The Outer Impartation Information Content of Rules and Rule Sets

Dan Hu[1] and Yuanfu Feng[2]

1.Beijing Normal University, College of Information Science and Technology
100875 Beijing, China
hufengdd@163.com
2.Beijing Union University,Basic Courses Department
Beijing 100101,China
longlonghd@163.com

Abstract. The appraisement of rules and rule sets is very important in data mining. The information content of rules is discussed in this paper and is categorized into inner mutual information and outer impartation information. We put forward the viewpoint that the outer impartation information content of rules and rule sets can be represented by relations from input universe to output universe. Then, the interaction of rules in a rule set can be represented by the union and intersection of binary relations expediently. Based on the entropy of relations, the outer impartation information content of rules and rule sets are well measured. Compared with the methods which appraise rule sets by their overall performance (accuracy, error rate) on the given test data sets, the outer impartation information content of rule sets is more objective and convenient because of the absence of test data sets.

Key words: interestingness measure of rules, relations, outer impartation information content of rules, outer impartation information content of rule sets

1 Introduction

The assessment of discovered knowledge has become a key problem in the domain of knowledge discovery by the emergence of enormous algorithms for knowledge acquisition. Rule is an important form of discovered knowledge, and interestingness measure is a common tool for the evaluation of it. In these thirty years, there have been many literatures focused on interestingness measures[1,2,3]. The interestingness measures can be divided into objective measures and subjective measures based on the estimator, a computer or human user. The subjective measures, such as Unexpectedness[4,5]and actionability[6], evaluate the rules by the synthesis of cognition, domain knowledge, individual experiences [7]. Variously, the objective measures, such as Coverage, Support, Accuracy and Credibility[7,8],etc, only depend on the structure of a pattern and the underlying data used in the discovery process. Detailedly, the objective interestingness can be categorized into some groups with the criterions and the theories for evaluation. The main criterions for the classification are Correctness[9], Generality[10], Uniqueness[11], and Information Richness[12,13]. We focus on Information Richness in this paper. Whichever criterion is used, there is a fact that almost all of the objective interestingness measures are determined by the quantitative information table of rules. Just because of this fact, most objective interestingness measures can not distinguish the rules having the same quantitative information table, even if these rules are obviously different. (The quantitative information table of a rule $r : c \to d$ is shown as table 1, where $m(c)$ denotes the set of elements which satisfy the condition expressed by c, and $|\cdot|$ denotes the cardinality of a set.)

Table 1. The quantitative information table of r

	d	$\neg d$	Total
c	$\|m(c) \cap m(d)\|$	$\|m(c) \cap m(\neg d)\|$	$\|m(c)\|$
$\neg c$	$\|m(\neg c) \cap m(d)\|$	$\|m(\neg c) \cap m(\neg d)\|$	$\|m(\neg c)\|$
Total	$\|m(d)\|$	$\|m(\neg d)\|$	$\|U\|$

example 1. Table 2 represents a database S.

Table 2. $S = (U,\ \{Temperature,\ Noise\} \cup \{Speed\},\ f,\ V)$

Temperature	*Noise*	*Speed*	*number of records*
Low	*Low*	*Low*	20
Medium	*Low*	*Medium*	30
Medium	*Normal*	*High*	10
High	*Normal*	*Medium*	10
High	*Normal*	*High*	30

r_1 and r_2 are rules induced from S, where

$r_1 :$ If Temperature is Low and Noise is Low, Then Speed is Low.

$r_2 :$ If Temperature is Low, Then Speed is Low..

The quantitative information table of r_1 is the same as r_2. There is no difference between the interestingness value of r_1 and r_2, if we estimate the interestingness of r_1 and r_2 by the measures which are absolutely decided by the quantitative information table(such as Support, Coverage, Recall, χ^2 measure, J-measure, Yao's interestingness measures, etc.[7,8]). But the fact is that r_2 is more general and interesting than r_1. So another new measure, which is not solely based on the quantitative information table, is needed to estimate the interestingness of rules.

On the other side, although lots of work have been done to discuss the interestingness of single rule, there are few literatures related to the appraisement of rule sets. Rule sets are usually assessed in terms of overall performance(accuracy, error rate) on the given test data sets. But the overall performance is not objective enough, because the result of evaluation depends on the choice of test data set. Aggregating the interestingness measures of single rule to appraise rule set is not a good way too, because the existed interestingness measures of single rule can not represent the interaction of rules in the rule set.

In this paper, we propose new measures for the estimation of rules and rule sets. At first, we classify the information content of rules to inner mutual information and outer impartation information. Then, the outer impartation information content of rules is represented by relations from input universe to output universe and measured based on the entropy of relations. Different from the objective measures proposed before, this new measure focuses on the corresponding relation between explanatory attributes and class attributes, and evaluates the information that helps human or receptor to make decision. Then, the union and intersection of relations are used to represent the interaction of rules in a rule set. A measure for the estimation of rule sets is well defined based on the information content of rules, and we named it as the outer impartation information content of rule sets.

2 The outer impartation information content of rules

Let $S =< U,\ A =\ C \cup D,\ V,\ f >$ be the database. U is the universe. $C = \{C_1,\ C_2, \cdots,\ C_n\}$ is the set of condition-attributes, and D is the set of decision-attributes (suppose that there is only one attribute included in D). $V = (\bigcup_{i=1}^{n} V_{C_i}) \cup V_D$, where $V_{C_i} = \{c_{i1},\ c_{i2},\ \cdots,\ c_{im_i}\}(i = 1, 2 \cdots, n)$, $V_D = \{d_1,\ d_2,\ \cdots,\ d_l\}$. $f :\ U \times A \to V$ is the evaluation function. A classification rule is defined as

$$r:\ \text{If } C_{k_1} \text{ is } c_{k_1 v_1} \text{ and } C_{k_2} \text{ is } c_{k_2 v_2} \text{ and } \cdots \text{and } C_{k_j} \text{ is } c_{k_j v_j},\ \text{Then } D \text{ is } d_v,$$

where $C_{k_1}, C_{k_2}, \cdots, C_{k_j} \in C$, $c_{k_t v_t} \in V_{C_{k_t}}$, $t = 1,\ 2,\ \cdots,\ j$, $v_t \in \{1,\ 2,\ \cdots,\ m_{k_t}\}$. For convenience, a classification rule can be abbreviated to the form $r:\ c \to d$, where c is the conjunction of condition-attribute values and d is the predicted class.

Concerning the information content of rules, we think that it should be categorized to two classes:

. **inner mutual information content**: How much information does the antecedent contribute to the consequent. The strength of the connection between the antecedent and consequent of rules.

. **outer impartation information content**: How much information is conveyed to the receptor and help the receptor to decide which is the predicted class in different conditions.

Symth and Goodman first define J-Measure[13] to measure the information content of rules,

$$J(r) = P(c)(P(d|c)\log(\frac{P(d|c)}{P(d)}) + (1 - P(d|c))\log(\frac{(1 - P(d|c))}{(1 - P(d))})) = P(c)j(d;c). \tag{1}$$

$j(d;c)$ measures the information that the antecedent c contributes to the consequent d[14]. The other part of J-Measure, $P(c)$, can be viewed as a preference for generality or simplicity of the rule $c \to d$. So, we can find the fact that J-Measure does not focus on the outer impartation information but inner mutual information content of rules.

There are other useful measures which use information theory to discuss the interestingness of rules[7]. Such as Normalized Mutual Information, Yao's interestingness and K-Measure. Just as the analysis of J-Measure, all of these measures focus on the inner mutual information content of rules.

To the best of our knowledge, there is no reported work related to measuring the outer impartation information content of rules. In fact, rules help the receptor to construct relations from condition-attributes to class-attributes, and the relations do help users to make decision and represent the outer impartation information content of rules. That is to say, the outer impartation information content of rules can be represented by relations from input universe to output universe. We have defined the entropy of relations in [15] to estimate the information conveyed by relations.

Definition 1.[15] Let U be the finite universe. $X, Y \in \wp(U)$, R is a relation from X to Y. $\forall x \in X$, $R(x) = \{y \in Y | (x, y) \in R\}$. R' is a relation defined by R,

$$R'(x_i) = \begin{cases} R(x_i) & x_i \in R^{-1}(Y) \\ Y & x_i \overline{\in} R^{-1}(Y), x_i \in X \end{cases} \tag{2}$$

The entropy of R restricted on X is denoted by $H(R \downarrow X)$ and defined as follows:

$$H(R \downarrow X) = -\sum_{x_i \in X} \frac{|R'(x_i)|}{\sum_{x_i \in X} |R'(x_i)|} \log \frac{|R'(x_i)|}{|Y|}. \tag{3}$$

The base of logarithm is 2, and $0 \log 0 = 0$.

Based on the entropy of relations, the outer impartation information content of rules can be easily defined and measured.

Definition 2. Let $S =< U, \ A = \ C \cup D, \ V, \ f >$ be the given database. $C = \{C_1, \ \cdots, \ C_n\}$, $V_{C_i} = \{c_{i1}, \ \cdots, \ c_{im_i}\}$, D contains only one element, $V_D = \{d_1, \ \cdots, d_l\}$. Suppose that we have induced a rule r_k from S,

$$r_k: \text{ If } C_{k_1} \text{ is } c_{k_1v_1} \text{ and } C_{k_2} \text{ is } c_{k_2v_2} \text{ and } \cdots \text{and } C_{k_j} \text{ is } c_{k_jv_j}, \text{ Then } D \text{ is } d_v,$$

where $C_{k_1}, \ C_{k_2}, \ \cdots, C_{k_j} \in C$, $c_{k_tv_t} \in V_{C_{k_t}}$, $t = 1, \ 2, \ \cdots, \ j$, $d_v \in V_D$. A relation R_{r_k} from $\prod_{i=1}^n V_{C_i}$ to V_D is defined to represent the outer impartation information content of r_k.

$$R_{r_k} = Re_{r_k} \cup Ru_{r_k}, \tag{4}$$

$$\begin{aligned} Re_{r_k} = \{(< c_{1h_1}, \cdots, c_{k_1v_1}, \cdots, c_{ih_i}, \cdots, c_{k_2v_2}, \cdots, c_{k_jv_j}, \cdots, c_{nh_n} >, d_v)| \\ h_1 \in \{1, \ \cdots, \ m_1\}, h_i \in \{1, \ \cdots, \ m_i\}, h_n \in \{1, \ \cdots, \ m_n\}\}, \end{aligned} \tag{5}$$

$$\begin{aligned} Ru_{r_k} = \{(< c_{1h_1}, \cdots, c_{k_1s_1}, \cdots, c_{ih_i}, \cdots, c_{k_2s_2}, \cdots, c_{k_js_j}, \cdots, c_{nh_n} >, d_q)| \\ h_1 \in \{1, \ \cdots, \ m_1\}, h_i \in \{1, \ \cdots, \ m_i\}, h_n \in \{1, \ \cdots, \ m_n\}, \\ c_{k_ts_t} \in V_{k_t}, t \in \{1, \ \cdots, \ j\}, \vee_{t=1}^{j} (c_{k_ts_t} \neq c_{k_tv_t}), \ q = 1, \cdots, l\}, \end{aligned} \tag{6}$$

Re_{r_k} and Ru_{r_k} represent the expanded information and unknown information of rule r_k, and they all relations from $\prod_{i=1}^n V_{C_i}$ to V_D.

Then, the outer impartation information content of r_k is measured by $IC(r_k)$,

$$IC(r_k) = Precision(r_k) \cdot H(R_{r_k} \downarrow \prod_{i=1}^{n} V_{C_i}), \tag{7}$$

where $Precision(r_k)$ is the Precision of r_k, $Precision(r_k) = P(d_v | \wedge_{t=1}^{j} c_{k_t v_t})$, and $H(R_{r_k} \downarrow \prod_{i=1}^{n} V_{C_i})$ is the entropy of R_{r_k} restricted on $\prod_{i=1}^{n} V_{C_i}$.

Remark 1. In definition 2, We do not use $H(R_{r_k})$[15] but $H(R \downarrow \prod_{i=1}^{n} V_{C_i})$ to measure the outer impartation information content of r_k, because classification rule has a direction from antecedent to consequent. If r_k is a rule without direction, such as association rule, we should use $H(R_{r_k})$ to substitute $H(R_{r_k} \downarrow \prod_{i=1}^{n} V_{C_i})$ in equation (7).

Remark 2. If you want to use some new measures to substitute one of the terms, $Precision(r_k)$ and $H(R_{r_k} \downarrow \prod_{i=1}^{n} V_{C_i})$ in equation (7), you must take care of your choice. The key of substitution is the trade-off between accuracy and generality. For example, Lift$(P(d_v|\wedge_{t=1}^{j} c_{k_t v_t})/p(d_v))$ or Relative Risk$(P(d_v|\wedge_{t=1}^{j} c_{k_t v_t})/P(d_v|\neg(\wedge_{t=1}^{j} c_{k_t v_t})))$ can be used to substitute $Precision(r_k)$.

Example 1.(continued) Compare r_1 and r_2 by J-Measure and $IC(*)$. The The outer impartation information content of r_1 and r_2 are shown in table 3 and table 4, where $\{L,\ M,\ H\}$ represents that the corresponding antecedent can lead to Low, Medium or high speed.

Table 3. The outer impartation information content of $r_1(R_{r_1})$

	Temperature	*Noise*	*Speed*
Original form of r_1	*Low*	*Low*	*Low*
Expanded Information(Re_{r_1})	*Low*	*Low*	*Low*
Unkonwn Information (Ru_{r_1})	*Low*	*Normal*	$\{L,\ M,\ H\}$
	Medium	*Low*	$\{L,\ M,\ H\}$
	Medium	*Normal*	$\{L,\ M,\ H\}$
	High	*Low*	$\{L,\ M,\ H\}$
	High	*Normal*	$\{L,\ M,\ H\}$

Table 4. The outer impartation information content of r_2 (R_{r_2})

	Temperature	*Noise*	*Speed*
Original form of r_2	*Low*		*Low*
Expanded Information (Re_{r_2})	*Low*	*Normal*	*Low*
	Low	*Low*	*Low*
Unkonwn Information(Ru_{r_2})	*Medium*	*Low*	$\{L,\ M,\ H\}$
	Medium	*Normal*	$\{L,\ M,\ H\}$
	High	*Low*	$\{L,\ M,\ H\}$
	High	*Normal*	$\{L,\ M,\ H\}$

Table 5. $J(*)$ and $IC(*)$

	r_1	r_2	Result
J(*)	$\frac{1}{5}\log 5$	$\frac{1}{5}\log 5$	r_1 is the same as r_2
$IC(*)$	$\frac{1}{16}\log 3$	$\frac{1}{7}\log 3$	r_2 is better than r_1

$IC(*)$ can distinguish r_1 and r_2, because the outer impartation information content of r_1 is less than r_2.

Proposition 1. $S =< U,\ A =\ C \cup D,\ V,\ f >$ is the given database, $C = \{C_1,\ \cdots,\ C_n\}$, $V_{C_i} = \{c_{i1},\ \cdots,\ c_{im_i}\}$, $V_D = \{d_1, \cdots, d_l\}$. r_k is the rule induced from S,

$$r_k:\ \text{If } C_{k_1} \text{ is } c_{k_1v_1} \text{ and } C_{k_2} \text{ is } c_{k_2v_2} \text{ and } \cdots \text{and } C_{k_j} \text{ is } c_{k_jv_j},\ \text{Then } D \text{ is } d_v.$$

Let $MIN = \min\{m_{l_1}, \cdots, m_{l_{n-j}}\}$, $(l_1, \cdots, l_{n-j} \in \{1, \cdots, n\} \backslash \{k_1, \cdots, k_j\})$. $K = \frac{\prod_{i=1}^{n} m_i}{\prod_{t=1}^{j} m_{k_t}}$,

(1). r_{k_s} is specialization of r_k. If the Precision of r_k satisfies the following inequality

$$Precision(r_k) = P(d_v | \wedge_{t=1}^{j} c_{k_tv_t}) \geqslant \frac{K + (\prod_{i=1}^{n} m_i - K) \times l}{K + (MIN \cdot \prod_{i=1}^{n} m_i - K) \times l}, \quad (8)$$

Then, $IC(r_{k_s}) \leq IC(r_k)$.

(2). Suppose r_{k_g} is the generalization of r_k, which is gotten by deleting the condition "C_{k_1} is $c_{k_1v_1}$" from r_k

$$r_{k_g}:\ \text{If } C_{k_2} \text{ is } c_{k_2v_2} \text{ and } C_{k_3} \text{ is } c_{k_3v_3} \text{ and } \cdots \text{and } C_{k_j} \text{ is } c_{k_jv_j},\ \text{Then } D \text{ is } d_v.$$

If

$$Precision(r_{k_g}) \leq Precision(r_k) \cdot \frac{m_{k_1} \cdot K + l \cdot \prod_{i=1}^{n} m_i - m_{k_1} \cdot K \cdot l}{m_{k_1} \cdot K + m_{k_1} \cdot l \cdot \prod_{i=1}^{n} m_i - m_{k_1} \cdot K \cdot l}, \quad (9)$$

then

$$IC(r_{k_g}) \leq IC(r_k). \quad (10)$$

Remark 3. When we want to use the way of specialization to capture more information, we should use proposition 1 to estimate if there is a chance to increase the outer impartation information content. On the other side, we always generalize a rule to make it suit for more situation. But we can't generalize the rule blindly. Otherwise, the outer impartation information may be lost in the process of generalization because of the decrease of Precision. Proposition 1 shows us there's a boundary for the Precision of r_{k_g}.

3 The outer impartation information content of rule sets

In section 2, the outer impartation information of a rule is represented by relations. Then, by the union and intersection of relations, we can deal with the interaction of rules in a rule set.

Definition 3. Suppose that we have induced a rule set $Q = \{r_k | k = 1, \cdots, h.\ h \geq 2\}$ from the database S. The expanded information of rule set Q can be defined as

$$Re_Q = \bigcup_{k=1}^{h} Re_{r_k}, \quad (11)$$

The unknown information of rule set Q can be defined as

$$Ru_Q = \bigcap_{k=1}^{h} Ru_{r_k}. \tag{12}$$

Thus the outer impartation information content of Q is represented by R_Q,

$$R_Q = Re_Q \bigcup Ru_Q. \tag{13}$$

$\forall \sigma \in R_Q$, we define the Precision of σ under Q and the Precision of Q as

$$Precision_Q(\sigma) = \max\{Precision(r_k) | r_k \in Q,\ \sigma \in Re_{r_k}\}. \tag{14}$$

$$Precision(Q) = \frac{\sum_{\sigma \in Re_Q} Precision_Q(\sigma)}{|Re_Q|}. \tag{15}$$

The outer impartation information content of rule set Q is measured by $IC(Q)$,

$$IC(Q) = Precision(Q) \cdot H(R_Q \downarrow \prod_{i=1}^{n} V_{C_i}). \tag{16}$$

Remark 4. If $Q = \{r_0\}$, then$IC(Q) = IC(r_0)$.

Remark 5. When there is $'\vee'$ exiting in the antecedent of a rule, we can transfer the rule to a rule set and deal with it by definition 3.

Example 1.(Continued) Suppose $Q = \{r_3,\ r_4\}$ is a rule set induced from database S, where

$$r_3: \quad \text{If Noise is Low, Then Speed is Low.},$$

$$r_4: \text{If Temperature is Medium, Then Speed is Medium..}$$

The outer impartation information content of Q is shown in table 6.

$$Precision(Q) = (0.4 + 0.4 + 0.4 + 0.75 + 0.75)/5 = 0.54;$$

$$IC(Q) = Precision(Q) \cdot (\frac{3}{10}\log\frac{3}{1} + \frac{1}{10}\log\frac{3}{2} + \frac{1}{10}\log\frac{3}{3}) = 0.2884.$$

Table 6. The outer impartation information content of Q (R_Q)

	Temperature	*Noise*	*Speed*
Original Form		*Low*	*Low*
	Medium		*Medium*
Expanded Information	*Low*	*Low*	*Low*
	Medium	*Low*	*Low*
	High	*Low*	*Low*
	Medium	*Normal*	*Medium*
	Medium	*Low*	*Medium*
Unknown Information	*Low*	*Normal*	$\{L,\ M,\ H\}$
	High	*Normal*	$\{L,\ M,\ H\}$

Definition 4. Let r_k and r_N be rules induced from database S. Let $W = \prod V_{C_i}$, $R \subseteq W \times V_D$. The antecedents of the elements in R is denoted by $R|_W$ and defined as

$$R|_W = \{\omega \in W | \exists d_v \in V_D,\ (\omega,\ d_v) \in R\}.$$

r_N is independent of r_k if and only if

$$(Re_{r_N}|_W) \cap (Re_{r_k}|_W) = \emptyset. \tag{17}$$

Proposition 2. Let $W = \prod V_{C_i}$, $|V_D| = l$. Suppose that we have induced a rule set $Q = \{r_k | k = 1, \cdots, h.\ h \geq 2\} (\neq \emptyset)$ from the database S. Q satisfies

$$\forall r_k \in Q, \quad IC(r_k) \neq 0. \tag{18}$$

then

(1).
$$0 \leq IC(Q) \leq \log l. \tag{19}$$

If $|R_Q| = |W| \cdot l$, then $IC(Q) = 0$. If $R_Q|_W = W$, $|R_Q| = |W|$ and $Precision(Q) = 1$, then $IC(Q) = \log l$ holds.
(2). If all rules in Q are independent from each other, then

$$IC(Q) > \sum_{i=1}^{m} IC(r_k). \tag{20}$$

Remark 6. In equation (19), the information content of a rule set may be zero. When we put rules together, the right result seems to be

$$IC(Q) \geq \max\{IC(r_1), \cdots, IC(r_h)\}. \tag{21}$$

But equation (21) is wrong at some time. There may be inconsistent information among the expanded information of rules. In table 6, there is inconsistency in the expanded information of Q:

$$\{(< Medium, Low >, Low), (< Medium, Low >, Medium)\}.$$

Based on the definition of the entropy of relations, it is easily seen that Adding a rule which has conflict with the former rule set will reduce the effective information.

Remark 7. The outer impartation information content of a rule set may be bigger than the sum of the outer impartation information content of each rule in the rule set.

When we estimate the relation between $IC(Q)$ and $\sum_{i=1}^{h} IC(r_k)$, the right result seems to be

$$IC(Q) \leq \sum_{i=1}^{h} IC(r_k), \tag{22}$$

But equation (23) does not always hold, and at sometime, we have

$$IC(Q) > \sum_{i=1}^{h} IC(r_k), \tag{23}$$

just as the result in proposition 2. This result is too strange to be trusted, but it is true. We can use an example which is more direct to explain this phenomenon. Suppose that we want to decide the occurrence of an event E_1. There are two relative evidences e_1 and e_2,

$$e_1: \quad \text{an event } E_2 \text{ has occurred,}$$

$$e_2: \quad E_2 \text{ must occur with } E_1.$$

If we want to solely use e_1 or e_2 to decide the occurrence of E_2, we'll get no information and can't reach the decision. But when we use the evident set $\{e_1,\ e_2\}$, the result is obvious.

The strength of collectivity is much bigger than the sum of individual's at sometime. This is a well known theory and is also held by the society of rules. Different rules interact with each other and remedy each other. This is just the reason of the result shown in equation (23).

4 Experiment

The outer impartation information content(OIIC) of rule sets can be easily incorporated with knowledge discovery algorithms to appraise and compare them without test data sets. The algorithm with maximal OIIC value is the best.

The well-known iris data set proposed by R.A.Fisher is used to compare the function of accuracy and OIIC measure of rule set. In [16], Iris database was transferred to a discrete database by fuzzy partition. We use different measures(SIG,Add,J-measure,Natual Order,ICR) to learn maximal structure rules for the data after discretization. 'Accuracy' and 'OIIC' are used to appraise the rule sets induced by different measures. In table 7, the accuracy of the induced rule set is estimated by Leaving-one-out method.

Table 7. The comparison of accuracy and OIIC measure of rule set

	SIG	$Add\ Value$	$J-measure$	$Natual\ Order$	ICR
$Accuracy$	0.8529	0.9412	0.8919	0.9118	0.9611
$OIIC\quad value$	0.8389	0.8721	0.8476	0.8945	1.1767

The appraisement result of accuracy and OIIC are

$$ICR > Add\ Value > Natural\ order > J-measure > SIG$$

and

$$ICR > Natural\ order > Add\ Value > J-measure > SIG$$

respectively, where the meaning of '>' is 'better than'. We find the fact that the appraisement result of accuracy is almost the same as OIIC. But it is obvious that the information content measure of rule sets is better because we need not to test the rule sets under different test sets, and the results independent of test data sets are more objective.

Acknowledgments. This research is supported in part by National Natural Science Foundation of China (Grant No. 60774049; No. 40672195);Natural Science Foundation of Beijing(Grant No. 4062020); the National 973 Fundamental Research Project of China (Grant No. 2002CB312200) and the Youth foundation of Beijing Normal University.

References

1. Yao, Y. Y. and Zhong, N.:An analysis of quantitative measures associated with rules. Proceedings of Pacific-Aisa Conference on Knowledge Discovery and Data Mining,(1999)479-488
2. Hilderman, R. J. Hamilton H J.:Knowledge Discovery and Measures of interest. Kluwer Academic Publishers(2001)
3. Ohsaki, M. and Sato,Y.:Investigation of rule interestingness in medical data mining. Lecture Notes in Artificial Intelligence, 3430(2005)174-189
4. Balaji, P and Alexander, T.:Unexpectedness as a measure of interestingness in knowledge discovery.Decision Support Systems 27(1999)303-318
5. Liu, B. and Hsu, W.:Post-Analysis of learned rules. In proceedings of AAAI,(1996) 828-834
6. Ailberschatz, A and Tuzhilin, A.:What makes patterns interesting in knowledge discovery systems. IEEE Transact on knowledge and Data Engineer, 8(6)(1996)970-974
7. Ohsaki,M. and Sato, Y.: Evaluation of rule interestingness measures with a clinical data set on hepatitis. Lecture Notes in Artificial Intelligence, 3202(2004)362-373
8. Hamilton, H. J. and Shan N.:Machine Learning of credible classifications. Proceedings of Australian Conference on Artificial Intelligence,(1997)330-339
9. Shapiro,G. S.:Discovery, analysis and presentation of strong Rules. Knowledge Discovery in Databases, AAAI/MIT press,(1991) 229-248
10. Gago, P. and Bento, C.:A metric for selection of the most promising rules. In Proceedings of European Conference on the Principle of Data Ming and Knowledge Discovery,(1998)19-27
11. Zhong, N. and Yao, Y. Y.:Peculiarity oriented multi-database mining. In Proceedings of European Conference on Principles of data Mining and Knowledge Discovery, (1999)136-146
12. Hamilton, H. J. and Fudger, D. F.: Estimationg DBLearn's potential for knowledge discovery in databases. Computational Intelligence, 11(2)(1995) 280-296
13. Symth,P and Goodman, R. M.:Rule induction using information theory. Knowledge Discovery in Databases,AAAI/MIT press(1991)
14. Shore,J. E. and Johnson R. W.:Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. IEEE transactions on Information Theory,26(1)(1980)26-37
15. Hu D.and Li H. X.:The entropy of relations and a new approach for decision tree learning, Lecture Notes in Artificial Intelligence,3614,(2005) 378-388
16. Castro,J.L. and Castro-Schez,J.J.:Learning maximal structure rules in fuzzy logic for knowledge acquisition in expert systems, Fuzzy Sets and Systems,101(1999),331-342