Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 36 (2009) 3630–3640
Expert Systems with Applications
www.elsevier.com/locate/eswa

Association rules applied to credit card fraud detection

D. Sánchez [a,*], M.A. Vila [a], L. Cerda [a], J.M. Serrano [b]

[a] Department of Computer Science, A.I., University of Granada, E.T.S.I. Informática y Telecomunicación, C/Periodista Daniel Saucedo Aranda s/n, 18071 Granada, Spain
[b] Department of Informatics, University of Jaen, Spain

Abstract

Association rules are considered to be the best studied models for data mining. In this article, we propose their use in order to extract knowledge so that normal behavior patterns may be obtained in unlawful transactions from transactional credit card databases in order to detect and prevent fraud. The proposed methodology has been applied on data about credit card fraud in some of the most important retail companies in Chile.

Keywords: Association rules; Data mining; Credit card fraud; Fraud detection; Fraud prevention

1. Introduction

Competitiveness in the retail industry is continuing and it is becoming increasingly aggressive as revealed by recent events in the sector and specialist studies such as (Zarufe, 2005). One of the leading businesses in this sphere is hire purchase and one of the main commercial strategies is the emission of department store credit cards to clients, as evident from the publications (Zarufe, 2005) which indicate that three companies lead the retail industry in the Latin American Southern Cone (Argentina, Chile, Peru, Colombia). In Chile, they compete for 95% of retail industry sales, which in 2003, according to these same publications, exceeded the three thousand, three hundred million dollar mark with market shares of 60.29%, 18.26% and 15.63%, respectively. By the end of 2003, they had issued 3.0, 2.7 and 2.6 million credit cards, and in Chile alone 16 million credit cards had already been issued by the different retail distribution chains. This form of payment was 7 times higher than the number of bank-issued credit cards, which were responsible for on average 65% of the sales of their issuing houses, represented almost 20% of Chile's consumption debt in total, and there were over 11,000 establishments who did not issue this type of card but which traded with them. Within the retail industry, the predominant trade component are financial services and the distribution of clothing and furnishings through department stores with sales in 2005 reaching US $3,194, sales which are distributed among the four major chain stores with the profile in Table 1.

For years, both from the academic and the technological advisory or consultancy perspective, it has been observed and in this case confirmed that in three of these four large companies, the level of technological support used as part of their computerized management systems in their decision-making processes is very different from the level of use of information technologies used in the operational transaction systems, and these computerized management systems may therefore be classified according to the following levels of computing maturity:

Level Zero (non-computerized systems): No incorporation of computer technology in computerized management systems, i.e. information for decision-making is obtained manually.

* Corresponding author. Tel.: +34 958 246397; fax: +34 958 243317.
E-mail addresses: daniel@decsai.ugr.es (D. Sánchez), vila@decsai.ugr.es (M.A. Vila), lcerda1511@hotmail.com (L. Cerda), jschica@ujaen.es (J.M. Serrano).

0957-4174/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.02.001

Table 1
Distribution of sales between the four major chain stores in Chile

Indicator	Falabella	Ripley	Paris	La Polar	Total
No. of stores	30	31	19	26	106
Sales surface area m^2	177,538	227,909	154,544	81,000	640,991
Surface area/stores	5,917	7,352	8,134	3,115	6,048
Sales (US$M)	1,178	871	782	360	3,194
% market	36.90%	27.30%	24.50%	11.30%	100
Sales/m^2	3.55	2.6	2.6	2.3	
Cards issued (M)	3.3	2.6	3.0	1.9	10.8
Active cards (M)	2.6	1.4	1.2	1.4	6.6
% card sales	67%	63%	67%	80%	
Projected investment (US$M)	1130	551	1200	100	

Level One (semi-computerized systems): Incipient use of technological resources in computerized management systems for decision-making, e.g. the use of spreadsheets to prepare the information.

Level Two (departmental computerized systems): Use of departmental information systems as the nucleus of the computerized management system for decision-making, and data is usually gathered from transactional processes associated to a specific function within the value chain.

Level Three (integrated computerized systems): Use of the company's integrated administrative information system (ERP) as the main element to supply the computerized management system (CMS) for decision-making. These CMS are supplied by data from all the processes supporting the company's value chain.

Level Four (controlled or synchronized computerized systems): These computerized management systems (CMS) integrate the use of control boards or control panels, enabling decisions to be made as the need arises by having online information for the fulfillment level of the objectives associated with the management indicators.

Level Five (predictive computerized systems): Data mining models are incorporated into the previous level to extract non-explicit information from the records of transactions from daily operation. These enable new behavior patterns to be determined which reinforce, confirm or modify management indicators and allow trends to be recognized for decision-making.

Level Six (automatic computerized systems): The previous level of computerized management systems is reinforced and combined with daily operation through the use of expert systems with knowledge bases and inference engines to support decision-making, thereby incorporating intelligence into the operational systems so that given certain management parameters and indicators they activate, restrict or modify business rules.

From the critical or strategic decision-making processes for the business in question, two areas were chosen, which are the most representative in this industry: operational risk control, and corporate management and planning.

In this article, we present the first work into operational risk control, whereby we worked with the area of the company with available data and with a Level Two computerized management system (the others were ruled out on account of them not having any available data for confidentiality reasons or having a Level Zero or Level One computerized management system). In our next publication, we will present our work into the area of corporate management, studying the case of one of these leading companies which also has a Level Two computerized management system.

The objective of both these works is to transform the computerized management systems of these decision-making processes from their current computerization levels with their reactive decision-making processes to computerization levels with proactive decision-making processes.

This first publication therefore presents the result of applying cutting edge information technologies to one of the operational risk control processes and transforming it from Level Two to Level Five or Six. In particular, the work focuses on the process for controlling the risk of fraud through the use of corporate credit cards as a form of payment.

In this respect, the selected process supports one of the widest used differentiation and sustained growth strategies in this industry for obtaining client loyalty. While it is true that the mass issue of credit cards by department stores has been successful as a marketing project, it is equally true that this has increased the risk of exposure to illegal activity, as demonstrated by the growing tendency for fraud which is highlighted in specialist publications (e.g. the latest Cybersource report, Sponsored by CyberSource Corporation Conducted by Mindwave Research, 2006). Diversification of the client portfolio with this mass issue of credit cards and aggressive marketing plans which encourage the diverse use of this payment method, combined with the lack of efficient techniques and intelligent systems to enable effective detection and prevention of their illegal use, without inconveniencing genuine credit card users, involve the challenge of seeking more efficient methods. This effort is reflected in various articles, in particular in specialist publications which offer different approaches for detecting and preventing this illegal behavior. Nevertheless, all of these concur with Bhatla's observations in 2002 (Bhatla, Vikram, & Dua, 2003) that the evaluated systems are prone to guaranteed effectiveness and that none of the reviewed tools and technologies can alone eliminate fraud. Furthermore, since each technique contributes to the ability to detect fraud, he believes that the most successful option would be a combination of several of these techniques, since the results of the Cybersource survey (Sponsored by CyberSource Corporation Conducted by Mindwave Research, 2006) seem to indicate that manual control is still the most used method for detecting and preventing fraud.

A summary of the state of the art in the techniques and methods used in fraud detection and prevention, and a review of various relevant publications over the last three years confirms the effort employed to obtain useful knowledge from the transaction databases or repositories using different techniques and methodologies. In the light of the results obtained, we have decided to use fuzzy logic-based data mining techniques for these purposes and to apply some of the soft computing methodology explained in the publications (Delgado, Sánchez, & Vila, 2000; Delgado, Marín, Sánchez, & Vila, 2003) and the concepts explained in Au and Chan (1998), Berzal, Blanco, Sánchez, and Vila (2001, 2002), Berzal, Cubero, Sánchez, Serrano, and Vila (2003), Sánchez, Serrano, Blanco, Martín-Bautista, and Vila (2008), Bra and Paredaens (1983), Calero et al. (2004), Chan and Au (1997), Chen and Wei (2002), Cubero, Medina, Pons, and Vila (1994), Delgado et al. (2003), Dubois, Hullermeier, and Prade (2003), de Graaf, Kosters, and Witteman (2001), Gyenesei (2001), Hullermeier (2001), Kaya, Alhajj, Polat, and Arslan (2002), Kivinen and Mannila (1995), Kuok, Fu, and Wong (1998), Hong, Kuo, and Chi (1999), Luo and Bridges (2000), Pedrycz (1998), Pfahringer and Kramer (1995), using in particular the fuzzy association rules presented by Sánchez (1999) and the logarithms presented by Serrano (2003).

This article extracts a set of fuzzy association rules from a data set containing genuine and fraudulent transactions made with credit cards and compares these results with the criteria which risk analysts apply in their fraud analysis processes.

The methodology used overcomes the difficulties relating to minimum support and confidence and optimizes the execution time and the excessive generation of rules, with more intuitive results than the methodologies analyzed in Bhatla et al. (2003).

In the second section of this article we present the main concepts related to the methodology of these fuzzy association rules used in this work. Section 3 explores the results of the first data mining stages, i.e. selection, organization, exploratory analysis of the sample and its results. Section 4 details the process to obtain the association rules with fuzzy logic, by defining the linguistic labels obtained with the methodology presented in Serrano (2003) and by searching for the association rules by applying the FuzzyQuery 2+ (Serrano, 2003) tool first to the client information table (achieving certainty factors of 92.66%) and then to both the client and transaction tables (achieving certainty factors of 80.06%). Finally, we conclude by analyzing the work carried out and we propose the future challenges which have emerged in the light of the good results obtained.

2. Methodology

In the following section, we will summarize the methodology used in this work and the main concepts involved.

The objective of data mining is to obtain useful, non-explicit information from data stored in large repositories and for our work (as indicated in Frawley, Piatetsky-Shapiro, & Matheus 1992), it represents a methodology which offers an excellent alternative solution to the problem being studied since knowledge extraction is considered to be a basic (and in our case specific) need for determining trends and patterns of illicit behavior in transactions using department store credit cards.

In order to obtain non-explicit, useful information from large repositories, the fuzzy set theory certainly contributes since it is widely accepted that many of the relations which occur in the real world are intrinsically fuzzy and any decisions made use incomplete or imprecise information; in our case, the criteria used by risk experts have this connotation.

In the context of data mining, one of the most studied knowledge extraction models is that of rule association (Agrawal, Imielinski, & Swami, 1993), which assumes that the basic object of interest is an item and that the piece of information appears in the form of an itemset called *transaction*. The association rules are implications which relate the item presence in the transaction and the transaction is the basic structure from where the association rules are obtained (in our case the transactions are the purchases paid for by credit card and the items we select are all those identifying this event and the person using this form of payment).

Zadeh (1965, 1971) and Sánchez et al.'s research (Sánchez, 1999) provides the conceptual support necessary for this work and Serrano's implementation in Serrano (2003) enables this work to be carried out. The main related concepts are drawn from Sánchez, Sánchez, Serrano, and Vila (2004) and reviewed in the following sections.

2.1. Association rules

Given an item set I and a transaction set T (also known as the T-set), where each transaction is a subset of I, an association rule (Agrawal et al., 1993) is said to be an "implication" of the form $A \Rightarrow C$ denoting the presence of itemsets A and C in some of the T transactions, assuming that A, $C \subset I$, $A \cap C = \emptyset$; and A, $C \neq \emptyset$.

The usual measures proposed in Agrawal et al. (1993) for establishing an association rule's fitness and interest are the *confidence* (Conf($A \Rightarrow C$), the conditional probability $p(C|A)$), the *support* (Supp($A \Rightarrow C$), and the joint probability $p(A \cup C)$).

Some authors, however, highlight certain disadvantages of the confidence, which is why an alternative solution appears in Berzal, Blanco, Sánchez, and Vila (2001, 2002), where Shortliffe and Buchanan's certainty factor (Shortliffe & Buchanan, 1975) is proposed instead (Pfahringer and Kramer (1995)). The certainty factor of rule $A \Rightarrow C$ is calculated as

$$\mathrm{CF}(A \Rightarrow C) = \frac{\mathrm{Conf}(A \Rightarrow C) - \mathrm{Supp}(C)}{1 - \mathrm{Supp}(C)} \tag{1}$$

if $\mathrm{Conf}(A \Rightarrow C) > \mathrm{Supp}(C)$ and

$$\mathrm{CF}(A \Rightarrow C) = \frac{\mathrm{Conf}(A \Rightarrow C) - \mathrm{Supp}(C)}{\mathrm{Supp}(C)} \tag{2}$$

otherwise. The certainty factor takes values in $[-1, 1]$, indicating the degree to which our belief that the consequent is certain varies when the antecedent is also certain. It ranges from 1 (the maximum increase: i.e. if A is certain, then so is C) to -1, indicating the maximum decrease.

By definition, association rules are defined on transaction sets. Given that it is more common to work with tuples rather than transactions in a database, various solutions to this problem have been proposed. When working with relational databases, it is usual is to consider each item to be a pair (attribute, value) and each transaction to be a tuple in a table. For example, let us say that the item $\langle A, a0\rangle$ is in the transaction associated to a tuple t `iff` $t[A] = a0$.

This variant does, however, have various disadvantages. In the case of numerical attributes, the number of items associated to each possible pair $\langle$attribute, value$\rangle$ could be very high. Some proposed solutions are the *generalized association rules* (Sánchez, 1999), which allow taxonomies to be defined between the attribute values, or the quantitative association rules (Sánchez, 1999), for which an $\langle$attribute, value$\rangle$ pair is made to correspond to each item, or in other words, the attribute domain is partitioned into intervals. Another possibility is to give the values a certain degree of imprecision and to extract fuzzy association rules.

2.2. Fuzzy association rules

Various proposals for fuzzy association rules can be found in the literature such as a generalization of association rules when initial data are fuzzy or if they have been previously processed to provide them with imprecision (Berzal et al., 2001; Delgado et al., 2003; Hong et al., 1999; Kuok et al., 1998). Although the majority of these approaches have been applied on relational databases, almost all the measures and algorithms proposed can be used in a more general setting. An interesting in depth study (with references) into the extensions to quantitative attribute cases or item hierarchies can be found in Delgado et al. (2003) and other additional approaches to the problem appear in Chan and Au (1997), Chen and Wei (2002), de Graaf et al. (2001), Gyenesei (2001), Hullermeier (2001), Kaya et al. (2002), Luo and Bridges (2000), Pfahringer and Kramer (1995). In Dubois et al. (2003), various fuzzy data mining measures are studied.

In Delgado et al. (2003), the authors first define fuzzy transactions as fuzzy item subsets, and we will use this idea as a basis further on. Therefore, let $I = \{i_1, \ldots, i_m\}$be an itemset and T' a fuzzy transaction set, in which each fuzzy transaction is a fuzzy subset of I Given the transaction $t \in T'$, we will use $t(i)$ to denote the membership degree of item i in the transaction t. A fuzzy association rule is an implication of the form $A \Rightarrow C$, such that A, $C \subset I$ and $A \bigcap C = \emptyset$.

It is immediate that the transaction set where a certain item appears is a fuzzy set called the item *representation*. For the item I in T', we have the following fuzzy subset of T' in

$$\tilde{\Gamma}_i = \sum_{t \in T} t(i) \Big/ t \tag{3}$$

This representation can be extended to an itemset in the following way: given an itemset $J \subset I$, its representation is:

$$\tilde{\Gamma}_J = \bigcap_{i \in J} \tilde{\Gamma}_i \tag{4}$$

using the minimum. To measure the fitness and interest of a fuzzy association rule, approximate reasoning tools must be applied because of the imprecision which may affect the fuzzy transactions. As Delgado et al. (2003) indicates, a semantic focus can be used based on the evaluation of quantified sentences (Zadeh, 1983). In the usual way, we shall use the quantifier $\mathrm{QM}(x) = x$, since it is the only quantifier providing a direct extension of classical support and confidence. Then

- The support of an itemset J in the FT-set T' is the result of evaluating the quantified sentence Q of T' are Γ_J
- The support of the association rule $A \Rightarrow C$ in the FT-set T', $\mathrm{Supp}(A \Rightarrow C)$, is the support of the itemset $A \cup C$.
- The confidence of the association rule $A \Rightarrow C$ in the FT-set T', $\mathrm{Conf}(A \Rightarrow C)$, is the evaluation of the quantified sentence Q of Γ_A are Γ_C.

The certainty factor $\mathrm{CF}(A \Rightarrow C)$ can be obtained from confidence and $\mathrm{Supp}(C)$ as in the crisp case. The previous sentences will be evaluated by means of the GD method defined in Delgado et al. (2000); following this method, the evaluation of the sentence Q of F are G is obtained as in:

$$\mathrm{GD}_Q(G/F) = \sum_{\alpha_i \in \Delta(G/F)} (\alpha_i - \alpha_{i+1}) Q\left(\frac{|(F \cap G)_{\alpha_i}|}{|F_{\alpha_i}|}\right) \tag{5}$$

where $\Delta(G/F) = \wedge(G \cap F) \cup \wedge(F)$, with $\wedge(F)$ being the set of levels of F, and $\Delta(G/F) = \{\alpha 1, \ldots, \alpha p\}$ with $\alpha i < \alpha i + 1$ for each $i \in \{1, \ldots, p\}$. Let us suppose that the F set is normalized. If this is not the case, F is normalized and the normalization factor is applied to $G \cap F$.

2.3. Use of semantic concepts for fraud detection

In order to process the transaction sample, linguistic labels are first established for each attribute comprising the t transactions belonging to T. By applying Delgado's algorithm, sets of rules can be obtained for each of the relations among the attributes to be taken into account. Given a concept of these attributes Ck belonging to C and a transaction set T, the corresponding rule set is known as $\mathrm{Rul}(T, Ck)$. It is worth mentioning that we are only interested in the CF rules in the range $(0, 1]$, which implies the presence or not of fraudulent concepts and so $\mathrm{Rul}(T, Ck)$ only contains those rules with a certainty factor which is greater than zero, $\mathrm{CF} > 0$.

We can use the model obtained in this way to check for the presence or not of fraudulent transactions, and given a set of lawful and unlawful transactions, our aim is to determine which transactions t comply with the characteristics of illicit transactions Ck. The certainty of transaction t verifying characteristics Ck will be $\mathrm{Cer}(t, Ck)$. Initially, we will consider that the transaction t verifying the characteristics Ck with a $\mathrm{CF} = 0$ is lawful.

We will use minCer to discriminate between a lawful transaction and one which is not, and each transaction with a $\mathrm{Cer}(t, Ck)$ which is greater than or the same as minCer will be reported.

By discriminating between the transactions in terms of the $\mathrm{Cer}(t, Ck)$ value, the idea is to obtain two groups (clusters) corresponding to lawful transactions and fraudulent transactions, by means of an automatic computational process.

Finally, storing the rules will enable online differentiation of the transactions, thereby avoiding processing time and the need for significant memory resources.

3. Application to a specific problem: the case of fraud in multinational department stores

In the financial industry in general, where regulations require for there to be minimum levels of transactional security, there is the tendency to validate the authenticity of the transaction, and there are no online systems enabling unlawful activity to be detected and prevented. This is fundamentally due to a lack of confidence that current systems are able to detect unlawful activity. The ratio of fraudulent transactions to lawful transactions in the sample under study is 3 in every one thousand, which represents a high probability of false positives being obtained. This would produce an undesired effect since the department stores' main concern by issuing credit cards is to cultivate customer loyalty. The effect that this illicit activity has on the loyalty process is very destructive, given that those affected would lose confidence in this form of payment and change their purchase behavior.

3.1. Data organization

As in many companies, daily operation data are spread over various information processing and representation systems on different hardware and software platforms. This naturally represents a disadvantage when it comes to extracting knowledge and requires the typical preliminary work of the data mining process which consists in selecting and refining data. For our purposes, we have resorted to the client control system data to obtain the items relating to the person making the transaction and to the sales vsystem data for the transaction details, and to the stock control system data for the product-related items. As a result of these first traditional data mining steps, and by only considering those relevant for establishing a test set, the obtained data are organized as described in Tables 2 and 3.

Table 2
Description of the client table

ID attribute	T. data	Length	Details
CL_ID	Number	11	Number identifying real client
CL_NCD	Number	5	Number of same-day purchases
CL_LCD	Number	5	Number of stores where same-day purchases made
CL_CUPO	Number	8	Amount authorized for purchases
CL_COACTI	Number	11	Customer activity code
CL_COMUNA	Character	11	Municipality code where account was opened
CL_SEXO	Character	1	"M": male; "F": female
CL_TICREDITO	Character	6	"L": credit; "C": installment; "R": refunded
CL_SUCU	Character	14	Branch where account was opened
CL_REGION	Number	8	Region where account was opened
CL_EDAD	Number	3	Customer's current age
CL_ANT_CTA	Number	3	Number of years account held
CAT_ CL	Number	3	Customer category VIP 1, normal 0

Table 3
Description of the transaction table

ID attribute	T. data	Length	Details
CL_ID	Number	11	Number identifying real client
Tx_FechaTx	Date	8	Transaction date
Tx_LC_ID	Number	6	Store where purchase was made
Tx_POS_ID	Number	11	POS where purchase was made
Tx_POS_BOLETA	Number	11	Slip number issued by POS
Tx_BOLETA	Number	11	Unique slip number = LC_ID + POS_ID + POS_BOLETA
Tx_tipo	Number	4	Type of transaction made
Tx_TITUL	Number	4	Identification if Tx was made by a holder or other
Tx_CANT	Number	4	Units purchased in the Tx
Tx_CUOTAS	Number	4	Number of installments
Tx_M_LIQ	Number	4	Total amount of purchases made in transaction
Tx_LPROD	Number	4	Product line for each transaction, e.g. (1) electronics, (2) down payments, (3) white goods, (4) furnishings, (5) clothing, (6) footwear, (7) other
Tx_NEGOCIO	Number	4	Store code associated to the Tx
Tx_EN DISP	Number	1	0 = NO; =UNDER DISCUSSION

3.2. Exploratory analysis of the selected sample

A first examination of the sample obtained, by comparing annual activity, reveals that fraud has increased rapidly by 15%, and that a larger proportion of transactions are being made outside the sphere of the sales outlets of the companies in the sample, in other words, they are being made in the sales outlets of these companies' commercial partners. The results of the exploratory analysis obtained from the data sample are the following:

- The product or service with the largest amount of fraud in 2002 (a situation which was again repeated in 2003) was gas sales.
- The four municipalities with the greatest numbers of customers affected by fraud were La Florida with 11.21%, followed by Maipú with 8.44%, Las Condes with 7.80% and Puente Alto with 7.16%. On the other hand, the four municipalities with the lowest numbers were Colina with 0.09%, followed by Lampa with 0.13%, Padre Hurtado 0.26% and Lo Espejo with 0.55%.
- Of the total number of fraudulent transactions, 55% were made on the cards of female clients.
- Around 34% of the fraud affected the 18–30 age group, which in comparison with the other groups indicates a high level of fraud.
- The majority of fraudulent transactions occur in gas stations (61.6%), drug stores (14.3%), fast food joints (4%), and for pay-as-you-go top-ups (15.3%), all of which are products which can easily be converted into cash.
- The amount of fraudulent transactions have increased by an annual rate of 15%.

From the exploratory analysis, the following situations emerge which risk analysts consider to be usual when detecting the problem:

- Women are most likely to be affected by credit card fraud.
- Young people are most likely to be affected by credit card fraud.
- The main products involved in fraud can easily be converted into money.
- A large proportion of the cards involved in fraud have recently been issued.
- Certain businesses are more prone to fraud.

Other pointers which are a good indication of fraud is a change in the client's purchase behavior, and these include:

- change in normal place of purchase,
- purchases of products which do not correspond to client's profile,
- purchases in abnormal quantities,
- increase in frequency of product purchases,
- purchases in different stores within a very short time period,
- no-money-down purchases.

4. Development of the process to obtain association rules with fuzzy logic

4.1. Exploration of the obtained database

This data mining process comprises the following three stages:

1. Establishment of the set of linguistic labels which enable the test transaction set to be defuzzified, using Serrano's software tools (Serrano, 2003).
2. Incorporation of the linguistic labels and their membership degrees as items in the client and transaction tables.
3. Application of the FuzzyQuery 2+ software tool Serrano et al. (2003). This tool allows experiments to be carried out with different levels of support, confidence, certainty factor and number of items to be considered in the process.
4. Analysis of the results, selecting from the fuzzy association rule set those rules with a certainty factor within a certain threshold and which are not so obvious or which are common sense.
5. Repeat from Step 2, varying the support, confidence and certainty factor, in order to optimize this process.

For the computational process, we used an HP Compaq nx9020 notebook with a 1.4 GHz Intel Celeron processor and 256 MB of RAM and Microsoft Windows XP Professional 2001 operating system with Service Pack 1, and Oracle 9i database engine.

4.2. Definition of linguistic labels

The preliminary exploratory analysis enabled us to determine the following attributes which had the greatest impact on fraud:

- product type
- place of purchase
- purchaser's age
- purchaser's sex
- years account held
- no-money-down purchase
- purchase in maximum installments
- purchase period

Using the K-means algorithm and expert knowledge, the following trapezoidal labels were obtained for Age:

- young person $(18, 18, 24, 28)$
- young adult $(24, 28, 38, 40)$
- adult $(38, 40, 60, 64)$
- elderly adult $(60, 64, 75, 75)$

For the number of years the account has been held (in months) the *K*-means algorithm is used, generating the following trapezoidal labels:

- recent (0, 0, 6, 12)
- normal (6, 12, 24, 36)
- old (24, 36, 48, 48)

In the same way, for the purchase period in number of installments, the *K*-means algorithm is used, generating the following trapezoidal labels:

- short (0, 0, 2, 4)
- medium (2, 4, 8, 10)
- long (8, 10, 12, 12)

Finally, for the purchase amount (in thousands of pesos):

- low (1, 1, 10, 30)
- medium (10, 30, 40, 50)
- high (40, 50, 80, 80)

4.3. Search for association rules

As we have already mentioned, once the linguistic labels have been defined, both the label and the membership degree of the attribute to the label are included in the sample as attributes. After the items comprising the data sample have been defuzzified, the transaction set is processed with the Fuzzy Query 2+ data mining tool. The results are shown in Tables 4 and 5.

The experiment which was carried out on the client table with 1959 rows, applying a support level of 0.015 with minimum confidence and certainty factor of 0.6% and with a maximum of 5 attributes in the antecedent, returned a total of 8008 rules, and the process indicated that there were 436 rules with two elements, 1009 with three elements, 1184 with four elements and 730 with five elements. From these, we select the 379 which gave positive fraud as the consequent (state fraud = 1), and from these, we select more than 50 with a certainty factor of over 73%.

The process produces various rules which constitute a novelty for experts in that it breaks two paradigms: firstly, that most fraud affects women, given that the experiment indicates that for this sample the combination sex, region and recent card (with a 2.6% presence in the sample) have a 92.8% confidence and a 92.66% certainty factor, and that there is a greater prevalence of fraud in Central Santiago than Mall Plaza Vespucio.

The experiment carried out with the client table with its 1959 calculations is combined with the transaction table with its 12,107 rows, applying a minimum support level of 0.015 with a minimum confidence and certainty factor of 0.6% and with a maximum of 4 attributes in the antecedent. This experiment indicates that there can be 600 rules with two elements, 2429 rules with three elements, and 5715 with four elements. Finally, the process generates 19,116 rules, of which 84 have positive fraud as a consequent (Tx in dispute = 1), and we select the 50 rules with the highest certainty factor (certainty factor over 66%).

As with the client table experiments, the rule with the greatest certainty factor (80.08%) indicates that young males are most affected by fraud (a result which contradicts the usual criterion that young females were most affected), presenting a certainty factor of 74.29% for the rule normally used by risk experts.

Having analyzed the results of the experiments carried out on the client table, we can conclude that of the 50 fuzzy association rules selected from the set returned by the software under study, the majority of these are the ones most used by the risk analysts responsible for the sample and which they consider to be relevant, and the methodology used offers significant non-explicit knowledge which contributes to the purpose of their work.

These results over a total of 1959 clients who made 12,107 transactions, indicate the great contribution that the application of fuzzy association rules can make to preventing credit card fraud, and they confirm that the soft computing application represents a significant contribution to the extraction of useful knowledge of the major repositories and positively contributes to the automation of this process. This work also demonstrates that the methodology overcomes not only the problems relating to the interest and fitness of the rules but also the communication barrier, returning intuitive results for the specialists. The use of the computational platform also proves that the performance restrictions of the algorithms have been overcome.

5. Conclusions and future work

5.1. Conclusions

1. The main requirements for obtaining useful knowledge from large repositories can be resolved by using soft computing techniques, contributing with practical solutions to the new competitive scenario which requires online actions in order to mitigate the risk of undesired activities or actions using such forms of payment as credit cards.
2. Intelligent tools are prevented from obtaining good results due to the heterogeneity of the systems for processing and representing the information and because of the imperfections in the data as a result of the traditional model supporting these systems. Fuzzy logic-related data mining techniques enable this risk to be reduced.
3. The applied methodology overcomes the difficulties of minimum support and confidence, optimizes the execution times, reduces the excessive generation of rules, and helps make the results more intuitive, thereby facilitating the work of fraud analysts.

Table 4
Results for the client table

Background	Support * 100	Confidence * 100	CF
[“SEX MALE”, “REGION METROPOLITAN”, “YOUNG”, “CARD RECENT”]	2,657128	92,85716	0,9266219
[“SEX MALE”, “REGION METROPOLITAN”, “YOUNG”, “CLIENT NORMAL”]	2,503832	92,45293	0,9225912
[“Santiago Center”, “REGION METROPOLITAN”, “YOUNG”]	2,452732	87,27267	0,8695265
[“Santiago Center”, “REGION METROPOLITAN”, “YOUNG”, “CARD RECENT”]	2,452732	87,27267	0,8695265
[“Santiago Center”, “REGION METROPOLITAN”, “YOUNG”, “CLIENT NORMAL”]	2,350537	86,7925	0,8647457
[“Santiago Center”, “YOUNG”, “CARD RECENT”, “CLIENT NORMAL”]	2,350537	86,7925	0,8647457
[“SEX FEMALE”, “Santiago Center”, “REGION METROPOLITAN”, “YOUNG”]	1,94175	86,3637	0,8609368
[“SEX FEMALE”, “Santiago Center”, “YOUNG”, “CARD RECENT”]	1,94175	86,3637	0,8609368
[CLERICAL WORKER, “REGION METROPOLITAN”, “YOUNG”, “CARD RECENT”]	2,606032	86,44076	0,8607795
[STUDENT, “REGION METROPOLITAN”, “YOUNG”, “CARD RECENT”]	3,270315	86,48655	0,8602967
[STUDENT, “REGION METROPOLITAN”, “YOUNG”, “CLIENT NORMAL”]	3,270315	86,48655	0,8602967
[“SEX FEMALE”, “Santiago Center”, “YOUNG”, “CLIENT NORMAL”]	1,890648	86,04658	0,8577768
[“Mall Plaza Vespucio”, “REGION METROPOLITAN”, “YOUNG”]	2,14614	85,71423	0,8540091
[“Mall Plaza Vespucio”, “YOUNG”, “CARD RECENT”]	2,14614	85,71423	0,8540091
[“Mall Plaza Vespucio”, “REGION METROPOLITAN”, “YOUNG”, “CARD RECENT”]	2,14614	85,71423	0,8540091
[“Mall Plaza Vespucio”, “YOUNG”, “CARD RECENT”, “CLIENT NORMAL”]	2,095043	85,41664	0,8510457
[CLERICAL WORKER, “REGION METROPOLITAN”, “YOUNG”, “CLIENT NORMAL”]	2,401637	85,45455	0,8509663
[“REGION METROPOLITAN”, “YOUNG”, “CARD RECENT”]	7,613704	86,12721	0,8498393
[“REGION METROPOLITAN”, “YOUNG”, “CLIENT NORMAL”]	7,409296	85,79871	0,846623
[“REGION METROPOLITAN”, “YOUNG”, “CARD RECENT”, “CLIENT NORMAL”]	7,409296	85,79871	0,846623
[STUDENT, “SEX FEMALE”, “REGION METROPOLITAN”, “YOUNG”]	2,299438	84,90572	0,8455048
[STUDENT, “SEX FEMALE”, “REGION METROPOLITAN”, “CLIENT NORMAL”]	2,401637	83,92862	0,8353314
[CLERICAL WORKER, “SEX FEMALE”, “REGION METROPOLITAN”, “YOUNG”]	1,584056	83,78374	0,8352273
[STUDENT, “REGION METROPOLITAN”, “CARD RECENT”, “CLIENT NORMAL”]	3,474707	83,95073	0,8337299
[STUDENT, “REGION METROPOLITAN”, “CARD RECENT”]	3,5769	83,33335	0,8271508
[STUDENT, “SEX FEMALE”, “REGION METROPOLITAN”, “CARD RECENT”]	2,503832	83,05081	0,8261554
[“SEX FEMALE”, “REGION METROPOLITAN”, “YOUNG”, “CARD RECENT”]	4,956565	82,90605	0,8201459
[“SEX FEMALE”, “REGION METROPOLITAN”, “YOUNG”, “CLIENT NORMAL”]	4,905464	82,75861	0,8186921
[“SEX MALE”, “YOUNG”, “CARD RECENT”]	4,138993	81,81816	0,8103313
[“SEX MALE”, “YOUNG”, “CARD RECENT”, “CLIENT NORMAL”]	3,985692	81,25	0,8047166
[STUDENT, “SEX MALE”, “YOUNG”, “CARD RECENT”]	1,53296	78,94743	0,7861967
[CLERICAL WORKER, “YOUNG”, “CARD RECENT”]	3,270315	79,01231	0,7830274
[STUDENT, “SEX FEMALE”, “YOUNG”, “CLIENT NORMAL”]	3,01482	78,66668	0,7800353
[BLUE-COLLAR, “REGION METROPOLITAN”, “CARD RECENT”, “CLIENT NORMAL”]	1,839552	78,26082	0,7785342
[STUDENT, “YOUNG”, “CARD RECENT”, “CLIENT NORMAL”]	4,547782	78,76106	0,7774913
[WHITE-COLLAR, “YOUNG”, “CARD RECENT”, “CLIENT NORMAL”]	3,065919	77,92207	0,7722377
[BLUE-COLLAR, “REGION METROPOLITAN”, “CARD RECENT”]	1,94175	77,55102	0,7710648
[STUDENT, “YOUNG”, “CARD RECENT”]	4,547782	78,07013	0,7702529
[STUDENT, “SEX FEMALE”, “YOUNG”, “CARD RECENT”]	3,01482	77,63162	0,7693629
[“YOUNG”, “CARD RECENT”, “CLIENT NORMAL”]	10,52631	78,3271	0,7577735
[WHITE COLLAR, “SEX FEMALE”, “REGION METROPOLITAN”, “CLIENT VIP”]	1,78845	76,08695	0,7565149
[WHITE-COLLAR, “SEX FEMALE”, “YOUNG”]	1,94175	75,99997	0,7552472
[WHITE-COLLAR, “SEX FEMALE”, “YOUNG”, “CARD RECENT”]	1,94175	75,99997	0,7552472
[STUDENT, “SEX FEMALE”, “CARD RECENT”, “CLIENT NORMAL”]	3,219216	75,90363	0,7510211
[WHITE-COLLAR, “SEX FEMALE”, “YOUNG”, “CLIENT NORMAL”]	1,890648	75,51016	0,7503822
[“SEX FEMALE”, “YOUNG”, “CARD RECENT”, “CLIENT NORMAL”]	6,540629	76,64672	0,7501237
[“SEX FEMALE”, “YOUNG”, “CARD RECENT”]	6,591716	75,8824	0,7418045
[STUDENT, “SEX FEMALE”, “CARD RECENT”]	3,32141	74,71264	0,7384389
[STUDENT, “YOUNG”, “CARD RECENT”, “CLIENT NORMAL”]	4,854365	74,80315	0,735176
[WHITE-COLLAR, CL_COMU = ‘M019’, “Mall Plaza Vespucio”, “CARD RECENT”]	1,584056	73,8095	0,7338796

4. Finally, the sharp increase in the complexity of commercial management as a result of globalization requires for there to be intelligent tools which integrally solve the problems of extracting knowledge from operational databases to support decision-making. For this, there must be intuitive and efficient interfaces such as those presented in this article. Fuzzy logic is presented as a valid current alternative for this purpose, particularly since the proposed methodology seeks to simplify the results, reduce the number of associations, using algorithms to discount the irrelevant ones and making the results more intuitive through the use of linguistic labels which the human expert finds more natural. We can therefore conclude that it is possible to provide a more comprehensive, proactive online solution to provide knowledge for commercial decision-making (e.g. as in the case of fraud prevention and detection) by extracting knowledge using fuzzy logic techniques which are applied to operational databases and other strategic decisions for the organization.

Table 5
Results for clients and transactions

Background	Support * 100	Confidence * 100	CF
[“PURCHASE AMOUNT LOW”, “SEX MALE”, “YOUNG”]	1,820765	80,4428	0,8008011
[“PURCHASE IN LOW NUMBER OF INSTALLMENTS”, “SEX MALE”, “YOUNG”]	2,138146	79,25697	0,7880376
[MEDIUM RISK BUSINESS TYPE, “SEX MALE”, “YOUNG”]	2,296832	78,79646	0,7829801
[“SEX MALE”, “YOUNG”, “RECENTLY ISSUED CREDIT CARD”]	2,296832	78,79646	0,7829801
[“LOW NUMBER OF DAILY PURCHASE STORES”, “SEX MALE”, “YOUNG”]	2,15485	77,71079	0,7721991
[“PURCHASE WITH HOLDER’S CARD”, “SEX MALE”, “YOUNG”]	1,812411	77,50008	0,7708476
[MEDIUM RISK PRODUCT LINE, “SEX MALE”, “YOUNG”]	1,503384	76,92307	0,7657085
[“MALL PLAZA TOBALABA”, “PURCHASE AMOUNT LOW”, “SEX FEMALE”]	1,578554	74,70354	0,7429782
[“MALL PLAZA TOBALABA”, MEDIUM RISK BUSINESS TYPE, “SEX FEMALE”]	1,712188	74,00716	0,7355437
[“LOW NUMBER OF DAILY PURCHASE STORES”, “MALL PLAZA TOBALABA”, “SEX FEMALE”]	1,687131	73,72259	0,7327164
[“MALL PLAZA TOBALABA”, “PURCHASE IN LOW NUMBER OF INSTALMENTS”, “SEX FEMALE”]	1,637018	73,40824	0,7296568
[“MALL PLAZA TOBALABA”, “PURCHASE WITH HOLDER’S CREDIT CARD”, MEDIUM RISK BUSINESS TYPE]	1,561847	73,33331	0,729102
[“LOW NUMBER OF DAILY PURCHASE STORES”, “MALL PLAZA TOBALBA”, “PURCHASE WITH HOLDER’S CREDIT CARD”]	1,528441	72,90835	0,7248784
[A SAME-DAY PURCHASE, PURCHASE IN LOW NUMBER OF INSTALLMENTS, “YOUNG”]	3,098637	72,46094	0,7158032
[A SAME-DAY PURCHASE, MEDIUM RISK BUSINESS TYPE, “YOUNG”]	3,240627	72,38803	0,7146326
[A SAME-DAY PURCHASE, “YOUNG”, “RECENTLY ISSUED CREDIT CARD”]	3,240627	72,38803	0,7146326
[A SAME-DAY PURCHASE, “LOW NUMBER OF DAILY PURCHASE STORES”, “YOUNG”]	3,106993	71,95358	0,7105423
[A SAME-DAY PURCHASE, PURCHASE, “PURCHASE AMOUNT LOW”, “YOUNG”]	2,714443	71,74386	0,7095546
[A SAME-DAY PURCHASE, “MEDIUM RISK PRODUCT LINE”, “YOUNG”]	1,987804	71,47143	0,7089284
[“MEDIUM RISK PRODUCT LINE”, MEDIUM RISK BUSINESS TYPE”, “YOUNG”]	4,042433	71,70367	0,7051163
[“MEDIUM RISK PRODUCT LINE”, “YOUNG”, “RECENTLY ISSUED CREDIT CARD”]	4,042433	71,70367	0,7051163
[PURCHASE IN LOW NUMBER OF INSTALLMENTS, “MEDIUM RISK PRODUCT LINE, “YOUNG”]	3,875381	71,60493	0,7046014
[PURCHASE IN LOW NUMBER OF INSTALLMENTS, “MEDIUM RISK BUSINESS TYPE”, “YOUNG”]	5,913305	71,87812	0,7011068
[PURCHASE IN LOW NUMBER OF INSTALLMENTS, “YOUNG”, “RECENTLY ISSUED CREDIT CARD”]	5,913305	71,87812	0,7011068
[“MEDIUM RISK BUSINESS TYPE, “YOUNG”, “RECENTLY ISSUED CREDIT CARD”]	6,22234	71,84186	0,6997351
[“LOW NUMBER OF DAILY PURCHASE STORES”, “MEDIUM RISK PRODUCT LINE, “YOUNG”]	3,833627	71,05257	0,698986
[“LOW NUMBER OF DAILY PURCHASE STORES”, PURCHASE IN LOW NUMBER OF INSTALLMENTS, “YOUNG”]	5,63769	71,42861	0,6972161
[“LOW NUMBER OF DAILY PURCHASE STORES”, “MEDIUM RISK BUSINESS TYPE, “YOUNG”]	5,946716	71,41421	0,6960681
[“LOW NUMBER OF DAILY PURCHASE STORES”, “YOUNG”, “RECENTLY ISSUED CREDIT CARD”]	5,946716	71,41421	0,6960681
[“PURCHASE AMOUNT LOW”, “MEDIUM RISK PRODUCT LINE, “YOUNG”]	3,098637	70,53225	0,6958996
[“PURCHASE WITH HOLDER’S CARD”, “MEDIUM RISK PRODUCT LINE, “YOUNG”]	3,390961	70,36395	0,6932373
[PURCHASE IN LOW NUMBER OF INSTALLMENTS, “PURCHASE AMOUNT LOW”, “YOUNG”]	4,63543	70,70058	0,692764
[“PURCHASE AMOUNT LOW”, “MEDIUM RISK PRODUCT LINE, “YOUNG”]	4,79412	70,68967	0,6921374
[“PURCHASE AMOUNT LOW”, “YOUNG”, “RECENTLY ISSUED CREDIT CARD”]	4,79412	70,68967	0,6921374
[“PURCHASE WITH HOLDER’S CARD”, PURCHASE IN LOW NUMBER OF INSTALLMENTS, “YOUNG”]	4,576968	70,61851	0,6920922
[““PURCHASE WITH HOLDER’S CREDIT CARD”, MEDIUM RISK BUSINESS TYPE, “YOUNG”]	4,835887	70,60973	0,6911622
[“PURCHASE WITH HOLDER’S CREDIT CARD”, “YOUNG”, “RECENTLY ISSUED CREDIT CARD”]	4,835887	70,60973	0,6911622
[“LOW NUMBER OF DAILY PURCHASE STORES”, “PURCHASE AMOUNT LOW”, “YOUNG”]	4,568619	70,21824	0,6879249
[“MALL PLAZA VESPUCIO”, “SEX FEMALE”, ADULT]	2,079678	69,16663	0,6851178
[“LOW NUMBER OF DAILY PURCHASE STORES”, “PURCHASE WITH HOLDER’S CREDIT CARD”, “YOUNG”]	4,610375	69,96197	0,6851016
[MEDIUM RISK PRODUCT LINE, “SEX FEMALE”, “YOUNG”]	2,539046	68,93422	0,681249
[“PURCHASE WITH HOLDER’S CARD”, “PURCHASE AMOUNT LOW”, “YOUNG”]	3,56636	68,64957	0,6749015
[“MALL PLAZA VESPUCIO”, “PURCHASE WITH HOLDER’S CARD”, “SEX FEMALE”]	3,357554	68,36736	0,6726837
[A SAME-DAY PURCHASE, “SEX FEMALE”, “YOUNG”]	1,93769	67,83629	0,6720074
[“PURCHASE IN LOW NUMBER OF INSTALLMENTS”, “SEX FEMALE”, “YOUNG”]	3,775163	68,27801	0,6703347
[“LOW NUMBER OF DAILY PURCHASE STORES”, “SEX FEMALE”, “YOUNG”]	3,791862	68,27061	0,6702005

Table 5 (*continued*)

Background	Support ∗ 100	Confidence ∗ 100	CF
[MEDIUM RISK BUSINESS TYPE, “SEX FEMALE”, “YOUNG”]	3,925495	68,31395	0,670193
[“SEX FEMALE”, “YOUNG”, “RECENTLY ISSUED CREDIT CARD”]	3,925495	68,31395	0,670193
[“MALL PLAZA VESPUCIO”, “PURCHASE AMOUNT LOW”, “SEX FEMALE”]	3,758455	68,07864	0,6683204
[“MALL PLAZA VESPUCIO”, “PURCHASE IN LOW NUMBER OF INSTALLMENTS”, “SEX FEMALE”]	4,009015	67,6056	0,6625266

5.2. Future work

Our future work consists of three stages. The first stage will be aimed at applying a similar procedure to the one presented in this article to the area of business management and planning. The second stage of this work consists in systematically incorporating this procedure into the strategic decision-making process, generating a methodology. The third stage envisages generating a methodology to enable this process to be systemized, allowing reactive decision-making processes to be transformed into proactive ones.

We are also considering integrating these methods and procedures into a methodology which incorporates up-to-date rules of current businesses into online transactional systems, which will enable operational risk to be reduced on one side, and profitability to be increased on the other, by improving corporate management and planning. The new online systems must be equipped with subsystems to maintain the validity of the association rules which keep the business rules valid, and therefore fuzzy logic techniques will be applied to generate fuzzy association rules from the knowledge extracted from the databases which incorporate linguistic labels and their membership degrees using the criteria reviewed in this article.

Also we plan

- To include data samples in the experiments from other areas which have not incorporated cutting edge technologies into their strategic decision-making processes (in particular clients and suppliers).
- To apply other data mining techniques to the sample set in order to carry out a comparative results analysis.
- To model an intelligent system which enables online systems to distinguish between those transactions which are considered to be within the normal ranges and those which are not using fuzzy association rules in order to provide risk and corporate analysts with sufficient intuitive information to enable them to reduce uncertainty in decision-making.
- To implement and maintain a fuzzy relational database which records the transactions and keeps the acquired knowledge valid by means of a process to extract knowledge about new conducts or behaviors.

5.3. Expected advantages of the future work

The future work proposed will allow the following advantages (among others):

- Intuitive results: the intuitive results shown in this publication are a clear contribution to the work of risk control analysts, and this is an immediate result of the inclusion of linguistic quantifiers which enable knowledge to be extracted in terms which are closer to the way human beings resolve their problems.
- Extraction of useful knowledge: the extraction of non-explicit knowledge from the transaction databases allows the association rule set to be updated and this helps to prevent and detect undesired activities with results which are easy for risk analysts to understand.
- Proactive work: the incorporation of linguistic labels and their membership degrees to the operational databases will provide the necessary data for detecting new, undesired behaviors and to update the online verification process about the progress of the corporate processes.

References

Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between set of items in large databases, In *Proceedings of the 1993 ACM SIGMOD conference* (pp. 207–216).

Au, W. H., & Chan, K. C. C. (1998). An effective algorithm for discovering fuzzy rules in relational databases. In *Proceedings of IEEE international conference on fuzzy systems* (Vol. 5, pp. 1314–1319).

Berzal, F., Blanco, I., Sánchez, D., & Vila, M. A. (2001). A new framework to assess association rules. In F. Hoffmann (Ed.), *Advances in intelligent data analysis. fourth international symposium. IDA'01. Lecture notes in computer science 2189* (pp. 95–104). Springer-Verlag.

Berzal, F., Blanco, I., Sánchez, D., & Vila, M. A. (2002). Measuring the accuracy and interest of association rules: A new framework. *Intelligent Data Analysis, 6*, 221–235.

Berzal, F., Cubero, J. C., Sánchez, D., Serrano, J. M., & Vila, M. A., (2003). Finding fuzzy approximate dependencies within STULONG data. In *Proceedings of the ECML/PKDD 2003, Workshop on Discovery Challenge* (pp. 34–46). 22–26 September.

Bhatla, T. P., Vikram, P., & Dua, A. (2003). Understanding Credit Card Frauds. *Cards Business Review 1*, Tata Consultancy Services.

Bra, P. D., & Paredaens, J. (1983). Horizontal decompositions for handling exceptions to functional dependencies. *Advances in Database Theory, 2*, 123–144.

Calero, J., Delgado, G., Sánchez-Marañon, M., Sanchez, D., Serrano, J. M., & Vila, M. A. (2004). An experience in management of imprecise soil databases by means of fuzzy association rules and fuzzy approximate dependencies. *ICEIS 2004*, Oporto (Portugal).

Chan, K. C. C., & Au, W.-H. (1997). Mining fuzzy association rules. *CIKM*, 209–215.

Chen, G., & Wei, Q. (2002). Fuzzy association rules and the extended mining algorithms. *Information Sciences, 147*(1–4), 201–228.

Cubero, J. C., Medina, J. M., Pons, O. & Vila, M. A. (1994). The Generalised Selection: An alternative way for the quotient operations in fuzzy relational databases. *IPMU'94* (pp. 489–495).

de Graaf, J. M., Kosters, W. A., & Witteman, J. J. W. (2001). Interesting fuzzy association rules in quantitative databases. *PKDD 2001* (pp. 140–151).

Delgado, M., Marín, N., Sánchez, D., & Vila, M. A. (2003). Fuzzy association rules: General model and applications. *IEEE Transactions on Fuzzy Systems, 11*(2), 214–225.

Delgado, M., Sánchez, D., & Vila, M. A. (2000). Fuzzy cardinality based evaluation of quantified sentences. *International Journal of Approximate Reasoning, 23*, 23–66.

Dubois, D., Hullermeier, E., & Prade, H. (2003). A note on quality measures for fuzzy association rules. *Fuzzy sets and systems – IFSA 2003, LNAI 2715* (pp. 346–353).

Frawley, William J., Piatetsky-Shapiro, Gregory, & Matheus, Christopher J. (1992). *Knowledge discovery in databases: An overview*. AAAI/MIT Press.

Gyenesei, A. (2001). Interestingness measures for fuzzy association rules. *PKDD 2001* (pp. 152–164).

Hong, T. P., Kuo, C. S., & Chi, S. C. (1999). Mining association rules from quantitative data. *Intelligent Data Analysis, 3*, 363–376.

Hullermeier, E. (2001). Implication-based fuzzy association rules. *PKDD 2001* (pp. 241–252).

Kaya, M., Alhajj, R., Polat, F., & Arslan, F. (2002). Efficient automated mining of fuzzy association rules. *DEXA 2002* (pp. 133–142).

Kivinen, J., & Mannila, H. (1995). Approximate dependency inference from relations. *Theoretical Computer Science, 149*(1), 129–149.

Kuok, C.-M., Fu, A., & Wong, M. H. (1998). Mining fuzzy association rules in databases. *SIGMOD Record, 27*(1), 41–46.

Luo, J., & Bridges, S. (2000). Mining fuzzy association rules and fuzzy frequency episodes for intrusion detection. *International Journal of Intelligent Systems, 15*(8), 687–703.

Pedrycz, W. (1998). Fuzzy set technology in knowledge discovery. *Fuzzy Sets and Systems, 98*, 279–290.

Pfahringer, B., & Kramer, S. (1995). Compression-based evaluation of partial determinations. In *Proceedings of the first international conference on knowledge discovery and data mining (KDD'95)* (pp. 234–239).

Sánchez, D. (1999). Adquisición de relaciones entre atributos en base de datos relacionales, Ph.D. Thesis.

Sánchez, D., Sánchez, J. R., Serrano, J. M., & Vila, M. A. (2004). Reglas de Asociación sobre Dominios Imprecisos aplicando Relaciones Difusas de Similitud. Actas del XII Congreso Español sobre Tecnologías y Lógica Fuzzy, ESTYLF 2004, Jaén (España), 15–17 de Septiembre del 2004 (pp.129–134).

Sánchez, D., Serrano, J.M., Blanco, I., Martín-Bautista, M.J., & Vila, M.A. (2008). Using association rules to mine for strong approximate dependencies. *Data Mining & Knowledge Discovery*, doi:10.1007/s10618-008-0092-3.

Serrano, J. M. (2003). Knowledge fusion in relational database: Aggregation and summarization measures, Ph.D. Thesis.

Shortliffe, E., & Buchanan, B. (1975). A model of inexact reasoning in medicine. *Mathematical Biosciences, 23*, 351–379.

Sponsored by CyberSource Corporation Conducted by Mindwave Research (2006). Seventh Annual ONLINE FRAUD REPORT Credit Card Fraud Trends and Merchants' Response (pp. 2–17).

Zadeh, L. (1965). Fuzzy sets. *Information and Control, 8*, 338–353.

Zadeh, L. A. (1971). Similarity relations and fuzzy orderings. *Information Sciences, 3*, 177–200.

Zadeh, L. A. (1983). A computational approach to fuzzy quantifiers in natural languages. *Computing and Mathematics with Applications, 9*(1), 149–184.

Zarufe, Rina M. (2005). Tiendas por Departamentos en Chile. Fitchratings (pp. 1–7).