

Reducción de problemas de adherencia en procesos de galvanizado mediante técnicas de minería de datos^(·)

F. J. Martínez-de-Pisón*, J. Ordieres*, A. Pernía*, F. Alba* y V. Torre**

Resumen

En este trabajo, se muestra un ejemplo de aplicación de técnicas de minería de datos para la obtención de conocimiento oculto a partir de los históricos de un proceso de galvanizado. El objetivo consistió en encontrar reglas que pudieran servir para mejorar la calidad del producto final y reducir los fallos del proceso. Para ello, se partió de los históricos correspondientes a la etapa de ajuste de una línea de galvanizado donde surgieron bobinas con problemas en la adherencia del recubrimiento de zinc aplicado. A partir de la base de datos de dicho proceso, se aplicó la metodología clásica de minería de datos para generar y analizar diversos árboles de decisión que clasificaban dos tipos de clases: las bobinas con adherencia correcta y las que presentaban adherencia irregular. De estos árboles se extrajeron las variables y sus valores, que más podían influir en la calidad del recubrimiento. Además, se establecieron reglas que podían ser aplicadas para reducir el número de bobinas con fallos de adherencia.

Palabras clave

Árboles de decisión. Línea de galvanizado. Minería de datos. Búsqueda de conocimiento en bases de datos. Mejora de procesos industriales. Análisis de los parámetros del proceso.

Reduce of adherence problems in galvanised processes through data mining techniques

Abstract

This paper presents an example of the application of data mining techniques to obtain hidden knowledge from the historical data of a hot dip galvanizing process and to establish rules to improve quality in the final product and to reduce errors in the process. For this purpose, the tuning records of a hot dip galvanizing line where coils with adherence problems in the zinc coating had been identified were used as a starting point. From the database of the process, the classical data mining approach was applied to obtain and analyze a number of decision trees that classified two types of coils, i.e. those with the right adherence and those with irregular adherence. The variables and values that might have influenced the quality of the coating were extracted from these trees. Several rules that may be applied to reduce the number of faulty coils with adherence problems were also established.

Keywords

Decision trees. Hot dip galvanizing line. Data mining. Database knowledge discovery. Industrial process improvement. Process parameter analysis.

1. INTRODUCCIÓN

La Minería de Datos (MD) es una disciplina que engloba un conjunto de metodologías y herramientas que son útiles para la búsqueda de conocimiento oculto en bases de datos. En estos últimos años, esta disciplina se ha aplicado a multitud de áreas con un éxito sin precedentes: banca, negocios, bolsa, seguros, energía, industria, etc.

El arranque de nuevas plantas de producción, el procesado de nuevos tipos de productos o el reajuste

de las condiciones originales de producción suele implicar gran cantidad de esfuerzo humano y mucho tiempo y dinero. En estos casos, MD puede ayudar a obtener conocimiento que sirva para comprender mejor el sistema a optimizar y, de esta forma, reducir costos y tiempos.

En este artículo, se reflexiona sobre el uso de MD para mejorar los procesos de fabricación y transformación del acero no solo mediante la creación de modelos de control o predicción sino también mediante la búsqueda de conocimiento oculto. Como

(·) Trabajo recibido el día 1 de junio de 2006 y aceptado en su forma final el día 25 de enero de 2007.

* Grupo EDMANS (<http://www.mineriadatos.org>), Área de Proyectos de Ingeniería, Departamento de Ingeniería Mecánica, Universidad de La Rioja, C/ Luis de Ulloa, 20, 26004, Logroño (España).

** Aceralia. Grupo ARCELOR MITTAL ESPAÑA, I+D+i. Centro de Desarrollo Tecnológico. Ap. 90, 33480 Avilés. España.

caso práctico, se muestra una experiencia de MD donde se realizó una búsqueda de las causas que originaban pérdidas de calidad en el recubrimiento de nuevos aceros dentro de una planta de galvanizado. Del estudio se obtuvieron reglas de conocimiento que sirvieron para detectar qué parámetros afectaban a la calidad del recubrimiento de las bobinas y cuáles eran los márgenes de control que se podían establecer como medida de seguridad para reducir los problemas aparecidos. De esta forma, se consiguió reducir ostensiblemente los tiempos necesarios de ajuste del proceso ante el nuevo producto reduciendo los costes y aumentando la seguridad y conocimiento del proceso.

En el presente trabajo se describirán las pautas utilizadas para la obtención de dicho conocimiento, las cuales pueden usarse en cualquier otro proceso de fabricación del acero. Para ello, primeramente se realizará una descripción del problema para pasar después a enumerar brevemente las técnicas de minería de datos y su aplicación en la industria del acero. Posteriormente, se presentará la metodología usada, los resultados obtenidos y las conclusiones finales.

2. PLANTEAMIENTO DEL PROBLEMA

El planteamiento del trabajo consistió en identificar la influencia de los parámetros del proceso de galvanizado en la adherencia del recubrimiento de zinc. La calidad, espesor y uniformidad del recubrimiento, son factores de los que depende la resistencia a la corrosión del producto galvanizado^[1-4] y se ve influenciado por múltiples factores^[1 y 5]: preparación de la superficie del metal, composición y temperatura del baño de zinc, velocidad de la banda, control de las cuchillas de aire que regulan el espesor del recubrimiento, temperatura de la banda, calidad del ciclo de recocido, composición de la atmósfera, etc.

El ajuste de todos los parámetros correspondientes a cada tipo de bobina según el tipo de acero y dimensiones de la misma (espesor y anchura), hacen esta tarea muy laboriosa. Además, los problemas se acentúan, aún más, cuando se procesan nuevos aceros o nuevos espesores que no han sido tratados anteriormente y de los que no se dispone de información veraz ni de modelos matemáticos adecuados.

Ante estos nuevos productos, los ingenieros de planta deben realizar estimaciones y múltiples ajustes hasta conseguir que la adherencia y homogeneidad del recubrimiento de zinc sea el adecuado. Esto supone un consumo muy elevado de tiempo y material desechado.

Muchas veces, el conocimiento implícito que se genera en estos trabajos no queda plasmado en ningún

sitio pues, al ser tantas las variables a modificar y ajustar, resulta muy difícil determinar cuáles son realmente cruciales en la calidad del producto final.

Es, en este momento, donde las técnicas de minería de datos pueden ayudar a descubrir reglas ocultas que “expliquen” los problemas de adherencia en las bobinas mediante el análisis de los históricos del proceso. De este modo, los expertos de la planta pueden establecer nuevas estrategias de control que permitan reducir el número de bobinas con adherencia irregular.

El problema no es fácil, pues además del enorme número de variables que influyen en el proceso, el número de bobinas defectuosas existentes en las bases de datos a analizar no suele ser muy elevado, lo que dificulta mucho el análisis con técnicas clásicas de estadística multivariante.

En este trabajo se demostrará que el uso de árboles de decisión combinados permite abordar este tipo de problemas de una forma bastante eficiente. A diferencia de otras técnicas de minería de datos los árboles de decisión permiten trabajar con gran cantidad de atributos (alta dimensionalidad) y pocos casos (baja densidad).

2.1. La línea de galvanizado en continuo

A grandes rasgos, el proceso dentro de una línea de galvanizado en continuo se puede describir de la siguiente manera (Fig. 1):

- El primer paso, consiste en la formación de una banda continua a partir de las bobinas de acero que proceden de los trenes de laminación. Para ello, se despunta la cabeza y la cola de las mismas y se sueldan a solape. El resultado final es una banda de acero continua formada por las bobinas entrantes.
- A continuación, la banda atraviesa una zona de precalentamiento en atmósfera no oxidante donde se eliminan las impurezas, se volatilizan los aceites de laminación y se reduce el óxido superficial.
- Posteriormente, se somete la banda a un ciclo de calentamiento y enfriamiento que se denomina “recocido” (Fig. 2). Este tratamiento es esencial para la mejora de las propiedades del acero y del recubrimiento final. El objetivo es recristalizar el metal endurecido que sale de la laminación en frío y homogeneizar la estructura cristalina.
- El horno donde se realiza este tratamiento, suele estar dividido en varias zonas: una zona de calentamiento formada por ocho subzonas,

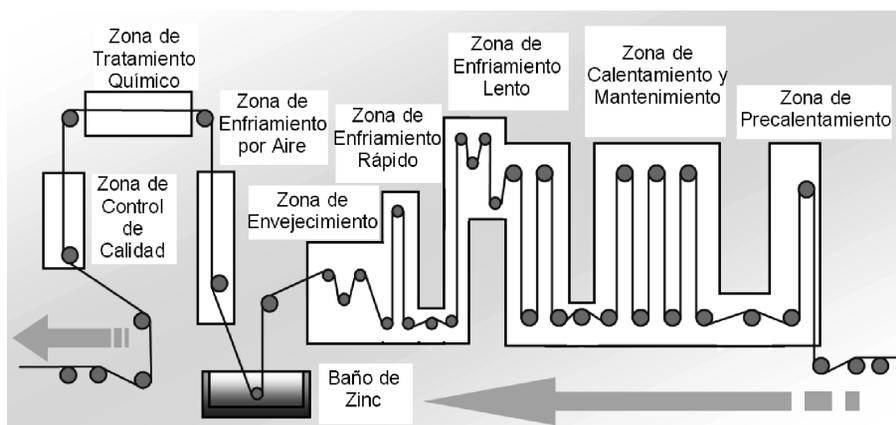


Figura. 1. Esquema básico de una línea de galvanizado en Caliente (LGC).

Figure. 1. Basic scheme of an hot dip galvanizing Line (HDGL).

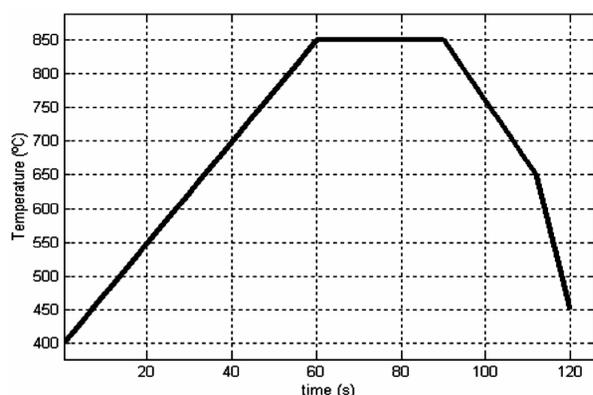


Figura. 2. Curva de tratamiento térmico.

Figure. 2. Thermal treatment curve.

una zona de mantenimiento, una de enfriamiento lento y otra de enfriamiento rápido. Al final del enfriamiento rápido se produce un envejecimiento o “igualación” con el fin de garantizar la precipitación del carbono y así minimizar los efectos de envejecimiento del acero.

- A continuación, la banda se sumerge en un pote de zinc fundido a temperatura constante para revestirla de dicho metal. De este baño, la banda sale verticalmente, pasando entre cuchillas de aire que regulan el espesor del recubrimiento.
- Después, atraviesa una serie de procesos auxiliares de tratamientos químicos donde se aplica una leve película de ácido crómico para prevenir la oxidación blanda.
- Por último, se realiza un aplanado hasta obtener el producto final bien en forma de bobinas o chapas cortadas.

— Esta descripción, con modificaciones menores, suele ser válida para la mayoría de líneas de galvanizado en continuo por inmersión instaladas en todo el mundo.

3. METODOLOGÍA: LAS TÉCNICAS DE MINERÍA DE DATOS

Las técnicas de MD más comúnmente usadas son^[6]:

- Modelización estadística paramétrica y no paramétrica.
- Redes neuronales.
- Árboles de decisión y sistemas de reglas.
- Máquinas de vectores soporte.
- Métodos bayesianos.
- Reglas de asociación.
- Métodos relacionales y estructurales.
- Algoritmos evolutivos y reglas difusas.
- Métodos basados en casos y vecindad.

Además, se deberían incluir todas aquellas técnicas y métodos necesarios para el análisis exploratorio y preprocesado de la información, pues son muy importantes y suelen desarrollarse antes de la etapa de modelizado^[3] (Fig. 3).

La mayoría de los trabajos publicados en la optimización de procesos industriales utilizan los dos primeros grupos de técnicas: métodos estadísticos y redes neuronales.

Una lectura muy interesante para comprender las posibilidades de dichas técnicas se encuentra en Hernández-Orallo *et al.*^[6], donde se realiza una muy buena introducción en la metodología y algoritmos de minería de datos.

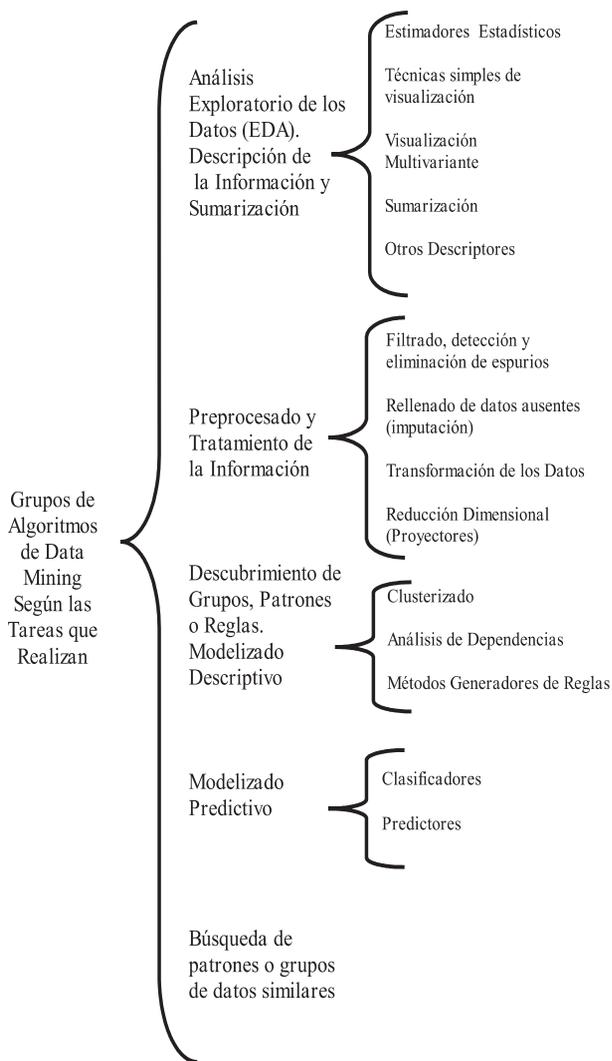


Figura. 3. Clasificación de las técnicas de minería de datos según la tarea que realizan.

Figure. 3. Data mining techniques classification according to the task they fulfil.

3.1. Los árboles de decisión y algoritmos generadores de reglas

Los árboles de decisión y los algoritmos generadores de reglas son unos de los algoritmos clasificadores más conocidos y usados en las tareas de MD. Éstos generan modelos formados por conjuntos de condiciones que se plasman en forma de árboles o reglas fácilmente comprensibles.

Los árboles, a diferencia de los algoritmos generadores de reglas, muestran las condiciones organizadas en una forma jerárquica, de forma que la clase final o decisión a tomar se puede obtener siguiendo las condiciones que se cumplen desde la raíz del árbol (Fig. 4).

Cabe destacar, que los algoritmos generadores de reglas y árboles son muy similares entre si. Muchas veces, se pueden extraer reglas directamente de los árboles o, al contrario, generar árboles de decisión a partir de un sistema de reglas.

La diferencia fundamental es que los sistemas de reglas no exigen exclusión ni exhaustividad, por lo que puede aplicarse más de una regla a la vez o ninguna de ellas. Además, se diferencian en la forma de búsqueda de las condiciones de creación: por partición del espacio de instancias o por cobertura.

Generalmente, los árboles de decisión realizan una búsqueda dentro del espacio de instancias de la variable (y su valor) que mejor particionan dicho espacio en clases disjuntas. Este proceso, denominado "búsqueda por partición", se realiza en fases sucesivas hasta que las clases son separadas adecuadamente (Fig. 4).

Otros algoritmos, en cambio, realizan una búsqueda de la regla que mayores ejemplos cubre. Una vez encontrada, se eliminan los ejemplos cubiertos y se siguen buscando nuevas reglas que cubran los

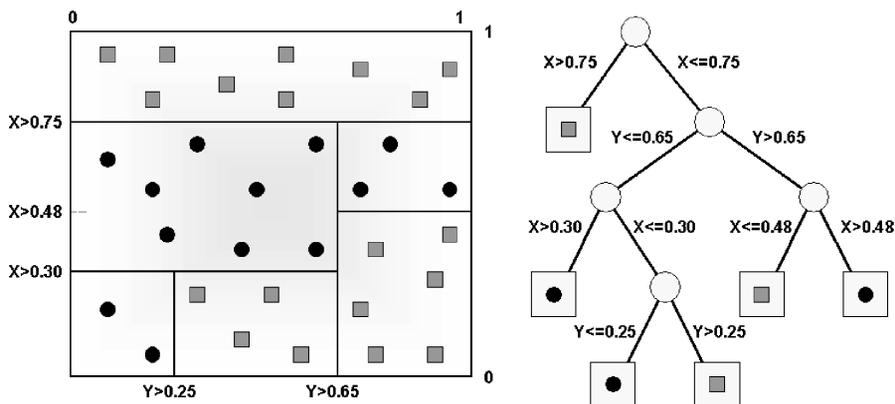


Figura. 4. Ejemplo de árbol de decisión para dos clases.

Figure. 4. Example of a decision tree for two classes.

ejemplos todavía no explicados. Este proceso se denomina "búsqueda por cobertura" y es muy común en los algoritmos generadores de reglas.

Los algoritmos^[6] actuales generadores de árboles más importantes como CART, ID3, ASSISTANT y C4.5 se diferencian fundamentalmente en los criterios de partición utilizados. Uno de los más robustos y eficientes, corresponde con el algoritmo C4.5^[7] o su versión comercial C5.0. Este algoritmo utiliza criterios de partición basados en el cálculo de la entropía aunque con características de robustez mayores ante ruido y valores anómalos o inexistentes.

Hoy en día, la mayoría de herramientas comerciales de MD como SAS, Enterprise Miner, Clementine, KXEN, etc.; o herramientas libres como WEKA, YALE, etc.; incorporan este algoritmo o variantes del mismo.

La ventaja fundamental de los árboles de decisión es que son muy fáciles de utilizar, son robustos, pueden trabajar con un elevado número de variables de entrada y no necesitan un elevado número de ejemplos para trabajar. Estas cuatro características los hacen adecuados para el análisis de información de procesos industriales donde el número de parámetros a estudiar es elevado y el conjunto de ejemplos erróneos suele ser pequeño.

Cabe destacar, como desventaja fundamental, que el grado de precisión de los mismos no es tan elevado como otros algoritmos y que, si los grupos de casos son muy dispersos, pueden generarse árboles muy grandes y difíciles de interpretar.

3.2. La Minería de Datos en la industria del acero

En la industria del acero es muy común el uso de técnicas clásicas para desarrollar modelos matemáticos que expliquen el comportamiento de un producto o proceso^[8-11] pero, cada vez más, se utilizan técnicas de MD o de Inteligencia Artificial (IA) para obtener modelos más precisos que sean capaces de obtener mejores consignas de control, mayor información útil para otros procesos, conocimiento de lo que está pasando, etc.

Dentro de las herramientas de MD más utilizadas en procesos industriales del acero destacan las redes neuronales. Éstas son técnicas que permiten obtener modelos no lineales con precisión mayor que otras técnicas clásicas como regresión, modelos lineales, etc.

En la literatura aparecen muchas experiencias donde se usan redes neuronales. Por ejemplo, estas técnicas se utilizan para determinar las mejores consignas de un horno de una línea de galvanizado^[13], predecir temperaturas de arrabio de un horno alto^[14], determi-

nar las propiedades metálicas de la banda en procesos del acero^[15 y 16], predecir el *end-point* de un convertidor^[17], deducir las propiedades de un acero ante altas temperaturas^[18], determinar la rugosidad final de la banda de acero^[19], entre otras muchas aplicaciones.

Otras veces, se combinan redes neuronales con otro tipo de técnicas para optimizar los modelos obtenidos. Por ejemplo, en Martínez-de-Pisón *et al.*^[5] se usan algoritmos genéticos para optimizar los modelos de redes neuronales usados para determinar las consignas del proceso u otras veces se usan algoritmos genéticos para optimizar la arquitectura de la red neuronal de un proceso de laminado en caliente^[20], etc.

Si bien existen numerosos casos prácticos del uso de redes neuronales en la industria del acero, prácticamente es inexistente el uso de otras técnicas como las máquinas de vectores soporte, redes bayesianas, lógica fuzzy, etc. Solamente aparecen algunos pocos trabajos donde, por ejemplo, se usan redes bayesianas para clasificar defectos superficiales en bloques de acero^[21], lógica difusa o neuro-difusa para controlar la tensión de los rodillos en una línea de laminado^[22 y 23] o la velocidad de los mismos^[24], etc., aunque realmente el número de aplicaciones encontradas no es muy elevado.

Como podemos observar, la mayoría de las experiencias se basan en el desarrollo de modelos matemáticos para controlar diversas partes del proceso industrial. En cambio, el uso de estas técnicas para la búsqueda de conocimiento oculto es un campo prácticamente no explotado en el ámbito industrial del acero, al contrario que en otras áreas como la banca, los seguros o el comercio, donde aparece como uno de los usos principales de MD.

En los apartados siguientes proponemos el uso de las técnicas de MD para la búsqueda de conocimiento oculto en históricos de procesos del galvanizado del acero. Para ello, se mostrará el uso de árboles de decisión que permiten la generación de reglas que puedan ayudar a mejorar dichos procesos.

4. CASO PRÁCTICO

El caso práctico que se presenta, consistió en la búsqueda de conocimiento oculto que pudiera explicar los problemas de adherencia que surgían cuando se trataban aceros poco comunes dentro de una de las líneas de galvanizado del Grupo ARCELOR MITTAL sita en Avilés (España).

Este procedimiento estaba formado por un proceso iterativo de creación y análisis de árboles de decisión de forma que, para cada modelo obtenido, se analizaba la información correspondiente de aquellas ramas del árbol que mejor separaban las bobinas

con problemas de adherencia de las que no tenían. Posteriormente, se visualizaban los resultados obtenidos y se contrastaban éstos con las opiniones de los tecnólogos de la planta.

Obviamente, esta metodología puede ser extrapolada con pocos cambios a cualquier otra línea de galvanizado o proceso de fabricación del acero.

El estudio se realizó para una partida de bobinas de un nuevo tipo de acero que había presentado un cierto número de ellas con irregularidades en la adherencia de la capa de zinc. La base de datos se extrajo de las primeras pruebas de ajuste de la planta por lo que el porcentaje de bobinas con adherencia irregular fue bastante significativo.

Obviamente, el análisis de la base de datos proveniente de los primeros ajustes es muy útil pues la variedad de consignas diferentes debidas a los numerosos ajustes permite explorar más conocimiento que en procesos ya ajustados y en régimen permanente, donde dichas consignas prácticamente no varían.

En la tabla I se muestra el número y porcentaje de bobinas con adherencias de diferente naturaleza de la base de datos que se usó en el estudio. Para simplificar las explicaciones designaremos dos clases de bobinas según esta característica, teniéndose así bobinas “TIPO A” y “TIPO B”.

4.1. Selección de variables y preprocesado de la información

Para la aplicación de cualquier proceso de MD, es necesario realizar una selección de variables, preprocesado y transformación de la información (Fig. 5).

Las variables que se seleccionaron, correspondían con medidas de cada una de las zonas del proceso (Tabla II) efectuadas cada 30 o 100 m de banda. Para cada una de estas zonas, se obtuvieron valores correspondientes a la composición del aire (en porcentaje de H₂ y O₂), la temperatura de cada zona, la tempe-

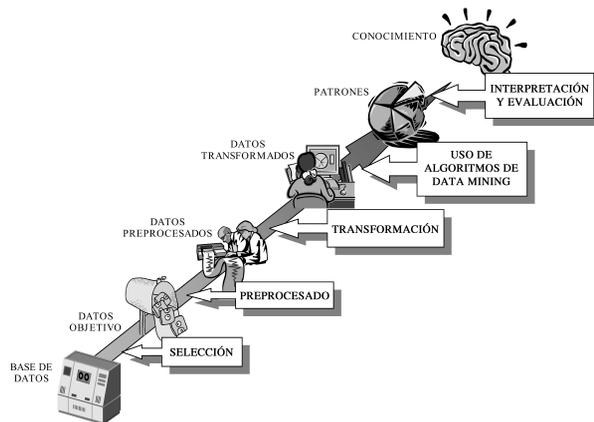


Figura. 5. Proceso clásico de minería de datos.

Figure. 5. Classic data mining process.

ratura de la banda de acero medida en varios sitios y obtenida mediante pirómetros, la velocidad de la banda, la temperatura de rocío en diversas zonas del proceso, etc. (Tabla III). Además, se incluyeron las dimensiones de cada bobina (ancho y espesor), composición química del acero de cada bobina procesada, así como la temperatura y composición del baño de zinc.

Después de la selección, se eliminaron los datos anómalos y se filtraron la gran mayoría de datos inexistentes para conseguir que cada variable estuviese formada, en su mayor parte, por datos consistentes.

Por último, se obtuvieron para cada bobina los siguientes estimadores estadísticos:

- El valor máximo (MX) y mínimo (MN) de cada variable.
- El valor medio (AV) y la desviación estándar (SD).

Los primeros daban una idea de los valores extremos mientras que los segundos aportaban una estimación del valor central y el grado de dispersión de la variable. Por ejemplo, un valor elevado en la desviación estándar de la temperatura de una zona de horno, para una bobina determinada, indicaba que el tratamiento térmico que se realizó en esa zona fue poco homogéneo.

Como variable de salida se definió un parámetro denominado “ADHERENCIA” que dividía a las bobinas en dos clases: “TIPO A” y las “TIPO B”, según la tipología en la adherencia del recubrimiento de zinc.

Para cada acero, y como paso previo al proceso de MD, se obtuvo una tabla con más de 200 variables de entrada y una de salida; siendo cada instancia de la misma, una de las bobinas procesadas.

Tabla I. Porcentaje de bobinas con adherencia tipo “A” e “B”

Table I. Percentage of coils with “A” and “B” adherence types

PORCENTAJE DE BOBINAS DE LA BASE DE DATOS		
Bobinas TIPO A (Adherencia tipo A)	Bobinas TIPO B (Adherencia tipo B)	Total
684 (94,6%)	39 (5,4%)	723

Tabla II. Zonas de medición dentro del HDGL

Table II. HDGL measure zones.

ZONAS PRINCIPALES DEL HDGL	
Zona	Descripción
Pre calentamiento (PRE)	Zona de pre calentamiento de la banda, donde se realiza la limpieza de la misma.
Calentamiento: subzonas 1 a 8 (CAL)	Subzonas 1 a 8 de la zona de calentamiento del horno donde se aumenta la temperatura de la banda entorno a los 850 °C.
Mantenimiento: subzonas 9 y 10 (MAN)	Zona donde se mantiene la temperatura de la banda entorno a los 850 °C: Subzonas 9 y 10.
Enfriamiento lento (ENL)	Zona donde se realiza un enfriamiento lento de la banda hasta la temperatura de 600-650 °C.
Enfriamiento rápido (ENR)	Zona donde se realiza un enfriamiento rápido de la banda hasta la temperatura de 400-450 °C.
Igualación (IGU)	Zona donde se homogeniza la temperatura de la banda antes de su inmersión en el pote de zinc.
Pote de zinc (POT)	Zona de inmersión en el baño con zinc fundido.
Trompa (TRM)	Zona de la trompa de igualación.
General (GRL)	Parámetros que son constantes en toda la HDGL.

Tabla III. Descripción de los grupos de variables más relevantes

Table III. Attributes description of the most relevant groups of variables

GRUPOS DE VARIABLES	
Nombre	Descripción
CODE	Código de la bobina.
CON_H2	Concentración de H ₂ en el aire presente en esa zona.
CON_O2	Concentración de O ₂ en el aire presente en esa zona.
TMP_PR	Temperatura del punto de rocío en esa zona.
Zxx_TMP	Temperatura de la zona xx.
TMP_Pxx	Temperatura de la banda, medida con el pirómetro xx, en cada zona del proceso.
ESP, ANCH	Espesor y anchura de la banda de acero.
SPD	Velocidad de la banda de acero.
POT_TMP	Temperatura del baño de zinc.
CMP_BAÑO	Composición química del baño de zinc. Porcentaje de los elementos químicos más relevantes del baño de zinc.
CMP_ACERO	Composición química del Acero. Porcentaje de los elementos químicos más relevantes: Fe, Mn, Al, Ni, etc.
ADHERENCIA	Si la adherencia del recubrimiento de zinc fue correcta o no (Valores: TIPO-A o TIPO-B)

siguiente forma: “SI la velocidad mínima de la banda de acero (GRL_SPD_MN) ha sido menor o igual a 80 m/min Y la temperatura máxima del baño de zinc (POT_TMP_MX) ha sido menor o igual que 457 °C Y la temperatura máxima en la zona de calentamiento 6 (CAL_Z06_TMP_MX) ha sido menor o igual que 912 °C, ENTONCES, se clasificaron 12,29 (12,29/1,27) bobinas, casi todas de TIPO-B. De éstas, solamente 1,27 fueron de TIPO-A”.

Es decir, si se cumplen esas condiciones, 11 bobinas presentaron adherencia irregular frente a 1 que presentó una adherencia correcta. Los decimales, que pueden producir confusión en la interpretación, son debidos a que el algoritmo C4.5 utiliza valores próximos a cero para rellenar los datos inexistentes en la información que maneja. Generalmente, a efectos prácticos, estos valores finales se suelen redondear hacia abajo para determinar el número de instancias clasificadas en cada rama del árbol de decisión.

De esta forma, el algoritmo realizó de forma automática una selección de las variables y los valores de corte de éstas que mejor dividían el espacio de instancias. Además, éste realizó un podado de las ramas del árbol para reducir el tamaño del mismo y hacerlo más fácil de interpretar.

El proceso de podado se realizó considerando dos factores: que la precisión de la regla obtenida de cada rama podada fuera mayor del 75 % y que el número de casos abarcados fuera mayor o igual a 2. La precisión de cada rama se define como el número de casos correctos dividido por el número de casos que abarca. El primer factor implica una restricción del podado, de forma que a mayor precisión exigida mayor longitud de rama. El segundo implica una restricción sobre el número de casos mínimo que puede explicar una rama.

Es importante destacar que la capacidad de generalización y el error admitido en el podado son parámetros que se deben modificar constantemente en el proceso de creación de árboles. Éstos deben ajustarse para obtener un árbol que no sea excesivamente grande pero que, a su vez, sea capaz de explicar con cierta precisión las clases que se le presentan.

4.3. Búsqueda de conocimiento útil

En los casos en que la relación entre el número de elementos de cada clase es muy dispar, como en este caso, conviene centrarse en las variables que arrancan de la base del árbol pues son las que dividen el mismo en grupos más amplios y que, por lo tanto, pueden aportar información más global que las variables de las ramas finales, pues éstas concentran información relativa a grupos reducidos de individuos.

Resulta interesante confirmar que, ante las numerosas variables del proceso (77), el algoritmo suele mostrar aquellos atributos que mejor explican las causas de fallos de adherencia, los cuales, suelen coincidir con el conocimiento propio de los tecnólogos de la planta (en este caso, fueron la velocidad de la banda y la temperatura del baño de zinc). Aún así, se aconseja comprobar cada uno de los resultados y seguir afinando los modelos obtenidos a medida que se obtienen conclusiones importantes.

Si se observa la última rama del árbol de la figura 6, se puede ver que, cuando la velocidad mínima de la banda (GRL_SPD_MN) fue mayor de 80 m/min se obtuvieron 462 bobinas TIPO-A y solo 3 de TIPO-B (un 0,7 % de bobinas TIPO-B, frente al 5,4 % de bobinas TIPO-B en toda la base de datos).

Analizando la figura 8, donde se mostraba la relación de las bobinas TIPO-A y TIPO-B según la velocidad mínima (GRL_SPD_MN) y la variación que se había producido en dicha velocidad para cada bobina (GRL_SPD_SD) se pudo deducir que un grupo elevado de las bobinas con adherencia irregular pudo ser debido a unos valores bajos de velocidad o variaciones bruscas de la misma. Seguramente, muchas de ellas correspondían a momentos de ajustes en arranques o paradas por lo que, obviamente, la explicación más plausible era que estas bobinas no habían trabajado

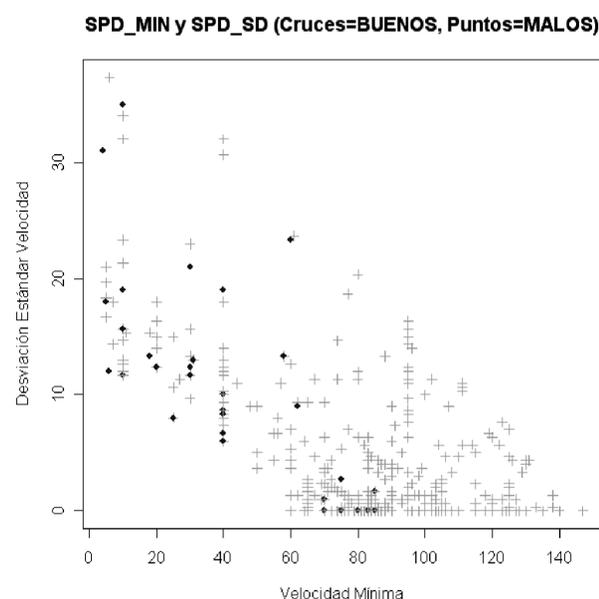


Figura 8. Relación de las bobinas TIPO-B (puntos) frente a las bobinas TIPO-A (cruces) según la velocidad mínima y desviación estándar de la velocidad de la banda.

Figure 8. TYPE-B coils (points) against TYPE-A coils (crosses) according to the minimal strip-speed and the strip-speed's standard deviation.

en régimen permanente y que, por lo tanto, su tratamiento térmico no había sido muy homogéneo.

A partir de la gráfica de la figura 8 y de las conclusiones obtenidas en el árbol de las figuras 6 y 7 se decidió eliminar las bobinas con velocidad mínima menor de 50 m/min y desviación estándar por encima de 15. De esta forma, el número de bobinas TIPO-B se redujo drásticamente a 18 y el de TIPO-A a 609.

El objetivo de esta reducción consistió en seguir estudiando aquellas bobinas que habían presentado irregularidades en la adherencia pero por causas que no fueran debidas a cambios bruscos de velocidad o velocidades bajas. En la figura 9 se muestra uno de los árboles que se obtuvo con la nueva base de datos.

De las 18 bobinas con problemas de adherencia, 11 fueron explicadas con la siguiente regla: “Si la temperatura media del baño de zinc (POT_TMP_AV) fue menor o igual que 456 °C Y la velocidad media (GRL_SPD_AV) fue menor o igual que 86 m/min y la temperatura media de la última zona de calentamiento del horno fue mayor de 885 °C, (CAL_Z08_TMP_AV) ENTONCES aparecieron 11 bobinas TIPO-B y 1 TIPO-A” (12,0/1,0).

Según esta regla, parecía que si la temperatura media del baño de zinc (Fig. 10) era menor o igual que 456 °C podían aparecer problemas de adherencia, siempre y cuando la temperatura de la subzona de calentamiento del horno última (subzona 8) fuera mayor de 885 °C. Además, se pudo observar que cuando la velocidad media superaba un valor de 86 m/min, solamente aparecían bobinas TIPO-A.

Esta nueva regla permitió a los tecnólogos confirmar lo que ya intuían, que la temperatura del baño de zinc debía superar un cierto valor. Pero, lo que es aún más importante, se les suministró dicho valor de corte. Además, el árbol mostraba una relación entre los atributos correspondientes a la velocidad de la banda, temperatura del pote de zinc y temperatura del horno, que podía ser muy interesante para comprender mejor el funcionamiento del proceso.

Si se observa la última rama del árbol de la figura 9, se puede ver que solamente 7 bobinas que



Figura. 9. Árbol de decisión obtenido después de eliminar las bobinas con velocidad mínima superior a 50 m/min y desviación estándar superior a 15.

Figure. 9. Decision tree remained after removing coils with minimal speed higher than 50m./min and standard deviation higher than 15.

POT_TMP y TMP_Z08 (Cruces=BUENOS, Puntos=MALOS)

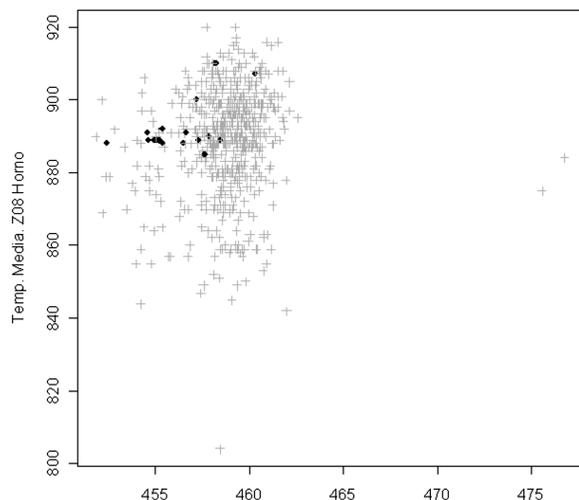


Figura. 10. Relación de las bobinas TIPO-B (puntos) frente a las bobinas TIPO-A (cruces) según la temperatura media del baño de zinc y la temperatura media de la subzona 8 (correspondiente a la última zona de calentamiento del horno).

Figure. 10. TYPE-B coils (points) against TYPE-A coils (crosses) according to the zinc-bath's mean temperature and to the 8th subzone's mean temperature (this one corresponding to the last heating zone of the furnace).

trabajaron en régimen permanente y cuya temperatura de la entrada del pote de zinc fue superior a 456 °C, presentaron irregularidades en la adherencia frente a unas 567 bobinas TIPO-A. De esta forma, y siguiendo con la misma metodología, se eliminaron las 11 bobinas TIPO-B ya explicadas previamente y se continuó buscando conocimiento oculto a partir de árboles creados con la base de datos de las 574 bobinas que correspondían con esta rama.

Del árbol obtenido (Fig. 11), tal y como se hizo en los anteriores modelos, se analizaron las variables más próximas a la raíz del mismo. En este caso, entraron en juego las concentraciones de hidrógeno de la parte principal de la zona de igualación (IGU_CON_H2_AV) y de la trompa de la zona de igualación del proceso (IGU_TRM_CON_H2_AV).

Se puede observar que, cuando las concentraciones fueron mayores del 2 % y 2,6 %, respectivamente, el número de bobinas con adherencia TIPO-A fue de (114+384=498) y una sola del TIPO-B. Es decir, un 87,8 % (498 de 567) de las bobinas TIPO-A frente a un 14 % de las TIPO-B (1 de 7).

En cambio, cuando el nivel de concentración de hidrógeno estuvo por debajo de los valores anteriores,

se obtuvieron 6 bobinas TIPO-B de las 7, es decir, un 86 % de las bobinas TIPO-B que quedaban, frente al 12,2 % de las TIPO-A.

Este árbol, permitía descubrir un posible grado de influencia de las concentraciones de hidrógeno del aire, el nivel mínimo de dichas concentraciones y las zonas del proceso donde debían ser vigiladas.

Como se puede observar, en el árbol también aparecía la concentración de oxígeno en la zona de mantenimiento del horno (MAN_CON_O2_AV) aunque el grado de influencia de esta variable y otras que aparecían en ramas finales tuvo que ser tomado con precaución, pues el número de elementos clasificados era demasiado pequeño para que pudieran ser tomadas como variables representativas.

5. CONCLUSIONES Y RECOMENDACIONES

Como se ha visto en este artículo, los algoritmos generadores de árboles de decisión fueron capaces de seleccionar las variables y el valor de éstas que mejor “discriminaban” las bobinas en las que aparecieron “irregularidades” de la adherencia del recubrimiento de zinc frente a las que no tuvieron.

A diferencia de otras técnicas de MD, estas técnicas pueden trabajar con un elevado número de atributos y muy pocos ejemplos. Ésta característica los hace muy útiles en la búsqueda de conocimiento oculto dentro de históricos de procesos industriales donde se dispone de una enorme cantidad de variables y pocos casos erróneos.

Generalmente, cuando se trabaja con este tipo de bases de datos (muchos atributos y pocos casos), los mejores resultados se obtienen cuando el análisis del árbol se centra en las reglas que más clases agrupan, aunque el error sea mayor, pues el grado de abstracción es más global. Muchas veces, es aconsejable comenzar el

análisis partiendo de la raíz del árbol y estudiar las reglas que surgen en el primer o según nivel del mismo.

Después de validado el conocimiento obtenido, habitualmente con los tecnólogos o expertos del proceso, se recomienda eliminar los casos “ya explicados” de la base de datos y crear un nuevo árbol con los casos restantes. De esta forma, el método de búsqueda se focaliza cada vez más en la parte del problema que aún no ha sido explicado.

Dentro del análisis, es conveniente ser cauto cuando bajamos muchos niveles dentro del árbol y obtenemos reglas muy complejas, pues el conocimiento particular de un pequeño grupo de casos puede ser falseado fácilmente debido a la elevada dimensionalidad o amplio número de atributos que se manejan y el escaso número de ejemplos de los que se suele disponer.

Por último, es muy importante destacar que los árboles de decisión solamente muestran las variables y sus valores que mejor separaran el espacio de instancias para clasificar una clase u otra pero, al final, es el experto el que debe corroborar los resultados obtenidos y encontrar las causas físicas o lógicas que los producen, pues muchas veces el conocimiento obtenido puede ser obvio, inútil o falso.

Agradecimientos

Los autores agradecen a la Unión Europea bajo el Research Fund for Coal and Steel (RFCS) la financiación de los proyectos con referencia RFS-CR-04023 y RFS-CR-04043. También, ha sido parcialmente financiado por el Ministerio de Ciencia y Tecnología de España a través de la Dirección General de Investigación bajo los proyectos DPI2004-07264-C02-01, DPI2006-03060 y DPI2007-61090. Por último, queremos agradecer la ayuda recibida a través del 2º Plan Riojano de I+D+i del Gobierno de La Rioja.

REFERENCIAS

- [1] O. DOMÍNGUEZ, J. M. MIRANDA-ARIZ Y L. SALVADOR. *Rev. Metal. Madrid* 38 (2002) 108-116.
- [2] A. GORNI, *The modelling of hot rolling processes using neural networks*, http://www.gorni.eng.br/el/neural_1998.html, 1998.
- [3] F. J. MARTÍNEZ-DE PISÓN, Tesis Doctoral, Escuela Técnica Superior de Ingeniería Industrial, Universidad de La Rioja, 2003.
- [4] J. A. GONZÁLEZ-RODRÍGUEZ, Tesis Doctoral, Universidad de Oviedo, 1994.
- [5] F. J. MARTÍNEZ-DE PISÓN, F. ALBA-ELÍAS, M. CASTEJÓN-LIMAS Y J. A. GONZÁLEZ-

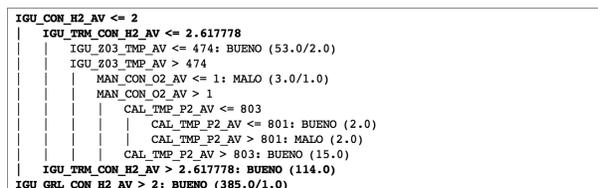


Figura. 11. Árbol de decisión obtenido después de eliminar las bobinas con temperatura media del baño de zinc menor o igual a 456 °C

Figure. 11. Decision tree remained after removing coils with zinc-bath's mean temperature less or equal to 456 °C.

- RODRÍGUEZ, *Ironmaking Steelmaking* 33 (2006) 1-9.
- [6] J. HERNÁNDEZ-ORALLO, M. J. RAMÍREZ-QUINTANA Y C. FERRI-RAMÍREZ, *Introducción a la Minería de Datos*, Pearson Educación S.A., Prentice Hall, Madrid, España, 2004.
- [7] J. R. QUINLAN, *Machine Learning* 1 (1986) 81-106.
- [8] J. ZUECO Y F. ALHAMA, *Rev. Metal. Madrid* 41 (2005) 227-232.
- [9] J. H. WEI Y H. T. HU, *Ironmaking Steelmaking* 32 (2005) 427-434.
- [10] A. MONSALVE, A. ARTIGAS, D. CELENTANO Y F. MELÉNDEZ, *Rev. Metal. Madrid* 40 (2004) 247-258.
- [11] D. M. JONES, J. WATTON Y K. J. BROWN, *Ironmaking Steelmaking* 32 (2005) 435-442.
- [12] C. CAMURRI. *Rev. Metal. Madrid* 39 (2003) 132-139.
- [13] A. PERNÍA-ESPINOZA, M. CASTEJÓN-LIMAS, A. GONZÁLEZ-MARCOS Y V. LOBATO-RUBIO, *Ironmaking Steelmaking* 32 (2005) 418-426.
- [14] C. CANTERA, J. JIMÉNEZ, I. VARELA Y A. FORMOSO. *Rev. Metal. Madrid* 38 (2002) 243-248.
- [15] J. ORDIERES-MERÉ, A. GONZÁLEZ-MARCOS, J. A. GONZÁLEZ, Y V. LOBATO-RUBIO, *Ironmaking Steelmaking* 31 (2004) 43-50.
- [16] Z. STERJOVSKI, D. NOLAN, K. R. CARPENTER, D. P. DUNNE Y J. NORRISH, *J. Mater. Process. Technol.* 170 (2005) 536-544.
- [17] R. VALENTINI, V. COLLA Y M. VANNUCCI, *Rev. Metal. Madrid* 40 (2004) 416-419.
- [18] J. LIU, H. CHANG, T. Y. HSU, Y R. RUAN, *J. Mater. Process. Technol.* 103 (2) (2000) 200-214.
- [19] K. P. SURJYA Y C. DEBABRATA, *Neural Computing Applications* 14 (2005) 319-324.
- [20] J. SON, D. LEE, I. KIM Y S. CHOI, *J. Mater. Process. Technol.* 153/154 (0) (2004) 643-648.
- [21] F. PERNKOPF. *Pattern Analysis Applications* 7 (2005) 333-342.
- [22] J. JUNG Y Y. IM, *J Mater. Process. Technol.* 96 (1999) 163-172.
- [23] F. JANABI-SHARIFI, *Control Eng. Practice* 1 (2005) 1-14.
- [24] J. JUNG U Y. IM, *Int. J. Mechanical Sci.* 42 (2000) 249-255.