MINERÍA DE DATOS DIFUSA:
DESCUBRIMIENTO DE REGLAS ASOCIATIVAS DIFUSAS

JOHN HARVEY GARCÍA[@]

JORGE EDUARDO ORTIZ[@@]

Resumen

Este documento centra su interés en el problema de descubrir reglas asociativas en datos que se encuentran clasificados de forma difusa. Uno de los procesos más importantes en el descubrimiento de conocimiento en bases de datos es la minería de datos. El objetivo primordial de un algoritmo de minería de datos es el de encontrar regularidades en grandes cantidades de datos, estas regularidades entre muchas formas, pueden ser tomadas como reglas asociativas, una regla asociativa es una expresión que establece una relación entre dos conjuntos de elementos pertenecientes a una base de datos. Clásicamente el descubrimiento de reglas asociativas generalizadas, es el proceso de descubrir relaciones entre atributos de datos pertenecientes a todos los niveles de una estructura de clasificación llamada taxonomía, está taxonomía es exacta. En muchas aplicaciones reales, la estructura de clasificación puede ser difusa. Este documento centra su interés en el problema de descubrir reglas asociativas generalizadas en estructuras taxonómicas difusas. Para ello se hace una extensión de los conceptos de grado de soporte y grado de confianza, que son nociones tradicionales en el proceso de descubrimiento de reglas asociativas. Posteriormente se hace una descripción del proceso de minería de datos asociado para finalizar con los posibles estudios futuros en el área.

Palabras Clave: Minería de Datos, Conjuntos Difusos, Reglas Asociativas Generalizadas, Taxonomía.

[@] Departamento de Ingeniería de Sistemas e Industrial. Universidad Nacional de Colombia. johgarciam@yahoo.com

[@@] Ingeniero de Sistemas. Magíster en Estadística de la Universidad Nacional. Profesor Asistente Departamento de Ingeniería de Sistemas e Industrial. Universidad Nacional de Colombia. jorgeo@ing.unal.edu.co

Abstract

This paper focused in the issue to mining association rules in fuzzy data. One of the processes most important in the Knowledge Discovery in Databases is data mining. The main goal of a data mining algorithm is to find regularities in large amount of data, these regularities can be viewed like association rules, and an association rule is an expression that establishes a relationship between two sets of elements in a database. Classically mining generalized association rules, is the process to discover the relationships between data items at all levels of related taxonomic structure. This paper focuses on the issue mining generalized association rules with fuzzy taxonomic structures. Fuzzy extensions are made to the notions of the degree of support and the degree of confidence; these measures are traditional notions in the process of find association rules. Then, the data mining process is show and finally further studies in data mining are suggested.

Key Words: Data Mining, Fuzzy Sets, Generalized Association Rules, Taxonomy.

1. INTRODUCCIÓN

La búsqueda de reglas asociativas en grandes cantidades de datos provee una forma eficiente para descubrir conocimiento en bases de datos, estas reglas son útiles e interesantes en el momento de tomar decisiones.

En el año 1993 Agrawal [1] propuso un algoritmo para descubrir reglas asociativas que representaban relaciones entre ítems básicos de datos (presentes en los registros de ventas). Un ejemplo de tales reglas puede ser: “el 20% de los clientes compran manzanas y carne de cerdo; además, de todos los clientes que compran manzanas el 80% también compran carne de cerdo”, de esta forma tenemos la regla representada de la siguiente manera: Si Manzanas $\Rightarrow$ Carne de Cerdo, con un grado de confianza del 80% y un grado de soporte del 20%. Este tipo de reglas representan información importante para el administrador ya que podría tomar decisiones que le permitan obtener mayores beneficios.

Este algoritmo ha sido implementado satisfactoriamente por muchos investigadores en el área, que interesados en la eficiencia y eficacia del algoritmo lo han mejorado de diversas formas. En el año de 1995 Srikant y Agrawal [2] hicieron una extensión del algoritmo para descubrir reglas asociativas generalizadas, que representan reglas entre niveles de una taxonomía. En la mayoría de los casos esta taxonomía es una estructura jerárquica sobre los ítems. Esta estructura es exacta y cada ítem pertenece a uno o a un conjunto de ítems claramente definidos, pero en muchos casos de la vida real las clasificaciones no son determinísticas y para ello lo mejor es utilizar técnicas que permitan un tipo de abstracción más cercano al pensamiento natural y por tanto a la vida real. Una de las formas más adecuadas de hacer este tipo de aproximaciones es haciendo uso de conjunto difusos, que permiten estados intermedios en la clasificación de los ítems.

Para lograr un correcto proceso de minería de datos en estructuras de clasificación difusa, se debe hacer una extensión de los conceptos básicos de descubrimiento de reglas asociativas y por tanto una adaptación a los algoritmos respectivos. Éste básicamente es el objetivo del presente documento.

2. CONJUNTOS DIFUSOS Y MINERÍA DE DATOS

A continuación se presentan unos conceptos básicos para el desarrollo del tema; en primera instancia tendremos algunas aproximaciones conceptuales de conjuntos difusos y posteriormente una revisión al proceso clásico de minería de datos.

2.1 Conjuntos difusos

El concepto de Lógica Difusa fue concebido por Lofti Zadeh un profesor de la Universidad de California en Berkley, quien presentó un artículo en 1965 llamado *Fuzzy Sets* [3]; en este documento se presentaba una forma alternativa al concepto de pertenencia de los conjuntos, en la que un elemento pueda pertenecer a un conjunto de manera parcial.

En la teoría clásica un conjunto es una colección bien definida de elementos, la cual es posible determinar para cualquier objeto, en un universo dado, de tal modo que se ve si éste pertenece o no al conjunto.

Ejemplos de conjuntos bien definidos:

1) *El conjunto de enteros no-negativos, mayores que* 0 *y menores que* 5*. Este un conjunto finito con cuatro elementos:* 1,2,3 *y* 4*.*
2) *El conjunto de dinosaurios vivos en las montañas de Colombia. Este es un conjunto sin elementos, y es llamado conjunto vacío.*
3) *El conjunto de medidas mayores a* 100 *metros. Aunque este conjunto es infinito, es posible determinar si una medida dada es un elemento o no del conjunto.*

En un ***conjunto difuso***, cada elemento del universo tiene asociado un grado de pertenencia, que es un valor entre 0 y 1 a ese conjunto; este grado de pertenencia de aquí en adelante será representado mediante la letra griega μ. De esta forma podemos ver un conjunto difuso como una función que hace corresponder cada elemento del universo con un grado de pertenencia. Así, se tiene una función cuyo dominio es el universo y cuyo rango es el intervalo [0, 1]. Con esta definición es posible representar el conjunto difuso mediante la gráfica de la función que lo caracteriza.

Ejemplo de conjunto difuso:

El conjunto de gente joven, se podría representar de manera más formal mediante la notación:

B = {gente joven}

Como la edad comienza su inevitable incremento a partir del nacimiento, es posible definir límite inferior en el número 0*, pero el límite superior para el conjunto de gente joven es más complicado de definir; como primera aproximación se podría establecer en* 20*, de esta forma tenemos el intervalo* [0, 20] *para la gente joven, pero ¿qué hace que una persona sea joven el día de su cumpleaños número* 20 *y al día siguiente no?.*

La forma de solucionar esto sería suavizando la separación exacta entre el joven y el no joven. En los primeros ejemplos de conjuntos convencionales a los elementos del Universo de discurso se les hacía corresponder un valor de 0 *o* 1 *para evaluar su pertenencia al* *conjunto de interés. Si ahora se asigna un valor real entre* 0 *y* 1 *para el Universo de discurso edades relacionándolo de forma directa al conjunto de gente joven, tenemos una función que nos permite representar al conjunto difuso* B *tal como se ve en la Figura No. 1*.

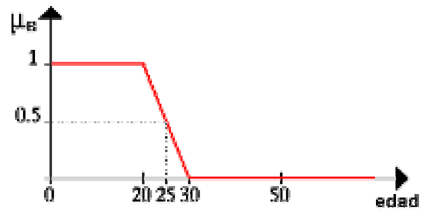

Figura No. 1. Función de Pertenencia

Así se puede suavizar la pertenencia o no al conjunto de gente joven, y una persona que tiene 25 *años es joven con un grado de* 0.5*.*

En el ejemplo anterior se definió el conjunto gente joven, esto permitió hacer representaciones más cercanas al pensamiento natural. Con el mismo universo de discurso se puede definir los conjuntos de gente muy joven, vieja, más o menos vieja, no tan joven, que también pueden ser representados mediante sus funciones de pertenencia.

Cada elemento en el universo de discurso es un miembro del conjunto difuso en algún grado, puede ser 0. El conjunto de elementos qué tienen una pertenencia mayor que 0 es llamado el *soporte* del conjunto difuso. La función que hace corresponder un número a cada elemento del universo es llamada la ***función de pertenencia*** $\boldsymbol{\mu(x)}$.

La función de pertenencia es uno de los componentes más importantes de un conjunto difuso. Es por lo tanto natural definir operaciones sobre conjuntos difusos mediante sus funciones de pertenencia.

Definición: Sea A y B conjuntos difusos en un universo mutuo:

(a) La intersección de A y B es:

$$A \cap B \equiv \boldsymbol{a} \ \textbf{min} \ \boldsymbol{b}$$

La operación ***min*** es la comparación mínimo de los grados de pertenencia, elemento a elemento entre los elementos correspondientes en A y B.

(b) La unión entre A y B es:

$$A \cup B \equiv \boldsymbol{a} \ \textbf{max} \ \boldsymbol{b}$$

Dónde ***max*** es la operación máximo, elemento a elemento.

(c) El complemento de A es:

$$\neg A \equiv \boldsymbol{1 - a}$$

Ejemplo (Zimmerman, 1993 [4]*):*

Una familia de cuatro personas quiere comprar una casa. Una indicación de qué tan confortable la quieren es por el número de cuartos en la casa. Pero ellos también quieren una casa grande. Sea **u** = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10) *el número de casas disponibles definidas por su número de cuartos. Luego el conjunto difuso **C** (para* Confortable*) puede ser descrito como*

$$\mathbf{C} = [0.2 \ \ 0.5 \ \ 0.8 \ \ 1 \ \ 0.7 \ \ 0.3 \ \ 0 \ \ 0 \ \ 0 \ \ 0]$$

*Sea **I** el conjunto difuso* Grande *definido como*

$$\mathbf{I} = [0 \ \ 0 \ \ 0.2 \ \ 0.4 \ \ 0.6 \ \ 0.3 \ \ 0 \ \ 0 \ \ 0 \ \ 0]$$

La intersección de Confortable *y* Grande *es*

$$\mathbf{C} \cap \mathbf{I} = [0 \ \ 0 \ \ 0.2 \ \ 0.4 \ \ 0.6 \ \ 0.3 \ \ 0 \ \ 0 \ \ 0 \ \ 0]$$

Esto quiere decir qué cinco cuartos es óptimo, pero sólo satisfactorio en un grado 0.6. La segunda mejor solución es cuatro cuartos.

La unión de Confortable *y* Grande *es*

$$\mathbf{C} \cup \mathbf{I} = [0.2 \;\; 0.5 \;\; 0.8 \;\; 1 \; 0.7 \;\; 0.8 \; 1 \;\; 1 \;\; 1 \;\; 1]$$

Aquí cuatro cuartos es totalmente satisfactorio porque éste es confortable, y 7 – 10 cuartos también, porque significaría una casa grande.

2.2 Minería de datos

La minería de datos es parte de un proceso llamado descubrimiento de conocimiento en bases de datos; el algoritmo de búsqueda de reglas asociativas corresponde a una de las muchas técnicas de minería de datos y fue propuesto originalmente por Agrawal, [1] como parte de una investigación en los laboratorios de IBM. La idea principal de este algoritmo consistía en encontrar todas la reglas asociativas significativas entre ítems en la base de datos. Un año después este algoritmo fue optimizado por los mismos autores y se denominó algoritmo ***Apriori*** [5], a partir de ese momento han existido varios enfoques basados en este algoritmo para obtener optimizaciones y resultados más eficaces. En el año de 1995, Agrawal y Srikant [2] publicaron un nuevo artículo, pero esta vez sus investigaciones estuvieron dirigidas al descubrimiento de reglas asociativas generalizadas, que se trata de un algoritmo basado en ***Apriori*** pero extendido a la búsqueda de reglas asociativas entre ítems que se encuentran clasificados en una estructura taxonómica jerárquica; por ejemplo en una taxonomía dada podríamos encontrar que un *gabán* es ropa de calle y *ropa de calle* es *ropa*, así estaríamos interesados en inferir reglas de la forma, *"la gente que compra **ropa de calle** tiende a comprar **zapatos**".*

Siendo de alguna forma más formales, se va a definir el modelo clásico de búsqueda de reglas asociativas generalizadas, basados en el documento de Agrawal y Srikant.

2.2.1 Modelo de búsqueda de reglas asociativas generalizadas

Dado un conjunto de transacciones, donde cada una es un conjunto de ítems, una regla de asociación es una expresión de la forma $X \Rightarrow Y$, dónde X y Y son conjuntos de ítems. El significado de tal regla es que las transacciones en la Base de Datos, que contiene los ítems en X, tiende también a contener los ítems en Y.

En la mayoría de los casos las taxonomías (jerárquicas) sobre los ítems son disponibles. Un ejemplo de taxonomía es el que se muestra en la Figura No. 2.

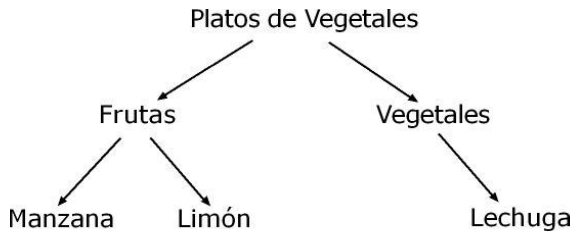

Figura No. 2. Estructura Taxonómica Exacta

En el caso de la Figura 2, una Manzana ***es una*** Fruta, una Lechuga ***es un*** Vegetal, etc., así se podrían inferir reglas significativas entre los niveles de la taxonomía, de la forma siguiente, "la gente que compra ***manzanas*** tiende a comprar ***vegetales***", este tipo de reglas pueden representar un tipo de información más significativa para la persona que toma decisiones en el supermercado.

Modelo Formal

Sea $I = \{i_1, i_2, \ldots i_m\}$ un conjunto de literales llamado ítems.

Sea G un grafo acíclico dirigido sobre los literales. Un lado en G representa una relación ***es-un(a)***, y G representa un conjunto de clasificaciones. Si existe un camino en G desde p hasta c, llamamos p un padre de c y a c un hijo de p. Modelar la taxonomía como un grafo permite mayores y diferentes clasificaciones que en un árbol.

Sea D un conjunto de transacciones, donde cada transacción T es un conjunto de ítems tal qué $T \subseteq I$.

Una transacción soporta un ítem $x \in I$, si x está en T o x es un ancestro de algún ítem en T. Una transacción T soporta $X \subseteq I$ si T soporta cada ítem en X.

Una regla de asociación generalizada es una implicación de la forma $X \Rightarrow Y$, dónde $X \subset I$, $Y \subset I$, $X \cap Y = \emptyset$, y ningún ítem en Y es ancestro de algún ítem en X.

La regla $X \Rightarrow Y$ se mantiene en el conjunto de transacciones D con un grado de confianza c si $c\%$ de las transacciones en D que soportan X también soportan Y.

La regla $X \Rightarrow Y$ tiene grado de soporte s en el conjunto de transacciones D si $s\%$ de las transacciones en D soportan $X \cup Y$.

El problema de extraer reglas generalizadas de asociación es descubrir todas las reglas que tienen grados de soporte y confianza mayores a un mínimo soporte y a una mínima confianza, respectivamente, estos grados mínimos son especificados por el usuario.

En este sentido el proceso de descubrimiento de reglas asociativas puede ser descompuesto en 3 partes:

1. Buscar todos los conjuntos de *ítems* (*itemsets* de aquí en adelante) cuyo soporte es mayor o igual que el mínimo soporte especificado. Los *itemsets* con soporte mínimo son llamados *itemsets* frecuentes.
2. Usar los *itemsets* frecuentes para generar las reglas deseadas. La idea general es que si, decimos *ABCD* y *AB* son *itemsets* frecuentes, entonces podemos determinar si la regla $AB \Rightarrow CD$ mantiene por cálculo la proporción *conf = soporte (ABCD)/soporte (AB)*. Si $conf \geq minconf$, entonces la regla se mantiene.
3. Eliminar todas las reglas no interesantes del conjunto. Es decir aquellas reglas cuya información aportada sea igual o menor a reglas ya extraídas.

3. ESTRUCTURAS TAXONÓMICAS DIFUSAS

Como se discutió anteriormente, una estructura taxonómica asume que un ítem hijo pertenece a su ancestro con grado 1. Pero en una taxonomía difusa, esta suposición no es del todo verdadera.

Sea $I = \{i_1, i_2, i_3,.., i_m\}$ el conjunto de literales llamados ítems.

Sea *FG* un grafo acíclico dirigido sobre los literales. Un lado en *FG* representa una relación difusa ***es-un(a)***, lo cual quiere decir que existe un grado parcial μ con el cual el nodo-hijo pertenece al nodo-padre en el lado del grafo correspondiente, donde $0 \leq \mu \leq 1$. Si existe un lado en *FG* desde *p* hasta *c*, *p* es llamado un padre de *c* y *c* es llamado un hijo de *p* (*p* representa una generalización de *c*).

Sea *T* el conjunto de todas las transacciones, *I* el conjunto de todos los ítems, y *t* es una transacción en *T* tal que $t \subseteq I$. Entonces una transacción *t* soporta un ítem $x \in I$ con grado 1 si *x* está en *t*, o con grado μ, si *x* en un ancestro de algún ítem *y* en *t* tal que *y* pertenece a *x* en un grado μ. Una transacción *t* soporta $X \subseteq I$ con grado β:

$$\beta = \min_{x \in X}(\mu)$$

donde μ es el grado con el cual *x* (en *X*) es soportado por *t*, y $0 \leq \mu \leq 1$.

Un ejemplo de una estructura taxonómica difusa es el mostrado en la Figura No. 3:

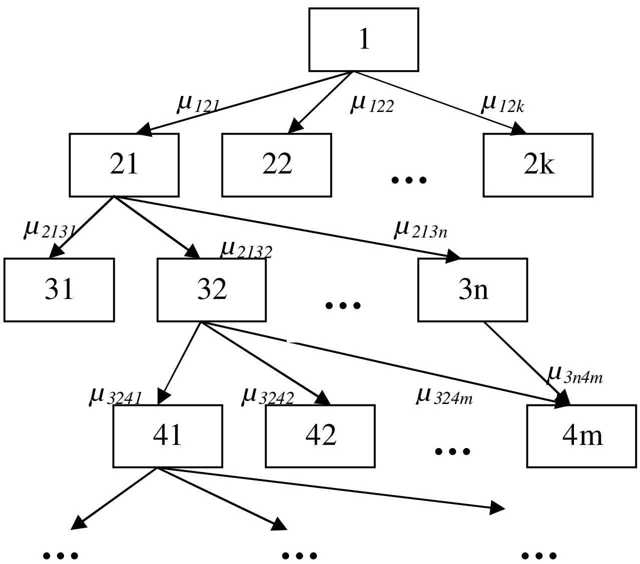

Figura No. 3. Estructura Taxonómica Difusa

En la Figura No. 3 cada nodo pertenece a su nodo padre con un grado μ_{yx} entre 0 y 1. Los nodos hoja de la estructura son atributos, que a su vez son los valores en los registros de la transacción. Cada nodo no hoja es denominado un nodo atributo, el cual es considerado como un conjunto cuyos elementos son los nodos hoja con sus respectivos grados de pertenencia. Como la estructura taxonómica difusa representa grados parciales de los lados del grafo, es necesario encontrar los grados entre nodos hoja y nodos atributos en *FG*. Este es un problema ya estudiado en la teoría de grafos difusos [6] [7], específicamente se tiene:

$$\mu_{xy} = \max_{\forall l:x \to y} (\min_{\forall eonl} \mu_{le}) \tag{1}$$

donde l: $x \to y$ es uno de los caminos de acceso entre los atributos x y y, *c sobre l* es uno de los lados del grafo que pertenece a l, μ_{yx} es el grado en el lado c sobre l.

3.1 Determinar el grado de soporte y el grado de confianza

Ahora es necesario considerar el cálculo del grado de soporte en una estructura taxonómica difusa. Si a es un atributo valor en una cierta transacción $t \in T$, T es el conjunto de transacción, y x es un atributo en cierto *itemset* X, entonces el grado μ_{xa} con el cual a pertenece a x puede ser obtenido de acuerdo a la fórmula (1). Así μ_{xa} puede verse como el grado en que la transacción $\{a\}$ soporta x. Además, el grado en que t soporta X puede ser obtenido de la siguiente forma:

$$\mu_{tX} = Soporte_{tX} = \min_{x \in X}(\max_{a \in t}(\mu_{xa})) \tag{2}$$

En este sentido, el grado en que la transacción t en T soporta a cierto *itemset* X es calculado. Además, en términos de cuántas transacciones en T soportan X, el operador $\Sigma count$ es usado para sumar todos los grados que están asociados con las transacciones en T:

$$Dsoporte(X) = \sum_{\forall t \in T} count(Soporte_{tX}) / |T| = \sum_{\forall t \in T} count(\mu_{tX}) / |T| \tag{3}$$

Donde |T| es el número de transacciones presentes en *T*. Para una regla asociativa generalizada $X \Rightarrow Y$, sea $X \cup Y = Z \subseteq I$, el grado de soporte $(X \Rightarrow Y)$ puede ser obtenido de la siguiente forma:

$$Dsoporte(X \Rightarrow Y) = \sum_{\forall t \in T} count(\mu_{tZ}) / |T| \tag{4}$$

De una manera análoga, el grado de confianza, $Dconfianza(X \Rightarrow Y)$ se puede calcular de la siguiente manera:

$$Dconfianza(X \Rightarrow Y) = \sum_{\forall t \in T} count(\mu_{tZ}) / \sum_{\forall t \in T} count(\mu_{tX}) \tag{5}$$

3.2 Ejemplo

Considerar la estructura de clasificación difusa mostrada en la Figura No. 4:

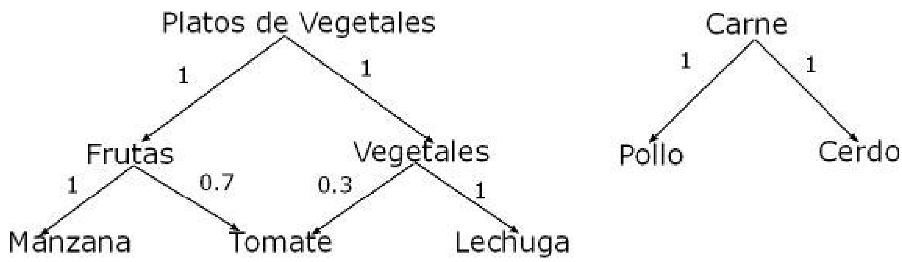

Figura No. 4: Estructura de Clasificación Difusa

Y la lista de ítems comprados por los usuarios como se muestra en la Tabla No. 1:

Id. de Transacción	**Elementos comprados**
100	*Manzana*
200	*Tomate, Pollo*
300	*Lechuga, Pollo*
400	*Tomate, Cerdo*
500	*Cerdo*

600	*Lechuga, Cerdo*

Tabla No. 1: Ítems Comprados

Según la fórmula (1) se pueden calcular los grados de los nodos hoja con sus ancestros como se muestra en la Tabla No. 2, por ejemplo:

$$\mu(Tomate \in \text{Platos de Vegetales}) = \mathbf{max}(\mathbf{min}(1, 0.7), \mathbf{min}(1, 0.3)) = 0.7.$$

Nodos Finales	**Los grados de los ancestros y sus hijos**
Manzana	*1/Manzana, 1/Fruta, 1/Platos de Vegetales*
Tomate	*1/Tomate, 0.3/Vegetales, 0.7/Frutas, 0.7/Platos de Vegetales*
Lechuga	*1/Lechuga, 1/Vegetales, 1/Platos de Vegetales*
Cerdo	*1/Cerdo, 1/Carne*
Pollo	*1/Pollo, 1/Carne*

Tabla No. 2: Grados de los nodos hijos y sus ancestros

El soporte mínimo es 30% y el grado de confianza es 60%. De acuerdo a la fórmula obtenida en (3), los valores de Σ*count* son calculados para obtener los *itemsets* frecuentes, los cuales son listados en la Tabla No. 3.

***Itemsets* Frecuentes**	**Valores para Σ*count***
{Lechuga}	2
{Tomate}	2
{Cerdo}	3
{Pollo}	2
{Fruta}	2.4
{Vegetales}	2.6
{Platos de Vegetales}	4.4

{Carne}	5
{Lechuga, Carne}	2
{Tomate, Carne}	2
{Vegetales, Carne}	2.6
{Platos de Vegetales, Carne}	3.4

Tabla No. 3: *Itemsets* Frecuentes

En la Tabla No. 3 se mostraron los valores de conteo para los *itemsets*, específicamente para el *itemset* {Vegetales, Carne} el valor $\Sigma count$ se calcula de la siguiente manera:

$$min(0.3, 1) + min(1,1) + min(0.3,1) + min(1, 1) = 2.6$$

Con base en los *itemsets* frecuentes es posible obtener las reglas que cumplen con los grados de soporte y confianza, estas reglas se muestran en la Tabla No. 4.

Reglas Interesantes	***Dsoporte***	***Dconfianza***
Vegetales⇒Carne	43%	100%
Platos de Vegetales⇒Carne	57%	77%
Carne⇒Platos de Vegetales	57%	38%

Tabla No. 4: Descubrimiento de Reglas Interesantes

4. PROCESO DE DESCUBRIMIENTO DE REGLAS ASOCIATIVAS GENERALIZADAS DIFUSAS

En síntesis, la tarea de descubrimiento de reglas asociativas se puede descomponer en 4 partes fundamentales:

1. Determinar el grado de pertenencia en el que cada nodo atributo pertenece a cada uno de sus ancestros. Existen varias formas que pueden ser usadas para determinar estos

grados de pertenencia, teniendo en cuenta que es un problema clásico de búsqueda de rutas óptimas.

2. Basados en los grados de pertenencia derivados del paso 1, encontrar todos los *itemsets* cuyo valor de conteo sea mayor que *min-soporte* $\times$ $|T|$. Estos *itemsets* son llamados *itemsets* frecuentes.
3. Usando los *itemsets* frecuentes generar las reglas cuyos grados de confianza son mayores que el mínimo grado de confianza especificado por el usuario.
4. Filtrar aquellas reglas que no sean interesantes.

5. CONCLUSIONES Y ESTUDIOS FUTUROS

La inclusión de los conceptos de la teoría de conjuntos difusos en los procesos de minería de datos permite obtener resultados más cercanos al pensamiento natural. En ese sentido este documento ha sido un intento de presentación de las variadas formas de hacer clasificaciones difusas para así obtener reglas asociativas generalizadas. Para ello se trabajó una forma alternativa para el cálculo de los grados de soporte y confianza. De está manera basados en el algoritmo clásico de búsqueda de reglas asociativas se muestra un modelo para el proceso de minería datos con una taxonomía de clasificación difusa. Así el proceso de minería de datos clásico se convierte en un caso especial del proceso descrito anteriormente.

Los estudios futuros deben intentar buscar métodos para el proceso de la *fuzzificación* de los datos, intentando mantener las relaciones apropiadas con el mundo real, así una forma interesante de estudio de esta técnica sería buscando aplicaciones de interés y sus resultados satisfactorios; otro camino de estudio venidero es el relacionado con la posibilidad de utilizar estás técnicas de búsqueda de reglas asociativas con los enfoques de sistemas bio-inspirados para lograr optimizaciones de los algoritmos y resultados en cierta forma más interesantes en el sentido computacional.

Referencias Bibliográficas

[1] AGRAWAL, Rakesh; IMICLINSKI, Tomasz; SWAMI, Arun. Mining Association Rules between Sets of Items in Large Databases, Proceedings of the 1993 ACM SIGMOD Conference Washington DC, USA, 1993.

[2] SRIKANT, Ramakrishnan; AGRAWAL, Rakesh. Mining Generalized Association Rules, Proceedings of the 21st VLDB Conference Zurich, Swizerland, 1995.

[3] ZADEH A, Lofti. Fuzzy sets in Information and Control, 1965.

[4] ZIMMERMANN H., -J, (1993). Fuzzy Set - Theory and Applications, 2 ed. Kluwer, Boston. (1 ed. 1991).

[5] AGRAWAL, Rakesh; SRIKANT, Ramakrishnan. Fast Algorithms for Mining Association Rules, Proceedings of VLDB Conference, Santiago, Chile, Sept. 1994. Expanded version available as IBM Research Report RJ9839, June 1994.

[6] M, Blue; B, Bush; J, Puckett. Applications of Fuzzy Logic to Graph Theory, Los Alamos National Laboratory (1997).

[7] CHEN, Guoqing. Fuzzy Logic in Data Modeling: semantics, constraints and database design, Boston: Kluwer Academic Publishers, 1998.

[8] AGRAWAL, Rakesh; MANNILA, Heikki; SRIKANT, Ramakrishnan; TOIVONEN, Hannu; VERKAMO A., Inkeri. Fast Discovery of Association Rules in Advances in Knowledge Discovery and Data Mining, AAAI Press/The MIT Press, 1996.

[9] CHEN, Guoqing, WEI, Qiang, KERRE, Etienne. Fuzzy Data Mining: Discovery of Fuzzy Generalized Association Rules. In Bordagna & Pasi (eds.), Recent Research Issues on Management of Fuzziness in Databases. Physica-Verlag (Springer), 2000.

[10] KERRE E., Etienne. Introduction to Basic Principles of Fuzzy Set Theory and Some of Its Applications. 2 ed. Gent, Belgium: Communication & Cognition, 1993.

[11] WEI, Qiang; CHEN, Guoqing. Mining Generalized Association Rules with Fuzzy Taxonomic Structures, Proceedings of the North America Fuzzy Information Processing Society (NAFIPS99). New York, 1999. p. 477-481.