Utilización de algoritmos de meta-clasificación para la mejora de los Modelos Predictivos de Control

Javier Nieves — Dr. Ingeniero Informático
Igor Santos — Dr. Ingeniero Informático
Pablo G.Bringas — Dr. Licenciado en Informática

UNIVERSIDAD DE DEUSTO. DeustoTech Computing, S3Lab. Avda. de la Universidades, 24 - 48007 Bilbao. Tfno: +34 944 139000 ext. 2040. jnieves@deusto.es

Recibido: 09/10/2012 • Aceptado: 08/04/2013
DOI: http://dx.doi.org/10.6036/5426

ENHANCING THE PREDICTION STAGE OF A MODEL PREDICTIVE CONTROL SYSTEMS THROUGH META-CLASSIFIERS

ABSTRACT

• A Model Predictive Control (MPC) is a system which allows to control a production plant. Thanks to this type of system, it is possible to make a production close to zero defects. To achieve its main goal, this kind of systems is composed of several phases. One of the most critical one is the phase that predicts the plant situation for a given time. Currently, the majority of the research in this field is related to linear MPCs, although the process may not be. Previously, we presented several experiments that prove that the forecast phase, usually represented by a single mathematical function, can be represented by machine-learning models. Nevertheless, the employment of standalone classifiers raises some limitations. In this paper, we extend our previous research and we propose a general method to foresee all the defects, building a meta-classifier using the combination of di erent methods and removing the need of selecting the best algorithm for each objective or dataset. Finally, we compare the obtained results, showing that the new approach obtains better results, in terms of accuracy and error rates.

• Keywords: model predictive control, machine learning, fault prediction, data mining, process optimisation.)

RESUMEN

Un Modelo Predictivo de Control (MPC) es un sistema que permite controlar una planta de producción. Gracias a este tipo de sistemas, es posible hacer que la producción llegue a cero defectos. Para alcanzar este objetivo, este tipo de sistemas está compuesto por varias fases. Una de las más críticas es la que se encarga de la predicción en un instante futuro. Actualmente, la investigación en este campo está relacionada con los MPC lineales, aunque el proceso pueda no serlo. Anteriormente, se presentaron diferentes experimentos que probaban que la fase de predicción, normalmente representada por una fórmula matemática, podía ser representada mediante modelos de aprendizaje automático. Sin embargo, la utilización de clasificadores unitarios introduce algunas limitaciones. En este trabajo, se extienden las investigaciones anteriores para proponer un método general para la predicción de defectos o características a través de la utilización de la combinación de diferentes métodos con un meta-clasificador, eliminando la necesidad de selección del mejor de ellos. Finalmente, se comparan los resultados obtenidos mostrando que esta nueva aproximación permite obtener mejores resultados tanto en términos de precisión como en tasas de error.

Palabras clave: modelo predictivo de control, aprendizaje automático, predicción de defectos, minería de datos, optimización de procesos.

1. INTRODUCCIÓN AL ESCENARIO DE TRABAJO

Los procesos productivos datan de épocas ancestrales. En la antigüedad, éstos se desarrollaban de una forma muy primitiva, sin embargo, actualmente, y

gracias a la aplicación de técnicas como la del control estricto del proceso, han conseguido ser mejorados. También, hechos como (i) la globalización y el aumento drástico de la competencia y (ii) las nuevas limitaciones medioambientales impuestas, hacen que los procesos de producción adquieran una complejidad extra.

Para permitir desarrollar los procesos productivos de una forma mucho más precisa, en la década de los 60 surge una nueva tecnología, los «*Modelos Predictivos de Control*» (a partir de ahora MPC) [1]. Éste es uno de los pocos métodos avanzados que tiene un impacto significativo en la ingeniería de control industrial. Los MPC se aplican a la industria de los procesos ya que permite [2–4] (i) el manejo de problemas de control multivariables de una forma natural, (ii) la gestión del fallo en los actuadores, (iii) determinar situaciones relacionadas con fases problemáticas del proceso y (iv) tener en cuenta la incertidumbre del modelo. Esta tecnología tuvo una gran acogida debido a tres factores fundamentales [5]:

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

La evolución de los MPC ha sido dirigida tanto desde las empresas como desde la comunidad académica [6]. Concretamente, los MPC con la denominación DMC, QDMC, IDCOM-M o HIECON nacieron de desarrollos de empresas como *Shell Oil*, *Adersa* o *Setpoint*. En relación a los modelos comerciales, Qin et ál. [6] realizaron un estudio de los diferentes MPC existentes en el mercado. De esta forma, la Tabla I muestra sus resultados, indicando las diferentes casas comerciales y las industrias en las que se han aplicado.

Asimismo, existen diferentes MPC dependiendo de su naturaleza (lineales y no lineales) y de cómo fueron creados (empíricos y basados en principios fundamentales) [2]. Aunque los procesos de fabricación son intrínsecamente no lineales, la mayoría de las soluciones desarrolladas se basan en modelos lineales. Este hecho se debe a que las mejoras producidas por los MPC no lineales no son comparables a los esfuerzos necesarios para generarlos y gestionarlos [4]. Además, la complejidad o los bajos tiempos de ciclo necesarios, hacen tener que plantearse la utilización de sistemas subóptimos, acabando de nuevo en un MPC lineal [7].

Independientemente del tipo de MPC a desarrollar, es necesario contar con personal experto en el dominio, en su construcción; y para los empíricos, del control de la planta durante varios días. Así, el proceso de creación y afinamiento consta de los siguientes pasos [8–10]:

- De los objetivos de control establecidos, se define el

CONCEPTOS SOBRE APRENDIZAJE AUTOMÁTICO Y LA SOLUCIÓN DE PROBLEMAS DE CLASIFICACIÓN

El aprendizaje automático se trata de una de las ramas de la inteligencia artificial que, partiendo del análisis de datos, intenta desarrollar un comportamiento predictivo. Concretamente, para desarrollar ese carácter los problemas suelen ser resueltos de dos formas diferentes: (i) regresión, donde la predicción se hace sobre variables que tienen un valor continuo, por ejemplo, el valor para las medidas de la resistencia a la tracción en MPa; o (ii) la clasificación, donde el objetivo a predecir es un valor discreto, por ejemplo, la probabilidad de aparición (alta, media, baja) de un defecto.

Para conseguir solucionar todos estos problemas, los métodos utilizados dependerán de los datos de entrada y la forma en la que se encuentran. Específicamente, nos encontramos con los siguientes: (i) supervisados, en los que todas las evidencias indican fielmente el resultado que tuvieron; (ii) semi-supervisados, sus evidencias cuentan tanto con resultados etiquetados como resultados sin estar evaluados; y (iii) no supervisados, aquellos que no cuentan con evidencias con etiqueta o valor asignado.

En el presente artículo nos centramos en la utilización del aprendizaje supervisado para la solución de los problemas de clasificación. Pero para conseguirlo, se hace uso de lo que se conoce como meta-clasificadores. Este tipo de clasificadores son una serie de métodos que permiten crear una colmena de algoritmos que trabajan unidos para realizar la clasificación final. La razón de su utilización es que los diferentes algoritmos existentes, por sí solos, se adecuan mejor a unas situaciones concretas que a otras. Sin embargo, al utilizarlos de forma conjunta se consigue una mayor robustez en las etapas de clasificación. Los métodos de meta-clasificación se utilizan para evitar el desarrollo de extensos experimentos que lleven a identificar el método idóneo. Gracias a su combinación son capaces de aproximarse, o incluso superar, a ese método óptimo gracias a que se llega a un consenso para obtener el resultado final de la clasificación.

Los diversos métodos de meta-clasificación permiten, por una parte, la generación de lo que se conoce como «ensemble», la unión de múltiples clasificadores débiles del mismo tipo para la creación de uno más preciso; o, por otra parte, la combinación de diferentes métodos que dotan al meta-clasificador de las características híbridas que todos los métodos simples utilizan.

Área	Aspen Technology	Honeywell Hi-Spec	Adersa	Ivensys	Continental Controls	DOT Controls	Pavilion Tech.	SGS	Total
Refinería	1200	480	280	25			13		1998
Petroquímicas	450	80		20					550
Químicas	100	20	5	21	15		5		166
Pulpa y papel	18	50					1		69
Aire y gas		10			18				28
Utilidades	5	6		4	2				17
Minería / Metalurgia	8		7	16					31
Procesamiento alimenticio			41	10			9		60
Polímeros	18					5	15		38
Hornos			42	3					45
Aeroespacial / Defensa			13						13
Automoción			7						7
Sin clasificar	40	40	1046	26	1			450	1603
Total	1839	686	1441	125	36	5	43	450	4625

Tabla I: Expansión de los MPC en la industria y las compañías que los producen

amaño del problema y se determinan las variables de control, las variables a manipular y las variables que no se pueden controlar.
- Para el caso de una generación empírica, la planta es probada de forma sistemática, modificando las variables a capturar y almacenando los datos en tiempo real que indican cómo se comporta el proceso. En caso de una generación basada en principios fundamentales, se extraen las reglas básicas que definen el proceso productivo.
- Se obtiene el modelo dinámico que rige el funcionamiento de la planta de producción.
- El controlador MPC es configurado, estableciendo los parámetros iniciales.
- Se desarrollan las pruebas de rendimiento del controlador a través de simulaciones.
- El controlador configurado es implantado, poniendo a prueba las predicciones que realiza.
- Se afina la puesta a punto según sea necesario.

A partir de este momento, el sistema de control del MPC es el encargado de dirigir el proceso desde el estado estacionario actual a otro diferente que mejore la situación predicha de la planta, manteniendo la producción dentro de lo que es considerado como normal. De esta forma, las etapas de control generales que debe seguir un MPC, recogidas de forma esquemática en la Fig. 1, son las siguientes:

Fig. 1: Flujo de trabajo general que siguen los controladores MPC para cada una de sus ejecuciones.

- **Lectura de los valores del proceso**. El controlador realiza las lecturas de todas las variables (tanto sobre las que puede actuar como sobre las que no).
- **Estimación del estado**. En este momento, el controlador se encarga de determinar el estado en el que se encontrará el proceso en un instante de tiempo t+1 de seguir la producción con los mismos valores del proceso.
- **Determinación del estado al que llegar**. Aquí, el MPC se encarga de establecer el estado óptimo en el que se pretende dejar a la planta funcionando, evitando grandes cambios en el proceso a la hora de redirigirlo
- **Determinación del camino de optimización**. Una de las etapas más importantes. Su resultado será el método a seguir para modificar la planta.

• **Salida de los valores a modificar**. Finalmente, en la última de las fases se lleva a cabo el *feedback* a la planta.

A pesar de las ya citadas mejoras que incorporan los MPC al proceso de control de una planta, su proceso de construcción incorpora una serie de limitaciones. De forma resumida, algunas de éstas son las siguientes [11]:

- Limitadas opciones entre las que elegir el modelo.
- Realimentación de la planta ineficiente.
- No se aborda la eficiencia estadística.
- Existe una falta de métodos de validación de los modelos encargados de la predicción.
- No existe un enfoque sistemático para la construcción de los MPC.
- Es difícil la formulación básica del controlador.
- Existen grandes problemas para la adaptación de los MPC al cambiante proceso de las plantas.
- El trabajo con modelos de predicción no lineales es muy limitado.

Así, se pueden identificar tres puntos clave sobre los que trabajar: (i) la etapa de generación de modelos y la predicción de estados estacionarios, (ii) la actualización de esos modelos y (iii) el cálculo de los parámetros que deben retornarse a la planta. Concretamente, este trabajo se centra en la primera de estas etapas, presentando un nuevo enfoque para la generación de los modelos mediante la utilización de la combinación de métodos de clasificación basados en aprendizaje automático y que permiten evolucionar el trabajo desarrollado en esta línea [7].

2. DEFINICIÓN DEL PROBLEMA, OBJETIVOS Y MEDIOS DISPUESTOS

2.1. DEFINICIÓN Y ALCANCE DEL TRABAJO

Con la investigación detallada en este artículo queremos responder las siguientes preguntas que surgieron como punto de partida:

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

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

Así, el objetivo principal de este trabajo es el de avanzar la etapa de predicción con la que cuentan los sistemas MPC

Fig. 2: Representación del proceso productivo de la fundición

actuales. Y de este objetivo se derivan los siguientes:

- Aumentar las opciones para la generación de modelos simplificando su proceso.
- Incorporar modelos estadísticos a los MPC.
- Incorporar un enfoque sistemático a la generación de MPC basado en el enfoque de aprendizaje automático introducido por los métodos utilizados.
- Permitir la combinación de múltiples métodos de predicción de naturaleza diferente (lineal y no lineal) acercándose a una solución de predicción híbrida.

2.2. CASO DE USO

Aunque en diversos sectores como la refinería o las petroquímicas este tipo de sistemas está muy extendido, no lo es tanto en el sector de la metalurgia, y concretamente en la fundición de hierro. Así, nos centraremos en un proceso concreto de moldeo automático que emplea arena verde y de hierro nodular en el que se producen piezas que formarán parte de otro sistema mucho más complejo (las etapas del proceso se detallan en la Fig. 2). La fundición seleccionada cuenta con una producción cercana a las 45.000 toneladas al año con piezas destinadas a la industria del automóvil.

De todas las variables del proceso se ha realizado una selección, obteniendo las 24 variables más representativas. Éstas pueden ser divididas en:

- **Variables relacionadas con el metal:**
  - Composición, inoculación y tipo de tratamiento.
  - Calidad de la mezcla y potencial de nucleación: Se obtiene por medio de un programa de análisis térmico.
  - Vertido: Duración del proceso de colada de moldes y la temperatura de colada.
- **Variables relacionadas con el molde:**
  - Arena: Tipo de aditivos utilizados, las características específicas de arena y realización de pruebas anteriores o no.
  - Moldeado: la máquina de moldeo utilizada y parámetros.
- **Variables relacionadas con la pieza:**
  - Geometría y dimensión.
  - Velocidad de enfriamiento.
  - Tratamiento térmico.

Dentro del caso de uso seleccionado, se ha trabajado en la optimización de la producción a través de controlar la aparición de defectos. Concretamente, los defectos sobre los que se ha trabajado son los siguientes:

- **Microrrechupes:** Se trata de pequeñas cavidades agrupadas en un elevado número. La aparición de esta irregularidad se debe a que los metales son menos densos como líquidos que como sólidos, y durante el enfriamiento, la densidad del metal crece mientras el volumen decrece.
- **Resistencia a la tracción:** Ensayo perteneciente a pruebas de las propiedades mecánicas de las piezas. Concretamente, se trata de medir el comportamiento de una probeta normalizada ante una fuerza axial de tracción. El no conseguir sobrepasar un límite especificado puede implicar problemas de seguridad.

La selección de estos defectos se debe a que ambos tienen que ser analizados y detectados una vez las piezas han sido producidas, incluso uno de ellos podría implicar la destrucción de la misma. Por consiguiente, una detección previa a la fabricación de las piezas induciría en una mejora del proceso.

2.3. MEDIOS Y HERRAMIENTAS DISPUESTOS EN EL TRABAJO

Los medios utilizados para la obtención de los datos que permitiesen desarrollar el experimento fueron los siguientes:

- **Sistemas automatizados de captura de información:** dado el grado de automatismo presente en la fundición seleccionada, y dados los sistemas actuales de recogida de datos, se mantiene una gran fuente de información para la trazabilidad del proceso y que ha sido tomada como fuente de datos para este experimento.
- **Sistemas manuales de recolección de datos:** los datos registrados por la fundición han sido completados con un sistema de recogida de datos manual, a través de las hojas de los partes de trabajo que se mantienen en la fundición a lo largo del proceso completo.
- **Resultados de análisis de piezas:** una selección de especímenes fueron analizados para comprobar el resultado final de las piezas en relación a los defectos comentados anteriormente. Los análisis de rayos X y las mediciones de la resistencia a la tracción son llevados a cabo en una entidad externa a la fundición y remitidos a la misma, pasando a formar parte de la información de trazabilidad de la planta.

En experimentos anteriores, se desarrolla un sistema de predicción basado en modelos de clasificación unitarios [7]. No obstante, esta aproximación también cuenta con una serie de limitaciones: no se puede asegurar que un modelo en concreto sea el idóneo para la predicción de la situación de la planta. De esta forma, y siguiendo los estudios desarrollados para la combinación de métodos de clasificación se evoluciona esta tarea a través de la utilización de las siguientes herramientas para la combinación de clasificadores heterogéneos:

- **Combinación por Voto:**
  - *La mayoría decide:* asumiendo que las salidas etiquetadas de los predictores son dadas como vectores binarios c-dimensionales $[d_{i,1},\ldots,d_{i,c}]^T \subset \{0,1\}^c$, $i = 1,\ldots, L$ donde $d_{i,j} = 1$ si el predictor D_i categoriza la instancia **x** como la clase w_j, ó 0 en el resto de los casos. La pluralidad de los votos resulta en un conjunto de clasificación para la clase w_k tal que

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

  - *Producto de probabilidades:* ahora teniendo en cuen-

ta las probabilidades, $p(x_1,\ldots,x_R|w_k)$ representa la distribución de probabilidad conjunta de las mediciones extraídas de los predictores. Asumimos que las representaciones son estadísticamente independientes y agregando la teoría de decisión bayesiana [12], el método asigna $Z \rightarrow w_j$ si

$$P^{-(R-1)}(\omega_j)\prod_{i=1}^{R}P(\omega_j \mid x_i) = máx_{k=1}^{m}P^{-(R-1)}(\omega_k)\prod_{i=1}^{R}P(\omega_k \mid x_i) \quad (2)$$

- *Media de probabilidades:* Para obtenerla comenzaremos generando la regla de la suma para, posteriormente, realizar una división empleando como denominador el número de predictores *R* utilizados. Así, la regla de la suma asigna $Z \rightarrow w_j$ si

$$(1-R)P(\omega_j)+\sum_{i=1}^{R}P(\omega_j \mid x_i) = máx_{k=1}^{m}[(1-R)P(\omega_k)+\sum_{i=1}^{R}P(\omega_k \mid x_i)] \quad (3)$$

- *Probabilidad máxima:* Partiendo de la regla de la suma, dejando de lado el producto de las probabilidades a posteriori y asumiendo igualdades en las precedencias, el método asigna $Z \rightarrow w_j$ si

$$máx_{i=1}^{R}P(\omega_k \mid x_i) = máx_{k=1}^{m}máx_{i=1}^{R}P(\omega_k \mid x_i) \quad (4)$$

- *Probabilidad mínima:* Para este método, partiendo de la regla del producto, dejando de lado las probabilidades a posteriori y asumiendo igualdades en las precedencias, se asignará $Z \rightarrow w_j$ si

$$med_{i=1}^{R}P(\omega_k \mid x_i) = máx_{k=1}^{m}med_{i=1}^{R}P(\omega_k \mid x_i) \quad (5)$$

• **Grading:** Los predictores base que vayan a ser combinados [13] son evaluados previamente mediante la validación cruzada [14], asegurándose que todas las instancias son utilizadas para el aprendizaje. El método construirá n_k conjuntos de datos de entrenamiento, uno por cada predictor *k*, al que le añadirá los resultados de las predicciones como la nueva clase de predicción. Posteriormente, se construirán unos predictores de primer nivel que, usando los nuevos conjuntos de datos, determinarán cuáles deben tenerse en cuenta para cada una de las predicciones finales. Los conflictos serán resueltos mediante la técnica del voto de la mayoría.

• **Stacking:** es otra generalización de la combinación basada en la validación cruzada [15]. Para ello, utilizaremos el conjunto de datos original q, dividido en dos grupos disjuntos. A estos nuevos conjuntos de datos los etiquetaremos con los nombres q_{ij}, donde $1 \leq i \leq r$ y $j \subset \{1,2\}$. Entonces, para cada una de las particiones generadas obtendremos un conjunto de *k* números. Estos puntos serán tomados como los valores de entrada para un segundo espacio apilado donde existirán otra serie de predictores. El proceso de predicción es el siguiente. Primero, todos los predictores realizan sus predicciones. Segundo, los resultados son transformados y llevados a los predictores del segundo nivel y éstos tomarán la decisión final. Por último, la decisión es transformada de nuevo al espacio inicial para darla a conocer. Este proceso podría ser más complejo si añadimos nuevos niveles.

• **Multiesquema:** Es un método que emplea una regla basada en el cálculo de la mediana de cada instancia y el ratio de error del predictores *G* alcanzado al predecir la salida relacionada con el conjunto de datos de prueba y el proceso de aprendizaje se llevó a cabo con el resto de los datos.

3. METODOLOGÍA EMPLEADA EN LA EXPERIMENTACIÓN

Con el fin de realizar contestar a las preguntas de la investigación, se ha llevado a cabo la siguiente metodología de experimentación:

• **Adquisición de datos:** Proceso mediante el cual se extrae el conocimiento de la fundición. Para los microrrechupes, se contó con un conjunto de datos de la producción y análisis mediante rayos X de 952 evidencias diferentes. Éstas engloban 2 referencias y el 27,44% de ellas fueron motivo de rechazo. El número de variables utilizadas para su representación es de 24. El conjunto de datos para la resistencia a la tracción cuenta con la información de producción y evaluación de 889 evidencias diferentes, de las cuales, el 37,82% fueron motivo de rechazo. Dentro del conjunto de datos se identifican 11 referencias diferentes, las cuales quedan representadas a través de 24 variables.

• **Validación cruzada:** Se ha realizado una validación cruzada de *k* partes [14] donde $k = 10$. De esta manera, el conjunto de datos para el aprendizaje se divide 10 veces en 10 diferentes partes. En cada posible división se seleccionan 9 de esas partes para la fase de aprendizaje y sólo una de las partes para la fase de pruebas.

• **Aprendizaje del modelo:** Se realiza la etapa de aprendizaje de los modelos que van a ser combinados posteriormente. Estos algoritmos son los mismos que se emplearon anteriormente [7]. Concretamente, se tratan de *redes bayesianas* (con algoritmos de aprendizaje estructural como K2, escalar colinas, *tree augmented naïve* o la simple red bayesiana ingenua), *K vecinos más próximos* (con valores de *k* entre 1 y 5), *máquinas de soporte vectorial* (con diferentes funciones de núcleo como la polinomial, la polinomial normalizada, RBF y Pearson VII), *árboles de decisión* (con el algoritmo C4.5 y la generación de bosques aleatorios) y *redes neuronales artificiales* (en concreto las *MultiLayer Perceptron*).

• **Aprendizaje del método de clasificación:** Una vez se dispone de los clasificadores, se aprenderá la forma en la que deben ser combinados. Específicamente, los métodos de combinación son los siguientes: *por voto* (combinaremos los resultados mediante las reglas de la mayoría, el producto, la media, el máximo y el mínimo), *grading* (utilizando como clasificadores de primer nivel una red bayesiana ingenua, una red bayesiana con *tree augmented naïve* y los k vecinos más próximos con valores de $k \subset [1,5]$), *stacking* (utilizando para el segundo espacio los mismo que para el método de *grading*), y *multiesquema*.

• **Pruebas del modelo.** La primera de las medidas sobre las que se ha trabajado es el nivel de precisión alcanzado por el clasificador. El segundo es la evaluación de las tasas de error. Se ha evaluado la tasa de error entre el conjunto de valores predichos X y el conjunto de valores reales Y mediante el error medio absoluto (MAE, *«Mean Absolute Error»*).

$$MAE(X,Y) = \sum_{i=1}^{m} \frac{|X_i - Y_i|}{m} \quad (6)$$

Del mismo modo, se ha utilizado la medida de la raíz del error cuadrado medio (*«Root Mean Square Error»*, RMSE).

$$RMSE(X,Y) = \frac{1}{m} \cdot \sqrt{\sum_{i=1}^{m} (X_i - Y_i)^2} \quad (7)$$

4. RESULTADOS

Una vez aplicada la metodología de investigación se obtuvieron los resultados que se muestran a continuación. Para facilitar la legibilidad de los mismos, éstos quedan divididos según el objetivo de clasificación seleccionado.

4.1. MICRORRECHUPES

Con el ánimo de poder comparar los resultados alcanzados con los de los predictores unitarios [7], hemos realizado las mediciones en los mismos términos, y éstos quedan recogidos en la Tabla II.

Predictor	Precisión (%)	MAE	RMSE
Stacking (TAN)	94,47	0,0602	0,2115
Stacking (Naïve Bayes)	94,12	0,0589	0,2357
Stacking (KNN k=5)	94,12	0,0801	0,2148
Grading (KNN k=5)	93,94	0,0606	0,242
MultiScheme	93,94	0,1562	0,234
Grading (KNN k=4)	93,93	0,0650	0,2425
Grading (KNN k=3)	93,87	0,0613	0,2439
Grading (J48)	93,86	0,0614	0,2435
Votación (Regla de la Mayoría)	93,85	0,0615	0,2439
Grading (Naïve Bayes)	93,81	0,0619	0,2449
Grading (TAN)	93,78	0,0622	0,2453
Stacking (KNN k=3)	93,72	0,0809	0,228
Stacking (KNN k=4)	93,66	0,0804	0,2201
Grading (KNN k=2)	93,49	0,0651	0,2519
Grading (KNN k=1)	93,48	0,0652	0,2517
Votación (Regla de la Media)	93,33	0,1364	0,2254
Stacking (J48)	93,26	0,0838	0,2462
Stacking (KNN k=2)	92,46	0,0794	0,2386
Stacking (KNN k=1)	92,18	0,0792	0,2754
Votación (Regla de Máxima Probabilidad)	73,29	0,2595	0,3306
Votación (Regla del Producto)	68,17	0,0435	0,1995

Tabla II: Resultados en la predicción del estado estacionario t+1 en la detección de mircorrechupes

En términos de precisión, podemos observar que el mejor de los predictores ha sido el método de *stacking* en el que los predictores del segundo espacio fueron creados a través de una red bayesiana ingenua ampliada con un árbol. La precisión conseguida con las técnicas de meta-clasificación ha sido de un 94,47%. Pero los comportamientos de todos ellos han sido muy similares. Concretamente, 19 de los 22 meta-clasificadores alcanzaron una precisión superior al 92%.

Cabe destacar el resultado obtenido con métodos de combinación muy simples como el multiesquema y la combinación por voto a través de la regla de la mayoría. Dado que los resultados son altos, 93,94% y 93,85% respectivamente, estos métodos pueden utilizarse con el objetivo de minimizar los cálculos y la complejidad computacional que tienen los demás métodos.

En cuanto a las tasas de error, podemos observar que no se alcanza el mismo comportamiento que en la precisión. En lo que a MAE se refiere, los métodos de combinación que mejor tasa de error han alcanzado han sido aquellos que obtuvieron peores resultados en la precisión. No obstante, el comportamiento alcanzado por la gran mayoría ha sido similar. De esta forma, eliminando excepciones, los meta-clasificadores han alcanzado unos valores de MAE que van desde las 0,0435 unidades a las 0,0838, lo que no supone una variación de más de una décima.

Para el RMSE, el comportamiento es bastante parejo. Los resultados que obtienen hacen pensar en que el comportamiento de todos ellos para la gestión de los errores es similar. Como se puede observar, los valores van desde las 0,1995 unidades (alcanzadas por los métodos de votación con las reglas del producto y de la mínima probabilidad) hasta las 0,3306 (alcanzadas por el método de votación con la regla de máxima probabilidad). La diferencia entre ellos no supera las 0,1311 unidades. Por consiguiente, las deducciones obtenidas para la tasa de error MAE son aplicables para la tasa RMSE.

En resumen, estas técnicas de clasificación mediante métodos de combinación de predictores unitarios no sólo han aproximado los resultados individuales, sino que han conseguido superarlos. En definitiva, puede considerarse que son una solución para incorporar una nueva forma de predicción de los estados estacionarios t+1.

4.2. RESISTENCIA A LA TRACCIÓN

Para este caso, la Tabla III ilustra con los resultados obtenidos en el experimento. En términos de precisión, el método que mejores resultados obtuvo fue el *grading* que emplea un predictor bayesiano ingenuo aumentado con un árbol. La máxima precisión alcanzada ha sido de un 86,63%. Aunque el meta-clasificador que mejores resultados obtuvo no es el mismo que el que lo consiguió para los microrrechupes, ese método sigue estando dentro de los 10 mejores meta-clasificadores con un resultado muy similar a su vencedor, un 85,73% de precisión.

Muy cerca se encuentran los métodos multiesquema y *grading* construido a través de un árbol C4.5. La diferencia en la precisión de estos métodos no es superior a las 0,29 unidades. También, dos predictores sencillos como el método multiesquema y el de votación (usando la regla de la mayoría) se encuentran entre los mejores. No obstante, existen otros predictores que se desvían de la tendencia marcada por la mayoría (p.ej., el resto de tipos de combinación por votación).

Predictor	Precisión (%)	MAE	RMSE
Grading (TAN)	86,63	0,1337	0,3625
MultiScheme	86,37	0,2134	0,3143
Grading (J48)	86,34	0,1366	0,3663
Votación (Regla de la Mayoría)	85,95	0,1405	0,3722
Grading (Naïve Bayes)	85,91	0,1409	0,3725
Votación (Regla de la Media)	85,86	0,2129	0,3187
Stacking (TAN)	85,73	0,1566	0,3356
Grading (KNN k=5)	85,71	0,1429	0,375
Stacking (Naïve Bayes)	85,48	0,1456	0,3752
Stacking (KNN k=5)	85,33	0,1875	0,335
Grading (KNN k=4)	85,23	0,1477	0,3816
Grading (KNN k=3)	85,15	0,1485	0,3825
Grading (KNN k=1)	85,05	0,1495	0,384
Votación (Regla de Máxima Probabilidad)	84,92	0,3137	0,3698
Stacking (J48)	84,79	0,1923	0,3619
Grading (KNN k=2)	84,77	0,1523	0,3875
Stacking (KNN k=3)	84,17	0,1866	0,3526
Stacking (KNN k=4)	83,69	0,188	0,3426
Stacking (KNN k=1)	81,43	0,1864	0,4278
Stacking (KNN k=2)	79,57	0,1868	0,3741
Votación (Regla del Producto)	69,8	0,0936	0,3023

Tabla III: Resultados en la predicción del estado estacionario t+1 en la detección de la resistencia a la tracción

En cuanto a las tasas de error (ver la Tabla III), se observa que no se mantiene el mismo comportamiento que marcan los niveles de precisión, pero la mayoría de los métodos se encuentran muy parejos. Así, en lo relativo a los valores del MAE, los mejores métodos han sido los que obtuvieron la peor precisión. Pero como ya hemos comentado, el comportamiento de la mayoría de los meta-clasificadores es similar. El rango en el que se mueven los predictores va desde las 0,3137 unidades a las 0,0936, es decir, 0,2201 unidades de diferencia. Eliminando las excepciones, como ya se hizo anteriormente, se comprueba que 19 de los métodos tienen una desviación de una décima.

El meta-clasificador basado en *grading* construido a base de un predictor bayesiano ingenuo aumentado con un árbol se encuentra el siguiente en el ranking. Éste es el que mejor precisión obtuvo, con lo que se postula como el mejor para la predicción de los estados estacionarios $t+1$ cuando trabajamos para la mejora de la resistencia a la tracción. Por otra parte, el segundo de ellos en lo que respecta a la precisión no es capaz de manejar los errores tan bien. A pesar de ello, aún queda la posibilidad de utilizar el método de votación basado en la regla de la mayoría, que sí maneja los errores de una forma correcta, como método que reduzca la complejidad computacional.

En lo referente al RMSE, se repite que los peores predictores en términos de precisión son los mejores en el control de errores. Obviando esta circunstancia, y centrándose en los predictores que sí obtuvieron resultados altos en la precisión, observamos que todos se comportan de manera similar. La desviación entre todos ellos ronda la décima. Por consiguiente, el meta-clasificador más adecuado será aquel que alcanzó mejor porcentaje de precisión.

Para concluir, este experimento muestra que la utilización de estos métodos en la predicción de la situación de la planta es factible. En el caso de la resistencia a la tracción no se ha logrado superar la precisión obtenida por el predictor unitario óptimo [7]. A pesar de ello, los resultados generales superan a los predictores unitarios al aproximar la precisión y reducir las tasas de error.

5. CONCLUSIONES

A pesar de que la combinación de clasificadores ha obtenido un buen rendimiento al enfrentarse a la predicción de los microrrechupes y a la resistencia a la tracción, hay una serie de aspectos a discutir en relación a la viabilidad de esta solución.

Primero, con estos métodos se soluciona el problema de la selección de un clasificador. Sin embargo, aparece un nuevo problema, los métodos de combinación también se comportan de forma diferente para la predicción de cada defecto. Por lo que el problema ha sido desplazado desde el primer nivel de clasificación al segundo de ellos. Nótese que a pesar de esta desventaja, los beneficios que aportan son mayores, concretamente, puede generarse un sistema MPC capaz de

ajustarse al proceso utilizando una predicción híbrida manteniendo, o incluso mejorando, la precisión y reduciendo las tasas de error.

Por último, los algoritmos de aprendizaje automático supervisado pueden ser un problema en sí mismo. El conjunto de datos puede ir creciendo según obtengamos nuevas inspecciones. Esto implica incrementar las posibilidades de almacenamiento, traduciéndose en una mayor complejidad computacional a la hora de generar nuevos modelos. Por consiguiente, la solución a este problema es la reducción del conjunto original de entrenamiento a través del concepto de *«Data Reduction»*.

No obstante, dadas las limitaciones de los sistemas MPC actuales, generados mediante una fórmula matemática, se ha partido de experimentos anteriores [7] para mejorarlo al utilizar los métodos de combinación de clasificadores. Los resultados alcanzados demuestran que estos métodos no sólo permiten aproximarse al mejor de los clasificadores unitarios, sino mejorarlos y reducir las tasas de error, alcanzando unos niveles de precisión del 94,97% en la predicción de microrrechupes y del 86,63% para la resistencia a la tracción.

El enfoque aquí alcanzado puede suponer un cambio para los sistemas MPC, permitiendo evolucionarlos al proveer de un proceso de predicción basado en un aprendizaje automático e híbrido, lo que les hace capaces de ajustarse tanto a procesos productivos lineales, como a los no lineales.

7. BIBLIOGRAFÍA

[1] Camacho EF and Bordons CA. Model Predictive Control in the Process Industry. S.l.: Springer-Verlag New York, Inc., 1997.
[2] ANSARI RM and TADÉ MO. Non-linear model-based process control: applications in petroleum refining. In: Measurement Science and Technology. 2000, Vol. 11, pp. 1830. http://dx.doi.org/10.1088/0957-0233/11/12/713
[3] Bekker JG, Craig IK and Pistorius PC. Model predictive control of an electric arc furnace off-gas process. In: Control Engineering Practice. 2000, Vol. 8, no. 4, pp. 445–455. http://dx.doi.org/10.1016/S0967-0661(99)00163-X
[4] Abou-Jeyab RA, Gupta YP, Gervais JR et al. Constrained multivariable control of a distillation column using a simplified model predictive control algorithm. In: Journal of Process Control. 2001, Vol. 11, no. 5, pp. 509–517. http://dx.doi.org/10.1016/S0959-1524(00)00029-9
[5] Camacho EF and Bordons C. Model predictive control. S.l.: Springer Verlag, 2004.
[6] Qin SJ and Badgwell TA. A Survey of Industrial Model Predictive Control Technology. In: Control engineering practice. 2003, Vol. 11, no. 7, pp. 733–764. http://dx.doi.org/10.1016/S0967-0661(02)00186-7
[7] Nieves J, Santos I and Bringas PG. Model Predictive Control on High Precision Foundries, a New Approach for the Prediction Phase. In: Revista Metalurgia. 2011, Vol. 47, no. 4, pp. 197–204. DOI 10.3989/revmetalm.1059.
[8] Maciejowski JM. Predictive control: with constraints. S.l.: Pearson education, 2002.
[9] Goodwin G, Seron MM and De Doná J.A. Constrained control and estimation: an optimisation approach. S.l.: Springer Verlag, 2005.
[10] Qin SJ and Badgwell TA. An Overview of Nonlinear Model Predictive Control Applications. In: Nonlinear Model Predictive Control. 2000, Vol. 26, pp. 369–392.
[11] Kassmann DE, Badgwell TA and Hawkins RB. Robust Steady-State Target Calculation for Model Predictive control. In: AIChE Journal. 2000, Vol. 46, no. 5, pp. 1007–1024. http://dx.doi.org/10.1002/aic.690460513
[12] KUNCHEVA Ludmila I. Combining Pattern Classifiers: Methods and Algorithms. S.l.: Wiley-Interscience, 2004. ISBN 0471210781. http://dx.doi.org/10.1002/0471660264
[13] SEEWALD A and FÜRNKRANZ J. An Evaluation of Grading Classifiers. In: Advances in Intelligent Data Analysis. 2001, pp. 115–124. DOI http://dx.doi.org/10.1007/3-540-44816-0_12. http://dx.doi.org/10.1007/3-540-44816-0_12
[14] Bishop CM. Pattern recognition and machine learning. S.l.: Springer New York, 2006.
[15] WOLPERT DH. Stacked Generalization. In: Neural networks. 1992, Vol. 5, no. 2, pp. 241–259. http://dx.doi.org/10.1016/S0893-6080(05)80023-1