INICIACIÓN A LA INVESTIGACIÓN
REVISTA ELECTRÓNICA
UNIVERSIDAD DE JAÉN

Ini Inv, e5: a5 (2013)

Apoyo a la toma de decisiones mediante el empleo de resúmenes lingüísticos

[1]Rita Castillo-Ortega, [1]Nicolás Marín, [1]Daniel Sánchez, [2]Carlos Molina

[1]Departamento de Ciencias de la Computación e I.A. Universidad Granada. Avda. del Hospicio s/n, 18071, Granada.
[2]Departamento de Informática. Universidad de Jaén. Campus Las Lagunillas s/n, 23071, Jaén.
rita@decsai.ugr.es

Resumen

En una sociedad en la que el conocimiento es esencial nos surge la necesidad de manejar adecuadamente ingentes cantidades de datos. Lo podemos ver en nuestra vida diaria, pero cobra una mayor importancia en el ámbito empresarial.

Empresas y organizaciones generan y consumen grandes volúmenes de datos con el fin de llevar a cabo sus actividades. Pero los datos en sí no tienen un impacto en el desempeño de las actividades si no somos capaces de obtener información a partir de ellos que al mismo tiempo sea útil e inteligible para nosotros.

La posibilidad de generar resúmenes lingüísticos de series de datos en el ámbito empresarial se presenta como una poderosa herramienta que facilitará el proceso de toma de decisiones.

INTRODUCCIÓN

La capacidad de manejar grandes volúmenes de datos se perfila cada vez más necesaria en una sociedad que sin duda alguna está basada en el conocimiento. El importante número de grandes empresas, así como organizaciones u organismos públicos, que generan y consumen ingentes cantidades de datos con el fin de llevar a cabo sus actividades, son un buen ejemplo de ello. La mayoría de estos datos están relacionados con la dimensión temporal de una u otra manera.

Pero no sólo el manejo de los datos es útil; del mismo modo, el proceso que permite realizar la extracción de la información a partir de conjuntos de datos, se está volviendo cada vez más importante para nuestro entorno. La importancia de este proceso se debe al hecho de que permite a los usuarios realizar tareas tan importantes como el análisis en la toma de decisiones, pronóstico o predicción [2].

Como se ha comentado anteriormente, la posibilidad de realización de pronóstico y predicción basados en el estudio de las series de datos es muy importante. En este trabajo, nosotros nos vamos a centrar en otra utilidad como es la descripción de la información que alberga la serie temporal. Existen diversas formas de presentar la información que se obtiene a través del análisis de las series de datos. En particular, en nuestro caso nos interesa que la presentación de esta información se realice en forma de resumen.

Los receptores de los resultados obtenidos son los usuarios con formación experta o no, de modo que es muy conveniente realizar un resumen de la información usando el lenguaje natural. Este tipo de resúmenes reciben el nombre

de resúmenes lingüísticos y gracias a ellos es posible ofrecer a los usuarios una información más comprensible y que al mismo tiempo les sea de más utilidad a la hora de tomar determinadas decisiones.

En este trabajo se presentan algunos de los enfoques que han sido desarrollados con el fin de realizar resumen lingüístico sobre las series de datos temporales obtenidas de almacenes de datos. Dichos enfoques usan como herramientas para resolver este problema conceptos de conjuntos difusos (FS - Fuzzy Sets) [28]. En la literatura se pueden encontrar diversos enfoques basados en el uso de *Soft Computing* para la realización de resúmenes de datos. De entre estas propuestas, el resumen lingüístico posee un especial interés debido a la ya mencionada necesidad de obtener sentencias cercanas al lenguaje natural al describir grandes conjuntos de datos.

El resumen lingüístico de series de datos temporales es de mucha utilidad en el ámbito referente a Sistemas de Información (IS - Information Systems), debido a la importancia de la dimensión temporal en el desempeño del negocio. Las soluciones basadas en *Business Intelligence* permiten a los administradores o gerentes de empresas, organizaciones u organismos, obtener un conocimiento más preciso del proceso que desempeñan y las operaciones comerciales o sociales que llevan a cabo, con el fin de apoyar una mejor toma de decisiones. En general se puede esperar que las compañías que hacen uso de algún tipo de tecnología para facilitar la toma de decisiones tengan un mayor rendimiento que las compañías que prescinden de ella. Una parte muy importante de las herramientas relacionadas con Business Intelligence está basada en el uso del modelo de datos multidimensional y las operaciones OLAP (OnLine Analytical Processing) llevadas a cabo sobre él y que permiten la consulta de grandes cantidades de datos [23].

El modelo de datos multidimensional es un modelo ampliamente extendido que se basa en el uso de los llamados cubos de datos (también conocidos como hipercubos o data cubos) los cuales están orientados al análisis de datos. Cada cubo de datos almacena una colección de hechos numéricos, llamados medidas, que se encuentran descritos por un conjunto de dimensiones. El cubo alberga en cada una de sus celdas los datos relacionados con los elementos, agregados en cada una de las dimensiones.

Los cubos de datos, como norma general, contienen una dimensión tiempo debido al importante papel de ésta en general y referente a la actividad comercial en particular. Muchas de las operaciones OLAP que se aplican sobre la dimensión temporal de los cubos de datos producen series de datos temporales. Un gran número de autores han dedicado sus esfuerzos e investigación en realizar minería de datos (DM - Data Mining) sobre estas series temporales. En [1], los autores presentan una extensa visión general de algunos enfoques a este respecto que son de interés.

Cada vez son más las voces que relacionan las tareas llevadas a cabo en los procesos de Data Mining con el soporte a la toma de decisiones [21]. En el presente trabajo consideramos el resumen lingüístico de datos como un caso particular de Data Mining ya que la entrada es la misma y como resultado obtenemos información no trivial, novedosa y potencialmente útil que antes no se conocía, pero con la cualidad añadida de que las salidas se realizan usando lenguaje natural.

EL RESUMEN LINGÜÍSTICO

De acuerdo con el diccionario de la Real Academia Española de la Lengua, *resumen* es una “exposición resumida en un asunto o materia”. Se trata también de la “acción y efecto de resumir o resumirse”. Y lo que hacemos al *resumir* es

"reducir a términos breves y precisos", o, "considerar tan sólo y repetir abreviadamente lo esencial de un asunto o materia".

Siguiendo las ideas e investigaciones de Pinto [13], diremos que el "resumen es el documento referencial más completo y por consiguiente el que mejor representa la información original, ofreciendo una visión global del contenido del documento".

El hecho de que un resumen sea breve no implica que en él se refleje lo *esencial* acerca de algo, aunque esta situación sería la más deseable. En muchas ocasiones estas dos palabras son usadas como sinónimos cuando en realidad no lo son. Es cierto que existen diversas circunstancias o situaciones en las que no es necesaria tanta precisión al hablar o definir términos, ya que existen patrones de conocimiento común respecto a un cierto tema, o que son compartidos por una determinada comunidad. Deberemos de tener muy en cuenta las diferencias existentes entre los términos *breve* y *esencial*.

La situación ideal a la hora de enfrentarnos a la realización de un resumen sería que fuésemos capaces de ofrecer la información esencial de una forma breve y concisa. Pero en la mayoría de los casos, las situaciones no se caracterizan por ser ideales. Podemos encontrarnos, por ejemplo, con que toda la información esencial no pueda ser reflejada de forma breve; aunque lo que realmente representa un problema son las distintas percepciones que tienen diferentes personas de los conceptos breve o esencial. Esta concepción vendrá marcada por los diferentes intereses de cada uno o la utilidad que se le vaya a dar al resumen obtenido. Lo que para un individuo o grupo puede resultar interesante o esencial, puede no serlo para otros, y viceversa, incluso se pueden presentar visiones diferenciadas acerca de qué extensión puede ser considerada o no como breve.

El proceso de resumir es una actividad inherente al ser humano. Las personas continuamente reciben información del exterior que someten a una serie de procesos (transformación, reducción, almacenaje, recuperación o utilización) según sus capacidades y necesidades, con vistas a una futura aplicación. Podemos considerar pues que resumir implica una actividad de reducción natural de información en la mente humana. En este proceso se pasa a fijar los conceptos más importantes o significativos de entre todos los datos percibidos. En general el resumen pretende ser lo mismo pero en tamaño más pequeño y no una parte arbitraria de lo que tenemos que resumir.

La obtención de un buen resumen es algo muy importante. No basta con tener un resumen, dicho resumen debe de satisfacer las necesidades del usuario y debe de hacerlo con una cierta calidad. La selección de características es una acción que determinará qué partes son las que queremos destacar o a cuáles prestar mayor atención. Una vez hecho esto y obtenido el resumen final, se deberá de medir la calidad de resumen, a fin de que podamos conocer cómo de bueno es el producto obtenido.

La tarea de intentar que las máquinas produzcan resúmenes comparables a los realizados por humanos, es una tarea de gran interés en la sociedad actual. El punto fuerte de los ordenadores en esta tarea lo aportan su capacidad para almacenar datos y su potencia de cálculo que puede realizar millones de operaciones en un tiempo muy inferior al que nosotros emplearíamos. En cambio, el punto débil es el de conseguir que la información que produce sea comprensible para los humanos, que en la mayoría de las ocasiones son los receptores habituales.

Pero algo que para los seres humanos es tan natural y que incluso se realiza sin reparar en ello, no es una tarea tan fácil como puede parecer. Los humanos adaptamos inconscientemente nuestra forma de actuar y desenvolvernos

según en entorno y la situación en la que nos encontremos. Cuando nos enfrentamos al reto de intentar que las máquinas realicen resúmenes comparables a aquellos que producimos nosotros mismos, nos damos cuenta de ello.

Debido a la dificultad y utilidad del tema un gran número de investigadores está centrando sus estudios en mejorar el proceso de comunicación entre humanos y máquinas. En nuestro caso, son de especial interés los esfuerzos de investigación realizados en el ámbito del resumen de series de datos temporales.

LA IMPORTANCIA DE LA DIMENSIÓN TEMPORAL

Desde los tiempos más remotos el ser humano ha medido el paso del tiempo con diferentes métodos y herramientas, algunos de ellos de gran precisión. A pesar de ello, el estudio de las series de tiempo en sí, posee un origen relativamente reciente. Se piensa que fue hace aproximadamente 1000 años cuando se produjo la primera representación gráfica de los eventos dividiendo un eje horizontal en intervalos de igual amplitud para representar iguales periodos de tiempo.

En este trabajo seguiremos las ideas de Peña cuando afirma que “una serie temporal es el resultado de observar los valores de una variable a lo largo del tiempo en intervalos regulares (cada día, cada mes, cada año, etc.)” [24]. Las primeras series temporales estudiadas correspondían a datos astronómicos y meteorológicos.

Las series de datos temporales pueden ser representadas mediante una sucesión de las medidas tomadas. Si estas medidas no se han tomado en intervalos regulares o necesitamos obtener más información acerca del momento de tiempo al que corresponde, dichas medidas pueden ser acompañadas por el instante de tiempo concreto con mayor o menor nivel de detalle dependiendo de nuestras necesidades. Sin embargo, en ocasiones estas formas de representación, bien sea en texto plano o ayudados mediante tablas, no suelen ser muy intuitivas.

En algunas ocasiones puede que el usuario que recibe la información no posea conocimiento experto en el tema específico. Otras veces, puede que la cantidad de datos sea tan elevada o la diferencia entre sus valores tan notable, que hagan complicado el proceso de análisis de los datos. Sea como fuere, incluso con las series de datos más sencillas, en muchos de los casos la representación gráfica aporta una buena herramienta de representación de las series de datos temporales. Por desgracia, la representación gráfica de las series no siempre es fácil de interpretar, ya que en ocasiones las series son muy complicadas o incluso tenemos varias series relacionadas entre sí presentadas en el mismo gráfico. A esto se suman problemas como la necesidad de equipos apropiados y las diferentes percepciones dependiendo de la granularidad usada.

En todos los casos, con independencia de la complejidad de la serie o series, el resumen lingüístico de series de datos temporales es un herramienta potente que permite presentar a usuarios no expertos información acerca de la serie en un formato comprensible y fácil de interpretar. Otra buena característica atribuible a los resúmenes lingüísticos es la posibilidad de uso de un sintetizador de voz, de especial utilidad en aquellos casos en los que la visualización de las gráficas no es adecuada o no es posible (por ejemplo, para poder presentarse los resultados a personas con algún tipo de problema visual).

La importancia del tiempo es crucial cuando se analizan datos en las empresas, por ejemplo, para asesorarse en el proceso de toma de decisiones. De hecho, el tiempo juega un papel muy importante en los almacenes de datos, o

data warehouses, debido a que esta dimensión permite almacenar la información histórica concerniente a las operaciones diarias de una organización. De este modo, el personal encargado de la toma de decisiones puede realizar el análisis de la evolución de diferentes y variados aspectos del negocio a lo largo del tiempo. Por ello es por lo que la dimensión temporal suele estar presente casi siempre entre las dimensiones de los cubos de datos.

De forma breve y general se puede describir un data cubo o cubo de datos como un conjunto multidimensional de celdas. Cada una de las dimensiones que conforman dicho cubo pueden ser vistas como un conjunto de miembros que pueden organizarse a su vez haciendo uso de una o varias jerarquías. Gracias a la organización jerárquica de las dimensiones, la granularidad en el cubo de datos puede variar en cada dimensión dependiendo del nivel seleccionado en cada momento. Las operaciones OLAP permiten a los usuarios aprovechar la ventaja de esta particularidad y consultar de diferentes formas al data cubo con el fin de obtener los datos deseados en cada momento.

Podemos obtener series de datos temporales mediante observación directa pero es más usual hacerlo mediante recogida automática de datos con volcado de la información en sistemas de almacenamiento. Si hacemos consultas sobre este tipo de estructuras de almacenamiento seremos capaces de obtener series de datos temporales. Si además las estructuras son multidimensionales como las mencionadas anteriormente, mediante la aplicación de operaciones OLAP seremos capaces de obtener series temporales con diferentes granularidades.

La importancia de la dimensión temporal y la necesidad de obtener resultados entendibles por los usuarios humanos, hacen de este tipo de series candidatas principales para beneficiarse de técnicas de resumen lingüístico en su estudio.

El propósito del estudio o análisis de las series de datos temporales puede ser dividido en dos grandes áreas según varios autores. La primera de ellas es “entender o modelar el mecanismo estocástico de una serie observada” y la otra “predecir o pronosticar los valores futuros de series basadas en la historia de esas series y, posiblemente, otras series o factores relacionados” (Cryer [11]).

De este modo, podemos decir que el análisis de series temporales comprende métodos que ayudan a interpretar los datos, extrayendo para ello información representativa, tanto referente a los orígenes o relaciones subyacentes como a la posibilidad de extrapolar y predecir su comportamiento futuro. De hecho uno de los usos más habituales de las series de datos temporales es su análisis para predicción y pronóstico. En nuestro caso, no vamos a centrar nuestra atención en áreas como la predicción o pronóstico, sino en la posibilidad de la descripción de las series.

La tarea de obtención de la información subyacente en las series de datos temporales es importante y permite conocer tendencias, eventos destacados o patrones, que permitirán una mejor toma de decisiones. Si esta nueva información se encuentra representada en formato textual con patrones similares a los usados por los humanos cuando se comunican entre ellos, la toma de decisiones se desarrollará de forma más amigable para el decisor.

RESUMEN LINGÜÍSTICO DE DATOS CON TÉCNICAS DIFUSAS

La lógica difusa tiende un puente entre los datos tal y como están almacenados en las máquinas, y las descripciones lingüísticas con las que los humanos están acostumbrados a tratar. La comunicación humano - máquina es

uno de los temas más candentes en la actualidad. Los resúmenes lingüísticos se sirven de las bondades de la lógica difusa para conseguir un proceso de comunicación más eficaz.

En trabajos como [20] de Mitra et al. y [9] de Chen et al. podemos ver completos estados del arte de técnicas de obtención de información a partir de grandes cantidades de datos mediante el uso de técnicas Soft Computing.

Dado el interés que posee la realización de este tipo de resúmenes existen distintos enfoques a la hora de afrontar la tarea de producir un resumen lingüístico de datos. Como denominador común cabe destacar que la gran mayoría de las técnicas hasta ahora desarrolladas hacen uso de etiquetas lingüísticas. Las etiquetas lingüísticas son esenciales ya que introducen la posibilidad de tratar con la imprecisión y vaguedad necesarias para trabajar con el lenguaje natural.

Son varios los autores que se han dedicado a trabajar en este campo. Ronald R. Yager lo aborda apoyándose en el uso de sentencias enriquecidas por conceptos difusos [25] introducidos años antes por Lofti A. Zadeh. Estas sentencias cuantificadas representan el grado satisfacción de una cierta cualidad por un grupo de objetos. En un primer momento, la cuantificación de las sentencias se realizaba usando el cardinal de Zadeh, pero posteriormente, y para la resolución de casos más complejos, desarrolló los denominados operadores OWA (Ordered Weighted Averaging) [26] y [27].

La investigadora Anne Laurent presentó sus técnicas en [19]. En su trabajo, se centra en la obtención de resúmenes lingüísticos a partir de bases de datos multidimensionales y operaciones OLAP, usando para ello una interfaz amigable para el usuario. En [29] de Zhang los resúmenes lingüísticos se realizan usando Degree Theory y FCA (Formal Concept Analysis). El autor diferencia entre el proceso a seguir para obtener lo que él denomina resúmenes simples y resúmenes complejos.

Anteriormente ya ha sido mencionado el fuerte vínculo que muchos autores comienzan a ver entre las técnicas de minería de datos y la obtención de resúmenes lingüísticos. Un trabajo más profundizando en esta reflexión lo encontramos en [17] de Janusz Kacprzyk and Slawomir Zadrozny.

Enfatizando la importancia de los resúmenes lingüísticos, pero sobre todo los resúmenes lingüísticos de series de datos temporales encontramos el trabajo de Ichiro Kobayashi y Naoko Okumura [18]. En el mencionado trabajo los autores muestran su interés en la verbalización de series de datos temporales mediante técnicas interesantes. En [10] Chiang et al. también expresan su interés en el resumen lingüístico de series de datos temporales.

El concepto de resumen lingüístico es también abordado en los trabajos de Díaz et al. 14, Triviño et al. [15] y Moreno [16].

HACIA UNA VERSIÓN LINGÜÍSTICA DE F-CUBE FACTORY

Durante esta sección se hará una pequeña introducción a los fundamentos de nuestro enfoque para el resumen lingüístico de series de datos temporales. Una vez realizado el repaso de las ideas más importantes, se presenta la introducción de nuestra herramienta en una plataforma web que permite el manejo de cubos de datos.

Nuestras contribuciones en el área pueden ser consultadas en [4], [5], [7], [6], [8], [3]. En nuestros trabajos tomamos como punto de partida una serie de datos temporales obtenida a través de la aplicación de operaciones OLAP sobre un

cubo de datos multidimensional. Como consecuencia encontramos que la dimensión tiempo se encuentra descrita con diferentes niveles de granularidad, dando lugar esto a una partición jerárquica difusa. Además de la nombrada partición, tenemos otra partición difusa en la dimensión del hecho que queremos describir. Dichas particiones cumplen una serie de características (consultar artículos para más información). Los resúmenes finales se componen de sentencias cuantificadas de tipo II, *Q de los D son A* donde *Q* es un cuantificador, *D* y *A* son conjuntos difusos representando el tiempo y el hecho. Un ejemplo: *La mayoría (Q) de los días en la estación cálida (D) las precipitaciones son escasas (A)*.

Vamos a verlo más claro un con ejemplo. La **Figura 1** muestra una serie de datos temporal y su contexto lingüístico. La serie representa la afluencia de pacientes a un centro de salud dado durante un periodo de tiempo igual a un año. Como se ha comentado anteriormente las dos dimensiones, tanto la de tiempo como la de la variable a estudiar, están particionadas usando lógica difusa. En este caso, las particiones pueden haber sido heredadas de un data cubo multidimensional o simplemente facilitadas por el usuario que desea obtener el resumen. La dimensión tiempo está definida mediante una jerarquía de particiones difusas. Cada una de las particiones con diferente granularidad, o lo que es lo mismo, con diferente grado de abstracción.

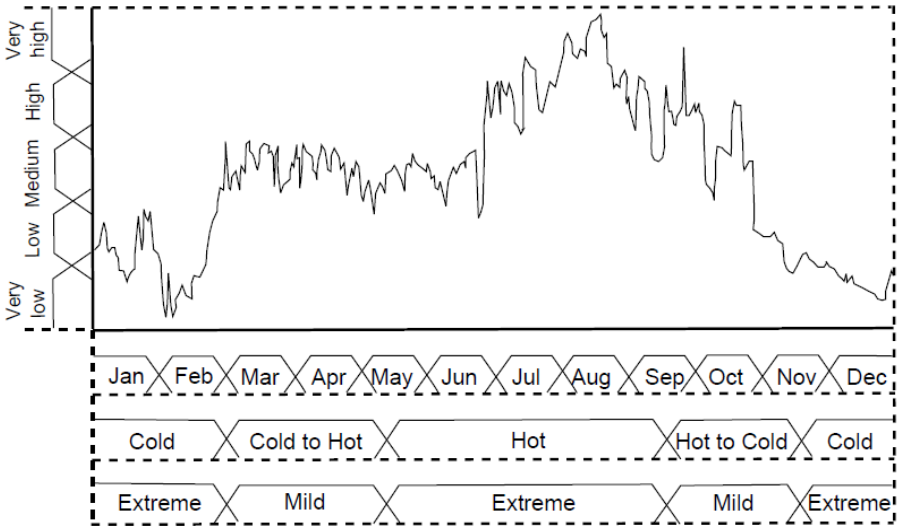

Figura 1. Ejemplo de serie de datos temporal y su contexto lingüístico.

A partir de una serie como la de la Figura 1 y considerando el contexto lingüístico, así como otros parámetros de configuración, nuestro método sería capaz de obtener resúmenes como el mostrado a continuación:

Most of the patient inflow is low or very low with cold weather; medium in seasons with cold to hot weather, May, and June; very high or high in July and August; and high or medium in September and October. Finally, at least 70% of patient inflow is low or very low in November.

F-Cube Factory es una plataforma web que nos permite la manipulación de grandes cantidades de datos almacenados siguiendo el modelo multidimensional ([12], [22]). En su versión inicial se implementaron diversas técnicas que ofrecían información novedosa obtenida a partir de los datos almacenados. Las capacidades de esta potente plataforma se están ampliando en la actualidad mediante la adición de la posibilidad de obtención de resúmenes de datos en general y de series temporales en particular.

A través de la plataforma se pretende que el usuario tenga la posibilidad de consultar cubos de datos multidimensionales de forma rápida y sencilla, presentando los datos con diferentes formatos (incluyendo la síntesis de audio a partir de los resúmenes), pudiendo consultar el conjunto de sentencias cuantificadas que conforman el resumen o un texto completo fruto del post-procesado de dichas sentencias.

En la **Figura 2** se puede ver el resultado obtenido al realizar resumen lingüístico de la variación de afluencia según la serie representada en la **Figura 1**.

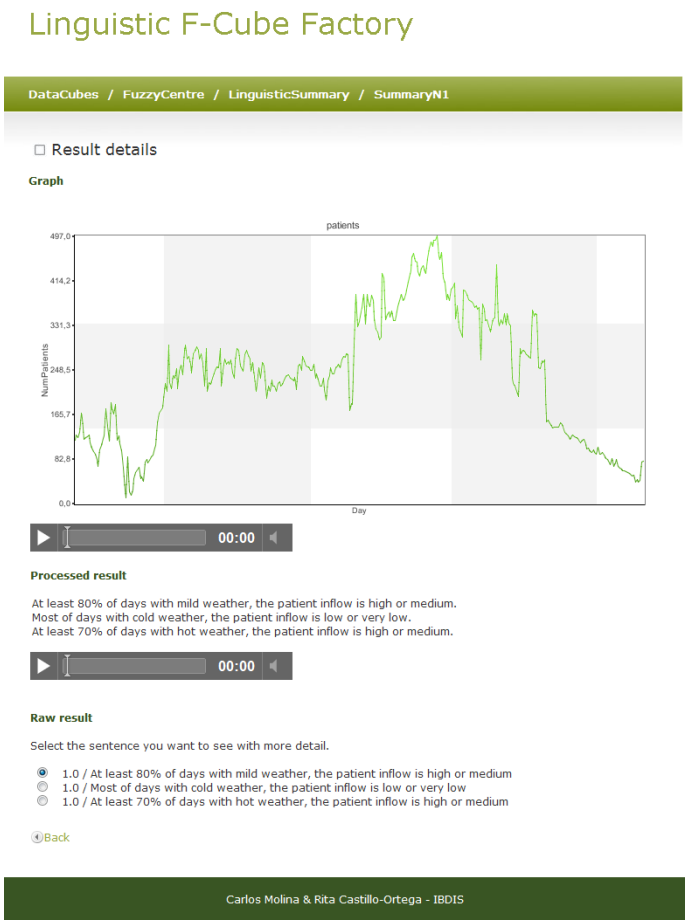

Figura 2. Captura de pantalla de un resumen lingüístico en F-Cube Factory.

La pantalla está dividida en tres grandes zonas de presentación de resultados. En primer lugar, se muestra al usuario una representación gráfica de la serie de datos temporal. A continuación, se presenta el resumen de texto post-procesado con la posibilidad de la reproducción del audio asociado. En tercer lugar se ofrece la posibilidad de analizar cada una de las sentencias que componen el resumen final. Al seleccionar una de las sentencias marcarán las zonas afectadas en el gráfico y se ofrecerá la posibilidad de reproducir el audio de la sentencia individual. Destacar que la posibilidad de asociar cada una de las sentencias a los datos que soportan la afirmación en el gráfico, ofrece al usuario una herramienta para ponerse en contexto y comprobar que la veracidad de la afirmación.

La posibilidad de contar con una herramienta visual y auditiva, sencilla de manejar que nos presente los resultados deseados siguiendo los patrones de lenguaje usado por los humanos, es un apoyo muy grande sobre todo en áreas tan importantes como el soporte a la decisión.

CONCLUSIÓN

En la actualidad es muy necesario ser capaz de manejar grandes conjuntos de datos. No es extraordinario el hecho de tener que enfrentarse a una situación como la descrita, si no frecuentemente, sí alguna vez en la vida. Las empresas, organizaciones e instituciones dependen en gran medida de llevar este proceso a cabo de forma eficaz.

Es vital el manejo de importantes volúmenes de datos, pero más lo es el hecho de ser capaz de extraer información útil de dichos datos. Una de las dimensiones más reflejadas en estos datos es la dimensión temporal. Las empresas mantienen datos de su producción, las ventas, los gastos, el stock de productos, y tantas otras facetas similares, que reflejan su funcionamiento y su evolución a lo largo del tiempo.

Se ha identificado la necesidad de introducir el resumen lingüístico en este campo con el fin de conseguir métodos rápidos, sencillos y que produzcan resultados informativos, de calidad y amigables para los usuarios. Los resultados obtenidos siguiendo patrones semejantes a los del lenguaje natural serán mucho más legibles y por lo tanto útiles para su posterior uso.

Los resúmenes lingüísticos de series de datos se presentan como una herramienta de gran utilidad en el ámbito del Business Intelligence en general y en el apoyo a la toma de decisiones en particular.

AGRADECIMIENTOS

Parte de la investigación ha sido subvencionada por la Junta de Andalucía, Consejería de Innovación, Ciencia y Empresa, bajo el proyecto P07-TIC-03175 Representación y Manipulación de Objetos Imperfectos en Problemas de Integración de Datos: Una Aplicación a los Almacenes de Objetos de Aprendizaje. Parte de la investigación ha sido subvencionada por el Gobierno Español, Ministerio de Innovación y Ciencia, bajo el proyecto TIN2009-08296.

BIBLIOGRAFÍA

[1] I. Z. Batyrshin y L. Sheremetov. Perception-based approach to time series data mining. Appl. Soft Comput., 8(3):1211–1221, 2008.

[2] I. Z. Batyrshin y T. Sudkamp. Perception based data mining and decision support systems. Int. J. Approx. Reasoning, 48(1):1–3, 2008.

[3] R. Castillo-Ortega, N. Marín, y D. Sánchez. Linguistic query answering on data cubes with time dimension. Special Topic Issue on Advances in Fuzzy Querying: Theory and Applications in International Journal of Intelligent Systems (IJIS), 26(10):1002-1021, 2011.

[4] R. Castillo-Ortega, N. Marín, y D. Sánchez. Fuzzy quantification based linguistic summaries in data cubes with hierarchical fuzzy partition of time dimension. En H. Yin y E. Corchado, eds., IDEAL'09, vol. 5788 de LNCS, págs. 578–585. Springer, Heidelberg, 2009.

[5] R. Castillo-Ortega, N. Marín, y D. Sánchez. Linguistic summary based query

answering on data cubes with time dimension. En T. Andreasen, R. R. Yager, H. Bulskov, H. Christiansen, y H. L. Larsen, eds., FQAS'09, vol. 5822 de LNAI, págs. 560–571. Springer, Heidelberg, 2009.

[6] R. Castillo-Ortega, N. Marín, y D. Sánchez. Linguistic local change comparison of time series. En Congreso Español de Informática. CEDI 2010, págs. 451–459, 2010.

[7] R. Castillo-Ortega, N. Marín, y D. Sánchez. Time series comparison using linguistic fuzzy techniques. In International Conference on Information Processing and Management of Uncertainty in Knowledge- Based Systems. IPMU 2010, págs. 330–339, 2010.

[8] R. Castillo-Ortega, N. Marín, y D. Sánchez. A fuzzy approach to the linguistic summarization of time series. Special Topic Issue en Soft Computing Techniques in Data Mining in Journal of Multiple-Valued Logic and Soft Computing (JMVLSC), 17(2,3):157–182, 2011.

[9] G. Chen, Q. Wei, y E. E. Kerre. Data Mining and Knowledge Discovery Approaches Based on Rule Induction Techniques, Massive Computing Series, chapter 14 Fuzzy Logic in Discovering Association Rules: An Overview, págs 459–493. Massive Computing Series. Springer, Heidelberg, Germany, 2006.

[10] D. Chiang, L. R. Chow, y Y. Wang. Mining time series data by a fuzzy linguistic summary system. Fuzzy Sets Syst., 112:419–432, Junio 2000.

[11] Jonathan D. Cryer y Kung-Sik Chan. Time Series Analysis with applications in R. Springer, 2008.

[12] M. Delgado, C. Molina, L. Rodríguez Ariza, D. Sánchez, y M. A. Vila Miranda. F-cube factory: a fuzzy olap system for supporting imprecision. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 15 (Supplement-1):59–81, 2007.

[13] María Pinto Molina et al. Aprendiendo a resumir. Prontuario y resolución de casos. TREA, Gijón, 1ª edición, 2005.

[14] A. Bugarín. y F. Díaz-Hermida. Linguistic summarization of data with probabilistic fuzzy quantifiers. En ESTYLF 2010, págs. 255–260, 2010.

[15] A. S. Montemayor J. J. Pantrigo R. Cabido G. Triviño, A. Sánchez y E. G. Pardo. Linguistic description of traffic in a roundabout. En IEEE International Conference on Fuzzy Systems - FUZZ-IEEE (WCCI 2010 IEEE World Congress on Computational Intelligence), págs. 2158–2165, 2009.

[16] J. Moreno García, J. J. Castro-Sánchez, y L. Jiménez. A fuzzy inductive algorithm for modeling dynamical systems in a comprehensible way. IEEE T. Fuzzy Systems, 15(4):652–672, 2007.

[17] J. Kacprzyk y A. Wilbik. Using fuzzy linguistic summaries for the comparison of time series: an application to the analysis of investment fund quotations. En U. Kaymak J. P. Carvalho, D. Dubois y J. M. C. Sousa, eds., IFSA-EUSFLAT 2009, págs. 1321–1326, 2009.

[18] I. Kobayashi y N. Okumura. Verbalizing time-series data: With an example of stock price trends. En IFSA/EUSFLAT Conf., págs. 234–239, 2009.