Mathware & Soft Computing 15 (2008) 285-295

A Logic Approach for Exceptions and Anomalies in Association Rules

M. Delgado, M.D. Ruiz and D. Sánchez
Dept. Computer Science and A.I., E.T.S. Ingeniería Informática,
University of Granada, 18071 - Granada (Spain)
mdelgado@ugr.es, mdruiz@decsai.ugr.es, daniel@decsai.ugr.es

Abstract

Association rules have been used for obtaining information hidden in a database. Recent researches have pointed out that simple associations are insufficient for representing the diverse kinds of knowledge collected in a database. The use of exceptions and anomalies deal with a different type of knowledge sometimes more useful than simple associations. Moreover exceptions and anomalies provide a more comprehensive understanding of the information provided by a database.

This work intends to go deeper in the logic model studied in [5]. In the model, association rules can be viewed as general relations between two or more attributes quantified by means of a convenient quantifier. Using this formulation we establish the true semantics of the distinct kinds of knowledge we can find in the database hidden in the four folds of the contingency table. The model is also useful for providing some measures for assessing the validity of those kinds of rules.

1 Introduction

One of the main objectives of the data mining field is to obtain the hidden associations, between two or more items (i.e., couples $\langle attribute, value \rangle$), in databases. These items usually appear forming sets called *transactions*. When the presence of items tends to occur together in most of the transactions, we tend to say that they are related in some way, and we can symbolize it by an *association rule*, for example "most of transactions that contain bread also contain butter", and it is usually noted $bread \rightarrow butter$. The intensity of the above association rule is measured most frequently by the *support* and the *confidence* measures.

Recent approaches are based on obtaining different kinds of knowledge, referred to as peculiarities, infrequent rules, exceptions or anomalous rules. The knowledge captured by these new types of rules is in many cases more useful than that obtained by simple association rules. They have several advantages. They provide a more comprehensive understanding of the information hidden in a database. Moreover, they are less numerous than simple association rules, and hence, the information obtained is more manageable.

This paper is an attempt to go deeper in the comprehension of the logical model first introduced by Hájek et al. in [10], and then developed in [5]. The model uses the simple notions of contingency table and quantifier. The contingency table called *four fold table* collects all the information about two chosen itemsets from a database, and the quantifier is a mathematical object that unifies two types of information: (1) it measures in some sense the relation between the two attributes and (2) it also says if the measure satisfies some predefined thresholds.

This model can manage more information than simple association rules. In [5] we saw how the use of different quantifiers is useful for dealing with more kinds of associations: implicational, double implicational, equivalence, etc. It is also helpful in order to generalize the notion of rule to many kinds of situations (see for example [15], [9], [12], [13] and [5]).

Our intention is to provide the true semantics of these kinds of knowledge using the model developed in [5]. In that paper we emphasize the use of the contrapositive of the rule for obtaining very strong rules. But here we center our attention in the frequencies provided when the items are not in the transactions. Some authors call them negative examples of a rule. But we need the specification of a third attribute (item) into the rule for explaining the semantics of exceptions and anomalies taking as reference the common sense rule fixed. We also use the known quantifier of founded implication (see [5]) for assessing the validity of exceptions and anomalies. Then we discuss about the frequencies used for classifying the distinct semantics of the exception and anomaly rules.

The paper follows with a brief review of definitions of peculiarity, exception and anomaly rules. Section 3 presents the logic model for association rules. Next section analyzes the semantics of exceptions and anomalies using the logic model. We conclude with a brief discussion about the contribution of the paper and we point out some interesting lines for future research.

2 Motivation and Related Work

Association rules have proved to be a practical tool in order to find associations in databases, and they have been extensively applied in many areas. Despite their proven applicability, association rules have serious drawbacks limiting their effective use. One of the main drawbacks stems from the large number of rules obtained even from small-sized databases. This disadvantage is a direct consequence of the type of knowledge the association rules try to extract, i.e, frequent and confident rules. Perhaps in some application domains the expert is interested in finding other kinds of knowledge. But the crucial problem here, is to determine the kind of knowledge useful in every context.

Many papers have addressed this task and proposed novel and useful kinds of knowledge that might be of interest for users. There are also subjective and objective approaches in order to obtain the knowledge of interest in every case. In

[7] there are a good survey about most of approaches performed until now. We are interested in the objective approaches. In this way we distinguish between different definitions of exception rules, peculiarity rules, and anomalous rules.

2.1 Peculiarity Rules

A peculiarity rule is discovered from the data by searching the relevance among the peculiar data [23]. Roughly speaking, a data is peculiar if it represents a peculiar case described by a relatively small number of objects and if it is very different from other objects in a data set. Although it looks like the exception rule from the viewpoint of describing a relatively small number of objects, the peculiarity rule represents the association between the peculiar data.

In order to discover the rules, the first step is searching the peculiar data in the relation. There are many ways of finding the peculiar data. In [23] the authors describe an attribute-oriented method. Let $X = \{x_1, x_2, \dots, x_n\}$ a data set related to an attribute in a relation, and n the number of different values in an attribute. The peculiarity of x_i can be evaluated by the Peculiarity Factor, $PF(x_i)$,

$$PF(x_i) = \sum_{j=1}^{n} \sqrt{CD(x_i, x_j)}$$

It evaluates whether x_i occurs relatively small number and is very different from other data x_j by computing the sum of the square root of the conceptual distance, CD, between x_i and x_j. The CD function can be given by $\|x_i - x_j\|$ or it can be obtained basing on the background knowledge specified by the user. After the evaluation for the peculiarity, the peculiar data are extracted by using a threshold value which depends on the peculiarity factor.

Once the peculiarity data is extracted, the process for discovering peculiarity rules is done using the peculiar data by searching the relevance (see [23]) among the peculiar data.

In [16] the authors generalize the definition of peculiarity measure in order to describe a record-oriented method for relational databases.

2.2 Exception Rules

There are many approaches about mining exception rules in databases. A general form of an exception rule is introduced in [14] as the table shows:

$A \to X$	Common sense rule (strong rule)	(high supp. and high conf.)
$A \wedge B \to \neg X$	Exception rule (weak rule)	(low supp. and high conf.)
$B \to \neg X$	Reference rule (not strong rule)	(low supp. or low conf.)

In fact the only requirement for the exception rule is to be a confident rule because the conditions imposed to the common sense rule and to the reference rule restrict it to have low support.

Depending on the translation of the third condition we could manage with different kinds of exception rules. In general terms, the kind of knowledge these exceptions try to capture can be interpreted as follows: A strongly implies X (and not B); but in conjunction with B, A imply $\neg X$ (or other item E in contraposition with X called the *exception*).

In particular, Hussain et al. in [14] take the reference rule as $B \rightarrow \neg X$ with low support or low confidence. But in fact to obtain this they impose that the rule $B \rightarrow X$ is a strong one, i.e. it has high support and confidence.

Suzuki et al. work in [19], [20], [21] and also in other papers, with a more restrictive approach. They consider the rule-exception pair

$$A \rightarrow x \qquad \text{(common sense rule)}$$
$$A \wedge B \rightarrow x' \qquad \text{(exception rule)}$$

where x and x' are items with the same attribute but with different values and $A = a_1 \wedge \ldots \wedge a_p$, $B = b_1 \wedge \ldots \wedge b_q$ are conjunctions of items.

They also propose several ways for measuring the degree of interestingness of a rule-exception pair. In [19], [21] they use an information-based measure to determine the interestingness of the above pair of rules. But they also impose the constraint of not to be confident (low confidence) to the reference rule $B \rightarrow x'$.

In other approaches [20], [22], Suzuki et al. use other criteria to evaluate the interestingness of the pair, and they also proposed a unified algorithm for dealing with many kinds of exception rule pairs.

2.3 Anomalous Rules

An *anomalous* association rule is an association rule that comes to the surface when we eliminate the dominant effect produced by a strong rule. In other words, it is an association rule that is verified when a common rule fails [1].

A formally definition of anomalous rule can be found in [1], and we present an scheme of it.

$A \rightarrow B$	Common sense rule (strong rule)	(high supp. and high conf.)
$A \wedge \neg B \rightarrow X$	Anomalous rule (confident rule)	(high conf.)
$A \wedge B \rightarrow \neg X$	Reference rule (confident rule)	(high conf.)

The semantics this kind of rules tries to capture is: A strongly implies B, but in those cases where we do not obtain B, then A confidently implies X, or in other words: when A, then we have either B (usually) or X (unusually).

It should be noted that anomalous rules have different semantic than exception rules; and the more confident the rules $A \wedge \neg B \rightarrow X$ and $A \wedge B \rightarrow \neg X$ are, the stronger the anomaly is.

3 A Logic Model for Association Rules

The logic model we present is based on a method developed in middle sixties by Hayek et al. [10]. The method is called GUHA (General Unary Hypotheses

Automaton) which has good logical and statistic foundations that help to a major understanding of both the nature of association rules and the basic properties of measures used for assessing the accomplishment of them.

The starting point is a data matrix M where the rows $O_1, \ldots, O_n$ are associated to the observed objects whereas the columns $A_1, \ldots, A_K$ are associated to the attributes (cualitative or categorial) which describe the objects. The entry (i, j) of M will be equal to 1 when the object O_i presents the attribute A_j and 0 otherwise.

For the rules association mining framework each matrix M represents a transaction and we represent all the matrices (transactions) into a so called database D (see table 1).

D	$< O_1, A_1 >$	$\ldots$	$< O_1, A_K >$	$< O_2, A_1 >$	$\ldots$	$< O_n, A_K >$
t_1	1	$\ldots$	0	0	$\ldots$	0
t_2	0	$\ldots$	1	0	$\ldots$	1
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
t_n	1	$\ldots$	0	1	$\ldots$	1

Table 1: Database D obtained from matrices like M.

To the logic method we are presenting, an attribute, denoted φ, will be a pair $\varphi =< O_1, A_1 >$ or more generally an aggregation of "atomic" attributes by means of the logic connectives $\wedge, \vee, \neg$. An example of attribute is $< O_1, A_1 > \wedge < O_3, A_2 >$ and $< O_1, A_1 > \wedge < O_3, A_2 > \approx < O_2, A_5 > \wedge < O_3, A_7 >$ is an example of association rule.

An association rule in the model proposed by [11] and after developed in [5] is an expression of the type $\varphi \approx \psi$ where φ and ψ are attributes (in the sense before) derived from matrix D, and the symbol $\approx$, called *quantifier*, is some assessment or condition depending on the measure used for the assessment of the rule. The intuitive meaning of the expression $\varphi \approx \psi$ is that attributes φ and ψ are associated in D with the strength that the quantifier $\approx$ marks.

For any pair of attributes φ and ψ the so called *four fold table* may be constructed from the database D as follow. Graphically

$\mathcal{M}$	ψ	$\neg\psi$
φ	a	b
$\neg\varphi$	c	d

The four fold table will be denoted by $\mathcal{M} = 4ft(\varphi, \psi, D) = \langle a, b, c, d \rangle$ where a, b, c and d will be non-negative integers such that a is the number of objects (i.e. the rows of D) containing both φ and ψ, b the number of objects having φ and not ψ, and analogously for c and d; obviously $a + b + c + d > 0$.

When the assessment underlying $\approx$ is made from a four fold table then $\approx$ is said to be a $4ft$-quantifier involved in the rule $\varphi \approx \psi$.

The association rule $\varphi \approx \psi$ is said to be true in the analyzed database D (or in the matrix $\mathcal{M}$) if and only if the condition associated to the 4ft-quantifier $\approx$ is satisfied for the four fold table $4ft(\varphi, \psi, D)$. We will write $Val(\varphi \approx \psi) = 1$, when the rule $\varphi \approx \psi$ is true in D, and $Val(\varphi \approx \psi) = 0$ otherwise.

Different kinds of association of the attributes φ and ψ can be expressed by suitable 4ft-quantifiers. We can find many examples of 4ft-quantifiers in [8], [9], [11], [12] and [17].

The classical framework of support-confidence for assessing association rules can be expressed by the founded implication quantifier [5] $\Rightarrow_{minconf,minsupp}$

$$\frac{a}{a+b} \geq minconf \wedge \frac{a}{n} \geq minsupp \tag{1}$$

where $0 < minconf, minsupp < 1$ denote the thresholds known as minimum confidence and minimum support respectively, and $n = a + b + c + d$ is the total number of transactions in the database D. In [5] we explain deeply the relation of quantifiers with the interest measures used for assessing the validity of rules.

4 Exceptions and Anomalies Analysis through the Logic Model

This section is devoted to present a deep analysis in the semantics and representation of the frameworks for extracting exceptions and anomalies from a database using the above logic model. For this we consider the three itemsets involved in the formulation of exceptions and anomalies. Let be A, B and X three itemsets (or attributes in the logic model) in a database D. We consider their frequencies of appearance in D as we show in the four fold table 2, where e is the number of

	B	$\neg B$	
A	$e + f$	$g + h$	
$\neg A$	$i + j$	$k + l$	
	$X \vert \neg X$	$X \vert \neg X$	n

Table 2: Four fold table for three attributes.

transactions in D satisfying A, B and X; f the number of transactions satisfying A, B and $\neg X$ and so on. The sums of these frequencies correspond to the a, b, c and d frequencies seen in section before, i.e. $a = e + f$, $b = g + h$, $c = i + j$ and $d = k + l$. We also use n for the total number of transactions of the database D.

The common part for the exceptions and anomalies extraction is the acceptance of the fulfilment of the confident rule $A \wedge B \to \neg X$, that is, the exception rule is the reference rule for the anomalies framework.

For assessing its validity, it can be expressed by the implication quantifier [5] $\Rightarrow_{minconf}$ using the frequencies in the $4ft$ seen in table 2:

$$\frac{i}{e+i} \geq minconf \tag{2}$$

where $0 < minconf < 1$ denotes the threshold known as minimum confidence.

For the exception, anomaly or reference rules we distinguish different approaches. Using again the classical support-confidence framework in each case we obtain the following results:

- **Exception for Suzuki et al.** For dealing with the exceptions in the approach of Suzuki et al., we impose the conditions for the common sense, exception and reference rules:

$$A \rightarrow X \quad \text{strong rule} \rightsquigarrow \frac{e+g}{n} \geq minsupp \wedge \frac{e+g}{e+f+g+h} \geq minconf$$

$$A \wedge B \rightarrow \neg X \quad \text{confident} \rightsquigarrow \frac{i}{e+i} \geq minconf$$

$$B \rightarrow \neg X \quad \text{not confident} \rightsquigarrow \frac{f+j}{e+f+i+j} \leq minconf$$

  where $0 < minsupp, minconf < 1$ are the minimum confidence and minimum support thresholds, and $n = e+f+g+h+i+j+k+l$ is the total number of transactions in the database.

- **Exception for Hussain et al.** For dealing with the exceptions in the approach of Hussain et al., we impose the conditions for the common sense, exception and reference rules:

$$A \rightarrow X \quad \text{strong rule} \rightsquigarrow \frac{e+g}{n} \geq minsupp \wedge \frac{e+g}{e+f+g+h} \geq minconf$$

$$A \wedge B \rightarrow \neg X \quad \text{confident} \rightsquigarrow \frac{i}{e+i} \geq minconf$$

$$B \rightarrow X \quad \text{strong rule} \rightsquigarrow \frac{e+i}{n} \geq minsupp \wedge \frac{e+i}{e+f+i+j} \geq minconf$$

  where $0 < minsupp, minconf < 1$ denote the thresholds known as minimum confidence and minimum support respectively, and $n = e+f+g+h+i+j+k+l$ is the total number of transactions in the database.

- **Anomaly for Berzal et al.** Dealing with anomalies is quite different from extracting exceptions. The first difference we find is the common sense rule taken into account. Moreover, the anomaly rule and the reference rule differ from the case of the exception framework. In this case, the conditions for the common sense, anomaly and reference rules are:

$$A \to B \quad \text{strong rule} \rightsquigarrow \frac{e+f}{n} \geq minsupp \wedge \frac{e+f}{e+f+g+h} \geq minconf$$

$$A \wedge \neg B \to X \quad \text{confident} \rightsquigarrow \frac{g}{g+h} \geq minconf$$

$$A \wedge B \to \neg X \quad \text{confident} \rightsquigarrow \frac{i}{e+i} \geq minconf$$

where $0 < minsupp, minconf < 1$ denote the minimum confidence and minimum support respectively, and n is the total number of transactions in the database.

If we go deeper in these reformulations of the conditions for each framework, we find the following facts:

- For the common sense rule in each case we need the frequencies we have noted by $a = e + f$ and $b = g + h$. Some authors [6], [18], call them positive examples (a) and counterexamples (b), and they are useful for example in the generalization of the support-confidence framework to the fuzzy case [6] and for explaining their meaning [18]. In [2] the authors refer to a and b as positive and negative examples respectively. They use them for removing the "mystery" of some suggested quality measures for fuzzy association rules.
- When dealing with exceptions besides using the frequencies a and b for the common sense rule, the two existent approaches need the value collected by $c = i + j$ join to the a for assessing the validity of the exceptions. Authors in [2], [6] and [18] doesn't take into account this value separately of d. But we make sense of this fact, and we distinguish between c and d because they are used for different tasks. Here we note that c is essential for extracting exceptions, and in [5] we remark that d is essential for the task of extracting very strong rules[1] [3].
- The anomalies however, center their attention in the frequencies collected by b and a, the frequencies known as negative and positive examples. But this is just the semantic we want to collect with the use of anomalies: 'when A, then we have either B (positive examples) or X (negative examples)'.

In summary, we have observed that depending on the frequencies in the four fold table we use, the different kinds of information collected are relevant for obtaining an appropriate 'resume' of the information hidden in the database.

5 Conclusions and Future Research

The search of a logic for dealing with exceptions or default values has received a lot of attention during many years. Several proposals were introduced as forms of default reasoning for building a complete and consistent knowledge base.

[1] An association rule $\varphi \to \psi$ is very strong if both $\varphi \to \psi$ and $\neg\psi \to \neg\varphi$ are strong.

The main topic here is how to deal with all the information that a database provides the easier as possible and how to distinguish between different kinds of knowledge. It is clear enough that the logic model developed in section 3 manage all the information that can be extracted from a database. There are not more information than the frequencies of the four possibilities between two items. When a third item is involved, the frequencies are decomposed each one into two quantities. But depending of the functionality of this third item, we deal with exceptions or with anomalies. We explain deeper this affirmation.

In last section we fixed the exception and anomaly rule for explaining their semantics through the logic model. Fixing the common sense rule we obtain a different vision. Starting with the same common sense rule, $\phi \rightarrow \psi$, an exception rule is defined as follows:

$\phi \rightarrow \psi$	Common sense rule (strong rule)	(high supp. and high conf.)
$\phi \wedge E \rightarrow \neg\psi$	Exception rule (weak rule)	(low supp. and high conf.)

where the third item E is called the exception. The role of E is that of an agent which interferes in the usual behavior of the common sense rule. We remove the reference rule because it does not apport meaning to the exception (see [4] for details). An example of exception rule will be: "with the help of *antibiotics*, the patient usually tends to *recover*, unless *staphylococci* appear", in such a case, antibiotics combined with staphylococci don't lead to recovery, even sometimes may lead to death.

For anomalies, if the common sense rule $\phi \rightarrow \psi$ is fixed, we can define them as:

$\phi \rightarrow \psi$	Common sense rule (strong rule)	(high supp. and high conf.)
$\phi \wedge \neg\psi \rightarrow A$	Anomalous rule (confident rule)	(high conf.)

where the third item A is called the anomalous behavior. The role of A is that of explaining the homogeneous deviation from the common sense rule (i.e. from the usual behavior). A is not an agent as E, but it is the alternative behavior when the usual fails. An example of anomalous rule will be: "if a patient have symptoms ϕ then he usually has the disease ψ or he has the disease A".

The knowledge provided by these two kinds of rules are complementary. If we are interested in the agent of the "strange" behavior, we will look for the exceptions. And if we are interested in what is the strange or unusual behavior, we will look for the anomalies.

We have observed that depending on the frequencies in the four fold table we use, the different kinds of information collected are relevant for obtaining an appropriate knowledge for users.

For future work we are interested in the study of new measures that better fit to the exceptional or anomalous knowledge we want to extract from the data set. We also plan to study an algorithm for extracting exceptions and anomalies together with their common sense rule associated.

6 Acknowledgements

We would like to acknowledge support for this work from the Dirección General de Investigación of Spain by the project grant TIN2006-15041-C04-01.

References

[1] F. Berzal, J.C. Cubero, N. Marín, and M. Gámez. Anomalous association rules. In *IEEE International Conference on Data Mining*, 2004.

[2] M. De Cock, C. Cornelis, and E.E. Kerre. Elicitation of fuzzy association rules from positive and negative examples. *Fuzzy Sets and Systems*, 149(1):73–85, 2005.

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

[4] M. Delgado, M.D. Ruiz, and D. Sánchez. Analyzing exception rules. In *Proceedings of IPMU'08, Málaga*, pages 433–440, 2008.

[5] M. Delgado, M.D. Ruiz, and D. Sánchez. Studying interest measures for association rules through a logical model. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Submitted, 2007.

[6] D. Dubois, E. Hüllermeier, and H. Prade. A note on quality measures for fuzzy association rules. In *IFSA*, pages 346–353, 2003.

[7] B. Duval, A. Salleb, and C. Vrain. On the discovery of exception rules: A survey. *Studies in Computational Intelligence*, 43:77–98, 2007.

[8] T. Havránek. Statistical quantifiers in observational calculi: An application in guha-methods. *Theory and Decision*, 6:213–230, 1974.

[9] P. Hájek. On generalized quantifiers, finite sets and data mining. *Intelligent Information Systems*, pages 489–496, 2003.

[10] P. Hájek, I. Havel, and M. Chytil. The guha method of automatic hypotheses determination. *Computing*, 1:293–308, 1966.

[11] P. Hájek and T. Havránek. *Mechanising Hypothesis Formation-Mathematical Foundations for a General Theory.* Springer-Verlag: Berlin, Heidelberg and New York, 1978.

[12] P. Hájek, J. Rauch, D. Coufal, and T. Feglar. The guha method, data preprocessing and mining. *LNAI*, 2682:135–153, 2004.

[13] M. Holeňa. Fuzzy hypotheses for guha implications. *Fuzzy Sets and Systems 98*, pages 101–125, 1998.

[14] F. Hussain, H. Liu, E. Suzuki, and H. Lu. Exception rule mining with a relative interestingness measure. In *Pacific Asia Conference on Knowledge Discovery and Data Mining*, pages 86–97, 2000.

[15] T. Karban, J. Rauch, and M. Šimunek. Sds-rules and association rules. In *ACM Symposium on Applied Computing*, pages 520–524, Nicosia, Cyprus, 2004.

[16] M. Ohshima, N. Zhong, Y.Y. Yao, and C. Liu. Relational peculiarity-oriented mining. *Data Mining and Knowledge Discovery*, 15(2):249–273, 2007.

[17] J. Rauch. Logic of association rules. *Applied Intelligence*, 22:9–28, 2005.

[18] T. Sudkamp. Examples, counterexamples, and measuring fuzzy associations. *Fuzzy Sets and Systems*, 149(1):57–71, 2005.

[19] E. Suzuki. Discovering unexpected exceptions: A stochastic approach. In *Proceedings of the fourth international workshop on RSFD*, pages 225–232, 1996.

[20] E. Suzuki. Undirected discovery of interesting exception rules. *International Journal of Pattern Recognition and Artificial Intelligence*, 16(8):1065–1086, 2002.

[21] E. Suzuki and M. Shimura. Exceptional knowledge discovery in databases based on information theory. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pages 275–278. AAAI Press, 1996.

[22] E. Suzuki and J.M. Zytkow. Unified algorithm for undirected discovery of exception rules. *International Journal of Intelligent Systems*, 20:673–691, 2005.

[23] N. Zhong, M. Ohshima, and S. Ohsuga. Peculiarity oriented mining and its application for knowledge discovery in amino-acid data. In *Proceedings of the 5th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 260–269, London, UK, 2001.