

## REGLAS DE ASOCIACIÓN EN PROCESOS CONTROLADOS POR REDES DE SENSORES INALÁMBRICAS

Roberto Fernández Martínez

Francisco Javier Martínez-de-Pisón Ascacibar

Alpha Pernía Espinoza

Carmen Bao Iturbe

Marina Corral Bobadilla

Grupo EDMANS. ([www.mineriadatos.com](http://www.mineriadatos.com)). Universidad de La Rioja. España

### Abstract

Obtaining useful information for decision-making based on the time series captured from different processes, as industrial, agricultural and environmental, will allow a better knowledge and management, as well as a bigger quality of the processes. Wireless sensor networks are responsible for gathering data that will form these time series. This kind of networks is formed by constant capture devices that allow register and transmit data in a wireless way to a central location.

In most mentioned processes it is very difficult, but impossible, to carry out a wired until the measure points, due to the high cost of making it or the impossibility of wiring until the localization to measure. With these networks the only limitation will be the characteristics of the sensors and not their location.

The process involves finding significant patterns in data streams and then applying techniques based on the search of frequent items in order to generate association rules that represent the repeated relationships between patterns of different time series and within a time window.

It is used the software tools CONOTOOL developed by the CONOSER project (DPI2006-03060) funded by Spain's Ministry of Science and Innovation.

**Keywords:** *Wireless sensor networks, association rules, improving processes, knowledge discovery in time series*

### Resumen

La obtención de un conocimiento útil que permita tomar decisiones a partir de series temporales, obtenidas en diferentes procesos, como pueden ser industriales, agrícolas o medioambientales, permitirá un mejor conocimiento y gestión, así como una mayor calidad de los procesos. La obtención de los datos que formarán dichas series temporales se realizará a partir de redes de sensores inalámbricos. Este tipo de redes están formadas por dispositivos de captura constante que nos permiten registrar y transmitir datos de forma inalámbrica hasta una localización central.

En la mayoría de los procesos mencionados es muy difícil, sino imposible, realizar un cableado hasta los puntos de medida, por el elevado costo de hacerlo o por la imposibilidad

de cablear hasta la localización a medir. Con estas redes la única limitación será las características de los sensores y no la situación.

La búsqueda de conocimiento oculto consiste en encontrar patrones significativos dentro de las series de datos, aplicar técnicas basadas en la búsqueda de ítems frecuentes y obtener reglas de asociación que representen relaciones concurrentes entre patrones de todas las series temporales.

Se utiliza la herramienta software CONOTOOL desarrollada dentro del proyecto CONOSER (DPI2006-03060) que ha sido financiado por el Ministerio de Ciencia e Innovación de España.

**Palabras clave:** *Redes de sensores inalámbricas, reglas de asociación, mejora de procesos, obtención de conocimiento en series temporales.*

## 1. Introducción

Los avances dentro de la monitorización de procesos, hacen que la tendencia actual sea la de obtener información de lo que está sucediendo, y a partir de ésta poder extraer conocimiento oculto que permita mejorarlos. La mayoría de los datos obtenidos son variables físicas medidas durante el proceso, obtenidas mediante redes de sensores y enviadas a una estación central donde son almacenadas y estudiadas.

Para realizar esta captación de datos, se usan redes inalámbricas de sensores formadas por un grupo de dispositivos con ciertas capacidades sensitivas y de comunicación inalámbrica que permiten formar redes sin infraestructura física preestablecida, ni administración central. Caracterizándose por su facilidad de despliegue y por ser autoconfigurables.

La rapidez de captación de datos y la amplia capacidad de almacenaje hace que las bases de datos de series temporales (TSDBs) tengan tal dimensión que desborde la capacidad humana de comprenderlas sin la ayuda de herramientas potentes. Para obtener un conocimiento de estas bases de datos se trabaja dentro de la Minería de Datos Temporales (TDM), y más concretamente sobre la búsqueda de reglas de asociación en series temporales. Este trabajo se centra en la presencia del atributo tiempo a la hora de obtener reglas, relaciones y patrones que ofrezcan un alto grado de explotación para el analista a la hora de entender, utilizar y predecir (Dong et al., 2004).

Uno de los campos de investigación en Gestión del Conocimiento, con un futuro más brillante dentro de los procesos industriales, agrícolas y medioambientales, corresponde al control de procesos de negocio. Obteniendo datos a través de redes de sensores, realizando una búsqueda de conocimiento oculto dentro de los datos obtenidos y las bases de datos de históricos almacenados, y realizando una operación que permita controlar o mejorar un proceso, como puede ser la calidad (Wang, 1999; Zheng, 2000; Taniguchi and Haraguchi, 2006; Shen y Chouchoula, 2000).

En la actualidad, el grupo de investigación EDMANS esta desarrollando herramientas que, gracias a algoritmos de inteligencia artificial, permitan ser de considerable utilidad a la hora de tomar decisiones o mejorar procesos productivos con grandes bases de históricos de datos.

En este artículo, se describe la metodología utilizada para encontrar interrelaciones entre datos obtenidos de un proceso agrícola que retraten estas relaciones de una manera que sea fácilmente entendible para el experto. Obteniendo reglas de asociación que permitan la extracción de conocimiento oculto desde series temporales multivariantes

## 2. Redes de Sensores Inalámbricas

Gracias a los avances en MEMS (Micro-Electro-Mechanical Systems), que nos permiten disponer de dispositivos para realizar mediciones de campo dentro de una explotación agraria, es posible obtener variables de cómo evolucionan diferentes parámetros ambientales del sistema, a través del uso de WSN (Wireless Sensor Networks) que nos permiten captar y gestionar estos datos en tiempo real.

Con esta tecnología, no solo se pueden recoger datos, sino que también se pueden procesar y transmitir hasta otro lugar, sin necesidad de realizar un desplazamiento hasta el lugar estudiado.

Los primeros pasos se realizaron en la Universidad de Berkeley, aunque actualmente un gran número de universidades investiga en diferentes aplicaciones. La obtención de información dentro de una explotación minera (Chehri A. et al. 2009), la detección de defectos estructurales de puentes (Shamim et al., 2005), o el control de riego de una explotación agraria (Kim Y. y Evans R.G., 2009) y varios más, que han demostrado que es posible crear una red ad-hoc con un elevado número de nodos.

Un ejemplo de WSN puede aplicarse a la captación de información de los viñedos pertenecientes a una bodega. Como se ve en la figura 1, los nodos de la WSN son distribuidos por los diferentes viñedos de la bodega, comunicándose hasta enviar los datos a la puerta de enlace, encargada de captar toda la información.

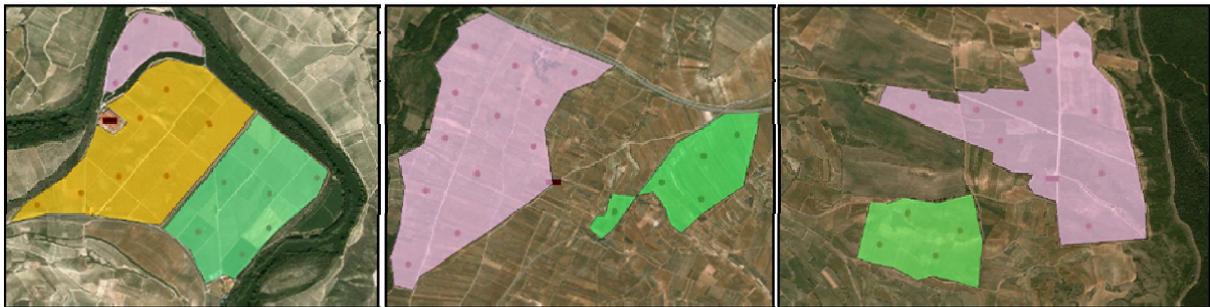


Figura 1. Situación de una WSN en los diferentes viñedos de la bodega.

Una red de sensores se compone de tres tipos de dispositivos: nodos inalámbricos, puertas de enlace y estaciones base. Los nodos sensores son pequeños dispositivos con capacidad limitada de cómputo y comunicación, cuyo tiempo de vida depende de una batería adjunta al dispositivo. Las puertas de enlace que permiten la interconexión entre la red de sensores y una red de datos estándar como TCP/IP. Y una estación base, que puede definirse como un recolector de datos basado en un ordenador común o en un sistema empujado.

En las redes de sensores, tener en cuenta el consumo de energía es muy importante, sobre todo si queremos que los periodos de vida de los dispositivos sean máximos. Puesto que los nodos de esta red no estarán conectados a una red eléctrica y su funcionamiento se basa principalmente en la utilización de baterías, se debe reducir y optimizar el consumo de energía dentro del nodo.

Realizar un estudio del consumo de energía en una red de sensores mediante medida directa es complicado, debido al gran número de nodos y al diferente nivel de consumo en cada uno de ellos. Las investigaciones realizadas sobre este consumo son más bien pocas y la mayoría están basadas en simulaciones de software (Joe H. et al., 2008). Se están realizando estudios en la utilización de los protocolos de enrutamiento y transmisión para hacer que el consumo de energía sea mínimo (Mikko K. et al., 2009; Doler M., 2008) ya que ante la limitación de la vida útil del dispositivo hay que realizar una gestión eficiente del consumo energético.

### 3. Búsqueda de patrones frecuentes en series temporales

Cuando se disponen de varios tipos de sensores captando información dentro de un proceso, se dispone también de las evoluciones de varias variables en el tiempo, como muestra la figura 2. En este caso vemos varias variables medioambientales captadas dentro de un viñedo y varias variables captadas de análisis realizados a las uvas de los viñedos supervisados.

Encontrar patrones que se repiten a lo largo del tiempo y las relaciones que existen entre ellos, puede ser de gran utilidad para el encargado del control de proceso.

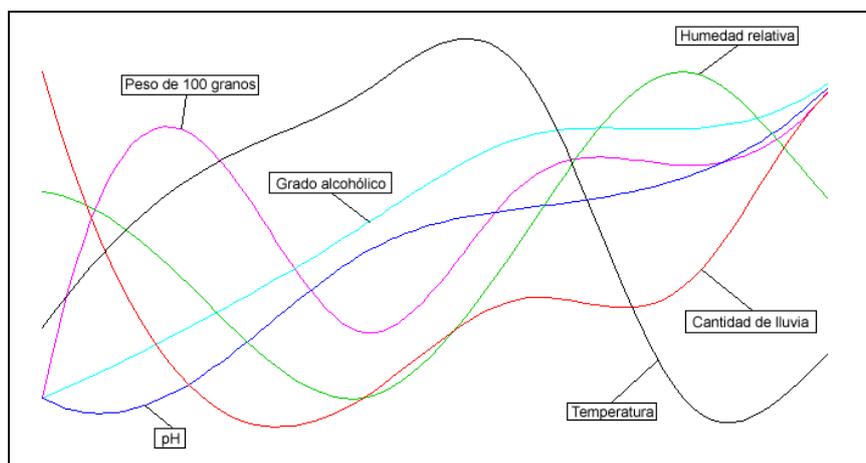


Figura 2. Evolución de diferentes variables temporales de un proceso.

Las dependencias que pueden aparecer suelen ser más complicadas, y en muchos casos no corresponden al mismo instante temporal, sino que pueden aparecer dependencias entre variables con importantes desfases en el tiempo. Por ejemplo, como puede verse en la figura al cabo de un tiempo de lluvias se aprecia un incremento en el peso de los granos de uva.

Puede deducirse de lo anterior, que esta tarea puede realizarse visualmente cuando la cantidad de información con la que se trabaja no es muy elevada, siendo prácticamente imposible cuando el número de variables y de lecturas crece bruscamente.

Estos últimos años han sido testigos del auge de las técnicas de Minería de Datos enfocadas hacia las series temporales (Brzezinska et al., 2007; Last et al., 2004; Charbonnier et al., 2004) cuyo propósito ha sido ayudar en el análisis de grandes bases de datos para obtener una utilidad de un conocimiento oculto que pueda ayudar en la toma de decisiones y generar nuevas predicciones de los modelos.

#### 4. Obtención de reglas de asociación a partir de series temporales

Los enfoques transaccionales (Dong et al., 2004; Lee et al., 2007; Tung et al., 2003) utilizados para la obtención de reglas de asociación no son lo suficientemente predictivos como para poder usarse adecuadamente en sistemas de soporte para tomas de decisiones. Para solucionar esta minusvalía se está investigando y utilizando el enfoque intertransaccional que da más posibilidades dentro de bases de datos de series temporales.

##### 4.1 El enfoque transaccional

Dentro del enfoque transaccional, las llamadas reglas de asociación, son del tipo:

$$X \rightarrow Y \quad (1)$$

donde  $X, Y$  ( $X$  implica  $Y$ ,  $X$  es llamada antecedente e  $Y$  consecuente) son grupos de artículos frecuentes dados en una base de datos tal que:

$$X \cap Y = \Phi \quad (2)$$

El soporte de la regla  $X \rightarrow Y$  representa el porcentaje de transacciones en la base de datos que contienen  $X$  e  $Y$ , esto es  $P(X \cup Y)$ . La confianza de la regla es el porcentaje de transacciones de la base de datos que contiene  $X$  y también contiene  $Y$ , eso es  $P\left(\frac{X}{Y}\right)$ .

La obtención de reglas de asociación es por lo tanto usada para encontrar relaciones entre los grupos de registros dentro de una base de datos, de tal manera que el soporte y la confianza de esas reglas mejoren los niveles de soporte y confianza establecidos anteriormente por el analista.

##### 4.2 El enfoque intertransaccional

El planteamiento intertransaccional trata de encontrar valores/relaciones a lo largo de diferentes registros de transacciones en la base de datos, y no asociaciones entre elementos pertenecientes a la misma transacción. Aunque algunos autores citan reglas de asociación intertransaccionales como extensiones del descubrimiento de patrones y episodios secuenciales, la tendencia es establecer la principal diferencia desde la perspectiva del proceso en la transacción.

El trabajo realizado por Mannila et al. (1997) es similar al concepto de dimensionalidad (tiempo) en reglas de asociación intertransaccionales: donde un episodio es una secuencia de eventos y las reglas de asociación entre episodios toman la forma:

$$P(V) \rightarrow Q(W) \quad (3)$$

donde  $P$  y  $Q$  son una secuencia de eventos y  $V$  y  $W$  son límites temporales. Aunque las subsecuencias son formadas usando límites temporales y los episodios que ocurren frecuentemente en estas subsecuencias son descubiertas a través de extensiones del algoritmo Apriori, el procedimiento parece indicar que los resultados obtenidos son intratransaccionales en oposición a los intertransaccionales.

De manera similar, Bettini et al. (1998) propuso el uso de estructuras de eventos para descubrir relaciones temporales entre eventos en una secuencia temporal. Esta estructura consiste en un número de variables que representan eventos y restricciones temporales entre estas variables, las cuales serán posteriormente catalogadas para la búsqueda de

reglas de asociación. Aquí de nuevo, no hay un concepto de transacción claro, y por lo tanto, estos avances están parcialmente en desacuerdo con el enfoque intertransaccional.

## 5. Caso práctico: Monitorización de un viñedo

Un ejemplo de la aplicación de estas técnicas puede ser la monitorización ambiental dentro del control de maduración de la uva. Donde a través de la obtención de diferentes medidas ambientales como la temperatura, la humedad, la radiación solar, la cantidad de luz ambiental, etc.; a través de una red de sensores inalámbrica podemos determinar ciertas circunstancias que afecten a la maduración de la uva y a la calidad final del mosto.

### 5.1 Tecnología utilizada.

Muchas son las posibles tecnologías disponibles para desarrollar una WNS, aunque no todas disponen de los sensores necesarios para poder monitorizar los parámetros deseados. Después de realizar un estudio de todas ellas, se concluyó que la familia que más accesorios tiene para realizar el trabajo de monitorización agraria es la familia Eko.



Figura 3. Tecnología utilizada para la monitorización de viñedos (1- Nodo sensor; 2- Puerta de enlace).

### 5.2 Descripción del problema

La maduración es el proceso fisiológico que en la vid se inicia con el envero y que incluye una serie de reacciones metabólicas en las que a su vez se implican una gama de componentes que cursan con una evolución diversa y que finalmente termina con lo que se denomina “índice de madurez”.

El conocimiento de la evolución del proceso bioquímico de transformación que se produce en el fenómeno de la maduración de la uva condiciona la calidad de los futuros vinos y prepara a las bodegas, para poder mejorar los procesos de fermentación de los mostos y conocer con antelación la situación técnica de la vendimia. Esto permite que la bodega pueda planificar la cosecha de uva de muy buena calidad, para la elaboración de grandes vinos.

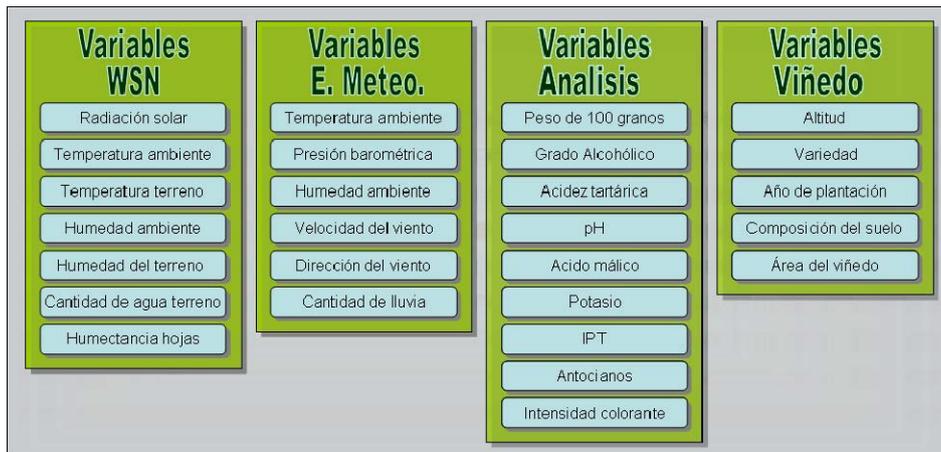


Figura 4. Variables estudiadas dentro del proceso de monitorización

Capturar información con nuestra red de sensores en diferentes zonas de varios viñedos como la temperatura ambiental, la humedad ambiental, la radiación solar, la humectación de las hojas, la temperatura del suelo, la humedad del suelo, etc.; junto con variables bioquímicas dentro del proceso de maduración como el peso de 100 granos, el grado alcohólico probable, la acidez total tártrica, el pH, el acido málico, el potasio, el IPT (índice de polifenoles totales), los antocianos, la intensidad colorante, etc.; nos puede permitir a través del estudio de sus series temporales obtener ciertas reglas de asociación entre ellas que nos aporten conocimiento oculto del proceso de maduración de la uva.

### 5.3 Preprocesado, segmentación de series temporales y búsqueda de reglas de asociación.

Las variables seleccionadas, ya sean obtenidas por análisis, a través de una WSN o una estación meteorológica, son series temporales que nos describen el proceso de maduración. El estudio de los datos obtenidos se realiza a través de la herramienta CONOTOOL, desarrollada por el grupo EDMANS de la Universidad de La Rioja. Esta herramienta permite de una manera más intuitiva y rápida el uso de todas las funciones de la librería KDSeries. La última versión disponible, de la librería y de la herramienta CONOTOOL, puede ser descargada desde la página Web del proyecto CONOSER: <http://api.unirioja.es/conoser>.

Siguiendo los diferentes pasos dentro de la herramienta CONOTOOL, obtención de datos desde la WSN o la base de datos de análisis, filtrado de datos, extracción de patrones característicos y búsqueda de reglas de asociación, obtenemos patrones significativos dentro de la evolución de la maduración de las uvas.

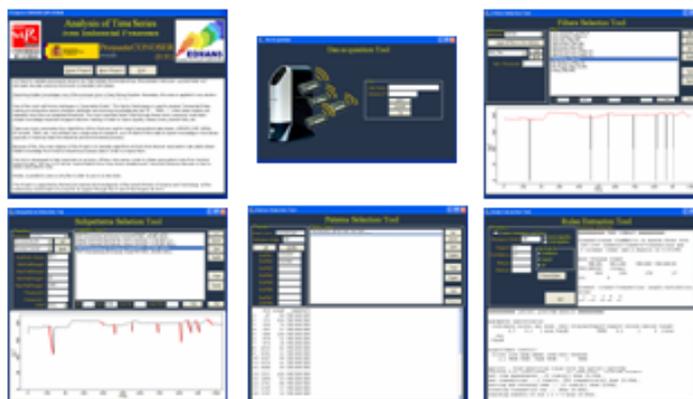


Figura 5. Herramienta CONOTOOL.

La metodología propuesta se compone de las siguientes etapas:

1. Obtención de datos desde una red de sensores a través de una puerta de enlace.
2. Filtrado de cada serie temporal para eliminar el ruido y obtener la forma básica de la misma.
3. Obtención de los máximos y mínimos significativos.
4. Extracción de los tramos característicos de cada serie temporal (“incrementales (INC)”, “decrementales (DEC)”, “horizontales (HOR)” o “de acuerdo a umbrales previamente establecidos (UMB)”).
5. Agrupamiento de los tramos “INC”, “DEC”, “HOR” y “UMB” creando patrones más complejos según las especificaciones realizadas por el experto.
6. Búsqueda de secuencias de patrones usando ventanas deslizantes.
7. Extracción de las reglas de asociación usando algoritmos de reglas de asociación intertransaccional.

Una vez todos los subpatrones y patrones significativos han sido encontrados para cada una de las series temporales analizadas, por medio de una ventana deslizante en el tiempo y con una anchura de ventana definida por el usuario, una matriz puede ser construida donde cada una de las columnas contiene los subpatrones y patrones que cayeron dentro de la ventana en el momento de tiempo dado (Figura 2).

Por lo tanto, cada una de las columnas de la matriz obtenida corresponde a los subpatrones y patrones que aparecen en un momento dado y dentro de una ventana de tiempo predefinida.

A partir de aquí podemos obtener reglas de asociación que nos ayuden a comprender el proceso de maduración y nos ayuden a predecir la mejor fecha de vendimia para obtener el mejor vino posible.

## 6. Conclusiones

El objetivo de este artículo ha sido mostrar que la aplicación de herramientas software para la adquisición, preprocesado y segmentación de series temporales unidas al uso de técnicas de minería de datos, como reglas de asociación, puede ser extremadamente útil dentro de la búsqueda de conocimiento oculto en los registros históricos de procesos industriales, agrícolas y medioambientales.

El proceso estudiado es una monitorización ambiental de viñedos durante la fase de maduración. Facilita el seguimiento del periodo de maduración, ayudando a la bodega a conocer como sus viñedos van evolucionando durante ese periodo y prediciendo cual será la mejor fecha para la vendimia.

En particular, la metodología que se presenta está basada en un preprocesado simple, en una segmentación y en la búsqueda de conocimiento oculto en series temporales básicas que pueden ser fácilmente aplicadas para mejorar procesos.

En un nivel practico, la conclusión obtenida es que a través de ajustes iterativos e interactivos de las funciones de preprocesado y segmentación (filtrado, detección de máximos y mínimos significativos, extracción de subpatrones y patrones) los expertos pueden obtener patrones significativos más fiables que en sistema de extracción de patrones totalmente automáticos.

## Referencias

- Bettini, C., Wang, X. S., Jajodia, S., "Mining temporal relationships with multiple granularities in time sequences", *Data Engineering Bulletin*, 1998, Vol. 21, pp. 32-38.
- Brzezinska, I., Grecob, S., Slowinsk, R., "Mining Pareto-optimal rules with respect to support and confirmation or support and anti-support". *Engineering Applications of Artificial Intelligence*, 2007, Vol. 20, pp. 587-600
- Charbonnier, S., Garcia-Beltan, C., Cadet, C., Gentil, S. "Trends extraction and analysis for complex system monitoring and decision support". *Engineering Applications of Artificial Intelligence*, 2004, Vol. 18, pp. 21-36.
- Chehri A., Portier P. and Tardif P. M., "UWB-based sensor networks for localization in mining environments", *Ad Hoc Networks*, Vol. 7, Issue 5, July 2009, pp 987-1000
- Dohler M., "Wireless Sensor Networks: The Biggest Cross-Community Design Exercise To-Date," *Bentham Recent Patents on Computer Science*, 2008, Vol. 1, No. 1.
- Dong Z., Li H. and Shi Z., "An efficient algorithm for mining inter-transaction association rules in multiple time series". *Journal of Computer Science*, 2004, Vol. 31, 108-111.
- Joe H., Park J., Lim C., Woo D., and Kim H., "Instruction-Level Power Estimator for Sensor Networks"; *ETRI Journal*, 2008, Vol.30, No.1, pp.47-58.
- Kim Y. and Evansa R.G., "Software design for wireless sensor-based site-specific irrigation", *Computers and Electronics in Agriculture*, May 2009, Vol. 66, Issue 2, pp 159-165
- Last, M., Kandel, A., Bunke, H., "Data Mining in Time Series Databases. Series in Machine Perception Artificial Intelligence". *World Scientific Publishing Co., London*, 2004.
- Lee, A., Wang, C., "An efficient algorithm for mining frequent inter-transaction patterns", *Information Sciences*, 2007, Vol. 177, pp. 3453-3476.
- Manilla, H., Toivonen, H., Verkamo, A., "Discovery of frequent episodes in event sequences", *Data Mining Knowledge Discovery*, 1997, Vol. 1, pp. 259-289.
- Mikko K., Jukka S., Mauri K., Ville K., Marko H. and Timo H., "Energy-efficient neighbour discovery protocol for mobile wireless sensor networks", *Ad Hoc Networks*, 2009, Vol. 7 pp. 24-41
- Pakzad S. N., Kim S., Fenves G. L., Glaser S. D., Culler D. E. and Demmel J. W., "Multi-Purpose Wireless Accelerometers for Civil Infrastructure Monitoring", *5th International Workshop on Structural Health Monitoring (IWSHM 2005)*, Stanford, CA, 2005.
- Shen Q. and Chouchoula A., "A modular approach to generating fuzzy rules with reduced attributes for the monitoring of complex systems". *Engineering Applications of Artificial Intelligence*, 2000. Vol. 13, pp. 263-278.
- Taniguchi T. and Haraguchi M., "Discovery of hidden correlations in a local transaction database based on differences of correlations". *Engineering Applications of Artificial Intelligence*, 2006, Vol. 19, pp. 419-428.
- Tung, A., Lu, H., Han, J., Feng, L., „Breaking the barrier of transactions: Mining Intertransaction Association Rules", *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, San Diego, USA, 1999, pp. 297-301
- Wang X. Z., "Data mining and knowledge discovery for process monitoring and control. Advances in industrial control", *Springer-Verlag, London*, 1999.
- Zheng Z., "Constructing conjunctive attributes using production rules", *Journal of research and practice in information technology*, 2000, Vol. 32, pp. 13-38.

## **Agradecimientos**

Los autores agradecen a la “Dirección General de Investigación” del Ministerio Español de Ciencia e Innovación por el apoyo financiero de los proyectos DPI2006-03060, DPI2006-14784, DPI2006-02454 y DPI2007-61090; y a la Unión Europea por el proyecto RFSPR-06035.

Finalmente, los autores quieren agradecer también al Gobierno Autónomo de La Rioja por su apoyo a través del 3º Plan Riojano de I+D+i.

## **Correspondencia** (Para más información contacte con):

Dr. Francisco Javier Martínez de Pisón Ascacíbar  
Grupo EDMANS ([www.mineriadatos.com](http://www.mineriadatos.com))  
Área de Proyectos de Ingeniería. Departamento de Ingeniería Mecánica  
Edificio Departamental. ETSII de Logroño  
C/ Luis de Ulloa, 20, 26004 Logroño (España).  
Phone: +34 941 299 232 Fax: + 34 941 299 794  
E-mail: [fjmartin@unririoja.es](mailto:fjmartin@unririoja.es). URL: <http://www.mineriadatos.com>