

Minería de datos en series temporales para la búsqueda de conocimiento oculto en históricos de procesos industriales

Francisco Javier Martínez de Pisón Ascacibar, Joaquín B. Ordieres Meré,
Alpha V. Pernía Espinoza, Fernando Alba Elías

Grupo EDMANS
Área de Proyectos de Ingeniería
Dept. de Ingeniería Mecánica
Universidad de La Rioja
Luis de Ulloa 20
26004 Logroño (La Rioja)

fjmartin@dim.unirioja.es, joaquin.ordieres@dim.unirioja.es, alpha.pernia@dim.unirioja.es,
fernando.alba@dim.unirioja.es

Resumen

Dentro de la Minería de Datos aplicada a la Optimización de Procesos Industriales, uno de los campos de investigación que más futuro tiene corresponde con la búsqueda de reglas o patrones ocultos en históricos que puedan ayudar en la toma de decisiones o en la mejora de los procesos productivos.

En este artículo, se plantea un tipo de metodología¹ para la búsqueda de reglas asociativas que permitan extraer conocimiento, no obvio, de bases de datos de históricos de procesos industriales.

La metodología se basa en la extracción de tramos dentro de cada una de las series temporales previamente filtradas, a partir de los máximos y mínimos de las mismas. Una vez realizada la segmentación, se agrupan éstos, mediante clisterizado jerárquico, según forma y longitud, etiquetando solamente aquellas secuencias o patrones que aparecen un número elevado de veces. Posteriormente, se combinan secuencialmente todos los patrones etiquetados de todas las variables a estudiar en una sola combinación o secuencia genética.

Por último, se busca dentro de una ventana temporal, reglas asociativas temporales mediante un algoritmo basado en el algoritmo APRIORI,

presentándose éstas en una forma comprensible por el experto.

En este artículo se explicará la metodología planteada y las etapas en que se divide la misma.

1. Introducción

Dentro de un proceso industrial, es usual encontrarse con bases de datos repletas de series temporales. Usualmente, éstas corresponden a medidas continuas o discretas que siguen un orden no aleatorio y que son capturadas con una frecuencia de muestreo constante dentro de un periodo cronológico relativamente largo. Estas medidas suelen obtenerse de forma automática a partir de sensores que indican el estado de algún dispositivo a lo largo del tiempo o corresponden a valores que son introducidos manualmente por algún operario. También pueden ser señales de consigna o valores objetivo para dispositivos o actuadores (motores, cilindros, hornos, etc.), o variables que corresponden a estados o valores objetivo del proceso relacionados con dichos intervalos.

Tradicionalmente, el experto que analiza bases de datos de históricos de un proceso industrial, se encuentra con multitud de series temporales almacenadas en diversas tablas con intervalos de tiempo diferentes o incluso, con variables que deben ser convertidas a series temporales para que puedan ser analizadas correctamente.

El primer paso lógico, consiste en desarrollar una tabla (*Vista Minable*) con las variables más

¹ Está metodología, en la que está trabajando el grupo de investigación EDMANS (*Engineering Data Mining And Numerical Simulations*), se está comenzando a aplicar en procesos de fabricación del acero y medio ambientales.

importantes del proceso adaptándolas dentro de la misma escala de tiempos.

1.1. Extracción de Conocimiento

Es bien conocido, que la capacidad del cerebro humano para segmentar y extraer patrones visuales es muy superior a la de cualquier sistema de visión artificial actual.

El experto que pretende extraer algún conocimiento útil que permita desarrollar estrategias de mejora de un proceso industrial, utiliza esta habilidad para descubrir visualmente patrones repetitivos y sus interrelaciones en el tiempo. Muchas veces, cuando lo escuchamos argumentar sobre el comportamiento de una serie temporal oímos términos del estilo “*este segmento crece linealmente*” o “*este segmento decrece exponencialmente*”, que demuestran claramente la forma en que el ser humano describe localmente una serie temporal. Este tipo de segmentación visual se realiza dividiendo la serie temporal en segmentos cuya forma o apariencia se aproxima a patrones ya conocidos (líneas, curvas crecientes o decrecientes, etc.) igual a como hace el cerebro para describir cualquier nuevo objeto que se le presenta [7].

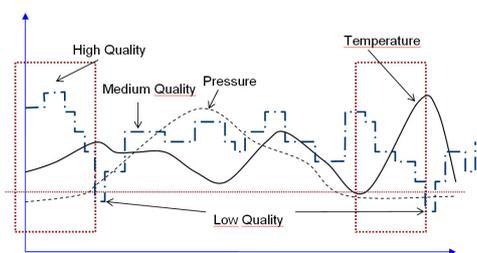


Figura 1. Detección de relaciones temporales entre variables de un proceso industrial.

Si el ser humano utiliza su conocimiento previo para describir nuevos objetos, así mismo, toda serie temporal puede ser descrita con mayor o menor precisión, mediante una serie de patrones ya conocidos.

El aspecto realmente interesante surge cuando se analizan aquellas secuencias de patrones que se repiten y se obtienen dependencias temporales entre diversas variables. Por ejemplo, analizando las series temporales correspondientes a un cierto proceso industrial (ver figura 1) podríamos

deducir la siguiente regla²: “*Cuando la temperatura sube linealmente y la presión permanece por debajo de un cierto nivel entonces se produce un descenso de la calidad del producto*”.

La búsqueda de este tipo de reglas puede complicarse aún más, porque este tipo de correlaciones locales no solamente puede corresponder al mismo instante temporal, sino que pueden aparecer dependencias entre variables con importantes desfases en el tiempo. Esto es muy común en sistemas con fuertes inercias como algunos procesos químicos o físicos, cuyas velocidades de respuesta son muy lentas e incluso varían según las condiciones del entorno.

El objetivo fundamental de este artículo, se centra en la descripción de una metodología que pretende descubrir patrones repetitivos y sus interdependencias dentro de series temporales de históricos de procesos industriales.

Primeramente, se pasará a describir la metodología propuesta describiendo cada uno de los pasos de la misma, y por último se plantearán las conclusiones y futuras líneas de investigación.

2. Metodología Planteada

La metodología propuesta se compone de las siguientes etapas.

1. Filtrado de cada serie temporal para eliminar el ruido y obtener la forma básica de la misma.
2. Obtención de los cruces por cero de la primera derivada (máximos y mínimos de la señal).
3. Extracción, para cada serie temporal, de los tramos “*incrementales (INC)*”, “*decrementales (DEC)*” y “*horizontales (HOR)*” según unos umbrales previamente establecidos³.
4. Agrupamiento de los tramos “*INC*”, “*DEC*” y “*HOR*” que tengan una forma y longitud similares. Etiquetado de los tramos obtenidos según la familia a la que pertenecen.
5. Búsqueda de secuencias de tramos parecidas (patrones) dentro de cada serie temporal.

² Lógicamente ésta será interesante si se repite un elevado número de veces.

³ Obviamente, pueden ser necesarios otros tipos de tramos. Por ejemplo: cuando una variable está por encima de un umbral o por debajo del mismo; o si una variable es cero, o si se produce una señal pulsante, etc.

Descripción de cada serie temporal únicamente con los patrones que se repiten asiduamente.

6. Creación de una secuencia única (secuencia genética) con la fusión de los patrones repetidos de todas las series temporales a estudiar.
7. Búsqueda de secuencias de patrones que se repitan entre las diversas series temporales.
8. Extracción de reglas asociativas de aquellos casos que aparecen un elevado número de ocasiones y tienen un elevado índice de aciertos.
9. Presentación de las reglas asociativas en una forma entendible por el usuario.

2.1. Filtrado de cada serie temporal

El primer paso, antes de realizar el proceso de segmentación de una señal, corresponde con el filtrado de la misma para obtener su forma básica.

Obviamente, si deseamos extraer patrones que indiquen el comportamiento de una serie temporal obtenida de un proceso industrial, debemos buscar la estructura general de la señal sin considerar el ruido.

En este caso, nosotros utilizamos un filtro de denominado “*filtro de ventana de gauss*” [8] porque el tipo de suavizado obtenido es muy estable para cada tipo de anchura de ventana utilizado.

Lógicamente, una vez decididos a utilizar un filtro de gauss, la siguiente duda que se nos plantea corresponde con el ancho de la ventana. Efectivamente, la idea fundamental es la de conseguir eliminar el ruido para obtener la forma básica de la señal, tal y como haría un experto visualmente, pero sin eliminar transitorios de la misma que realmente sean importantes. Esto es debido a que una ventana con una anchura demasiado grande (ver último caso de la figura 2) puede destruir la forma básica de la señal en las zonas con aquellos cambios bruscos de pendiente que deseamos considerar dentro del estudio, mientras que si la ventana es demasiado pequeña (ver segundo caso de la figura 2) el ruido puede seguir existiendo en la señal filtrada dificultando enormemente la caracterización de la señal.

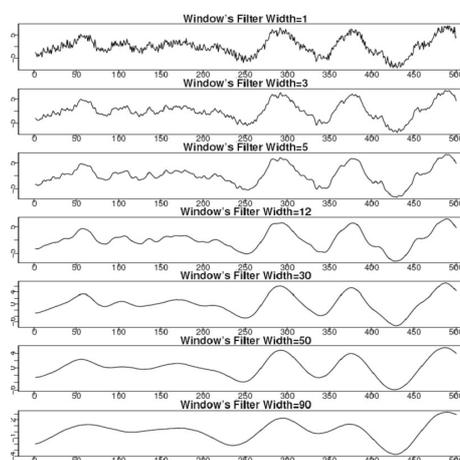


Figura 2. Filtros de diferentes anchuras.

Hay que tomar en cuenta, que cada una de las series temporales de un proceso industrial puede provenir de una variedad de fuentes considerable: ruido de sensores, fallos, discontinuidades, espúreos, distorsiones de señal, degradación de señales, perturbaciones locales, etc. Cada una de ellas puede tener un comportamiento en el dominio del tiempo y de la frecuencia completamente diferente, por lo que es necesario que sean estudiadas por el experto separadamente. Como cada tipo de serie temporal es muy diferente, no es fácil encontrar una técnica automática que permita obtener la anchura óptima de la ventana del filtro. Debido a esto, diversos autores han desarrollado diferentes estudios y técnicas que permiten buscar la anchura más adecuada, como por ejemplo: la Caracterización Multiescalar [7][8], el uso de Transformadas de Fourier o Wavelets [5].

2.2. Extracción de Tramos

Una vez aplicado el filtro adecuado y extraída la forma básica de cada serie temporal, el siguiente paso consiste en buscar e identificar tramos o segmentos dentro de la misma.

El problema de la segmentación de series temporales es bien conocido en la literatura científica. El número de técnicas es numeroso, aunque la mayoría de los trabajos se centran en las siguientes técnicas [5]: representación lineal por

trozos (*Piecewise Linear Representation (PLR)*), Transformadas de Fourier, Wavelets y Mapeado Simbólico (*Symbolic Mappings*). De entre ellas, cabe destacar la primera como una de las técnicas más utilizadas e interesantes a tomar en cuenta, ya que los porcentajes de compresión son muy elevados y los algoritmos son bastante rápidos.

Dentro de las técnicas PLR, se ha desarrollado una técnica que permite aproximarse al mismo razonamiento que emplearía un experto de una planta industrial que busca visualmente comportamientos repetitivos dentro de series temporales de históricos de un proceso. Es decir, se trata de simular el funcionamiento que tiene el cerebro cuando determina si dos trozos de serie temporal son similares o no.

Si observamos detenidamente cómo detectamos visualmente clases en series temporales, podemos ver que generalmente identificamos secuencias repetidas de tendencias dentro de la serie temporal. Dichas tendencias las describimos habitualmente como “subidas”, “bajadas” o “valores estables u horizontales” a las que se le añaden adjetivos del tipo: “corto”, “largo”, “brusco”, “pronunciado”, “suave”, “dilatado en el tiempo”, “fuerte”, “débil”, etc.; que suelen usarse para describir la magnitud y duración en el tiempo de las mismas.

Lógicamente, lo anteriormente expuesto es una visión demasiado simplista de la enorme capacidad que tiene el cerebro humano para extraer y comparar patrones de todo tipo, pero creemos que aún así, ésta puede ser una aproximación que puede resultar bastante útil para la descripción de subpatrones de la mayor parte de series temporales que provienen de procesos industriales.

Es por ello, que decidimos realizar una segmentación lineal considerando únicamente los cruces por cero de la primera derivada (máximos y mínimos de la señal), clasificando posteriormente dichos tramos según su longitud y forma. A partir de la localización de los máximos y mínimos, y definiendo “el umbral” de altura que diferencia un tramo horizontal de los tramos crecientes y decrecientes, se obtienen los segmentos o tramos (ver figura 3). De éstos se almacena: el tipo de tramo, longitud, altura y posición inicial del mismo.

Generalmente, es necesario realizar un post-procesado para fundir tramos seguidos que son del

III Taller de Minería de Datos y Aprendizaje

mismo tipo en uno sólo o convertirlos a otro tipo si, al unirlos, cambian sus características.

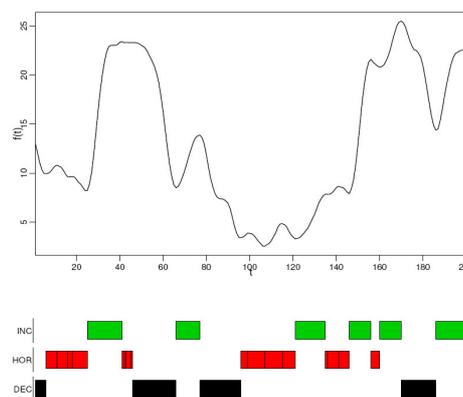


Figura 3. Obtención de los tramos INC, HOR y DEC.

2.3. Agrupamiento de los Tramos

Una vez identificados los segmentos crecientes, horizontales y decrecientes; se procede a agruparlos según características similares.

Fundamentalmente, la mayor dificultad estriba en poder realizar una clasificación robusta que pueda evitar que cambios leves en forma, magnitud o longitud de dos tramos que serían fácilmente clasificados visualmente por un humano como pertenecientes al mismo grupo, sean tomados como “diferentes” en un “sistema automático”. En la actualidad, esta dificultad se está intentando solventar de diversas formas [6].

En nuestro caso, la disimilitud entre dos tramos se determina estudiando separadamente su forma a lo largo del tiempo y su duración o longitud, de tal manera que podamos ajustar con mayor o menor restricción los umbrales impuestos que identifiquen dos tramos de igual forma indistintamente de su longitud o duración en el tiempo. Es decir, la comparación de la longitud o duración en el tiempo entre tramos se realiza separadamente de la de su forma. Esto permite dar más flexibilidad al experto a la hora de clasificar tramos. Por ejemplo: puede ser interesante considerar todos los tramos horizontales en un solo grupo, independientemente de su longitud temporal; o agrupar tramos crecientes bruscos que

corresponden a crecimientos de la señal en un breve periodo de tiempo; etc. Inicialmente, la disimilitud de la forma la obtenemos mediante la distancia Euclídea⁴ de los vectores formados por N puntos equidistantes entre si de cada uno de los tramos a clasificar (ver figura 4). Esta selección de puntos equidistantes sirve para realizar una comparación normalizada de los tramos sin que influya la longitud de los mismos.

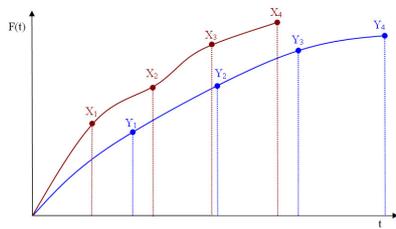


Figura 4. Distancia entre dos tramos.

De esta forma, dados dos tramos $X(t)$ e $Y(t)$ de longitudes L_x y L_y , el cálculo de la disimilitud de los mismos con respecto a la forma que toman a lo largo del tiempo, pero independientemente de la longitud de los mismos, es:

$$(1) \quad DIS_{Shape} = \frac{1}{N} \cdot \sqrt{\sum_{i=1}^N \left(X\left(i \cdot \frac{L_x}{N}\right) - Y\left(i \cdot \frac{L_y}{N}\right) \right)^2}$$

Lógicamente, cuanto más elevado es N mejor precisión se obtiene en el cálculo de la disimilitud y menos influencia individual tiene cada uno de los puntos en el cálculo final. Por otro lado, es lógico que el aumento de N exija un mayor tiempo de cómputo.

La disimilitud en longitud simplemente se obtiene mediante la obtención del valor absoluto de la diferencia de longitudes.

$$(2) \quad DIS_{Long} = |L_x - L_y|$$

Para poder llevar una clasificación lo más acertada posible, se realiza un clusterizado jerárquico de todos los patrones crecientes y decrecientes atendiendo a su disimilitud por forma.

Este tipo de técnicas⁵, además de que nos permite observar en una estructura en forma de árbol de distancias cómo están distribuidos los tramos crecientes y decrecientes según la forma que tienen, nos ayuda a clasificar las familias de tramos de cada tipo. Las herramientas de clusterizado jerárquico solamente necesitan que se les indique la altura del árbol por donde se va a realizar el corte o el número de grupos deseados (ver figura 5) para que se extraigan las familias del árbol de distancias jerárquicas previamente desarrollado.

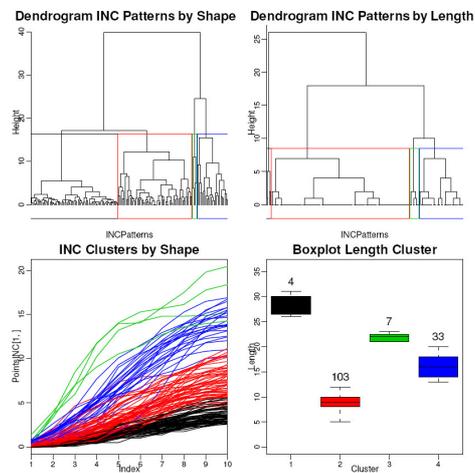


Figura 5. Clasificación de patrones crecientes mediante el uso de algoritmos jerárquicos.

Cada uno de los tramos va a venir caracterizado por dos letras. La primera indica su tipo general (I: Incremental (INC), D: Decremental (DEC) y H: Horizontal (HOR)) y la segunda, la familia a la que pertenece según su forma (A, B, C,...). De este modo, un tramo denominado DA, corresponde a un tramo decreciente cuya forma corresponde con la de la familia A.

Solamente queda por caracterizar los tramos según su longitud. Para ello, se puede proceder con la misma metodología siempre que interese clasificarlos según su longitud. Es decir, podemos apoyarnos en el uso del árbol jerárquico de las

⁴ Se están estudiando otro tipo de distancias que puedan resultar más robustas y generalizadoras.

⁵ Lógicamente, también pueden utilizarse otro tipo de técnicas de agrupamiento (k-medias, LVQ, etc.) o mediante el uso de proyectores lineales o no lineales.

distancias de los patrones para determinar el número de grupos que vamos a clasificar según la distancia para los tramos INC, DEC y HOR.

Finalmente, solamente queda por definir con un número la longitud tipo de cada tramo. De esta forma, un tramo del tipo IC3 corresponderá a un tramo INC de forma C y tipo de longitud 3.

2.4. Búsqueda de Secuencias Repetidas

Una vez definidos los tramos y clasificados convenientemente en familias según su tipo, forma y longitud, entramos en la búsqueda de secuencias repetidas de tramos dentro de la misma serie temporal, a los que llamaremos patrones.

El problema que se plantea, es similar al de una búsqueda de genes dentro de una secuencia genética. En este caso, la serie temporal está constituida por tramos que han sido etiquetados en el proceso anterior. Por ejemplo, una serie temporal numérica se convertiría en una secuencia de etiquetas parecida a esta: {DA2, H3, IB2, DF5, H6, DA1, H5, IA3, H8, ...}.

El proceso de búsqueda de patrones dentro de la misma se convierte en un problema bien conocido dentro de la Minería de Datos de Series Temporales. Generalmente, este problema se resuelve mediante la aplicación de algoritmos basados en el algoritmo APRIORI para la búsqueda de subsecuencias en series temporales. Para ello, se realiza una búsqueda de aquellos patrones repetidos formados por un solo miembro que superan un umbral o soporte (*support*) dado. Este corresponde con el número mínimo o porcentaje de patrones repetidos que se considera como significativo.

Una vez localizados, se almacenan en una base de datos así como su número de ocurrencias. Seguidamente, se procede a buscar aquellos patrones de dos miembros cuyo primer miembro corresponde con uno de los almacenados previamente. Se almacenan aquellos patrones que superan el número mínimo indicado por el soporte. Se repite la búsqueda para patrones de 3 miembros formados con patrones del grupo anterior, y así sucesivamente hasta que no se encuentre ningún patrón que supere el valor mínimo de apariciones (ver figura 6).

Cada una de las ramas del árbol, que corresponde con un patrón que aparece un número de veces superior al establecido por el usuario, es

III Taller de Minería de Datos y Aprendizaje

etiquetada para poder ser identificada en pasos posteriores.

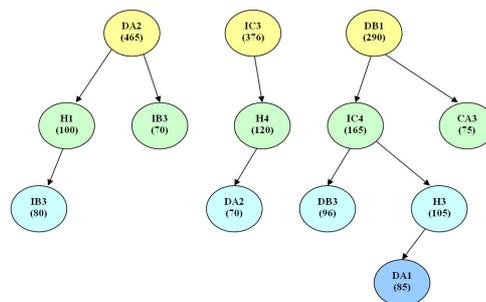


Figura 6. Árbol de búsqueda de patrones repetitivos.

Realmente, la potencia del uso de árboles estriba en la posibilidad de realizar una reducción del número de patrones mediante la fusión o podado de diversas ramas del árbol a través de la creación de reglas robustas que permitan identificar cuáles de ellos son parecidos.

2.5. Búsqueda de Reglas Asociativas

Una vez identificados los patrones repetitivos que aparecen en cada una de las variables del proceso, entramos en la etapa final de búsqueda de reglas.

La idea fundamental es la de facilitar el trabajo tedioso de búsqueda de correlaciones temporales entre variables y de mostrar de una forma fácilmente comprensible dichas relaciones locales. Esta búsqueda pretende detectar relaciones temporales entre variables que puedan ser la causa del comportamiento de una tercera variable (ver figura 7).

Es decir, en esta etapa se pretende buscar correlaciones locales entre series temporales que se repiten con un cierto grado de asiduidad y resumirlas en una serie de reglas asociativas del tipo:

$$(3) \quad \text{IF } \{P_1(t-d_1), P_2(t-d_2), \dots\} \text{ THEN } S(t) \\ \text{WITH } (\text{Support} = X, \text{Confidence} = Y)$$

Donde:

- $P_1(t), P_2(t), \dots$: corresponde al "Antecedente" de la regla e indica aquella secuencia de patrones de diferentes variables que aparecen

- repetidos antes o durante el patrón “Consecuente”.
- $S(t)$: es el patrón de salida denominado “Consecuente”.
- d_1, d_2, \dots : corresponde a los desfases entre los patrones Antecedentes con respecto al patrón Consecuente.
- Cobertura o Soporte (*Support*): que corresponde con el número o porcentaje de veces que se repite la regla en toda la base de datos. Indica el grado de generalización de la misma.
- Precisión o Confianza (*Confidence*): Porcentaje de veces que la regla se cumple cuando se puede aplicar. Es decir, cuando existe el Antecedente qué probabilidad existe que aparezca el Consecuente.

El problema no es fácil, pues debido a la enorme combinación de posibilidades, es necesario contar con potentes ordenadores y con una elevada cantidad de memoria para poder almacenar los árboles o tablas *hash* que se vayan creando. Además éste se puede complicar mucho más si se buscan reglas complejas con más de un patrón en el consecuente o si se toma en cuenta la relación temporal de cada patrón con respecto a los demás[4], es decir: si un patrón se produce durante, antes, después, inmediatamente después, etc.

Actualmente, varios investigadores en el mundo están desarrollando diversas técnicas para la búsqueda de reglas asociativas en series temporales que toman en cuenta estos y otros factores (Métodos MOWCATL, Gen-REAR, etc.) [1][2][3][4].

A efectos prácticos, nosotros estamos considerando en aplicar algunos de estos métodos pero tomando en cuenta algunas consideraciones importantes.

Debido a que estamos trabajando en procesos industriales, la búsqueda de relaciones debe realizarse en una ventana temporal deslizante de una anchura acorde con el proceso a estudiar. Es decir, es necesario estudiar las reacciones del proceso ante cambios en las variables para poder establecer el ancho de tiempo de la ventana en la que se van a buscar relaciones. Lógicamente, si ésta es demasiado pequeña, podemos no encontrar las relaciones buscadas y si ésta es demasiado grande, podemos necesitar demasiado tiempo y espacio en memoria para almacenar todas las posibles combinaciones. Es por ello, que una buena definición de la anchura de la ventana de búsqueda puede ayudar a ahorrar tiempo y a hacer más eficiente los algoritmos de búsqueda.

Además, otra forma de agilizar la búsqueda consiste en indicar cuál de las variables corresponderá al Consecuente y cuales al Antecedente, ya que muchas veces queremos que nos indique cuales son las relaciones de unas variables determinadas (por ejemplo: “presión”, “temperatura” y “velocidad”) con respecto a otra variable objetivo (por ejemplo: “calidad”).

Para la búsqueda de secuencias, nosotros proponemos desarrollar una única secuencia a modo de “secuencia genética” donde aparecen todos los patrones ordenados según su aparición en el tiempo (ver el cuadro verde de la figura 7), junto con la distancia relativa de cada uno de ellos con el anterior (indicado entre paréntesis después de cada patrón) y la longitud de cada uno de los patrones.

De esta forma, la búsqueda de secuencias dentro de esa cadena de patrones se puede realizar usando técnicas parecidas a las utilizadas en la búsqueda de genes en secuencias genéticas. Lógicamente, hay que tomar en cuenta que esta búsqueda debe realizarse dentro de una ventana temporal deslizante parametrizable por el usuario.

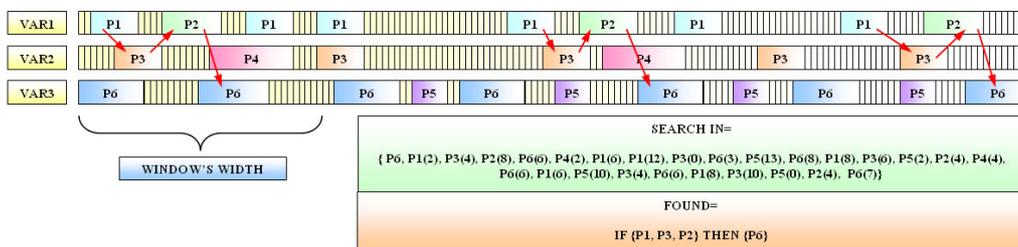


Figura 7. Búsqueda de reglas asociativas en varias series temporales.

El proceso de búsqueda que proponemos se basa en el algoritmo APRIORI y, tal como hacíamos en la fase anterior, comienza por el almacenamiento de las secuencias de patrones que forman un solo individuo en su Antecedente y uno solo en su Consecuente que superen la Cobertura y Precisión fijadas.

El algoritmo continúa buscando parejas con dos antecedentes, tres, cuatro, etc.; hasta que no se encuentra ninguna regla que cumpla con la Cobertura y Precisión establecidas.

2.6. Presentación de los Resultados

Las reglas obtenidas deben ser presentadas en orden ascendente según la Cobertura y Precisión que consigan. Lógicamente, en este caso la traducción de una regla como la obtenida en la figura 7:

$$(4) \text{ IF } \{P1, P3, P2\} \text{ THEN } P6 \\ \text{ WITH (Support = 3\%, Confidence = 80\%)}$$

debe ser interpretada por el experto de esta forma: “SI la variable VAR1 se comporta como P1 y después la variable VAR2 se comporta como P3 y después la variable VAR1 se comporta como P2 ENTONCES la variable VAR3 se comporta como P6. Esto se ha producido un 3% de las veces y esta regla se produce un 80% de las veces que se cumple el Antecedente”.

3. Conclusión

En este artículo se ha explicado una técnica que puede ayudar en la búsqueda de conocimiento oculto dentro de series temporales de procesos industriales. La posibilidad de extraer ese conocimiento de los históricos de procesos industriales y utilizarlos para la toma de decisiones, abre un amplio abanico de posibilidades para los ingenieros de procesos u otros expertos que trabajen con series temporales.

Lógicamente, éste es un problema muy complejo que necesita aún de mucho esfuerzo en investigación para solventar los numerosos problemas que surgen en cada una de las fases de esta metodología, pero creemos que los resultados que se pueden obtener podrían ser muy sustanciosos.

Agradecimientos

Este trabajo ha sido financiado por la Unión Europea bajo el *Research Fund for Coal and Steel* (RFCS) en los proyectos con referencia siguientes: RFS-CR-03012, RFS-CR-04023 y RFS-CR-04043; y el programa CEUTIC INTERREG IIIA. También ha sido parcialmente financiado por el Ministerio de Ciencia y Tecnología de España a través de la Dirección General de Investigación bajo el proyecto DPI2004-07264-C02-01.

Por último, queremos agradecer la ayuda recibida a través del 2º Plan Riojano de I+D+i del Gobierno de La Rioja.

Referencias

- [1] Adamo, Jean-Marc. Data Mining for Association Rules and Sequential Patterns. Sequential and Parallel Algorithms. Springer. New York, 2001.
- [2] Bettini, C.; Wang, X. S.; Jajodia, S. Discovering Temporal Relationships with Multiple Granularities in Time Sequences. IEEE Transactions on Knowledge and Data Engineering, 10(2), 222-237.
- [3] Harms, Sherri K.; Deogun, Jitender S. Sequential Association Rule Mining with Time Lags. Journal of Intelligent Information Systems, 22:1, 7-22, 2004.
- [4] Höppner, F. K. Knowledge Discovery from Sequential Data. Ph.D. Thesis. Fachhochschule BS/WF (University of Applied Sciences). Germany, 2003.
- [5] Keogh, E.; Chu, S.; Hart, D.; Pazzani, M. Segmenting Time Series: A Survey and Novel Approach. Series in Machine Perception and Artificial Intelligence. Vol. 57, 1-21, 2004.
- [6] Lie Hetland, Magnus. A Survey of Recent Methods for Efficient Retrieval of Similar Time Sequences. Series in Machine Perception and Artificial Intelligence. Vol. 57, 23-42, 2004.
- [7] Lindeberg, T. Effective scale: A natural unit for measuring scale-space lifetime. IEEE Trans. on Pattern Analysis and Machine Intelligence, 15(10):1068-1074, 1993.
- [8] Witkin, A. P. Scale space filtering. Proc. of the 8th International Conference on Artificial Intelligence. Germany, 1019-1022, 1983.