Extracción de reglas de asociación difusas a partir de datos de baja calidad

A.M. Palacios[1], **J. Alcalá-Fdez**[2]

[1] *Universidad de Oviedo, Departamento de Informática, 33204 Gijón, España*, palaciosana@uniovi.es
[2] *Universidad de Granada, Departamento de Ciencias de la Computación e Inteligencia Artificial, CITIC-UGR 18071 Granada, España*, jalcala@decsai.ugr.es

Resumen

Una de las técnicas más utilizas de la Minería de Datos consiste en inducir reglas de asociacion a partir de base de datos. Estas reglas pueden ayudarnos a analizar los datos y a tomar buenas decisiones dentro del ámbito del problema. Muchos de los trabajos han sido enfocados sobre bases de datos con valores precisos, sin embargo en muchas aplicaciones del mundo real los datos presentan un cierto grado de imprecisión. Además, en algunas ocasiones esta impresión puede llegar a ser significativa y no es natural modelarla con una distribucion de probabilidad. En este trabajo proponemos un nuevo algoritmo capaz de extraer conocimiento interesante a partir de bases de datos con datos imprecisos. Este algoritmo integra conceptos de imprecisión con el algoritmo Apriori difuso con el objetivo de obtener reglas de asociación difusas relevantes. Los resultados obtenidos en un problema real sobre el rendimiento de los deportistas en el atletismo muestran la efectividad del método propuesto.

Palabras Clave: Minería de Datos, Reglas de Asociación Difusas, Datos de Baja Calidad

1. Introducción

La Minería de Datos (MD) consiste en extraer conocimiento interesante a partir de conjuntos de datos del mundo real y es el paso central del proceso de Extracción de Conocimiento a partir de Base de Datos (BD). El descubrimiento de asociaciones es una de las técnicas de la MD que más ha sido utilizada para extraer conocimiento interesante a partir de BD. Las reglas de asociación son utilizadas para representar e identificar dependencias entre ítems en una BD [18]. Estas reglas son expresiones del tipo $X \rightarrow Y$, donde X e Y son conjuntos de ítems que cumplen $X \cap Y = \emptyset$, es decir, si todos los ítems de X existen en una transacción entonces todos los ítems de Y, con una alta probabilidad, están en la transacción, y donde X e Y no tienen ningún ítem en común [1, 2]. Últimamente, la teoría de los conjuntos difusos ha sido utilizada más frecuentemente para describir relaciones entre los datos [9]. El uso de conjunto difusos nos permite extender los tipos de relaciones que se pueden representar, facilitando la interpretación de las reglas en términos lingüísticos y eludiendo las fronteras no naturales en el particionamiento de los dominios de los atributos [5, 6].

Muchos investigadores han propuesto métodos para la extracción de reglas difusas a partir de datos cuantitativos [8, 12, 17]. Estos métodos han sido enfocados sobre BD con valores precisos y exactos, sin embargo en muchas aplicaciones reales los datos presentan un cierto grado de imprecisión. En ocasiones, esta imprecisión es lo suficientemente pequeña y puede ser ignorada sin peligro. En otras ocasiones, la incertidumbre de los datos puede ser modelada con una distribución de probabilidad. Sin embargo, existen otro tipo de problemas donde la imprecisión es relevante y no puede ser modelada mediante una distribución de probabilidad [3]. Por esta razón, diseñar algoritmos que sean capaces de extraer conocimiento interesante a partir de Datos de Baja Calidad (DBC) representa un desafio para los investigadores [15, 16].

La estadística difusa considera el uso de los conjuntos difusos para representar la información imprecisa de los datos. Recientes trabajos en estadística difusa sugieren el uso de una representación difusa cuando el dato es desconocido a través de una familia de intervalos de confianza [4], considerando una representación posibilística para modelar los datos imprecisos.

En este trabajo, integramos conceptos de DBC con el algoritmo Apriori difuso propuesto por Hong y otros en [8] con el objeto de obtener reglas de asociación difusas de alta calidad a partir de BD con DBC. Extenderemos este algoritmo considerando una representación posibilística para modelar datos de entrada con valores imprecisos. Así, una entrada imprecisa será representada por un conjunto difuso, el cual será identificado por la familia de todas las distribuciones de probabilidad [4, 16]. La función de pertenencia de esta entrada imprecisa será otro conjunto difuso de acuerdo con el Principio de Extensión [11], el cual es compatible con la interpretación posibilística de los conjuntos difusos [16]. Esto implicará, que la confianza de una regla de asociación difusa será definida a partir de un conjunto de probabilidades.

Los resultados obtenidos sobre un problema real con DBC, basado en el rendimiento de los atletas en las pruebas de atletismo de 100 metros lisos [14], muestran la efectividad del método propuesto.

La estructura del trabajo es como sigue. En la siguiente sección se describe el algoritmo Apriori difuso propuesto por Hong y otros. La Sección 3 introduce la interpretación y representación posibilística de los DBC. El método propuesto es descrito en detalle en la Sección 4. La Sección 5, muestra los resultados obtenidos por el método propuesto sobre una BD real con DBC. Finalmente, la Sección 6 muestran algunas conclusiones.

2. Algoritmo Apriori Difuso

El objetivo principal del algoritmo Apriori difuso, introducido en [8] por Hong y otros, es encontrar ítems relevantes así como reglas de asociación difusas en las instancias con valores cuantitativos, descubriendo interesantes patrones.

Este método consiste en transformar cada valor cuantitativo en un conjunto difuso de etiquetas lingüísticas asumiento que las funciones de pertenencia son conocidas de antemano. El algoritmo posteriormente calcula la cardinalidad de cada item difuso, a lo que denomina "cuenta". Si el valor de cuenta del item difuso es superior o igual que el valor del mínimo soporte este item será considerado un item difuso frecuente. A continuación combina los items frecuentes y vuelve a repetir el proceso. Finalmente, este método obtiene las reglas de asociación difusas mediante el criterior del algoritmo Apriori [2].

3. Datos de Baja Calidad: Representación e Interpretación

En la actualidad existen problemas reales donde la información viene representada por DBC. En estos DBC no se puede observar con precisión las propiedades de un objeto, por lo que no se percibe con exactitud el valor del objeto ni se tiene un conocimiento completo sobre su distribución de probabilidad.

Para modelar y tratar estos datos imprecisos, en este trabajo consideramos una representación posibilísitica. De esta forma, una entrada imprecisa será representada por un conjunto difuso identificado por las familias de distribuciones de probabilidades, siendo cada α-corte de este conjunto difuso un conjunto aleatorio que contiene el valor exacto y desconocido de la variable con una probabilidad de al menos 1-α [4, 16] (ver Figura 1). Resaltar que, esta representación incluye los valores intervalares y precisos como casos particulares.

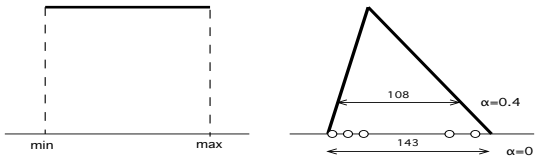

Figura 1: Representación difusa de datos imprecisos.

Esta representación proporciona un marco común para tratar con datos numéricos, palabras, intervalos, conjuntos difusos, valores perdidos (mediante un intervalo que agrupe todo el rango de la variable) o diferentes valores para un mismo atributo como se describe en [16], donde todos estos conceptos son agrupados bajo el término DBC.

En esta propuesta la representación y dominio de cada item de baja calidad puede ser definido mediante:

- Un intervalo $\overline{X} = [x_1, x_2]$ donde x_1 y x_2 están incluidos en el dominio del item. Por ejemplo, un item con dominio entre [0'0, 10'0] podrá ser definido como $\overline{X}$ = [1'5,3'6] o mediante un valor perdido $\overline{X}$ = [0'0, 10'0] (implicando que "x $\in \overline{X}$").
- Un conjunto difuso $\widetilde{X} = (x_1, x_2, x_3)$. Por ejemplo, un item definido mediante tres valores diferentes $\widetilde{X}$=(1'0,1'3,2'0).

4. Función de Pertenencia Difusa con Datos de Baja Calidad

En esta sección describiremos como se obtiene la función de pertenencia a partir de datos imprecisos. Si se tiene una percepción precisa "x" de las propiedades de un objeto y un conjunto difuso $\widetilde{A}$ con un conjunto finito de etiquetas lingüísticas, $L = \{l_1, \ldots, l_n\}$, donde "n" es el número de etiquetas, la función de pertenencia será:

$$pert(x)(l_i) = P_x(l_i) \mid \sum_{i=1}^{n} P_x(l_i) = 1. \quad (1)$$

Si el objeto observado es impreciso y toda la información que se tiene es que está en el conjunto $\overline{X}$ ($x \in \overline{X}$), la función de pertenencia de este conjunto de valores será un conjunto de funciones de pertenencias:

$$pert(\bar{X})(l_i) = \{pert(x)(l_i) \mid x \in \bar{X}\}. \quad (2)$$

Por ejemplo, imaginemos que disponemos de la entrada $\overline{X}$ = [2'1,3'0] y que se quiere calcular la función de pertenencia en la etiqueta "Baja". Toda la información de la que se dispone es que el valor real y desconocido (x_0) pertenece a $x_0 \in$ [2'1,3'0] y donde cada α-corte sobre $\overline{X}$ es un conjunto aleatorio que contiene el valor preciso del objeto con una probabilidad de al menos 1-α. En este caso, al tratarse de un valor intervalar, cada α-corte tiene una probabilidad del 100 % de contener el valor real x_0 ($P(x_0 \in$[2'1,3'0]$)$=1). Esto implica que $pert$([2'1,3'0])(Baja) será un conjunto de probabilidades obtenidas a partir de cada valor x, donde $x \in$ [2'1,3'0].

En el caso de que la entrada sea un valor impreciso $\widetilde{X}$, la función de pertenencia será un conjunto difuso de acuerdo con el Principio de Extensión [11], el cual es compatible con la interpretación posibilística de los conjuntos difusos [16]:

$$\begin{aligned} \text{pert}(\widetilde{X})(l_i)(\text{t}) = \max\{\alpha \mid \text{t} = \text{fuzz}(\text{x})(l_i) \ \text{ y} \\ \text{x} \in [\widetilde{X}]_\alpha\}_{\alpha \in [0,1]}. \end{aligned} \quad (3)$$

Como podemos observar, el conjunto difuso $pert(\widetilde{X}(l_i))$ está asociado al conjunto de familias $\{pert(\widetilde{X}_\alpha(l_i))\}_{\alpha \in [0,1]}$.

5. Extracción de Reglas de Asociación difusas con Datos de Baja Calidad

Esta sección describe nuestra propuesta para obtener reglas de asociación difusas a partir de BD con DBC. Los parámetros de entrada para esta propuesta son:

- Un conjunto de DBC, $\widetilde{D}$, compuesto de t instancias, donde cada una de ellas contiene m atributos, y donde $\widetilde{X}_j^i$ representa el ítem j ($1 \leq j \leq m$) en la instancia i ($1 \leq i \leq t$).
- El conjunto S=$\{L_1, \ldots, L_m\}$, donde L_j=$\{l_1, \ldots, l_n\}$ representa el conjunto finito de etiquetas lingüísticas del atributo j siento n el número de etiquetas lingüísticas.
- Un valor predeterminado del mínimo soporte α.
- Un valor predeterminado de la confianza λ.
- El número de cortes disponibles para obtener los posibles valores reales de la variable.
- El número de veces γ que barremos las probabilidades de cada uno de los posibles valores obtenidos (ver paso 10 del método).

A continuación se describe el algoritmo propuesto:

P1: Transformar cada DBC $\widetilde{X}_j^i$, mediante la función de pertenencia, en un conjunto difuso interpretado como un conjunto de probabilidades para cada etiqueta de L_j ($\overline{P_j^i}$=$\{[p_{*1}, p_1^*]_{l_1}^i, \ldots, [p_{*n}, p_n^*]_{l_n}^i\}$).

P2: Calcular la frecuencia de ocurrencias del ítem j en cada término lingüístico k de L_j (L_{jk}), donde $L_j = \{l_1,\ldots,l_k,\ldots,l_n\}$ y $L_{j_k} = l_k$.

$$\overline{Cuenta}_{L_{j_k}}(\,\%) = \tfrac{1}{t} \bigoplus_{i=1}^{t} [p_{*k}, p_k^*]_{l_k}^i \quad (4)$$

donde t representa el número de instancias y $\bigoplus$ es la suma aritmética difusa [10].

Todos los L_{j_k} son recolectados para formar el conjunto de candidatos C_r donde r representa el número de ítems relacionados en el conjunto de candidatos, inicialmente r=1:

$$\begin{aligned} C_r = \{\cup\{L_{j_k},\ \forall \text{k} \mid 1 \leq k \leq n\}, \\ \forall \text{j} \mid 1 \leq j \leq m\}. \end{aligned} \quad (5)$$

P3: Comprobar si $\overline{Cuenta}_{L_{j_k}}(\,\%)$ para todo L_{j_k} de C_r es mayor o igual que el mínimo soporte α. Si $\overline{Cuenta}_{L_{j_k}}(\,\%)$ satisface esta condición entonces el conjunto L_r contendrá a L_{j_k}. $\overline{Cuenta}_{L_{j_k}}(\,\%)$ esta representado por un conjunto de probabilidades lo que implica que, esta condición será satisfacida si el límite superior (p_k^*) es mayor o igual que α. De este modo, todas las posibles ocurrencias del conjunto de ítems serán consideradas. Por tanto, L_r será el conjunto:

$$L_r = \{L_{j_k} \mid p_k^* \geq \alpha,\ L_{j_k} \in C_r\}. \quad (6)$$

P4: Si L_r no es nulo, entonces continuar con el algoritmo; en otro caso, salir del algoritmo.

P5: Generar el conjunto de candidatos C_{r+1} a partir de la unión de los ítems que componen L_r. C_{r+1} se obtiene de forma similar a la indicada por el algoritmo Apriori [2], con la excepción de que dos ítems con el mismo atributo j no pueden formar parte de C_{r+1} [8].

P6: Realizar los siguientes acciones para cada "(r+1)-itemset" obtenido en C_{r+1}:

a) Calcular la función de pertenencia de cada "(r+1)-itemset", $\overline{P}^i_s = \overline{P}^i_1 \wedge \ldots \wedge \overline{P}^i_{(r+1)}$. La t-normal del producto generaliza la agregación o combinación entre el conjunto de probabilidades de los ítems de "(r+1)-itemset":

$$\overline{P}^i_s = \bigotimes_{j=1}^{(r+1)} \overline{P}^i_j. \quad (7)$$

b) Calcular la frecuencia de ocurrencias de cada "(r+1)-itemset":

$$\overline{Cuenta}_s(\%) = \frac{1}{t} \bigoplus_{i=1}^{t} \overline{P}^i_s \quad (8)$$

Si $max\{\overline{Cuenta}_s(\%)\}$, es mayor o igual que el mínimo soporte α entonces, "(r+1)-itemset" forma parte del conjunto L_{r+1}.

P7: Si L_{r+1} es nulo continuar con el siguiente paso; en caso contrario actualizar el número de ítems relacionados en el conjunto de candidatos (r=r+1) y repetir los pasos 5 y 6.

P8: Recolectar en R los items relacionados de cada L_i, donde $2 \leq i \leq (r+1)$.

P9: Construir todas las reglas de asociación difusas (X $\rightarrow$ Y) a partir del conjunto R.

P10: Determinar, mediante los siguientes dos pasos, si las reglas de asociación difusas obtenidas son relevantes y proporcionan patrones e información interesante de los DBC:

- Calcular la confianza de la regla.
- Comparar la confianza obtenida en el paso anterior con la mínima confianza λ.

La presencia de entradas imprecisas implica que la confianza de una regla vendrá definida por un valor intervalar entre [0,1]:

$$\begin{aligned} &\overline{Confianza}(X \rightarrow Y)_{(\bar{P}_{s_1},\ldots,\bar{P}_{s_q})} = \\ &= \{\text{Confianza}(X \rightarrow Y)_{(x_{s_1},\ldots,x_{s_q})} \mid \\ &\quad Cuenta_{anteced.} > 0,\ \forall x_{s_j} \in \bar{P}_{s_j}\} \end{aligned} \quad (9)$$

donde Confianza($X \rightarrow Y$) es definida en [8] y $Cuenta_{anteced}$ representa la frecuencia de ocurrencias del antecedente de la regla.

El coste computacional de la ec. (9) es muy elevado, y contiene el valor real y desconocido de la confianza de la regla que dependiendo de los valores de x_{s_j}, la regla podría ser o no considerada relevante. Por esta rarón, nosotros proponemos la siguiente aproximación. Si se considera la regla $(X \rightarrow Y)$ con $\overline{Cuenta}_s = [x_1, x_2] = \overline{X}$ y $\overline{Cuenta}_{anteced.} = [y_1, y_2] = \overline{Y}$, algunos "cortes" (c) son aplicados para obtener los posibles valores reales de $\overline{X}$ e $\overline{Y}$, donde a cada corte c (x_j) se le asigna una probabilidad aleatoria (P_{x_j}) de ser el valor real, y donde la suma de todas las probabilidades, de todos los cortes, tiene que ser 1. Estos posibles valores reales de $\overline{X}$ e $\overline{Y}$ son V_X=$\{\overline{X}_w$, w =1,. . .,c$\}$ y V_Y=$\{\overline{Y}_w$, w =1,. . .,c$\}$.

A partir de un valor de V_Y ($y_j \in V_Y$) y de un valor del conjunto V_X ($x_j \in V_X$), para determinar si una regla es relevante no solamente hay que satisfacer la condición $x_j \geq y_j * \lambda$, $x_j \leq y_j$, $y_j > 0$, sino que, además, hay que considerar que la probabilidad de que x_j sea el valor real tiene que ser superior a 0,5 ($P_{x_j} \geq 0{,}5$). Resaltar que, para cada valor de V_Y ($y_j \in V_Y$), todos los posibles valores de V_X han sido considerados y un nuevo conjunto C = $\{C_{y_j} \mid y_j \in V_Y\}$ es obtenido a partir de todas las probabilidades que determinan si una regla es relevante o no a partir de cada valor de V_Y ($y_j \in Y$):

$$\begin{aligned} C_{y_j} = \textstyle\sum_{i=1}^{c} P_{x_i} \geq 0{,}5 \mid & x_i \geq y_j * \lambda, \\ & x_i \leq y_j,\ y_j > 0. \end{aligned} \quad (10)$$

Este proceso es repetido γ veces para barrer las posibles probabilidades de cada posible valor de V_X. Así, para cada valor de V_Y se obtiene un conjunto de posibles probabilidades que dependen de las probabilidades asignadas a cada uno de los posibles valores de V_X:

$$\overline{C}_{y_j} = \{C_{y_j}, \forall C_i,\ 1 \leq i \leq \gamma\}.$$

Finalmente, para determinar si la regla es relevante estos conjuntos de posibles probabilidades son ordenados, de acuerdo a la dominancia uniforme definida en [13]. El conjunto que representa la mediana será seleccionado $\overline{C}_{y_j}$ para determinar a la regla como relevante siempre y cuando su probabilidad mínima sea superior o igual a 0,5.

6. Experimentos

En esta sección se muestra la efectividad del método propuesto a partir de un problema real con DBC basado en el rendimiento de los atletas en la prueba de atletismo de 100 metros lisos ("100mlI") [14].

En las siguientes subsecciones se analizarán las reglas de asociación difusas obtenidas según el valor del mínimo soporte y de la mínima confianza. Además, se analizará el conocimiento obtenido por los expertos a partir de las reglas generadas por el método propuesto.

Las particiones linguísticas consideradas están compuestas por tres términos lingüísticos, con funciones de pertenencia triangulares uniformemente distribuidas. Los valores considerados como parámetros del método son: 7 α-cortes y 2 barridos de las probabilidades (γ).

Cuadro 1: Nivel de conocimiento e interpretabilidad de las reglas de asociación difusas para "100mlI".

Reglas	Sorp-Regla	$\{\overline{C}_{yi}\}$	Mediana$\{\overline{C}_{yi}\}$
Si Salto Es Medio y Reacción Es Media Entonces Carrera Es Media	[0'229,0'324]	{[0'19,0'4],[0'19,0'4] [0'50,0'76],[0'933,0'97], [0'96,0'97],[0'97,0'97],[1'0,1'0]}	[0'933,0'971]
Si Salto Es Medio y Atleta Es Relevante Entonces Carrera Es Media	[0'15,0'24]	{[0'046,0'167],[0'096,0'261] [0'09,0'26],[0'6,0'87], [0'73,0'90],[0'83,0'95],[0'94,0'99]}	[0'6,0'87]
Si Reacción Es Baja Entonces Carrera Es Media	[0'09,0'14]	{[0'92,0'99],[0'92,0'99],[0'94,0'99] [1'0,1'0],[1'0,1'0],[1'0,1'0],[1'0,1'0]}	[1'0,1'0]
Si Salto Es Bajo y Carrera Es Media Entonces Atleta Es No Relevante	[0'11,0'179]	{[0'0,0'38],[0'43,0'44] [0'43,0'44],[0'71,0'99], [0'82,1'0],[0'82,1'0],[0'92,1'0]}	[0'718,0'999]
Si Reacción Es Baja Entonces Atleta Es Relevante	[0'09,0'13]	{[0'08,0'08],[0'15,0'47],[0'15,0'47] [0'52,0'81],[0'52,0'81],[0'99,0'99],[1'0,1'0]}	[0'524,0'816]

6.1. Análisis de las reglas de asociación difusas según soporte y confianza

Varios experimentos se han realizado para analizar el número de reglas de asociación difusas obtenidas a partir del algoritmo propuesto capaz de soportar DBC. La relación entre el número de reglas de asociación difusas con respeto a varios valores del mínimo soporte considerando distintas confianzas es mostrada en la parte superior de la Figure 2. Se aprecia como el número de

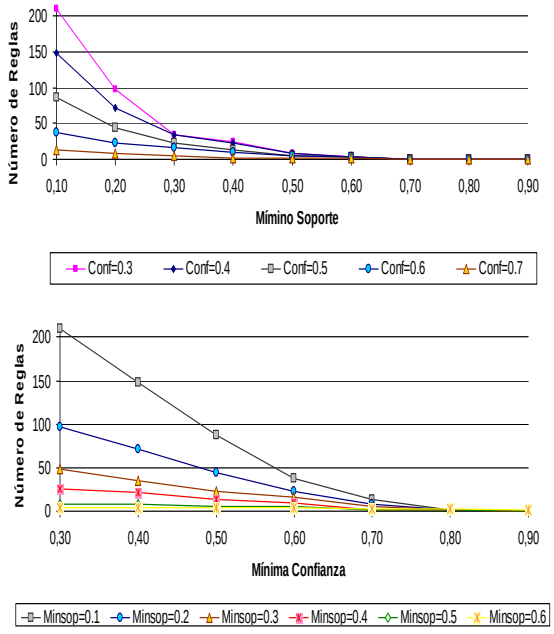

Figura 2: Relación entre reglas de asociación, α y λ.

reglas disminuye cuando el valor del mínimo soporte va aumentando. Además, se observa como las curvas obtenidas tienen forma similar y la separación entre ellas es más pequeña con valores superiores a 0,2 en el mínimo soporte. Sin embargo, se aprecia como la distancia entre las curvas es equidistante para distintas confianzas y un soporte fijo. Esto implica que el número de reglas, a partir de un soporte establecido, aumenta de una manera constante según va disminuyendo la confianza y por tanto, no se resalta la presencia de instancias especiales.

En la parte inferior de la Figure 2 se muestra la relación que existe entre el número de reglas de asociación difusas y varios valores de confianza con respecto a distintos valores del mínimo soporte. Se aprecia como el número de reglas aumenta según va disminuyendo el valor de la confianza. Resaltar como el valor de la confianza influye en el número de reglas cuando el mínimo soporte toma valores pequeños tales como 0,1 y 0,2.

6.2. Información extraida de "100mlI"

El análisis de las reglas de asociación difusas proporcionan altos niveles de conocimiento a partir del conjunto de datos "100mlI". El número de reglas obtenidas con un soporte de 0,2 y una confianza de 0,6 fue de 23, con respecto a las 38 obtenidas con un soporte de 0,1. Observamos como el valor predeterminado de la confianza implica la obtención de reglas de asociación relevantes debido al proceso de barrido de las posibles probabilidades de cada uno de los posibles valores que pueden tomar las variables (ver P10, Sección 5). Una muestra del nivel de interpretabilidad de la información obtenida es mostrada en la Tabla 1 obtenida a partir de varias reglas de asociación difusas relevantes. Estas reglas muestran como un atleta tiene características similares en las distintos parámetros que componen la prueba, es decir, un atleta tendrá una velocidad o carrera estimada como "Media" si su salto y su reacción también son "Medios". Además, gracias a la información obtenida se puede determinar que un

atleta no es relevante, para la prueba de 100ml si su salto es "Bajo", aunque tenga una velocidad considera "Media". Por el contrario, sí se considera relevante siempre que su reacción sea "Baja" y su carrera sea "Media". Esta información obtenida ha sido de gran utilidad para configurar el programa de entrenamiento de los deportistas de este equipo de atletismo.

7. Conclusiones

El objetivo principal de este trabajo es extraer reglas de asociación difusas a partir de DBC. Para ello, integramos conceptos de DBC con el algoritmo Apriori difuso propuesto por Hong y otros. Para alcanzar el objetivo, varios aspectos han sido considerados debido a la asusencia del valor real y preciso de los datos y a la interpretación posibilística aplicada. Esto ha afectado: a) al cálculo de la frecuencia de ocurrencias de los ítems, la cual es determinada a partir de las funciones de pertenencia que son interpretadas como conjuntos de probabilidades, b) al cálculo de la confianza de una regla la cual está contenida en un conjunto de probabilidades. El buen comportamiento y rendimiento de esta propuesta es mostrado a partir de un problema real con DBC, obteniendo como resultado información y patrones relevantes entre las dependencias y relaciones de los ítems.

Agradecimientos

Este trabajo está soportado por el Ministerio de Educación y Ciencia de España, TIN2008-06681-C06-{01 y 04} y TIN2011-28488 y por el proyecto andaluz TIC-2010-6858

Referencias

[1] R. Agrawal, T. Imielinski, A. Swami. Mining association rules between sets of items in large datasets. En: *SIGMOD*, Washington D.C. (USA), pp. 207-216, 1993.

[2] R. Agrawal, R. Srikant. Fast algorithms for mining association rules, En: *Int. Conf. on Very Large Data Bases*, pp. 487-499, 1994.

[3] C. Baudrit, D. Dubois, N. Perror. Representing parametric probabilistic models tainted with imprecision. *Fuzzy Sets and Systems*, vol. 15 (1), pp. 1913-1928.

[4] I. Couso, L. Sánchez. Higher order models for fuzzy random variables. *Fuzzy Sets and Systems*, vol. 159, pp. 237-258, 2008.

[5] M. Delgado, N. Marín, D. Sánchez, M.A. Vila. Fuzzy association rules: General model and applications. *IEEE Transactions on Fuzzy Systems*, vol. 11(2), pp. 214-225, 2003.

[6] D. Dubois, H. Prade, T. Sudamp. On the representation, measurement, and discovery of fuzzy associations. *IEEE Transactions on Fuzzy Systems*, vol. 13(2), pp. 250-262, 2005.

[7] J. Han, M. Kamber. *Data Mining: Concepts and Techniques.* Second Edition. M. Kaufmann, 2006.

[8] T.P. Hong, C.S. Kuo, S.C. Chi. Trade-off between time complexity and number of rules for fuzzy mining from quantitative data. *International Journal Uncertain Fuzziness Knowledge-Based Systems*, vol. 9(5), pp. 587-604, 2001.

[9] H. Ishibuchi, T. Nakashima, M. Nii. *Classification and Modeling with Linguistic Information Granules: Advanced Approaches to Linguistic Data Mining*, Springer-Verlag, 2004.

[10] A. Kaufmann, M. Gupta. *Introduction to Fuzzy Arithmetic: Theory and Applications.*, 1991

[11] G.J. Kulr, B. Yuan. Fuzzy Sets and Fuzzy Logic: Theory and Applications. Prentice Hall PTR, Upper Saddle River, NJ, 1995.

[12] Y.C. Lee, T.P. Hong, W.Y. Lin. Mining fuzzy association rules with multiple minimum supports using maximum constraints. *LNCS*, vol. 3214, pp. 1283-1290, 2004.

[13] P. Limbourg. Multi-objetive optimization of problems with epistemic uncertainty, 2005.

[14] A. Palacios, L. Sánchez, I. Couso. Future performance modelling in athletism with low quality data-based gfss. *J. of Mult.-Valued Logic and Soft Computing*, vol. 17(2-3), pp. 207-228, 2011.

[15] A. Palacios, L. Sánchez, I. Couso. Linguistic cost-sensitive learning of genetic fuzzy classifiers for imprecise data. *IJAR*, vol. 5(6), pp. 841-862, 2011.

[16] L. Sánchez, I. Couso, J. Casillas. Genetic learning of fuzzy rules based on low quality data.*Fuzzy Sets and Systems*, vol. 160(17), pp. 2524-2552, 2009.

[17] S. Yue, E. Tsang, D. Yeung, D. Shi. Mining fuzzy association rules with weighted items. *IEEE International Conference on Systems, Man and Cybernetics*, pp. 1906-1911, 2000.

[18] C. Zhang, S. Zhang. *Association Rule Mining: Models and Algorithms.* Springer-Verlag, 2002.