Analyzing Exception Rules

M. Delgado, M.D. Ruiz, D. Sánchez
Dept. Computer Science and Artificial Intelligence, University of Granada,
C/Periodista Daniel Saucedo Aranda s/n , 18071 Granada, Spain
mdelgado@ugr.es, {mdruiz, daniel}@decsai.ugr.es

Abstract

Recent researches have pointed out that associations rules are insufficient for representing the diverse kinds of knowledge collected in a database. Exception rules deal with a different type of knowledge sometimes more useful than simple associations. Moreover exception rules (and other kinds of rules) provide a more comprehensive understanding of the information provided by a database.

This work intends to go deeper in the logic model studied in [3]. 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 and the real formulation of exception rules. We also study the behavior of exceptions when we deal with double rules.

Keywords: Data mining, association rules, exception rules, double rules, logic model.

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 [5], and then developed in [3]. 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.

L. Magdalena, M. Ojeda-Aciego, J.L. Verdegay (eds): Proceedings of IPMU'08, pp. 433–440
Torremolinos (Málaga), June 22–27, 2008

This model can manage more information than simple association rules. In [3] 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 the references given in [3]).

Our intention is to provide a deep analysis of the formulation and the semantics of exceptions using the model developed in [3]. In [2] we used the logic model for studying some semantical aspects of exceptions and anomalies in association rule mining. Here, we discuss about the different definitions that exception rules have received, analyzing them semantically and unifying them in a single definition. We also present the definition of a double strong rule and the motivation for mining exceptions in these type of rules.

This paper begins with a brief review of definitions of peculiarity, exception and anomalous rules. Section 3 presents the logic approach for association rules. Next section analyzes the definitions proposed until now of exceptions. Then we follow studying exception rules using the logic model. Last section is devoted to present double strong rules and their associated exceptions. 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 disadvantages 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 [4] there are a good survey with most of approaches presented until now. We are interested in the objective approaches. In this way we can distinguish some approaches for finding new kinds of knowledge: peculiarity, exception and anomalous rules.

A *peculiarity rule* is discovered from the data by searching the relevance among the peculiar data [14]. Roughly speaking, a data is peculiar if it represents a peculiar case described by a relatively small number of objects and it is very different from other objects in a data set. Peculiarity rules represent the associations between the peculiar data. In [14] the authors describe an attribute-oriented method for extracting peculiarity rules.

Looking for *exception rules* consists of finding an attribute which interacting with another may change the consequent in a strong association rule [12], [7], [13]. In general terms, the kind of knowledge the exception rules try to capture can be interpreted as follows:

X strongly implies Y,
but, in conjunction with E,
X confidently does not imply Y.

Anomalous rules are in appearance similar to exception rules, but semantically different. 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 formal definition of anomalous rule can be found in [1]. The semantics this kind of rules tries to capture is:

X strongly implies Y,
but, in those cases where X does not imply Y,
then X confidently implies A,

or in other words: when X, then we have either Y (usually) or A (unusually).

The knowledge provided by the exception and the anomalous rules are (semantically) complementary. If we are interested in the agent of the "strange" behavior, we will look for the exceptions, and if we are interested however in what is the strange or unusual behavior, we will look for the anomalies.

We are going to center our attention in studying the distinct formulations of exception rules, analyzing their semantics, and then we will study them using the logic model developed in [3].

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. [5]. 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. Recently, some authors have implemented a good and fast algorithm [9] based on a bit string approach to mining association rules. This algorithm can be used to mining exception rules (with some modifications) by means of their representation in this logic model.

The starting point is a data matrix M where the rows $O_1, \dots, O_n$ are associated to the observed objects whereas the columns $A_1, \dots, A_K$ 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 rule 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). 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

Table 1: Database D obtained from matrices like M.

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

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 [6] 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, and it will depend on the *four fold table* associated to the pair of attributes φ and ψ.

The so called 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 $a+b+c+d > 0$. Graphically

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

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 $4ft(\varphi, \psi, D)$.

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 [6] and [8]. The classical framework of support-confidence for assessing association rules can be expressed by the founded implication quantifier [3]

$\Rightarrow_{minconf,minsupp}$, as follows

$$\frac{a}{a+b} \geq minconf \wedge \frac{a}{n} \geq minsupp \qquad (1)$$

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 [3] we explain deeply the relation of quantifiers with the interest measures used for assessing the validity of rules.

4 Exception Rules

There are many approaches about mining exception rules in databases. A general form of an exception rule is introduced in [7] as follows:

$X \to Y$	Common sense rule (strong)
$X \wedge E \to \neg Y$	Exception rule (confident)
$E \to \neg Y$	Reference rule (not strong)

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 called *reference* rule, we could manage with different kinds of exception rules.

4.1 Analyzing Semantics of Exception Rules

In particular, Hussain et al. in [7] take the reference rule as $E \to \neg Y$ with low support or low confidence. But in fact to obtain this they impose that the rule $E \to Y$ is a strong one, i.e. it has high support and confidence.

If we take into consideration this formulation of exception:

$X \to Y$ and $E \to Y$	Strong rules
$X \wedge E \to \neg Y$	Confident rule

We see that Hussain et al. present a restrictive definition of exception. They impose a *double meaning*. The former is the original semantic of exception rule associated to the common sense rule $X \to Y$, and the second is that X is also an exception to the common sense rule $E \to Y$.

Suzuki et al. work in [12], [11] and also in other papers, with a distinct approach. They consider the rule-exception pair

$$X \to y \qquad \text{(common sense rule)}$$
$$X \wedge E \to y' \qquad \text{(exception rule)}$$

where y and y' are items with the same attribute but with different values and $X = x_1 \wedge \ldots \wedge x_p$, $E = e_1 \wedge \ldots \wedge e_q$ are conjunctions of items. They also propose several ways for measuring the degree of interestingness of a rule-exception pair. In [12] 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 to the reference rule $E \to y'$.

Centering our attention in this formulation,

$X \to Y$	Strong rule
$X \wedge E \to \neg Y$	Confident rule
$E \to \neg Y$	Not confident

Suzuki et al. pretend to impose an exception in some sense objectively novel. But in consonance to the general semantics of exception rules it should not impose a restriction outside the dominance of the common sense rule. This choosing for the reference rule expresses that the major percentage of transactions which contain E and $\neg Y$ also contain X, because if it does not occur, $E \to \neg Y$ will be confident. Therefore, this approach intends to give importance to the fact that E appears almost always when dealing with X and $\neg Y$.

In [1] the authors pretend to collect the whole meaning of exceptions proposing this alternative approach:

$X \to Y$	Strong rule
$X \wedge E \to \neg Y$	Confident rule
$X \to E$	Not strong

But the third condition is not necessary. If the opposite occurs, i.e. $X \to E$ is a strong rule. Since $X \to Y$ is also a strong rule these two conditions lead to $X \wedge E \to \neg Y$ is not a confident rule because both E and Y are frequent in those transactions which contain X. This argument says that the third condition is not necessary.

The role of E in the exception rule is that of an agent which interferes in the usual behavior of the common sense rule. 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. Following this example the possible reference rules we have used are:

$E \to Y$ Strong	A patient with staphylococci tends to recover
$E \to \neg Y$ Not confident	In general, staphylococci may not lead to the recovering of the patient (low conf.)
$X \to E$ Not strong	The use of antibiotics in a patient does not imply the appearance of staphylococci

The approaches given by Hussain and Suzuki (first and second in the table) does not give any reasonably semantics to the reference rule, nevertheless in the Berzal et al's approach (last in the table) the reference rule says that antibiotics and staphylococci doesn't have a direct relation. This example shows that the reference rule in all these cases does not contribute to the semantics of exceptions even sometimes give a contradictory information. Therefore we only take the two first rules (common sense and exception rules) for defining exception rules in the following.

5 A General Vision of Exception Rules through the Logic Model

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.

The main advantage of dealing with exception rules is that of collecting useful information hidden in data. Besides, this kind of knowledge is very specific and the information encountered is expressed in a small set of rules. In [2] we present a first approach for explaining the true semantics of exception and anomalous rules using the logic model in the previous section. Moreover the logic model provides an easy way to manage all the information that can be extracted from a database.

Exceptions are defined for finding the agent that causes an exceptional behavior in a predefined strong rule. In this context, the strong rule is called common sense rule (as in the before sections). Following the approach given in [7], [10]-[13], and fixing the strong rule, $X \to Y$, an *exception rule* is a rule of the form $X \wedge E \to \neg Y$ which fulfils the following conditions:

$X \to Y$	Strong rule
$X \wedge E \to \neg Y$	Confident rule

where the third item E is called the *exception* associated to the common sense rule $X \to Y$. The role of E is that of an agent which interferes in the usual behavior of the common sense rule. We eliminate the reference rule because it does not contribute a meaning to the exception, although it intensifies the strength of the exception.

This section is devoted to offer a new representation for the concept of exception using the previous logic approach. For this we consider the three itemsets involved in the formulation of exceptions. Let be X, Y and E three itemsets (or attributes in the logic model) in a database D. We consider the frequency of appearance of E and Y when in the database D_X, where D_X are those transactions in D which contain the itemset X. The four fold table ($4ft$) associated is in next table:

D_X	Y	$\neg Y$	
E	e	f	
$\neg E$	g	h	
			$a+b$

where e is the number of transactions in D_X satisfying Y and E, and so on. The sums of these frequencies correspond to the a and b frequencies seen in the previous section, i.e. $a = e + g$, $b = f + h$. We also use n for the number of transactions in D.

Using the predefined $4ft$-quantifiers for association rules we can adapt them in the particular case of mining exception rules. We will denote by E-quantifier the quantifier associated to the three attributes X, Y and E which involves only the four frequencies e, f, g, h and the total number of transactions n.

We present the founded implication E-quantifier, $\Rightarrow^{E}_{minsupp,minconf}$, defined by

$$\frac{e+g}{n} \geq minsupp \wedge$$
$$\frac{e+g}{e+f+g+h} \geq minconf \wedge \frac{f}{e+f} \geq minconf$$

It should be noted that the last condition proceed from imposing that $X \wedge E \rightarrow \neg Y$ is a confident rule.

The algorithm presented in [9] can be adapted for mining exception rules using the founded implication E-quantifier previously defined.

6 Double Rules

When mining rules, sometimes we may find that two itemsets are very related and there is no difference about the direction of that relationship. We can take advantage of this situation for extracting a new kind of knowledge from the database. We propose a toy example for elucidating a prototypical ambient where this kind of "bidirectional" rules are useful.

Example 1. Imagine we have a database which collects information about the vertebrate animals and their characteristics in a national park. One of the strong rules we can extract is:

if the animal flies, it is a bird.

But this rule can also be extracted in the other direction, i.e. the rule

if the animal is a bird, it flies

is also a strong rule.

In this context, we need a new kind of rules which collects this new type of knowledge.

Definition 1. An association rule $X \rightarrow Y$ is *double strong* if both $X \rightarrow Y$ and $Y \rightarrow X$ are strong.

Here and subsequently the double directional arrow $X \leftrightarrow Y$ will denote double strong rules. We also consider that $X \rightarrow Y$ has more or equal confidence than $Y \rightarrow X$. According to this definition, we propose the analogous definition for a 4ft-quantifier.

Definition 2. A quantifier $\approx$ is called *double strong* if it is defined by the conditions:

$$\approx (a,b,c,d), \approx (a,c,b,d) \geq p \wedge \frac{a}{n} \geq minsupp$$

where $0 < p < 1$ and $0 < minsupp < 1$.

Note that the support of both rules $X \rightarrow Y$ and $Y \rightarrow X$ is the same, and p represents a threshold for the value of the quantifier. In our case, $p = minconf$.

We remember that a 4ft-quantifier $\approx$ is *symmetric* [6] if $\approx (a,b,c,d) = \approx (a,c,b,d)$. Then, it is easy to see that a symmetric strong quantifier is always a doble strong quantifier. This is resumed in next corollary.

Corollary 1. A symmetric quantifier will be *double strong* if it satisfies

$$\approx (a,b,c,d) \geq p \wedge \frac{a}{n} \geq minsupp$$

where $0 < p < 1$ and $0 < minsupp < 1$.

6.1 Exceptions for Double Rules

When there is a double strong rule $X \leftrightarrow Y$ in a database, we can discover several exceptions. Following with the previous example, the next one clarifies that it could happen.

Example 2. In the ambient of example 1 two kinds of exceptions can be discovered. The first type contains the exceptions for the consequent of the double strong rule, for example:

if the animal flies, it is a bird, *except* bats

and the second type is constituted by the exceptions for the antecedent of the double rule:

if the animal it is a bird, it flies,
except penguin, ostrich, cock and hen.

In general, we are going to deal with some different cases.

1. As the double strong rule could be divided into two strong rules, we can mine their exceptions separately. In this case, we can distinguish some situations.

-It could happen that there are no exceptions for the double strong rule.

-There are two symmetric situations: we find an exception in only one direction of the double rule.

$X \leftrightarrow Y$	Double strong rule
$X \wedge E \rightarrow \neg Y$	Confident rule

$X \leftrightarrow Y$	Double strong rule
$Y \wedge E \rightarrow \neg X$	Confident rule

The last cases are the most common. In the second and third situation we have found an "agent" which interferes in the usual behavior of one of the sides of the rule.

-A special situation is the one presented in the example 2 when we have exceptions in both directions of the double rule and they are different.

$X \leftrightarrow Y$	Double strong rule
$X \wedge E \rightarrow \neg Y$	Confident rule
$Y \wedge E' \rightarrow \neg X$	Confident rule

2. The double strong rule has the same exceptions in both directions of the rule, i.e. we have that $E = E'$:

$X \leftrightarrow Y$	Double strong rule
$X \wedge E \rightarrow \neg Y$	Confident rule
$Y \wedge E \rightarrow \neg X$	Confident rule

In order to illustrate this situation, let us consider the relation shown in table 2 containing twelve transactions. From this dataset, we obtain: $\text{supp}(X \leftrightarrow Y) \simeq 0.583$, $\text{Conf}(X \rightarrow Y) \simeq 0.78$, and $\text{Conf}(Y \rightarrow X) \simeq 0.78$, which show that $X \leftrightarrow Y$ is a double strong rule if we impose the *minsupp* threshold to 0.5 and the *minconf* threshold to 0.6. We also obtain: $\text{Conf}(X \wedge E \rightarrow \neg Y) \simeq 0.67$, and $\text{Conf}(Y \wedge E \rightarrow \neg X) \simeq 0.67$, thus we obtain that E is the same exception for both sides of the double rule.

Table 2: Database with a double exception.

X	Y	F	$\cdots$
X	Y	F	$\cdots$
X	Y	F	$\cdots$
X	Y	F	$\cdots$
X	Y	F	$\cdots$
X	Y	F	$\cdots$
X	Y	E	$\cdots$
X	Y'	E	$\cdots$
X	Y'	E	$\cdots$
X'	Y	E	$\cdots$
X'	Y	E	$\cdots$
X'	Y'	E	$\cdots$

Semantically we could say that we have found an "agent" which affects the both sides of the common sense rule, i.e. its presence disturbs the usual behavior of the double strong rule $X \leftrightarrow Y$.

The most important case is the last one which the interaction of an "strange" (in the sense of unfrequent) factor makes changing the normal behavior of the double strong rule. When this happen, we say that E is a *double exception*.

6.2 Discovering Exceptions of Double Rules

Given a database, mining the two kinds of exception rules consists of generating all the doble strong rules and then mining the exception rules considering the two strong rules hidden in the double strong rule. The mining process follows some simple steps:

1. Find all the strong rules.

2. Find the exception rules associated to those strong rules.

3. Group the double strong rules with their associated exceptions.

4. Distinguish those double rules which contain a double exception.

7 Conclusions and Future Research

We have studied the different proposals for defining exceptions in association rules analyzing them semantically. We use the logic approach for defining them more formally, and then using an existing approach for mining association rules based on the definition of the suitable quantifier. We also present the double strong rules and the different kind of exceptions associated to them. When extracting double exceptions the user manage with a useful and novelty knowledge not considered until now.

For future work we are interested in the study of new measures that better fit to the exceptional knowledge we want to extract from a data set. We also plan to study an algorithm for extracting exceptions for strong and double strong rules and compare them. We will continue our research with anomalous rules, another kind of interesting knowledge in some sense complementary to that of exception rules.

Acknowledgements

We would like to acknowledge support for this work from the Ministerio de Educación y Ciencia of Spain by the project grants TIN2006-15041-C04-01 and TIN2006-07262.

References

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

[2] M. Delgado, M.D. Ruiz, and D. Sánchez. A logic approach for exceptions and anomalies in association rules. *Mathware & Soft Computing*, Submitted, 2007.

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

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

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

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

[7] 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.

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

[9] J. Rauch and M. Šimunek. An alternative approach to mining association rules. *Studies in Computational Intelligence (SCI)*, 6:211–231, 2005.

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

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

[12] E. Suzuki and M. Shimura. Exceptional knowledge discovery in databases based on information theory. In *Proc. of the 2nd Int. Conf. on Knowledge Discovery and Data Mining*, pages 275–278, 1996.

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

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