Exception Rule Mining with a Relative Interestingness Measure

Farhad Hussain[1], Huan Liu[1] Einoshin Suzuki[2] and Hongjun Lu[3] *

[1]PRIS, School of Computing, National University of Singapore.
[2]Electrical and Computer Engineering, Yokohama National University, Japan.
[3]Computer Science, Hong Kong University of Science and Technology.
{farhad,liuh}@comp.nus.edu.sg, suzuki@dnj.ynu.ac.jp, luhj@cs.ust.hk

Abstract. This paper presents a method for mining exception rules based on a novel measure which estimates interestingness relative to its corresponding common sense rule and reference rule. Mining interesting rules is one of the important data mining tasks. Interesting rules bring novel knowledge that helps decision makers for advantageous actions. It is true that interestingness is a relative issue that depends on the other prior knowledge. However, this estimation can be biased due to the incomplete or inaccurate knowledge about the domain. Even if possible to estimate interestingness, it is not so trivial to judge the interestingness from a huge set of mined rules. Therefore, an automated system is required that can exploit the knowledge extractacted from the data in measuring interestingness. Since the extracted knowledge comes from the data, so it is possible to find a measure that is unbiased from the user's own belief. An unbiased measure that can estimate the interestingness of a rule with respect to the extractacted rules can be more acceptable to the user. In this work we try to show through the experiments, how our proposed relative measure can give an unbiased estimate of relative interestingness in a rule considering already mined rules.

1 Introduction

After performing data mining tasks that usually terminate with a large set of rules, there is a need to find some interesting rules that decision makers can use for advantageous actions. The number of discovered rules can, however, be so large that browsing the rule set and finding interesting rules from it can be rather difficult for the user. Moreover, it is much harder to know which of the discovered rules are really interesting. Interestingness is a relative issue since it always depends on the user's prior knowledge about the domain. However, user's belief can give a biased estimate for incomplete or inaccurate knowledge about the domain. True knowledge about the domain can be extracted form the data. Therefore, to provide an unbiased [1] measure, intuitively we can say,

* This author's work is partially supported by a grant from the National 973 project of China (No. G1998030414)

[1] In measuring interestingness user's biased belief is not incorporated

interestingness can be estimated relative to the common sense rules found in the data. Data mining facilitates understanding the large amount of data by generating some rules. These rules, considering as common sense knowledge, can be used to justify the interestingness in other mined rules.

We understand that something that contradicts user's common belief is bound to be interesting. Let's define exceptions as rules that contradict the common belief. Exceptions [4, 8, 12] can take an important role in making critical decisions. Exceptions and common sense rules point at opposite directions. Exceptions usually are minority, they are either not known (thus new) or omitted. Therefore, in many respects, for mining interesting rules, we are interested to mine exceptions for the common sense [11–14] rules. A common sense rule represents a common phenomenon that comes with high support and confidence in a particular domain. Subjectively, a common sense rule can be interesting if the user is not well aware about the problem domain or if he has a completely different view about the domain. It implies that a subjective measure of interestingness may be biased. This bias may not always be wanted in critical decision making.

Intuitively, exceptions contradict the common sense rules and they have a low support [12]. Therefore, exception rules are weak in terms of support, but having high confidence similar to those common sense rules. A weak rule of low support may not be a reliable. A user can specify minimum support for an exception to ensure mining reliable exception rules [4, 12].

If we mine exception rules, literally there would be a huge number of them because of their usual lower support threshold. In this paper, an exception rule is structurally defined as shown in Table 1. Note that this is a simplified version of [11–14]. Here A and B represent a single item or a set of items.

$A \to X$	- common sense rule (strong pattern)
	(high support, high confidence)
$A, B \to \neg X$	- exception rule (weak pattern)
	(low support, high confidence)
$B \to \neg X$	-reference rule
	(low support and/or low confidence)

Table 1. Rule structure for exceptions

It is clear from the rule structure that the reference item B that explains the cause of exception, with the given common sense $A \to X$, can be of several combinations of items satisfying all the support and confidence constraints. The more the number of items we allow in the reference B, the more the candidate exception rules we can be generated for a given common sense rule. The number of reference rules also depends on the number of items for each data attribute. This implies that for a particular common sense rule, a user usually finds more and more exceptions that ultimately mislead the user for mining exceptions in the data. Moreover, all the references that help form exceptions may not be meaningful or the user may not be aware of them due to their lower support or confidence for a particular common sense rule to be an exception. Therefore, it will be more meaningful to include those reference items about which the user is concerned. This implies that to satisfy the criteria (low support and/or low confidence) for a reference rule $B \rightarrow \neg X$ in Table 1, we simply ensure the existence of a common sense rule $B \rightarrow X$.

We are motivated to mine those exception rules that are interesting in nature and can be estimated by the knowledge from the other mined rules. Estimation of interestingness is necessary to identify the most interesting rules from the mined rules. It would be an unbiased estimate since it considers only the common sense rules in the data, not that of the user's belief. Before giving a measure of interestingness, it is necessary to understand the components of what and how they bring surprise or interestingness in a rule given the knowledge about some common sense rules.

The remainder of this paper is organized as follows. In Section 2, we talk about some preliminaries and related work on interestingness. In Section 3, we describe our approach and provide details of our measure analytically. Implementation of our algorithm for digging out the interesting rules is described in Section 4. Analysis of our experimental results are shown in Section 5. Finally, in Section 6, we present our conclusion.

2 Preliminaries and related work

In this paper we try to capture interestingness of a given rule through the amount of change in information relative to the common sense rules. A particular rule contains the information about its premise, predictive power and coverage. Therefore, in order to measure the relative interestingness, we calculate the difference between the information that a rule contains and the information needed to describe the given rule with common sense rules. The interestingness of the given rule is proportional to the the amount of difference between this two information. Bigger the difference in information the more interesting is the given rule. This resembles to the procedure, an expert uses to estimate the interestingness of a given rule using other rules.

We define the common sense knowledge κ for a rule, is the knowledge about some common sense rules that we can apply to estimate the interestingness in the given rule. Since, interestingness is a relative factor that depends on the knowledge about other rules, therefore, for a rule $AB \rightarrow X$, κ composed of the knowledge about the rules $A \rightarrow X$ and $B \rightarrow X$. Therefore, by our definition, the interestingness I, of a rule should be a function of its support(S), confidence(C) and the knowledge about common sense rules (κ)

$I = f(S, C, \kappa)$

We believe that lack of any one of these parameters may result in incomplete estimate of the interestingness. That means we are interested to mine those interesting rules where all these three components can be applied to make it more acceptable. The measure we propose, offers both subjective and objective interestingness. When κ comes from the user's belief instead of the other mined rules, then the same measure can be considered as a subjective measure of interestingness for a given rule.

Generally, there are two categories of finding interesting rules: subjective vs. objective. In a subjective case, a user directly applies his own knowledge, sometimes even without knowing the nature of the domain to find interestingness in a rule. That is why, subjective interestingness may be biased and may vary with different users. Different methods have been proposed to capture interesting rules subjectively. The main idea is to impose user's own belief about the domain. Users usually apply their knowledge in terms of rule templates [6] and then try to match the template by scanning the data satisfying some threshold parameters. This approach may be suitable to justify a particular user's own belief system but may fail to discover some surprising rules that they even don't know. One potential problem is that user's subjective judgement may be unrealistic while applying those rules in the competitive business environment.

In an objective measure where common sense rules are not applied ($\kappa = \phi$) to estimate the interestingness, one of the main focus is to find if there is any deviation in the rule from the norm. Deviations are powerful and useful in this case as they provide a simple way of identifying interesting patterns in the data. This has been shown in the *KEFIR* [2] application. This type of measure is likely to be unbiased since no user's preference is given while estimating the interestingness. However, a rule that deviates may not be interesting if that deviation can be explained by other rules in the data. This implies that one can still generate a large number of rules that are interesting objectively but of little interest to the user [16] as they might know other rules. Again, since for the existing objective measures, $\kappa = \phi$ so, two different rules with similar support and confidence usually come with the same interestingness. Despite their similarity in support and confidence, interestingness usually should depend on the prior knowledge about those two rules. For example, using *J-measure* [10], two rules $AB \rightarrow X$ and $PQ \rightarrow X$ for which $\Pr(XAB) = \Pr(XPQ)$ and $\Pr(X|AB) = \Pr(X|PQ)$ give identical interestingness irrespective to their prior knowledge. For a rule $AB \rightarrow X$, *J-measure* is defined as follows.

$$J(X, AB) = \Pr(XAB)\, \log_2 \frac{\Pr(X|AB)}{\Pr(X)} + \Pr(\neg XAB)\, \log_2 \frac{\Pr(\neg X|AB)}{\Pr(\neg X)}$$

Though for an automated system, objective measures are always reliable due to their unbiased nature, sometimes they are completely unable to justify a rule's interestingness as they cannot handle knowledge from common sense rules. Since user's true belief will eventually build upon the common sense rules in the data, it is worth proposing a measure that can manipulate the common sense rules to estimate the interestingness in other rules. That is why, we need an objective measure that takes into account κ - the already found rules. In [11], *Geometric mean of the Average Compressed Entropies (GACE)* was defined to estimate the interestingness of a rule if and only if the given rule is an exception to a common sense rule. *GACE* defines the interestingness in terms of *J-measure* from one rule pair (common sense and exception). For a common sense rule $A \rightarrow X$ and its exception $AB \rightarrow \neg X$, *GACE* is defined as follows

$$GACE(A \rightarrow X, AB \rightarrow \neg X) = \sqrt{J(X, A) J(\neg X, AB)}$$

When we wish to include more (such as the already extracted rule both common sense and reference) in evaluating a rule's interestingness, we need a new measure that can take into account that something more.

3 Our approach

In this work, we are interested to mine rules that are objectively interesting, but in the meantime we are able to measure the interestingness with respect to already mined rules. For example, a rule "*Sex = F and Age = Young* $\rightarrow$ *Credit = No*" with 50% confidence may not be interesting from both subjective and objective viewpoints. From the subjective viewpoint, it is not interesting if the user believes that the females are not usually given credit compared to the males. Objectively it is also not interesting since it is not very conclusive as it has 50% confidence. However, if two strong [2] common sense rules "*Sex = F* $\rightarrow$ *Credit = Yes*" and "*Age = Young* $\rightarrow$ *Credit = Yes*" are known then "*Sex = F and Age = Young* $\rightarrow$ *Credit = No*" with 50% confidence should be interesting to us. This suggests a need to have a relative interesting measure.

We have mentioned eralier that an objective measure evaluates a rule's interestingness with its support(S) and confidence(C). Now we wish to consider the relevant extracted rules (κ) in our new measure. Below we explain in detail how all these factors together define relative interestingness of a rule.

3.1 Confidence-based interestingness

When no other information is given, an event with lower probability to occur gives more information, than an event with higher probability. From the information theory [9], the number of bits required to describe the occurrence is defined as

$$I = -\log_2 P$$

where,

P = The probability of that event to occur.

Similarly, for a given rule $AB \rightarrow X$ with confidence $\Pr(X|AB)$, will require $-\log_2(\Pr(X|AB))$ and $-\log_2(\Pr(\neg X|AB))$ number of bits to describe the events X and $\neg X$ given AB. Note that in the case of analyzing the rules we not only consider the event X or $\neg X$ but also the premise that causes the events to occur. That is why in the case of rules, whenever we talk about the events X or

[2] High support and high confidence

$\neg X$, we always mention the premise which leads the events to occur. However, in information theory, we are not interested to know how the events X and $\neg X$ occur. Since the probability of occurrence of the events X and $\neg X$ given AB are $\Pr(X|AB)$ and $\Pr(\neg X|AB)$, so, the expected number of bits required to describe events X and $\neg X$ influenced by AB are $-\Pr(X|AB)\log_2(\Pr(X|AB))$ and $-\Pr(\neg X|AB)\log_2(\Pr(\neg X|AB))$ respectively. Thus the total number of bits required to describe the rule $AB \to X$ is

$$I_c^{AB_0} = (-\Pr(X|AB)\log_2 \Pr(X|AB)) + (-\Pr(\neg X|AB)\log_2 \Pr(\neg X|AB))$$

where,

$I_c^{AB_0}$ = Number of bits required to describe the rule $AB \to X$ when no other knowledge have been applied.

However, the difference in number of bits in describing the rule $AB \to X$ in terms of $A \to X$ and $B \to X$ can bring surprise. Bigger the difference in describing the rule $AB \to X$, the more it is interesting. Therefore, to estimate the relative interestingness in terms of rules $A \to X$ and $B \to X$, we need to know the number of bits required to describe the event X when the probability of that event to occur given A and B are $\Pr(X|A)$ and $\Pr(X|B)$ respectively. The Table 2 shows the number of bits required to describe the events X and $\neg X$ when A and B are given.

Events	Given	Number of bits
X	A	-$\log_2 \Pr(X\|A)$
¬X	A	-$\log_2 \Pr(\neg X\|A)$
X	B	-$\log_2 \Pr(X\|B)$
¬ X	B	-$\log_2 \Pr(\neg X\|B)$

Table 2. The events and the information

Since the rule $AB \to X$ describes the event X in terms of A and B so, to describe the similar event X, in terms of A and [3] B using the rules $A \to X$ and $B \to X$ we need $-\log_2(\Pr(X|A))$ and $-\log_2(\Pr(X|B))$ number of bits. Now, in rule $AB \to X$ probability of the event X to occur is $\Pr(X|AB)$. Therefore, the expected number of bits required to describe all the X events in rule $AB \to X$ in terms of A and B using the two rules is thus $-\Pr(X|AB)(\log_2(\Pr(X|A)) + \log_2(\Pr(X|B)))$. Similarly, for the event $\neg X$ in rule $AB \to X$, $-\Pr(\neg X|AB)$ $(\log_2(\Pr(\neg X|A)) + \log_2(\Pr(\neg X|B)))$ number of bits will be required. Thus the total number of bits required to describe the event X and $\neg X$ in the rule $AB \to X$ by rules $A \to X$ and $B \to X$ is

$$\begin{aligned} I_c^{AB_1} = & -\Pr(X|AB)[\log_2 \Pr(X|A) + \log_2 \Pr(X|B)] \\ & -\Pr(\neg X|AB)[\log_2 \Pr(\neg X|A) + \log_2 \Pr(\neg X|B)] \end{aligned}$$

[3] It is not **or** because, event X was influenced in rule $AB \to X$ by A and B

where,

$I_c^{AB_1}$ = Number of bits required when $AB \rightarrow X$ is described by $A \rightarrow X$ and $B \rightarrow X$.

Thus, the relative surprise or relative interestingness that comes from the difference between two descriptions for the given rule $AB \rightarrow X$ is

$$
\begin{aligned}
RI_c^{AB} &= I_c^{AB_1} - I_c^{AB_0} \\
&= -\Pr(X|AB)\log_2 \frac{\Pr(X|A)\Pr(X|B)}{\Pr(X|AB)} - \Pr(\neg X|AB)\log_2 \frac{\Pr(\neg X|A)\Pr(\neg X|B)}{\Pr(\neg X|AB)} \\
&= \Pr(X|AB)\log_2 \frac{\Pr(X|AB)}{\Pr(X|A)\Pr(X|B)} + \Pr(\neg X|AB)\log_2 \frac{\Pr(\neg X|AB)}{\Pr(\neg X|A)\Pr(\neg X|B)}
\end{aligned}
$$

where,

RI_c^{AB} = The relative surprise or interestingness of the rule, considering the confidence and the knowledge about other rules.

The interestingness of a rule that we have formulated in terms of the confidence gives the exact impression of relative entropy [15]. Here the entropy of a rule is calculated relative to the other rules. It is a measure of distance between two distributions. In statistics, it arises as an expected logarithm of the likelihood ratio. The relative entropy $D(p(x)||q(x))$ is a measure of the inefficiency of assuming that the distribution is $q(x)$ when the true distribution is $p(x)$. The relative entropy or *Kullback Leibler distance* between two probability function is defined as,

$$D(p(x)||q(x)) = \sum_{x \in X} p(x) log \frac{p(x)}{q(x)}$$

In estimating the interestingness of the rule $AB \rightarrow X$ with true confidence $\Pr(X|AB)$ we approximated its confidence from the rules $A \rightarrow X$ and $B \rightarrow X$.

3.2 Support-based interestingness

By support of a rule $AB \rightarrow X$, we mean the frequency of the rule's consequent evaluated as X by AB relative to the whole dataset. When we know the support of two common sense rules $A \rightarrow X$ and $B \rightarrow X$, we know the relative frequency of the consequent X and $\neg X$ evaluated by A and B separately. A similar relative entropy measure can be applied to estimate the surprise from support. Now, for the newly discovered rule $AB \rightarrow X$, the true distribution of the consequent X and $\neg X$ evaluated by A and B are $\Pr(ABX)$ and $\Pr(AB\neg X)$ respectively. From the knowledge of our one common sense rule $A \rightarrow X$, for which the relative frequency of X and $\neg X$ are $\Pr(AX)$ and $\Pr(A\neg X)$ respectively, can be used to find the distance between two distributions of consequent using relative entropy. The relative entropy of $AB \rightarrow X$ relative to the rule $A \rightarrow X$ in terms of their support is thus

$$D(AB \rightarrow X||A \rightarrow X) = \Pr(ABX)\log \frac{\Pr(ABX)}{\Pr(AX)} + \Pr(AB\neg X)\log \frac{\Pr(AB\neg X)}{\Pr(A\neg X)}$$

Similarly for rule $B \rightarrow X$, the relative entropy is

$$D(AB \rightarrow X||B \rightarrow X) = \Pr(ABX)\log \frac{\Pr(ABX)}{\Pr(BX)} + \Pr(AB\neg X)\log \frac{\Pr(AB\neg X)}{\Pr(B\neg X)}$$

Thus the total relative interestingness due to rule's support that comes from the relative entropy of $AB \to X$ for the two common sense rule is,

$$\begin{aligned} RI_s^{AB} &= D(AB \to X || A \to X) + D(AB \to X || B \to X) \\ &= \Pr(ABX) \log \frac{\Pr(ABX)}{\Pr(AX)\Pr(BX)} + \Pr(AB\neg X) \log \frac{\Pr(AB\neg X)}{\Pr(A\neg X)\Pr(B\neg X)} \end{aligned}$$

Hence, the total interestingness of a rule $AB \to X$ relative to $A \to X$ and $B \to X$ is

$RI = RI_c^{AB} + RI_s^{AB}$

This includes support, confidence and consideration of other rules to estimate the relative surpriseness.

4 Digging out the exceptions

Since exceptions are weak in terms of support, we are supposed to dig deeper into the data with lower support threshold to bring them out. Applying lower support threshold for mining exceptions is not a cost-effective solution. Moreover, in that case, a large number of rules will be generated where not all of them are exceptions. Actually we are going to mine those exceptions where the rules extracted as a common sense will be used to alleviate the problem of dealing with lower support threshold. In other words, we search for reliable exceptions starting from the common sense rules. To satisfy all the constraints defined in Table 1, we find exception $AB \to \neg X$ from two common sense rules $A \to X$ and $B \to X$ ($B \to X$ as common sense infers $B \to \neg X$ to be reference for its obvious low support or/and low confidence). By doing this we can estimate the amount of surprise the exception rule brings from the knowledge of the extracted rules. From Figure 1 we can visualize how exceptions are mined going deeper into the data. The threshold *CS support* is the minimum support to mine the common sense rules from the data and the *EX support* to assure the reliability of the exception rules. The following algorithm describes the way we mine interesting rules.

begin
$LI = \phi$ // list containing large item set
$LC = \phi$ // list containing common sense rules
$LR = \phi$ // list containing reference rules for a common sense
$LE = \phi$ // list containing candidate exception rules
$LI \leftarrow$ *GenerateLargeItemSet()* // running apriori [1]
$LC \leftarrow$ *GenerateAllCommonSense(LI)*
for each CS_i **from** LC **do**
    $A \leftarrow$ *GetAntecedent*(CS_i)
    $LR \leftarrow$ *GetReferences*(CS_i, LC)
    for each RR_j **from** LR **do**
        $B \leftarrow$ *GetAntecedent*(RR_j)
        if $(A \cup B)$ **is not in** LI
            insert$(A \cup B \cup \neg Consequent(CS_i), LE)$

end for
end for
$LE \leftarrow$ *GenerateExceptions(LE)* // Database scan once
EstimateInterestingness(LC, LE) // Output interesting rules
end. // according to the degree of surprise

The function *GetReferences(CS_i, LC)*, returns all the candidate reference rules for CS_i, from LC. The reference rules are those common sense rules in LC that have similar consequent as that of CS_i. Once we have inserted all the candidate exception rules into LE, we scan the database once to obtain the support and confidence of each candidate exceptions. We output those rules that satisfy the thresholds using *GenerateExceptions(LE)*. *EstimateInterestingness(LC, LE)* estimates the relative interestingness.

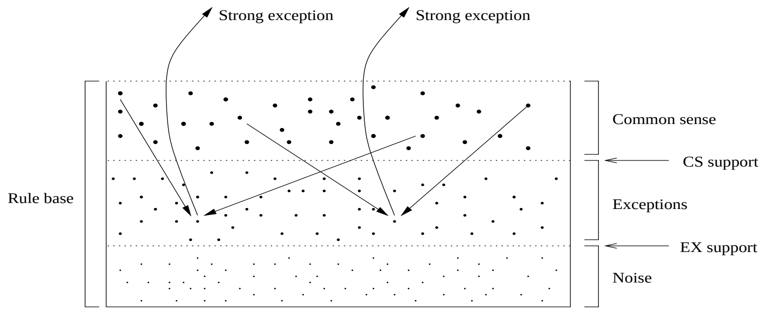

Fig. 1. Rules in the data

5 Experiments

In this section we explain our interesting rules obtained from Japanese credit data and mushroom data [7]. The credit data set has 10 attributes and 125 instances with a binary class. The two types of classes define when a credit is given to a particular person depending on other attribute values. Based on our approach we obtain the strong patterns that would eventually be considered as a common sense rule for that credit domain. We use 20% support threshold for common sense. To mine the reliable exceptions considering already extracted rules, we use 5% support. For both cases we mine those having more than or equal to 50% confidence.

We will show the effects of the proposed measure (RI) in two settings: (1) removing reference rules by assuming they are not available; and (2) including reference rules. As in the first case, GACE has been shown effective. Therefore we compare RI with GACE to see if RI can achieve the planned objective: they should identify the interesting exceptions as GACE does. In the second case, we show that the ranking order changes due to considering reference rules.

Common Sense (CS) Reference Rule (RR) Exception = CS + RR	Credit Given	Conf %	Supp %	RI With Ref.	RI Without Ref.	GACE
Sex = F, MPay = 0-20 (CS)	Yes	61	26			
Save = 0-100, Mnts = 0-10 (RR)	No	37	17			
Sex = F, MPay = 0-20, Save = 0-100, Mnts = 0-10	No	54	11	3.16	1.98	0.013
MPay = 0-20, Mnts = 0-10 (CS)	Yes	65	32			
Sex = F (RR)	No	38	17			
Sex = F, MPay = 0-20, Mnts = 0-10	No	52	10	3.00	1.93	0.006
Mnts = 0-10 (CS)	Yes	65	32			
Sex = F, Save = 0-100 (RR)	No	42	18			
Sex = F, Save = 0-100, Mnts = 0-10	No	52	10	2.96	1.93	0.006
Sex = F, Age = 0-40 (CS)	Yes	61	20			
Marry = Yes (RR)	No	29	15			
Sex = F, Age = 0-40, Marry = Yes	No	51	6	2.52	1.67	0.007

Table 3. Exception rules (credit) justified by the common sense rules

From the experimental result it is clear that RI without considering reference rules behaves similar as GACE. So, RI can identify reliable exceptions. When considering reference rules, RI has changed the rankings (2nd and 3rd rule sets in Table 3). An *actionable* exception rule should be reliable. An exception would be more reliable if it has higher confidence and support. Therefore, from the perspective of both interestingness and reliability, the first exception "*Sex = F, MPay = 0-20, Save = 0-100, Mnts = 0.10* $\rightarrow$ *Credit = No*" should have the highest ranking among the four exception rules.

For the mushroom data (22 attributes and 8124 tuples) we conduct a similar experiment with 15% support and 55% confidence for mining common sense rules. To mine the exceptions, we specify 5% support and 55% confidence. Table 4 shows some of the rules and their corresponding interestingness, when RI without reference cannot diferentiate the difference between the third and four rule sets, RI with reference can. Hence, when we have some knowledge about the rule, the knowledge should be used if we wish to find relative interestingness. GACE has made use of common sense rules and RI goes one step further to include reference rules in the measure.

6 Conclusion

In this work we define an objective measure of relative interestingness that ties up the with common sense and reference rules in the data when estimating interestingness. This opens the door to apply RI to measure subjective interestingness

Common Sense (CS) Reference Rule (RR) Exception = CS + RR	Edible/ Poisonous	Conf %	Supp %	RI With Ref.	RI Without Ref.	GACE
stalk-root = ? (CS)	P	71	22			
bruises = f, gill-size = b, stalk-shape = e, veil-type = p (RR)	E	27	6			
CS + RR	E	100	6	6.76	3.98	0.05
ring-type = e (CS)	P	64	22			
bruises = f, gill-size = b, ring-number = t (RR)	E	43	12			
CS + RR	E	100	9	4.71	3.58	0.04
gill-color = p, veil-type = p, veil-color = w (CS)	P	57	23			
stalk-root = ? (RR)	E	29	9			
CS + RR	E	94	6	4.19	2.28	0.02
gill-attachment = f, stalk-shape = e, veil-type = p (CS)	P	57	23			
cap-surface = s, veil-color = w (RR)	E	40	11			
CS + RR	E	94	6	3.31	2.28	0.02

Table 4. Exception rules (mushroom) justified by the common sense rules

if we can describe an expert's knowledge in rules. For example, we wish to evaluate the new rule in Table 5; it is not very interesting than those in Table 3. This should be the case as the new rule is not an exception.

Common sense Common sense New rule	Credit Given	Conf %	Supp %	RI
Jobless = No, Sex = F	Yes	70	28	
Married = No	Yes	64	30	
Jobless = No, Sex = F, Married = No (Not exception)	Yes	71	12	2.34

Table 5. A new rule (credit) justified by the common sense rules

In our approach we focus on relative interestingness that evaluates an exception rule with respect to its common sense and reference rules. Not much extra costs invloved in calculating RI: Using Apriori [1] to find the frequent itemsets, then we just need another database scan to estimate the support and confidence of the candidate exception items. This efficient new measure provides one more means in exception rule mining along with other interestingness measures.

References

1. R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In *Proc. 20th conference on Very Large Databases (VLDB)*, pages 478–499, 1994.
2. G. Piatetsky-Shapiro C. Matheus and D. McNeil. *Selecting and Reporting What is Interesting: The KEFIR Application to Healthcare Data.* AAAI Press/ MIT Press, 1996.
3. Sarawagi Chakrabarti. Mining surprising patterns using temporal description length. In *Proc. 24th on Very Large Databases (VLDB)*, pages 606–616, 1998.
4. Liu H. and Lu H. Efficient search of reliable exceptions. In *Proc. third Pacific-Asia conference on Knowledge Discovery and Data mining (PAKDD)*, pages 194–203, 1999.
5. W. Hsu Liu B. and Shu Chen. Using general impression to analyze discovered classification rules. In *Proc. third international conference on Knowledge Discovery and Data mining (KDD)*, pages 31–36, 1997.
6. H. Mannila M. Klemettinen. Finding interesting rules from large sets of discovered association rules. In *Third Intl. Conference on Information and Knowledge Management (CIKM)*, 1994.
7. C.J. Merz and P.M. Murphy. UCI repository of machine learning databases. `http://www.ics.uci.edu/~mlearn/MLRepository.html`. Irvine, CA: University of California, Department of Information and Computer Science, 1996.
8. B. Padmanabhan and A. Tuzhilin. A beleif-driven method for discovering unexpected patterns. In *Proc. fourth international conference on Knowledge Discovery and Data mining (KDD)*, pages 27–31, 1998.
9. C. Shannon and W. Weaver. *The Mathematical Theory of Information.* Urbana: University of Illinois Press, 1949.
10. P. Smyth and Goodman R. M. Rule induction using information theory. In *Knowledge Discovery in Databases, Piatetsky-Shapiro, G. AAAI Press / The MIT Pres*, pages 159–176, 1991.
11. E. Suzuki. Discovering unexpected exceptions: A stochastic approach. In *Proc. RFID*, pages 225–232, 1996.
12. E. Suzuki. Autonomous discovery of reliable exception rules. In *Proc. third international conference on Knowledge Discovery and Data mining (KDD)*, pages 259–262, 1997.
13. E. Suzuki and Y. Kodratoff. Discovery of surprising exception rules based on intensity of implication. In *Proc. second Pacific-Asia conference on Knowledge Discovery and Data mining (PAKDD)*, 1998.
14. E. Suzuki and M. Shimura. Exceptional knowledge discovery in databases based on information theory. In *Proc. second international conference on Knowledge Discovery and Data mining (KDD)*, pages 295–298, 1996.
15. C. Thomas M and J. Thomas A. *Elements os Information Theory.* Wiley-Interscience Publication, 1996.
16. A. Tuzhilin and A. Silberschatz. What makes patterns interesting in knowledge discovery systems. In *IEEE Trans. Knowledge Discovery and Data Engineering*, pages 970–974, 1996.