approach to zoom-lens camera calibration from data with outliers'. *Image and Vision Computing* **20**, 619–630. 1, 2, 3.2.3
- Bellman, R.: 1961, *Adaptive Control Processes: A Guided Tour*. Princeton University Press. 2
- Bishop, C. M.: 1995, *Neural Networks for Pattern Recognition*. Oxford University Press. 1
- Brachman, R. J. and T. Anand: 1996, *The Process of Knowledge Discovery in Databases*, Chapt. 2: Advances in knowledge discovery and data mining. AAAI Press/The MIT Press. 1
- Burdekin, M.: 1992, *Steel Designers' Manual*, Chapt. Applied metallurgy of steel. Cambridge: Blackwell Science. 3
- Callister, W.: 1996, *Introducción a la ciencia e ingeniería de los materiales*. Barcelona: Reverté. 2
- Castejón Limas, M., J. B. Ordieres Meré, F. J. Martínez de Pisón Ascacibar, and E. P. Vergara González: September, 2004, 'Outlier detection and data cleaning in multivariate *non*-normal samples. The *PAELLA* algorithm'. *Data Mining and Knowledge Discovery* **9**, 171–187. 1
- Chapman, P., J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth: 2000, 'CRISP-DM 1.0: Step-by-step data mining guide'. Technical

- Report CRISPWP-0800, CRISP-DM consortium / SPSS Inc. (document), 1.1, 1, 1
- Chen, D. and R. Jain: 1994, 'A robust backpropagation learning algorithm for function approximation'. *IEEE Transactions on Neural Networks* **5**, 467–479. 1, 2, 3.2.3
- Cherkassky, V. and F. Mulier: 1998, *Learning From Data: concepts, theory and methods*. John Wiley & Sons, Inc. 1
- Chuang, C., S. Su, and C. Hsiao: 2000, 'The annealing robust backpropagation (ARBP) learning algorithm'. *IEEE Transactions on Neural Networks* **11(5)**, 1067–1077. 1, 2, 3.2.3
- Cios, K., W. Pedrycz, and R. Swiniarski: 1998, *Data Mining Methods for Knowledge Discovery*. Kluwer Academic Publishers. 1
- Connor, J., R. Martin, and L. Atlas: 1994, 'Recurrent neural networks and robust time series prediction'. *IEEE Transactions on Neural Networks* **5**, 240–254. 2
- Dunteman, G.: 1989, *Principal components analysis*. Sage Publications. 3.2.3
- Fahlman, S.: 1988, 'An empirical study of learning speed in backpropagation networks'. Technical Report CMU-CS-88-162, Carnegie-Mellon University. 1
- Fayyad, U., G. Piatetsky-Shapiro, and P. Smyth: 1996, 'Knowledge discovery and data mining: Towards a unifying framework'. *Proceeding of The Second International Conference on Knowledge Discovery and Data Mining (KDD'96)* pp. 82–88. 1
- Fletcher, R.: 1987, *Practical methods of optimization*. Wiley. 2
- Funahashi, K.: 1989, 'On the approximate realization of continuous mapping by neural networks'. *Neural Networks* **2**, 183–192. 2
- Hampel, F., E. Ronchetti, P. Rousseeuw, and W. Stahel: 1986, *Robust statistics, the approach based of influence functions*. Wiley. 1
- Hawkins, D.: 1980, *Identifications of Outliers*. New York: Chapman and Hall. 2

- Haykin, S.: 1999, *Neural networks. A comprehensive foundation*. Prentice Hall. 2, 2
- Hornik, K., M. Stinchcombe, and H. White: 1989, 'Multilayer feedforward networks are universal approximators'. *Neural Networks* **2(5)**, 359–366. 2
- Huang, B., B. Zhang, and Q. Huang: 1998, 'Robust interval regression analysis using neural network'. *Fuzzy Sets Systems* pp. 337–347. 1, 2, 3.2.3
- Huber, P. J.: 1981, *Robust Statistics*. New York: John Wiley and Sons. 2
- Huber, P. J.: 2003, *Robust Statistics*. New York: John Wiley and Sons. 2
- LeCun, Y.: 1993, *Efficient learning and second-order methods. A tutorial at NIPS 93*. Denver. 2
- LeCun, Y., L. Bottou, Y. Bengio, and P. Haffner: 1998, 'Gradient-based learning applied to document recognition'. *Proceedings of the IEEE* **86**. 2
- Liano, K.: 1996, 'Robust error measure for supervised neural network learning with outliers'. *IEEE Transactions on Neural Networks* **7(1)**, 246–250. 1, 2, 3.2.3, 4
- Martínez de Pisón Ascacíbar, F. J.: 2003, 'Optimización mediante técnicas de minería de datos del ciclo de precocido de una línea de galvanizado'. Ph.D. thesis, Universidad de La Rioja. 1, 3.2.3
- McDowall, T. and F. Ham: 1997, 'Robust partial least-squares regression: A modular neural network approach'. *Proceedings: SPIE* **3077**, 344–355. 1, 2, 3.2.3
- Moody, J. and C. Darken: 1989, 'Fast learning networks of locally-tuned processing units'. *Neural Computation* **1**, 281–294. 1
- Ordieres Meré, J. B., A. González Marcos, J. A. González, and V. Lobato Rubio: 2004, 'Estimation of mechanical properties of steel strips in hot dip galvanizing lines.'. *Ironmaking and Steelmaking* **31(1)**, 43–50. 1, 2, 3, 3.2.3
- Pernía Espinoza, A. V., M. Castejón Limas, A. González Marcos, and V. Lobato Rubio: 2005a, 'Steel annealing furnace robust neural network model.'. *Ironmaking and Steelmaking* **32(5)**, 418–426. 1, 2, 2, 3, 3.2.3

- Pernía Espinoza, A. V., J. B. Ordieres Meré, F. J. Martínez de Pisón, and A. González Marcos: 2005b, 'TAO-robust backpropagation learning algorithm'. *Neural Networks* **18(2)**, 191–204. 1, 2
- Platt, J.: 1991, 'A resource-allocating network for function interpolation'. *Neural Computation* **3**, 213–225. 1
- Pyle, D.: 1999, *Data Preparation for Data Mining*. Morgan Kaufmann Publishers. 1, 1
- Ripley, B.: 1996, *Pattern Recognition and Neural Networks*. Cambridge University Press. 1
- Rousseeuw, P. and V. Yohai: 1984, *Robust regression by means of S-estimators*, pp. 256–272. Springer. 2
- Rousseeuw, P. J. and A. Leroy: 1987, *Robust Regression and Outlier Detection: Diagnostic Regression Analysis*. New York: John Wiley and Sons. 1, 2
- Rumelhart, D., J. McClelland, and the PDP research group: 1986, *Parallel Distributed Processing*. MIT Press. 1, 2
- Sammon, J.: 1969, 'A nonlinear mapping for data structure analysis'. *IEEE Transactions on Computers* **18**, 401–409. 3.2.3
- Tabatabai, M. and I. Argyros: 1993, 'Robust estimation and testing for general nonlinear regression models'. *Applied Mathematics and Computation* **57**, 85–101. 2, 2, 2, 2
- Tollenaere, T.: 1990, 'Supersab: Fast adaptive backpropagation with good scaling properties'. *Neural Networks* **3**, 561–573. 1
- Vergara González, E. P.: 1999, 'Modelo de control inteligente de espesor de recubrimiento en galvanizado continuo por inmersión'. Ph.D. thesis, Universidad de Oviedo. 4
- Walczak, B.: 1996, 'Neural networks with robust backpropagation learning algorithm'. *Analytica Chimica Acta* **322**, 21–29. 2
- Yohai, V. and R. Zamar: 1988, 'High breakdown-point estimates of regression by means of the minimization of an efficient scale'. *Journal of the American Statistical Association* **83(402)**, 406–413. 2, 2