

Técnicas de clasificación supervisadas para la detección de anomalías en el control de procesos industriales

Álvaro Michelena, Francisco Zayas-Gato, Esteban Jove, José-Luis Casteleiro-Roca, Héctor Quintián, Natalia Prieto-Fernández, Héctor Alaiz-Moretón and José Luis Calvo-Rolle
 University of A Coruña, CTC, CITIC,
 Department of Industrial Engineering, Ferrol, A Coruña, Spain
 {alvaro.michelena, f.zayas.gato, esteban.jove, jose.luis.casteleiro, hector.quintian, jlcalvo}@udc.es
 University of León,
 Department of Electrical Engineering and Systems and Automatic, León, Spain
 {npirf, hector.moreton}@unileon.es

Resumen

En la actualidad la detección de anomalías en procesos industriales es clave para optimizarlos y generar una mayor eficiencia en el proceso productivo, reportando unos mayores beneficios a las empresas. Por ello, en el presente artículo se implementan cinco técnicas de clasificación supervisadas para la detección de anomalías en sistemas industriales. Estas técnicas han sido entrenadas y validadas empleando un conjunto de datos que incluían datos etiquetados de funcionamiento normal y anómalo de una planta de control de nivel de líquido en un depósito. Finalmente, los resultados obtenidos fueron analizados y comparados para obtener el modelo con el que se obtiene un mayor rendimiento.

Palabras clave: Detección de anomalías, clasificadores supervisados y sistemas de control.

1. Introducción

El continuo avance tecnológico ha provocado que en la actualidad la mayoría de los procesos industriales se hayan automatizado, desarrollando complejos sistemas en los que intervienen una gran cantidad de sensores y actuadores que deben de funcionar con alta precisión y de forma completamente coordinada para un correcto funcionamiento de todo el proceso.

Por todo ello, para optimizar estos sistemas y obtener un mayor rendimiento del sistema productivo, y, por ende, mayores ganancias, la detección de anomalías o fallos en cualquiera de los componentes que conforma un sistema es un factor clave para las empresas.

Se entiende anomalía o comportamiento anómalo a aquellos datos que se distancian o diferencian del patrón que define el comportamiento normal del sistema. En la Figura 1 se muestra un ejemplo de anomalías en un conjunto de datos. El funcio-

namiento normal está conformado por la región definida por el conjunto de datos M , siendo los puntos a_1 y a_2 datos anómalos que se encuentran distanciados de la región de funcionamiento del sistema.

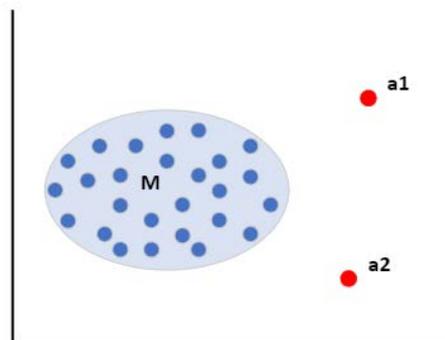


Figura 1: Ejemplo de anomalías

Debido a la creciente complejidad de los sistemas, así como al continuo desarrollo computacional, en los últimos años muchas líneas de investigación se han centrado en la detección de anomalías o fallos en una gran cantidad de sistemas y campos, tales como, la medicina [12, 19, 20, 22, 25, 13], la economía [24], la informática [23, 9] o la industria [2, 7, 8, 11, 3, 17, 18], entre otros.

1.1. Estado del arte

Dentro del proceso de detección de anomalías, y en función del conjunto de datos del que se disponga, se diferencian tres métodos distintos [4]:

- **Métodos supervisados para la detección de anomalías:** este tipo de técnicas utilizan un conjunto de datos, que contiene datos etiquetados tanto de funcionamiento normal como de anomalías. Este conjunto de datos se utiliza para entrenar un modelo de aprendizaje automático para así obtener un clasificador binario que clasifique datos normales y anómalos. Estos métodos son ampliamente

empleados cuando se dispone de un conjunto de datos etiquetados.

- **Detección de anomalías con métodos semi-supervisados:** más conocidos como métodos de una clase, *One-class*, estas técnicas requieren de un conjunto de datos con sólo datos de funcionamiento normal. En este caso, las anomalías se detectan cuando los nuevos datos tienen características diferentes a los datos normales.
- **Detección de anomalías con técnicas no supervisadas:** estos algoritmos no requieren datos de entrenamiento, y su funcionamiento asume que el comportamiento normal es más frecuente que el comportamiento anormal que ocurre con menos frecuencia.

En el ámbito de procesos de control muchas técnicas han sido empleadas. Por ejemplo, en [15], se emplean técnicas no supervisadas para la detección de anomalías en lazos de control industriales. Por otro lado, en [16] se emplean técnicas semi-supervisadas, *One-class* para detectar anomalías en el control del nivel de llenado de tanques.

Por ello, en el presente artículo se comparan el rendimiento de diferentes técnicas supervisadas para la detección de anomalías en lazos de control de procesos industriales. En este caso se emplean datos de funcionamiento de una planta de control de nivel de líquido almacenado en un tanque como la empleada en los estudios citados anteriormente.

Este artículo se estructura de la siguiente manera: después de la presente sección, se presenta el caso de estudio Sección 2. En la Sección 3, se describen los métodos de clasificación implementados. Por otro lado, la Sección 4 recoge la configuración de los experimentos, mientras que la Sección 5 presenta los resultados obtenidos. Por último, las conclusiones y trabajos futuros se recogen en la Sección 6.

2. Caso de estudio

En la presente sección se describirá de forma detallada tanto la planta de estudio como el conjunto de datos empleado.

2.1. Planta de control de nivel

El Laboratorio de Optimización y Control de la Escuela Politécnica de Ingeniería de Ferrol (EPIF) de la Universidad de A Coruña, dispone de diferentes plantas educativas que simulan procesos industriales.

Una de estas plantas didácticas es la planta de control de nivel. Este sistema se utiliza para controlar el nivel de llenado de agua en un depósito. Como se muestra en la Figura 2, este sistema consta de dos tanques que están colocados a diferentes alturas. El tanque superior es donde se controla el nivel de llenado mientras que el tanque inferior se utiliza para almacenar el agua descargada desde el tanque superior. Para propulsar el agua del tanque inferior al superior se utiliza una bomba centrífuga controlada por un variador de frecuencia (VFD). Además, para medir el nivel de llenado en el tanque superior se emplea un sensor de ultrasonidos.



Figura 2: Planta de control de nivel

Por otro lado, el sistema también incluye una válvula manual y servopilotada para controlar el caudal de vaciado del tanque superior.

En la Figura 3 se presenta de forma esquemática la estructura de la planta de control de nivel.

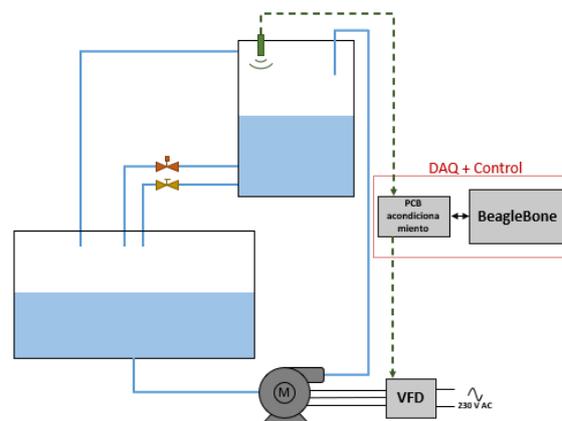


Figura 3: Esquema de la estructura de la planta de control de nivel

2.2. Implementación del lazo de control

Tanto para el proceso de adquisición de datos como para el propio proceso de control del sistema, se emplea una BeagleBone Black junto con una PCB de acondicionamiento que adapta las señales a las tensiones de trabajo del dispositivo. Esta placa de acondicionamiento es necesaria ya que la planta trabaja con señales de entrada y salida de 0/10V DC mientras que la BeagleBone Black trabaja con señales 0/3.3V DC.

Para el control de la planta se emplea un controlador virtual, diseñado en Python, que es ejecutado sobre la BeagleBone Black. El controlador recibe a través de la placa de acondicionamiento el nivel de líquido detectado por el sensor de ultrasonidos. En función del valor de consigna y del nivel medido, el controlador genera una señal de control con la que se regula la velocidad de la bomba centrífuga. Es importante destacar que, para simplificar el proceso de control, todas las señales implicadas en el controlador virtual se trabajan en valores porcentuales. La Figura 4 muestra de forma esquemática el lazo de control del sistema.

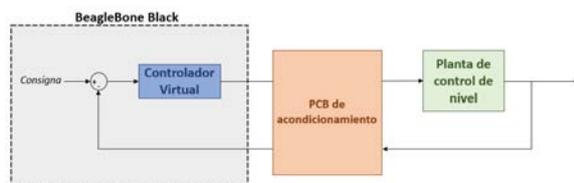


Figura 4: Esquema del proceso de control de la planta de nivel

La planta de control de nivel corresponde a un sistema no lineal, lo que significa que sus dependencias funcionales cambian según el valor de la consigna. Por esta razón, para controlar el sistema se implementa un controlador PID adaptativo, junto con un proceso de identificación en tiempo real basado en el método de mínimos cuadrados recursivos (RLS, por sus siglas en inglés) [6]. En este caso, la planta se identifica con una función de transferencia de segundo orden con retardo, que se define en la Ecuación 1:

$$G_p(z^{-1}) = \frac{b_0 z^{-k}}{1 - a_0 z^{-1} - a_1 z^{-2}} \quad (1)$$

en donde k es el valor del retardo y a_0 , a_1 y b_0 son los parámetros de la función de transferencia. A partir de la función de transferencia identificada, se ajusta un controlador PID autoajutable que sigue la Ecuación 2:

$$G_{controlador}(z^{-1}) = \frac{p_0 + p_1 z^{-1} + p_2 z^{-2}}{1 - z^{-1}} \quad (2)$$

en donde los coeficientes p_0 , p_1 y p_2 se determinan por medio de las Ecuaciones 3, 4 y 5:

$$p_0 = \frac{1}{b_0} \left(1 + a_0 - \frac{3}{a} \right) \quad (3)$$

$$p_1 = \frac{1}{b_0} \left(a_1 - a_0 - \frac{3}{a^2} \right) \quad (4)$$

$$p_2 = -\frac{1}{b_0} \left(a_1 + \frac{1}{a^3} \right) \quad (5)$$

2.3. Conjunto de datos empleado

Puesto que el objetivo de la presente investigación es el detectar anomalías en el funcionamiento del sistema descrito, el conjunto de datos está conformado tanto por datos de funcionamiento normal como por datos anómalos. Para ello se registraron las señales de control, de salida del sistema y del valor de consigna, así como los coeficientes de la función de transferencia obtenidos por el proceso de identificación durante el funcionamiento del sistema en diferentes configuraciones. Tanto para la obtención de los datos de funcionamiento normal como anómalo el tiempo de muestreo empleado fue de 0.5 segundos.

Para los datos normales, se puso en operación la planta para diferentes puntos de operación con las válvulas de descarga completamente abiertas. Los puntos de operación empleados fueron desde el 10% hasta el 90% con incrementos del 10% del nivel de llenado.

Por otro lado, el conjunto de datos anómalos corresponde al funcionamiento para distintos valores de consigna en el que la electroválvula se abre gradualmente. Esta válvula se abre del 10% al 90% con incrementos del 10%.

Finalmente, el conjunto de datos obtenido tiene un total de 72900 datos y un total de 7 variables definidas a continuación:

- SP: nivel deseado de llenado (*Set Point*).
- CP: señal de control (*Control Process*).
- PV: salida del sistema (*Process Value*).
- b_0 : coeficiente identificado de la función de transferencia.
- a_0 : coeficiente identificado de la función de transferencia.
- a_1 : coeficiente identificado de la función de transferencia.

- Anomalía: tipo de dato. Esta variable es binaria y tiene un valor 1 cuando los datos registrados corresponden a un funcionamiento anómalo.

Por último, destacar que el conjunto de datos recoge 48600 muestras de funcionamiento normal y 24300 de funcionamiento anómalo.

3. Métodos de clasificación supervisada

En esta sección se procederá a explicar las diferentes técnicas supervisadas empleadas.

3.1. Discriminante Lineal de Fisher

El discriminante lineal de Fisher, también conocido como LDA por sus siglas en inglés, *Linear Discriminant Analysis*, es un método de clasificación desarrollado por R.A. Fisher [10], ampliamente utilizado en estadística y en problemas de clasificación de aprendizaje automático.

La idea principal de este método es encontrar el hiperplano ideal que permita proyectar los datos sobre él separando lo máximo posible los datos de ambas clases. Para ello, el LDA busca el hiperplano en el que las clases tengan la mínima varianza entre sus datos y en el que las medias de cada clase estén lo más alejadas posible. El problema de optimización se basa en maximizar la función objetivo $J(\theta)$ definida por la Ecuación 6:

$$J(\theta) = \frac{(\mu_1 - \mu_2)^2}{\hat{s}_1^2 + \hat{s}_2^2} \quad (6)$$

donde μ_1 y μ_2 corresponden al valor medio de la clase 1 y 2, y \hat{s}_1 y \hat{s}_2 definen la varianza dentro de la clase 1 y 2.

3.2. Vecinos Próximos

La técnica de vecinos próximos, más conocida como k -NN de sus siglas en inglés k-Nearest Neighbor, es una técnica no lineal de clasificación supervisada [14]. Este método estadístico se basa en la idea de que los datos de la misma clase se encuentran cerca en el espacio. Por ello, la clase de un nuevo dato dependerá únicamente de la clase mayoritaria de los k datos más cercanos (conocidos como vecinos).

Por todo ello, el rendimiento de este algoritmo depende de forma directa del valor de k , el cual debe de ser ajustado de manera precisa. Para el ajuste de este valor existen diferentes técnicas que pueden ser aplicadas, siendo una de las más frecuentes la de comparar los resultados obtenidos con

diferentes valores de k y seleccionar aquel valor que mejor se ajuste al problema. Hay que tener en cuenta que valores grandes de k producen regiones de decisión muy simplificadas.

Esta técnica no genera ningún modelo durante el proceso de entrenamiento, sino que únicamente almacena los datos de entrada. De esta manera el proceso de entrenamiento es muy rápido. Por contra, el proceso de clasificación de un nuevo dato corresponde con un proceso bastante lento, pues es necesario calcular la distancia del nuevo punto a todos los datos almacenados, lo que dificulta su implementación en sistemas críticos donde se requiere un reducido tiempo de clasificación.

Para determinar cuáles son los vecinos de un nuevo dato se emplean diferentes medidas de distancia siendo la más común la distancia Euclídea.

3.3. Árboles Aleatorios

Los árboles de decisión (más conocidos como *Decision Trees*) son una técnica de aprendizaje automático no paramétrica y supervisada ampliamente empleada en problemas de clasificación y regresión [21]. Como su propio nombre indica este método construye un árbol de decisión con las variables del conjunto de datos de entrenamiento. Estos árboles de decisión están compuestos por dos tipos de nodos diferentes:

- Nodos de decisión: corresponden con los nodos interiores del árbol. Estos nodos están asociados a una de las variables del conjunto de datos y tienen 2 o más ramas que salen de ellos, cada una de las cuales representa los posibles valores que puede tomar esa variable.
- Nodos de respuesta: también conocidos como hojas, que son los nodos finales del árbol y que indican la clase a la que pertenece el nuevo dato.

De esta manera para clasificar un nuevo dato se recorre el árbol desde el nodo inicial (raíz) hasta llegar a uno de los nodos de respuesta. El camino que recorre el nuevo dato lo determinan las ramas pasando por los nodos de decisión que determinan la rama por la que tiene que avanzar el nuevo dato en función del valor de sus variables.

El rendimiento de este modelo depende de la selección de las variables en cada etapa del árbol. Para ello el algoritmo se basa en medir el grado de incertidumbre de las variables empleando para ello el cálculo de la entropía o del índice Gini. Obtenido el valor de incertidumbre se puede obtener la ganancia de información de cada variable, comparando la impureza del conjunto de datos antes

y después de la división de los nodos. Finalmente, el árbol de decisión se construye de arriba hacia abajo colocando las variables con mayor ganancia de información en las primeras capas del árbol.

3.4. Bosques Aleatorios

Los bosques aleatorios, más conocidos como Random Forest, es una técnica supervisada de aprendizaje automático muy empleada en problemas de clasificación, aunque también se puede emplear en problemas de regresión. Este método fue propuesto por Breiman [1], y se basa en la combinación de diferentes árboles de decisión y la técnica de agregación Bootstrap.

La técnica de agregación Bootstrap más conocida como Bagging, es una técnica utilizada para, a partir del conjunto de datos inicial, generar subconjuntos aleatorios con los que entrenar cada árbol de decisión implementado en el bosque aleatorio. De esta manera cada uno de los árboles de decisión, al ser entrenados con diferentes conjuntos de datos, serán distintos. El método Bagging corresponde a un proceso de selección por sustitución, por lo que la misma muestra de datos puede estar en más de un subconjunto.

Finalmente, con todos los árboles de decisión entrenados de forma independiente, la clasificación del bosque aleatorio corresponde a la clase obtenida en la mayoría de los árboles de decisión.

3.5. Máquina de Vector de Soporte

Las máquinas de vector de soporte, más conocidas como SVMs, por sus siglas en inglés *Support Vector Machines*, son un conjunto de algoritmos supervisados de aprendizaje automático introducidos por Cortes y Vapnik [5]. Este tipo de algoritmos se utilizan comúnmente tanto en problemas de clasificación como de regresión. El objetivo principal de las SVM es encontrar un hiperplano que maximice la distancia mínima entre el hiperplano y las muestras de cada clase más cercanas al hiperplano calculado (lo que se conoce como margen).

La definición anterior de las SVM supone que las clases pueden separarse linealmente. Sin embargo, la mayoría de los conjuntos de datos reales no pueden separarse linealmente. En estos casos, los algoritmos de SVMs implementan transformaciones de datos para mapearlos en un espacio de mayor dimensión en el que el conjunto de datos puede ser linealmente separable. La función de transformación $\phi(x)$ se determina mediante la selección de una determinada función conocida como *Kernel*.

4. Experimentos

Para la ejecución de los experimentos se desarrolló un algoritmo en Python donde se implementaron los métodos presentados en el apartado anterior. Para una mejor comprensión de los experimentos realizados este apartado se divide en tres secciones claramente diferenciadas.

4.1. Preprocesado de datos

La primera etapa que se ejecutó fue el preprocesado del conjunto de datos obtenido. En esta primera etapa se procedió a comprobar si había muestras que contenían datos faltantes o erróneos. Por último, y con el objetivo de obtener unos mejores resultados se procedió a normalizar el conjunto de datos con una normalización de media 0 y desviación estándar 1.

4.2. Ajuste de los clasificadores

Los clasificadores empleados fueron comprobados para diferentes valores de sus hiperparámetros. Las configuraciones que se probaron fueron las presentadas a continuación:

- Para el modelo del discriminante lineal de Fisher se probaron dos métodos distintos de resolución (*solver*).
- Para los árboles de decisión se ajustó la máxima profundidad, es decir, el número máximo de capas que puede tener el árbol (este valor puede no ser especificado).
- Para los bosques aleatorios se ajustó el número de árboles que se integran.
- Para el algoritmo de vecinos próximos se comprobó su funcionamiento para distintos valores de k (número de vecinos).
- Por último las máquinas de vector de soporte fueron configuradas con dos tipos de kernel diferentes y con diferentes valores de regularización.

4.3. Evaluación de los clasificadores

Las diferentes configuraciones de los modelos empleados fueron entrenadas por medio de un proceso de validación cruzada con $k = 10$. Con los modelos entrenados, se procedió a comparar los modelos basados en la misma técnica con el objetivo de conocer qué valores de sus hiperparámetros se ajustaban mejor al problema planteado. Finalmente, los modelos de cada tipo seleccionados se compararon entre ellos para saber que clasificador obtenía unos mejores resultados.

Es importante destacar que la métrica empleada para comparar y seleccionar el mejor modelo es el área bajo la curva ROC, ampliamente conocida como AUC por sus siglas en inglés, *Area Under the receiving operating Curve*. La selección de esta métrica se debe a que el conjunto de datos empleado está desbalanceado y esta métrica es insensible a la distribución de estos datos.

5. Resultados obtenidos

Siguiendo los pasos descritos en el apartado anterior los resultados obtenidos son expuestos. En los resultados se muestra tanto la métrica AUC obtenida como el tiempo de entrenamiento, con el que se puede determinar el coste computacional de cada uno de los modelos.

La Tabla 1 muestra los resultados obtenidos con las dos configuraciones del método del discriminante lineal de Fisher. En este caso, variar el *solver* no afecta a los resultados obtenidos pero si al coste computacional, donde el *solver* basado en mínimos cuadrados (*lsqr*) requiere un menor tiempo de entrenamiento.

Cuadro 1: Resultados discriminante lineal
Discriminante Lineal

Solver	AUC	T. entrenamiento (s)
SVD	0,734	0,183
LSQR	0,734	0,151

Por otro lado, la Tabla 2 muestra los resultados obtenidos para los diferentes árboles de decisión creados. En este caso se puede observar como el mayor valor de AUC, 97,3 %, se obtiene fijando una máxima profundidad de 15 capas. Así mismo el coste computacional es bastante reducido con un tiempo de entrenamiento de 0,35 segundos.

Para el caso del método de bosques aleatorios los resultados obtenidos reflejaron un alto porcentaje de AUC superior al 96 % en todas las configuraciones planteadas. Al obtener un valor de AUC muy similar en todos ellos, para la selección de la mejor configuración se tuvo también en cuenta el tiempo de entrenamiento. Por ello, la configuración seleccionada fue el bosque aleatorio compuesto por 15 árboles de decisión ya que era el que mayor AUC obtenía, 97,3 %, con un menor tiempo de entrenamiento, 1,525 segundos.

Por otro lado en el caso del método de vecinos próximos se obtuvo como el mejor modelo se obtenía para un valor de $k = 5$ llegando a alcanzar

Cuadro 2: Resultados árboles de decisión
Árbol de Decisión

Máx. prof.	AUC	T. entrenamiento (s)
3	0,709	0,129
5	0,843	0,190
7	0,893	0,234
10	0,945	0,287
12	0,969	0,312
15	0,973	0,350
None	0,970	0,405

Cuadro 3: Resultados Bosques Aleatorios
Bosques Aleatorios

Nº árboles	AUC	T. entrenamiento (s)
3	0,968	0,419
5	0,971	0,663
8	0,972	1,073
10	0,972	1,270
15	0,973	1,525
20	0,973	2,055
25	0,973	2,542
30	0,973	3,090
40	0,973	4,125

un 87,7 % del valor de AUC.

Por último la configuración de las máquinas de vector de soporte que mejor rendimiento obtuvo fue para un valor de regularización de 0,1 y un *Kernel* lineal. Sin embargo este método supone un coste computacional bastante alto, con un entrenamiento superior a 62 segundos para cualquiera de las configuraciones.

Comparando todos los métodos se puede concluir que el mejor método para la detección de anomalías en el proceso de control de la planta de nivel es el método de bosques aleatorios con un valor de AUC del 97,3 % y con un tiempo de entrenamiento de 1,525 segundos.

6. Conclusiones y trabajos futuros

Los resultados obtenidos mostraron una buena funcionalidad de los métodos del árbol de decisión y bosques aleatorios para la detección de anomalías en el funcionamiento de la planta de control de nivel mostrada. Con estos métodos se obtuvieron valores de AUC superiores al 97 % para las

Cuadro 4: Resultados Vecinos Próximos
k-Vecinos Próximos

k	AUC	T. entrenamiento (s)
5	0,877	0,893
10	0,857	0,990
15	0,851	1,086
20	0,840	1,071
30	0,829	1,183
40	0,819	1,275
50	0,810	1,364
80	0,792	1,591
100	0,782	1,723

Cuadro 5: Resultados Máquina de Vector de Soporte

Máquina de Vector de Soporte			
Kernel	C	AUC	T. entrenamiento (s)
linear	1,0	0,785	93,215
linear	0,1	0,783	62,463
linear	0,01	0,764	79,903
linear	0,001	0,738	94,725
poly	1,0	0,756	98,952
poly	0,1	0,736	96,493
poly	0,01	0,664	104,685
poly	0,001	0,553	111,326

mejores configuraciones de sus hiperparámetros.

Por otro lado, el método de vecinos próximos también mostró buenos resultados con un valor máximo de AUC del 87,7%. Por contra, las máquinas de vector de soporte así como el discriminante lineal de Fisher no arrojaron tan buenos resultados como las técnicas anteriores, no superando para ninguna configuración el 80% de AUC.

Como líneas futuras se implementarán otros algoritmos supervisados, así como otro tipo de tecnologías no supervisadas y semi-supervisadas para comparar los resultados obtenidos. Así mismo, también se comprobará la funcionalidad de los algoritmos propuestos con otras plantas que hay en el Laboratorio de Optimización y Control de la Escuela Politécnica de Ingeniería de Ferrol (EPIF) de la Universidad de A Coruña.

Agradecimientos

Este trabajo ha sido realizado gracias al apoyo del Centro de Investigación en Tecnologías de la Información y las Comunicaciones (CITIC), como Centro de Investigación del Sistema Universitario de Galicia, financiado por la Consellería de Educación, Universidade e Formación Profesional de la Xunta de Galicia a través del Fondo Europeo de Desarrollo Regional (FEDER) y la Secretaría Xeral de Universidades (Ref. ED431G 2019/01).

English summary

Abstract

Nowadays, detecting anomalies in industrial processes is key to optimizing them and generating greater efficiency in the production process, bringing more significant benefits to companies. Therefore, in this paper, five supervised classification techniques are implemented to detect anomalies in industrial systems. These techniques have been trained and validated using a dataset that included labeled normal and anomalous operation data from a liquid level control plant. Finally, the results obtained were analyzed and compared to obtain the model with the best performance.

Keywords: Anomaly detection, supervised classifiers and control systems.

Referencias

- [1] Biau, G.: Analysis of a random forests model. *The Journal of Machine Learning Research* **13**(1), 1063–1095 (2012)
- [2] Casado-Vara, R., Sittón-Candanedo, I., la Prieta, F.D., Rodríguez, S., Calvo-Rolle, J.L., Venayagamoorthy, G.K., Vega, P., Prieto, J.: Edge computing and adaptive fault-tolerant tracking control algorithm for smart buildings: A case study. *Cybernetics and Systems* **51**(7), 685–697 (2020)
- [3] Castelleiro-Roca, J.L., Javier Barragan, A., Segura, F., Luis Calvo-Rolle, J., Manuel Andujar, J.: Intelligent hybrid system for the prediction of the voltage-current characteristic curve of a hydrogen-based fuel cell. *Revista Iberoamericana de Automática e Informática industrial* **16**(4), 492–501 (2019)

- [4] Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: A survey. *ACM computing surveys (CSUR)* **41**(3), 1–58 (2009)
- [5] Cortes, C., Vapnik, V.: Support-vector networks. *Machine learning* **20**(3), 273–297 (1995)
- [6] Engel, Y., Mannor, S., Meir, R.: The kernel recursive least-squares algorithm. *IEEE Transactions on signal processing* **52**(8), 2275–2285 (2004)
- [7] Fernandez-Serantes, L.A., Casteleiro-Roca, J.L., Berger, H., Calvo-Rolle, J.L.: Hybrid intelligent system for a synchronous rectifier converter control and soft switching ensurement. *Engineering Science and Technology, an International Journal* p. 101189 (2022)
- [8] Fernandez-Serantes, L.A., Casteleiro-Roca, J.L., Calvo-Rolle, J.L.: Hybrid intelligent system for a half-bridge converter control and soft switching ensurement. *Revista Iberoamericana de Automática e Informática industrial* (2022)
- [9] Fernández-Serantes, L.A., Estrada Vázquez, R., Casteleiro-Roca, J.L., Calvo-Rolle, J.L., Corchado, E.: Hybrid intelligent model to predict the soc of a lfp power cell type. In: *International conference on hybrid artificial intelligence systems*. pp. 561–572. Springer (2014)
- [10] Fisher, R.A.: The use of multiple measurements in taxonomic problems. *Annals of eugenics* **7**(2), 179–188 (1936)
- [11] García-Ordás, M.T., Alaiz-Moretón, H., Casteleiro-Roca, J.L., Jove, E., Benítez-Andrades, J.A., García-Rodríguez, I., Quintián, H., Calvo-Rolle, J.L.: Clustering techniques selection for a hybrid regression model: A case study based on a solar thermal system. *Cybernetics and Systems* **0**(0), 1–20 (2022)
- [12] Gonzalez-Cava, J.M., Arnay, R., Mendez-Perez, J.A., León, A., Martín, M., Reboso, J.A., Jove-Perez, E., Calvo-Rolle, J.L.: Machine learning techniques for computer-based decision systems in the operating theatre: application to analgesia delivery. *Logic Journal of the IGPL* **29**(2), 236–250 (09 2020)
- [13] Gonzalez-Cava, J.M., Reboso, J.A., Casteleiro-Roca, J.L., Calvo-Rolle, J.L., Méndez Pérez, J.A.: A novel fuzzy algorithm to introduce new variables in the drug supply decision-making process in medicine. *Complexity* **2018** (2018)
- [14] Guo, G., Wang, H., Bell, D., Bi, Y., Greer, K.: Knn model-based approach in classification. In: *OTM Confederated International Conferences. On the Move to Meaningful Internet Systems*. pp. 986–996. Springer (2003)
- [15] Jove, E., Casteleiro-Roca, J.L., Quintián, H., Méndez-Pérez, J.A., Calvo-Rolle, J.L.: A fault detection system based on unsupervised techniques for industrial control loops. *Expert Systems* **36**(4), e12395 (2019)
- [16] Jove, E., Casteleiro-Roca, J.L., Quintián, H., Simić, D., Méndez-Pérez, J.A., Luis Calvo-Rolle, J.: Anomaly detection based on one-class intelligent techniques over a control level plant. *Logic Journal of the IGPL* **28**(4), 502–518 (2020)
- [17] Jove, E., Casteleiro-Roca, J.L., Casado-Vara, R., Quintián, H., Pérez, J.A.M., Mohamad, M.S., Calvo-Rolle, J.L.: Comparative study of one-class based anomaly detection techniques for a bicomponent mixing machine monitoring. *Cybernetics and Systems* **51**(7), 649–667 (2020)
- [18] Jove, E., Casteleiro-Roca, J.L., Quintián, H., Méndez-Pérez, J.A., Calvo-Rolle, J.L.: A new method for anomaly detection based on non-convex boundaries with random two-dimensional projections. *Information Fusion* **65**, 50–57 (2021)
- [19] Jove, E., Gonzalez-Cava, J.M., Casteleiro-Roca, J.L., Quintián, H., Méndez Pérez, J.A., Vega Vega, R., Zayas-Gato, F., de Cos Juez, F.J., León, A., Martín, M., Reboso, J.A., Wozniak, M., Luis Calvo-Rolle, J.: Hybrid Intelligent Model to Predict the Remifentanyl Infusion Rate in Patients Under General Anesthesia. *Logic Journal of the IGPL* **29**(2), 193–206 (09 2020)
- [20] Jove, E., Gonzalez-Cava, J.M., Casteleiro-Roca, J.L., Quintián, H., Méndez Pérez, J.A., Vega Vega, R., Zayas-Gato, F., de Cos Juez, F.J., León, A., Martín, M., Reboso, J.A., Wozniak, M., Luis Calvo-Rolle, J.: Hybrid Intelligent Model to Predict the Remifentanyl Infusion Rate in Patients Under General Anesthesia. *Logic Journal of the IGPL* **29**(2), 193–206 (09 2020)
- [21] Kotsiantis, S.B.: Decision trees: a recent overview. *Artificial Intelligence Review* **39**(4), 261–283 (2013)
- [22] Leira, A., Jove, E., Gonzalez-Cava, J.M., Casteleiro-Roca, J.L., Quintián, H., Zayas-

- Gato, F., Álvarez, S.T., Simic, S., Méndez-Pérez, J.A., Luis Calvo-Rolle, J.: One-Class-Based Intelligent Classifier for Detecting Anomalous Situations During the Anesthetic Process. *Logic Journal of the IGPL* (11 2020)
- [23] Michelena, Á., Zayas-Gato, F., Jove, E., Calvo-Rolle, J.L.: Detection of dos attacks in an iot environment with mqtt protocol based on intelligent binary classifiers. *Engineering Proceedings* **7**(1), 16 (2021)
- [24] Shahid, N., Shah, M.A.: Anomaly detection in system logs in the sphere of digital economy (2021)
- [25] Sipes, T., Jiang, S., Moore, K., Li, N., Karimabadi, H., Barr, J.R.: Anomaly detection in healthcare: Detecting erroneous treatment plans in time series radiotherapy data. *International Journal of Semantic Computing* **8**(03), 257–278 (2014)

