

Discovery of Surprising Exception Rules Based on Intensity of Implication

Einoshin Suzuki¹ and Yves Kodratoff²

¹ Electrical and Computer Engineering, Yokohama National University,
79-5 Tokiwadai, Hodogaya, Yokohama 240-8501, Japan

² Equipe Inference et Apprentissage, Université de Paris-Sud,
91405 Orsay Cedex, France

Abstract. This paper presents an algorithm for discovering surprising exception rules from data sets. An exception rule, which is defined as a deviational pattern to a common sense, exhibits unexpectedness and is sometimes extremely useful. A domain-independent approach, PEDRE, exists for the simultaneous discovery of exception rules and their common sense rules. However, PEDRE, being too conservative, have difficulty in discovering surprising rules. Historic exception discoveries show that surprise is often linked with interestingness. In order to formalize this notion we propose a novel approach by improving PEDRE. First, we reformatize the problem and settle a looser constraints on the reliability of an exception rule. Then, in order to screen out uninteresting rules, we introduce, for an exception rule, an evaluation criterion of surprise by modifying intensity of implication, which is based on significance. Our approach has been validated using data sets from the UCI repository.

1 Introduction

Rule discovery [1, 2, 6, 7, 8, 9, 10] is, due to its generality and simplicity, one of the most important research topics in Knowledge Discovery in Databases (KDD). In KDD, a rule can be classified into two categories: strong rules and weak rules. A strong rule is a description of a regularity for numerous objects with high confidence. On the other hand, a weak rule represents, for a relatively small number of objects, a regularity with high confidence. Typically, strong rule discovery, while useful in recognizing general trends in a data set, results in a set of overly general rules. Finding a widely known rule, such as “most people who buy butter also buy bread”, is uninteresting. On the other hand, weak rule discovery can produce extremely interesting results.

Various methods for weak rule discovery have been proposed in KDD community. Some of them [2, 7] try to capture interestingness by using a criterion. Others [6, 8, 9, 10] assume a knowledge representation for this purpose, and seek for chunks of knowledge which deviate from strong rules. Among such rules, an exception rule, which represents a deviation to a strong rule, exhibits unexpectedness and is often useful. For instance, the rule “using a seat belt is risky for a small child”, which represents exceptions to the well known regularity “using a seat belt is safe”, exhibited unexpectedness when it was discovered from

car accident data several years ago, and is still useful. Moreover, an exception rule is often beneficial since it differs from a common sense rule which is often a basis for people's daily activity. For instance, suppose a species of poisonous mushrooms has a small number of exceptions. The exact description of the exceptions is highly beneficial since it enables the exclusive possession of the edible mushrooms.

Simultaneous approach [8, 9, 10], which was proposed by one of us (E. Suzuki), discovers a set of rule pairs each of which is a common sense rule and its exception rule. This approach deserves special attention since it requires neither domain specific criteria nor background knowledge to obtain interesting rules. Especially, the most recent system PEDRE [10], obtains reliable rule pairs.

Careful observation of historical discovery, however, shows that surprise, which was not directly considered in the previous systems represents an important aspect in exception knowledge discovery. For instance, antibiotics are widely known to cure diseases, however, MRSA, a kind of staphylococci, is often resistant to antibiotics in a hospital. Although MRSA is not a dangerous bacteria, people in a hospital, often weakened by other causes, die by taking antibiotics which kill other bacteria and favor MRSA. This phenomenon, when it was observed in the U.S. for the first time, lacked reliability in its generality since only a small number of patients were found in this situation. However, it showed an astonishment to people because it is represented by the conjunction of two very rare events: dying from MRSA and dying by taking antibiotics.

Here, this rule is not easily discovered by PEDRE since a huge number of rule pairs are more reliable. In other words, PEDRE is too conservative in evaluating rules with low reliability to discover surprising exception rules. In order to discover such kind of knowledge, we propose a modified approach of PEDRE which is based on intensity of implication [3]: an evaluation criterion of surprise with several desirable properties. Preliminary results with data sets from the UCI repository [5] are promising.

2 Discovery of Reliable Exceptions: PEDRE [10]

2.1 Reliable Exception

Consider a data set D with n examples each of which is expressed by m discrete attributes. An event representing, in propositional form, a single value assignment to an attribute will be called an atom. We define a conjunction rule as the production rule of which premise is represented by a conjunction of atoms and conclusion is a single atom. Another interesting class of rules is an association rule [1], in which premise and conclusion are a set of examples and all attributes are binary. Conjunction rules have been chosen in PEDRE since they do not assume such restrictions on attributes.

In PEDRE, we consider the problem of finding a set of rule pairs each of which consists of an exception rule associated with a common sense rule. Let a_i, b_j, c and c' be a single atom, where c and c' have the same attribute with

different values, then a rule pair $r(\mu, \nu)$ is defined as a pair of conjunction rules as follows:

$$r(\mu, \nu) \equiv \begin{cases} A_\mu & \rightarrow c \\ A_\mu \wedge B_\nu & \rightarrow c'. \end{cases} \quad (1)$$

$$\text{where } A_\mu \equiv a_1 \wedge a_2 \wedge \dots \wedge a_\mu, B_\nu \equiv b_1 \wedge b_2 \wedge \dots \wedge b_\nu. \quad (2)$$

We hereafter call $A_\mu \rightarrow c$, $A_\mu \wedge B_\nu \rightarrow c'$ and $B_\nu \rightarrow c'$ a common sense rule, an exception rule, and a reference rule respectively.

In discovery of reliable exception rules, both a common sense rule and an exception rule should be general and accurate with some confidence level. It should be also noted that, in an exception rule, the extra condition B_ν should not contribute to the prediction of the conclusion c' , unless the exception rule is easily predicted from $B_\nu \rightarrow c'$ and is thus uninteresting. Let $1 - \delta$ be a user-provided confidence level, and $\theta_1^S, \theta_2^S, \theta_1^F, \theta_2^F, \theta_2^I$ and M be a user-provided threshold, then the constraints considered in PEDRE for a discovered rule pair are given as follows.

$$\Pr \{ p(A_\mu) \geq \theta_1^S, p(c|A_\mu) \geq \theta_1^F, p(A_\mu, B_\nu) \geq \theta_2^S, p(c'|A_\mu, B_\nu) \geq \theta_2^F, \\ p(c'|B_\nu) \leq \theta_2^I \} \geq 1 - \delta. \quad (3)$$

$$\mu, \nu \leq M. \quad (4)$$

For resolving (3), we should estimate the confidence region of the probabilities related to (3). Normal approximations of a multinomial distribution is a natural choice for this purpose. It is shown that the following equations are equivalent to (3) [10].

$$G(A_\mu) \hat{p}(A_\mu) \geq \theta_1^S, F(A_\mu, c) \hat{p}(c|A_\mu) \geq \theta_1^F, G(A_\mu, B_\nu) \hat{p}(A_\mu, B_\nu) \geq \theta_2^S, \\ F(A_\mu, B_\nu, c') \hat{p}(c'|A_\mu, B_\nu) \geq \theta_2^F, F(B_\nu, c') \hat{p}(c'|B_\nu) \leq \theta_2^I, \quad (5)$$

where,

$$G(x) \equiv 1 - \beta \sqrt{\frac{1 - \hat{p}(x)}{n\hat{p}(x)}}, F(x, y) \equiv 1 - \beta \sqrt{\frac{\hat{p}(x) - \hat{p}(x, y)}{\hat{p}(x, y)\{(n + \beta^2)\hat{p}(x) - \beta^2\}}}, \quad (6)$$

and β is a positive constant related to the confidence level [10].

2.2 Search Algorithm

In PEDRE, a discovery task is viewed as a search problem, in which a node of a search tree represents a rule pair $r(\mu, \nu)$. Let $\mu = 0$ and $\nu = 0$ represent the state in which the premises of a rule pair $r(\mu, \nu)$ contain no a_i and no b_i respectively, then we define that $\mu = \nu = 0$ holds in a node of depth 1, and as the depth increases by 1, an atom is added to the premise of the exception rule or the common sense rule. A node of depth 2 is assumed to satisfy $\mu = 1, \nu = 0$, and a node of depth l (≥ 3), $\mu + \nu = l - 1$ ($\mu \geq 1, \nu \geq 0$).

A depth-first search method is employed to traverse this tree, and the maximum value M of μ and ν is given by the user. To improve search efficiency, a node which satisfies at least one of the stopping criteria (7) in theorem 1 is not expanded without altering the algorithm's output.

Theorem 1. *Let the rule pair of the current node be $r(\mu', \nu')$. If the rule pair satisfies an equation in (7), no rule pairs $r(\mu, \nu)$ of the descendant nodes satisfy (5).*

$$\begin{aligned} G(A_{\mu'})\hat{p}(A_{\mu'}) &< \theta_1^S, \quad G(A_{\mu'})\hat{p}(c, A_{\mu'}) < \theta_1^S\theta_1^F, \quad G(A_{\mu'}, B_{\nu'})\hat{p}(A_{\mu'}, B_{\nu'}) < \theta_2^S, \\ G(A_{\mu'}, B_{\nu'})\hat{p}(c', A_{\mu'}, B_{\nu'}) &< \theta_2^S\theta_2^F, \quad G(A_{\mu'}, B_{\nu'})\hat{p}(B_{\nu'}) < \frac{\theta_2^S\theta_2^F}{\theta_1^S}. \end{aligned} \quad (7)$$

Proof Each equation can be proved by contradiction using that the function $G(x)$ increases monotonously for $\hat{p}(x)$ since $n, \beta > 0$. \square

3 Measures for Surprising Exceptions

3.1 Reformalization of Problem

As mentioned in Introduction, PEDRE is appropriate for the discovery of reliable exceptions, but fails to capture the surprise of an exception rule. For instance, table 1 [10] shows a rule pair discovered by PEDRE, where the edibility class is the only attribute allowed in the conclusions and the parameters were set to $M = 3$, $\delta = 0.1$, $\theta_1^S = 0.2$, $\theta_2^S = 0.05$, $\theta_1^F = 0.7$, $\theta_2^F = 1.0$ and $\theta_2^I = 0.5$.

Table 1. A rule pair with its associated reference rule discovered by PEDRE from the mushroom data set, where the edibility class is the only attribute allowed in the conclusions. Here, $\uparrow p$ and $\downarrow p$ represent the upper and lower bound of p respectively.

common sense rule	$\downarrow p(A_\mu)$	$\downarrow p(c A_\mu)$
exception rule	$\downarrow p(A_\mu B_\nu)$	$\downarrow p(c' A_\mu B_\nu)$
reference rule		$\uparrow p(c' B_\nu)$
$b=f, gs=b, sts=e \rightarrow class=p$	0.213	0.703
$b=f, gs=b, sts=e, str=v \rightarrow class=e$	0.051	1.000
$str=v \rightarrow class=e$		0.318

We see that the rule pair in the table shows interesting exceptions. However, the accuracy of the common sense rule is relatively small: its lower bound value is 0.703. We attribute this to the high reliability of the exception rule since it has a negative effect on the accuracy of a common sense rule. In PEDRE, a common sense rule cannot have a high accuracy unless the user assumes a low value for δ , θ_1^S and θ_1^F . In other words, PEDRE has lost surprise while gaining reliability.

In order to discover truly surprising exception rules, we reformalize the constraints (3), and consider point-estimated probabilities $\hat{p}(A_\mu, B_\nu)$ and $\hat{p}(c'|A_\mu, B_\nu)$ instead of true probabilities $p(A_\mu, B_\nu)$ and $p(c'|A_\mu, B_\nu)$.

$$\Pr \{p(A_\mu) \geq \theta_1^S, p(c|A_\mu) \geq \theta_1^F, p(c'|A_\mu, B_\nu) \leq \theta_2^I\} \geq 1 - \delta. \quad (8)$$

$$\hat{p}(A_\mu, B_\nu) \geq \theta_2^S, \hat{p}(c'|A_\mu, B_\nu) \geq \theta_2^F, \quad (9)$$

$$\mu + \nu \leq M. \quad (10)$$

While this modification makes our approach less tolerant to noise, it generates an increasing number of rule pairs due to the looser restrictions on the generality and accuracy of an exception rule. Constraints on the length of premises are slightly modified in order to keep the computation time reasonable. Similarly as in the previous section, the following equations are equivalent to (8).

$$G(A_\mu)\hat{p}(A_\mu) \geq \theta_1^S, F(A_\mu, c)\hat{p}(c|A_\mu) \geq \theta_1^F, F(B_\nu, c')\hat{p}(c'|B_\nu) \leq \theta_2^I. \quad (11)$$

3.2 Evaluation of Surprise

Since this reformalization of the problem increases the numbers of discovered rule pairs, an evaluation criterion of surprise is introduced to screen out rule pairs of minor interest.

Intensity of implication [3] is an evaluation criterion of surprise for a rule. It represents the degree of surprise that a rule $A_\mu \rightarrow c$ has so few counter examples. Let U and V be a randomly-selected set whose number of observation $|U|$ and $|V|$ in data set D is equal to those of $|c|$ and $|A_\mu|$ respectively

$$|U| = |c|, |V| = |A_\mu|, \quad (12)$$

then an intensity of implication $\varphi(A_\mu, c, D)$ for this rule is given as follows.

$$\varphi(A_\mu, c, D) \equiv 1 - \Pr(|\overline{UV}| \leq |\overline{cA_\mu}|). \quad (13)$$

Assuming that U and V are independent, the random variable $|\overline{UV}|$ follows the hypergeometric law. Poisson approximations can be applied when $|\overline{UV}|$ is small [4], which is often the case in rule discovery.

$$\varphi(A_\mu, c, D) = 1 - \sum_{k=\max(0, |A_\mu| - |c|)}^{|\overline{cA_\mu}|} H(A_\mu, c, D, k) \quad (14)$$

$$\approx 1 - \sum_{k=0}^{|\overline{cA_\mu}|} \frac{\lambda^k}{k!} e^{-\lambda}, \quad (15)$$

$$\text{where } H(A_\mu, c, D, k) \equiv \frac{\binom{|c|}{|A_\mu| - k} \binom{|\overline{c}|}{k}}{\binom{n}{|A_\mu|}}, \quad \lambda \equiv \frac{|A_\mu|(n - |c|)}{n} \quad (16)$$

Since intensity of implication is based on a statistical significance $\Pr(|\overline{UV}| \leq |\overline{c}A_\mu|)$, its definition is easier to be interpreted than an information-based criterion such as J measure [7]. This definition shows that intensity of implication increases as the number of counter examples $|\overline{c}A_\mu|$ decreases. Moreover, intensity of implication has the desirable property of evaluating the reliability of a rule according to the size of a data set: it increases as n increases. Also, it evaluates the rareness of the conclusion that a rule predicts: it increases as $|c|$ decreases. A detailed analysis of this criterion is given in [2].

Let consider a direct application of intensity of implication to the evaluation of an exception rule. First, we should replace n by $|A_\mu|$ since the premise of a common sense rule can be considered as the universe for an exception rule. Similarly, A_μ and c are replaced by $A_\mu B_\nu$ and $c' A_\mu$ respectively. Therefore, the intensity of implication for an exception rule in a rule pair $r(\mu, \nu)$ is given by $\varphi(A_\mu B_\nu, c' A_\mu, A_\mu)$.

$$\varphi(A_\mu B_\nu, c' A_\mu, A_\mu) = 1 - \sum_{k=0}^{|\overline{c'}A_\mu B_\nu|} \frac{\lambda'^k}{k!} e^{-\lambda'} \quad (17)$$

$$\text{where } \lambda' \equiv \frac{|A_\mu B_\nu| (|A_\mu| - |c' A_\mu|)}{|A_\mu|} \quad (18)$$

Let examine this criterion in the context of exception knowledge discovery. If we have a large number of exception rules each of which covers more than 20 examples $|A_\mu B_\nu| \geq 20$, then its λ' is also nearly 20. If these rules have few counter examples $|\overline{c'}A_\mu B_\nu|$, which is often the case, their intensity of implication is all equal to 1. Therefore, the intensity of implication $\varphi(A_\mu B_\nu, c' A_\mu, A_\mu)$ is inappropriate for ranking exception rules according to their degree of surprise.

To overcome this problem, we propose a modified version of this criterion. The idea is to take the logarithm of the probability of having so few counter examples in order to have a precise view when this probability is nearly 0. The criterion of surprise $\varphi'(A_\mu B_\nu, c' A_\mu, A_\mu)$ for an exception rule in this paper is given as follows.

$$\varphi'(A_\mu B_\nu, c' A_\mu, A_\mu) \equiv -\ln \left(\sum_{k=0}^{|\overline{c'}A_\mu B_\nu|} \frac{\lambda'^k}{k!} e^{-\lambda'} \right) \quad (19)$$

where $\ln(x)$ represents the natural logarithm of x . Note that this criterion is equal to λ' when an exception rule has no counter examples ($|\overline{c'}A_\mu B_\nu| = 0$).

3.3 Search Algorithm

Similarly to PEDRE, the proposed approach employs a depth-first search to traverse a search tree for obtaining rule pairs which satisfy (9) ~ (11) and which have the K highest $\varphi'(A_\mu B_\nu, c' A_\mu, A_\mu)$. A theorem similar to theorem 1 holds in this search problem.

Theorem 2. *Let the rule pair of the current node be $r(\mu', \nu')$. If the rule pair satisfies an equation in (20) or (21), no rule pairs $r(\mu, \nu)$ of the descendant nodes satisfy (9) and (11).*

$$G(A_{\mu'})\hat{p}(A_{\mu'}) < \theta_1^S, G(A_{\mu'})\hat{p}(c, A_{\mu'}) < \theta_1^S\theta_1^F, \quad (20)$$

$$\hat{p}(A_{\mu'}, B_{\nu'}) < \theta_2^S, \hat{p}(c', A_{\mu'}, B_{\nu'}) < \theta_2^S\theta_2^F. \quad (21)$$

Proof First, the proof for (20) is equivalent to theorem 1. Second, (21) is derived from $\hat{p}(A_{\mu'}, B_{\nu'}) \geq \hat{p}(A_{\mu}, B_{\nu})$ and $\hat{p}(c', A_{\mu'}, B_{\nu'}) \geq \hat{p}(c', A_{\mu}, B_{\nu})$. \square

An improvement to PEDRE is that a node represents not only a rule pair $\{A_{\mu} \rightarrow c \ \& \ A_{\mu}, B_{\nu} \rightarrow c'\}$, but also $\{A_{\mu} \rightarrow c' \ \& \ A_{\mu}, B_{\nu} \rightarrow c\}$, $\{B_{\nu} \rightarrow c \ \& \ B_{\nu}, A_{\mu} \rightarrow c'\}$, and $\{B_{\nu} \rightarrow c' \ \& \ B_{\nu}, A_{\mu} \rightarrow c\}$. Since these 4 rule pairs employ similar atoms in calculating (9) \sim (11) and (20) \sim (21), this fourfold representation improves time efficiency. A 4-bit flag is employed to represent which rule pairs are considered in a node of a search tree.

4 Application to Data Sets

The proposed method was implemented and tested with data sets from several domains. The results were quite successful. Here, we show the results using the mushroom data set and the shuttle data set from the UCI Repository [5].

The mushroom data set includes 22 descriptions and the edibility class of 8,124 mushrooms, each attribute having 2 \sim 12 values. Table 2 shows the rule pairs discovered by our approach, where the edibility class is the only attribute allowed in the conclusions and the parameters were set to $M = 3$, $\delta = 0.1$, $\theta_1^S = 0.2$, $\theta_2^S = 40/8124$, $\theta_1^F = 0.8$, $\theta_2^F = 1.0$, $\theta_1^I = 0.5$ and $K = 5$.

The discovered rule pairs in the table show very interesting exceptions. According to the first rule pair, with 90 % confidence level, at least 86.6 % of the mushrooms whose “gs” is “n” are poisonous but 100 % of them are actually edible if “sts” is “e” and “p” is “y”. The first rule holds for at least 2,388 mushrooms (29.4 %). Note that, from the reference rule, only at most 28.0 % of the mushrooms whose “sts” is “e” and “p” is “y” are edible. This shows that the discovered exception rule is truly unexpected. Compared with the rule pair discovered by PEDRE (table 1), each rule pair in this table has a stronger common sense rule, and thus a more surprising exception rule. These results confirm the validity of our approach.

The shuttle data set includes 10 descriptions of 58,000 examples, each attribute having 3 to 8 values. Table 3 shows the rule pairs discovered by our approach, where the parameters were set to $M = 3$, $\delta = 0.1$, $\theta_1^S = 0.2$, $\theta_2^S = 50/58,000$, $\theta_1^F = 0.7$, $\theta_2^F = 1.0$, $\theta_1^I = 0.5$ and $K = 5$.

The discovered rule pairs in the table also show very interesting exceptions. According to the second rule pair, with 90 % confidence level, at least 83.6 % of the examples whose “att9 = 2” have “att8 = 1” but 100 % of them actually have “att8 = 0” if “att2 = 1” and “att7 = 0”. The first rule holds for at least 19,836 examples (34.2 %). Note that, from the reference rule, only at most 27.7 % of

Table 2. The rule pairs discovered from the mushroom data set, where the edibility class is the only attribute allowed in the conclusions.

No.	common sense rule exception rule reference rule	$\downarrow p(A_\mu)$ $ c'A_\mu B_\nu $ $\varphi'(A_\mu B_\nu, c'A_\mu, A_\mu)$	$\downarrow p(c A_\mu)$ $\hat{p}(c' A_\mu B_\nu)$ $\uparrow p(c' B_\nu)$
1	gs = n \rightarrow class = p	0.294	0.866
	gs = n, sts = e, p = y \rightarrow class = e	72	1.000
	sts = e, p = y \rightarrow class = e	63.7	0.280
2	gs = n \rightarrow class = p	0.294	0.866
	gs = n, b = f, p = y \rightarrow class = e	72	1.000
	b = f, p = y \rightarrow class = e	63.7	0.156
3	gs = n \rightarrow class = p	0.295	0.868
	gs = n, gsp = c, ssb = f \rightarrow class = e	48	1.000
	gsp = c, ssb = f \rightarrow class = e	42.4	0.339
4	gs = n \rightarrow class = p	0.294	0.866
	gs = n, str = b, scb = n \rightarrow class = e	48	1.000
	str = b, scb = n \rightarrow class = e	42.4	0.175
4	gs = n \rightarrow class = p	0.295	0.868
	gs = n, spc = h \rightarrow class = e	48	1.000
	spc = h \rightarrow class = e	42.4	0.041

Table 3. The rule pairs discovered from the shuttle data set.

No.	common sense rule exception rule reference rule	$\downarrow p(A_\mu)$ $ c'A_\mu B_\nu $ $\varphi'(A_\mu B_\nu, c'A_\mu, A_\mu)$	$\downarrow p(c A_\mu)$ $\hat{p}(c' A_\mu B_\nu)$ $\uparrow p(c' B_\nu)$
1	time = 0 \rightarrow att7 = 2	0.678	0.729
	time = 0, att3 = 2 \rightarrow att7 = 3	2,400	1.000
	att3 = 2 \rightarrow att7 = 3	1766.3	0.343
2	att9 = 2 \rightarrow att8 = 1	0.342	0.836
	att9 = 2, att2 = 1, att7 = 0 \rightarrow att8 = 0	555	1.000
	att2 = 1, att7 = 0 \rightarrow att8 = 0	511.0	0.277
3	att9 = 2 \rightarrow att8 = 1	0.342	0.836
	att9 = 2, att2 = 1, Class = 4 \rightarrow att8 = 0	554	1.000
	att2 = 1, Class = 4 \rightarrow att8 = 0	510.1	0.276
4	att5 = 2 \rightarrow att7 = 2	0.861	0.736
	att5 = 2, att8 = 2 \rightarrow att7 = 3	94	1.000
	att8 = 2 \rightarrow att7 = 3	73.9	0.058
5	att7 = 2 \rightarrow time = 0	0.681	0.726
	att7 = 2, att2 = 1, att3 = 3 \rightarrow time = 2	59	1.000
	att2 = 1, att3 = 3 \rightarrow time = 2	46.0	0.127

the examples which satisfy “att2 = 1” and “att7 = 0” have “att8 = 0”. This shows that the discovered exception rule is truly unexpected and surprising.

5 Conclusion

This paper has described a novel approach for discovering surprising exception rules using a simultaneous estimation method and a modified version of intensity of implication. The approach depends on neither a subjective evaluation nor an inspection of the reliability by a human. An exception rule discovered in this approach deviates from a strong common sense rule, has few counter examples and predicts a rare conclusion. Consequently, our exception knowledge discovery system, compared with PEDRE [10], is oriented to the discovery of surprising exception rules rather than reliable exception rules. We believe that both approaches are valuable, and should be used according to the goal of discovery.

Our approach has been applied to several benchmark data sets in the machine learning community. Experimental results show that our system is promising for the efficient discovery of surprising exception rules. Ongoing research is focused on the comparison of various rule evaluation criteria and practical applications of our approach to large databases.

References

1. Agrawal, R., Mannila, H., Srikant, R. *et al.*: Fast Discovery of Association Rules, *Advances in Knowledge Discovery and Data Mining*, AAAI Press/The MIT Press (1996) 307–328
2. Fleury, L., Djeraba, C., Briand, H. and Philippé, J.: Rule Evaluations in a KDD System, *Database and Expert Systems Applications*, Springer-Verlag (1995) 405–414
3. Gras, R. and Lehrer, A.: L’Implication Statistique: une Nouvelle Methode d’Analyse de Données, *Mathematiques, Informatique et Sciences Humaines*, **120** (1993) 5–31.
4. Lerman, I. C., Gras, R. and Rostam, H.: Elaboration et Evaluation d’un Indice d’Implication pour Données Binaire, *Mathematiques, Informatique et Sciences Humaines*, **74** (1981) 5–35
5. Merz, C. J. and Murphy, P. M.: UCI Repository of machine learning databases, <http://www.ics.uci.edu/~mllearn/MLRepository.html>, Univ. of California, Dept. of Information and Computer Sci. (1998)
6. Silberschatz, A. and Tuzhilin, A.: On Subjective Measures of Interestingness in Knowledge Discovery, *Proc. of KDD-95* (1995) 275–281
7. Smyth, P. and Goodman, R. M.: An Information Theoretic Approach to Rule Induction from Databases, *IEEE Trans. on Knowledge and Data Eng.*, **4** (4) (1992) 301–316
8. Suzuki, E. and Shimura, M.: Exceptional Knowledge Discovery in Databases based on Information Theory, *Proc. of KDD-96* (1996) 275–278
9. Suzuki, E.: Discovering Unexpected Exceptions: A Stochastic Approach, *Proc. of RSFD-96* (1996) 225–232
10. Suzuki, E.: Autonomous Discovery of Reliable Exception Rules, *Proc. of KDD-97* (1997) 259–262