A Decision Analysis Method Based on Rough Set

Shenghai Chen[1, a], Zongyan Jiang[2], Chao Duan[1], Xuejing Dai[1], Xianwu Mi[1]

[1]College of Electrical and Information Engineering, Huaihua University, Huaihua 418000, China;

[2]College of Mathematics and Statistics, Hunan normal University, Changsha 410006, China.

[a]xtusch@163.com

Abstract. Rough sets theory provides a good method for getting excellent decision rules set with the conduct of reduction from the decision table. This paper describes many evaluation metrics of decision rules, analyzes the properties for these evaluation metrics. Then it presents the evaluation metrics for rule set, which globally shows the properties of a rule set. And the evaluation system plays an important role for one try to choose the right decision rules in the research of rough set theory.

Keywords: rough set; rules, rule set; evaluation metrics; decision analysis.

1. Introduction

A lot of useful information and knowledge is hidden in voluminous data and to discover useful information from those voluminous data called knowledge discovery. To discover the implied knowledge and disclose those implied rules from the voluminous data through analysis, classification and reasoning of these datas an important research method of knowledge discovery [1]. The Rough Sets theory [2] is a theory which classifies, simplifies and acquires knowledge till the forming of a rule from the raw data by the use of information tables. Therefore, rule generation is a significant application of Rough Sets theory [3,4].

It the paper [5][6], the authors have given some quantitative metrics associated with the single rule we generated. However, the sample datas in the decision table always interfere by noise and other factors, which makes the rules with uncertainty. So, it's necessary to analysis the measures of the rules. In this paper, we describe many metrics of decision rules, analyzes the properties for these metrics. Then it presents the metrics for rule set, which shows the properties of a rule set in general. And this will play an important role for one try to choose the right decision rules in the research of rough set theory.

2. Single Rule Evaluation

In order to describe the properties of decision rules, many quantitative metrics associated with rules have been studied [9]. We review some metrics for single rule evaluation.

2.1 Granularity

Definition 3.1 Let S= (U, Q=C$\cup$D, V, F) be a decision table, $X_i \in$U/IND(C), $Y_j \in$U/IND(D) then is referred to as the granularity of the condition class Xi and the decision class Y_j.

$$P(X_i) = \frac{|X_i|}{|U|}, 0 \le P(X_i) \le 1, i=1,2,...,n$$

$$P(Y_i) = \frac{|Y_i|}{|U|}, \ 0 \le P(Y_i) \le 1, j=1,2,...,m$$

For rules generated by the indiscernibility relation, the granularity of the entire data collection is the manifestation of the generalization ability of the decision table rules, the greater the granularity of a condition class the stronger generalization ability of the corresponding rules. Condition class

granularity reflects the probability of a rule that may occur from the point of view of the probability, the decision class granularity reflects the degree of the importance of a rule.

2.2 Support

Definition 3.2 Let $S= (U, Q=C\cup D, V, F)$ be a decision table, $X_i\in U/IND(C)$, $Y_j\in U/IND(D)$ then is called the support of the decision rule $X_i\to Y_j$, where $|X|$ denotes the cardinality of X.

$$\text{The number}\quad supp(X_i,\ \ Y_j) = \frac{\left|X_i\cap Y_j\right|}{\left|U\right|}, 0\le supp(X_i,Y_j)\le 1$$

The support means the strength of the decision rule $X_i\to Y_i$

The support also reflects the randomness degree of the decision table, when a rule has a small support, we can think that it is a random rule, and it has poor classification for new objects.

2.3 Confidence

Definition *3.3* Let $S= (U, Q=C\cup D, V, F)$ be a decision table, $X_i\in U/IND(C)$, $Y_j\in U/IND(D)$ then

$$\text{The number}\quad cer(X_i,Y_j) = \frac{\left|X_i\cap Y_j\right|}{\left|X_i\right|}, 0\le cer(X_i,\ \ Y_j)\le 1$$

is called the confidence of the decision rule $X_i\to Y_j$ in S, where $X_i\ne\varphi$.

When, $cer(X_i, Y_j)=1$, $X_i\to Y_j$ is a certainty decision rule;

When, $0<cer (X_i, Y_j) <1$, $X_i\to Y_j$ is an uncertainty decision rule.

In fact, when the positive region $POS(C,D)=U$, the rules are generated certainty ; otherwise it will create uncertainty rules. From the decision table we can straightforward see that the uncertainty rules is because of there contain objects that have the same condition attributes but the different decision attributes.

Confidence focuses on the classification of an equivalence class X_i. The higher the confidence, the stronger the rule is.

We can conclude that, granularity reflects the decision rule's generalization ability, support reflects the strength of the decision rules and confidence is the full extent of the rules.

2.4 Coverage

Definition *3.4:* Let $S= (U, Q=C\cup D, V, F)$ be a decision table, $X_i\in U/IND(C)$, $Y_j\in U/IND(D)$ then

$$\text{The number}\, cov(X_i,\ \ Y_j) = \frac{\left|X_i\cap Y_j\right|}{\left|Y_j\right|}, 0\le cov(X_i,\ \ Y_j)\le 1$$

is called the coverage of the decision rule $X_i\to Y_j$ in S, where $Y_j\ne\varphi$.

The coverage measure of a rule is defined as the ratio of the number of correctly classified objects in the decision class Y_j by a condition class X_i and the number of objects in the decision class Y_j, which shows an inverse relationship between confidence and coverage. And a rule with a higher coverage is more general with respect to the decision class Y_j.

3.　Rule Set Evaluation

Above, we analyzed the metrics of a single rule, described the probabilistic of a single decision rule. However, in fact, in order to measure the decision performance of the rule knowledge base from a sample datas and to compare the rule sets, it's more necessary to analyze the overall performance of the rule set from a decision table.

3.1 The Average Granularity

From the decision table, the average granularity is weighted sum of the multiply of the probability of each equivalence class and the granule of each equivalence class. Since, the probability of the equivalence class is the ratio of the equivalence class size of the universe, coincided with the granularity from the size of the form and value is the same.

Definition *4.1* Let $S= (U, Q=C \cup D, V, F)$ be a decision table, $X_i \in U/IND(C)$, $Y_j \in U/IND(D)$then
The average granularity of condition class:

$$\pi(C) = \sum_{i=1}^{n} \left[P(X_i) \cdot \frac{|X_i|}{|U|} \right] = \sum_{i=1}^{n} \left[\frac{|X_i|}{|U|} \cdot \frac{|X_i|}{|U|} \right] = \sum_{i=1}^{n} \frac{|X_i|^2}{|U|^2}$$

The average granularity of decision class:

$$\pi(D) = \sum_{j=1}^{m} \left[P(Y_j) \cdot \frac{|Y_j|}{|U|} \right] = \sum_{j=1}^{m} \left[\frac{|Y_j|}{|U|} \cdot \frac{|Y_j|}{|U|} \right] = \sum_{j=1}^{m} \frac{|Y_j|^2}{|U|^2}$$

From the formula we can see that when all the conditions (or decisions) classes share of the smallest. That is to say, there are only one object in every class, the average granularity has the smallest value $1/|U|$; when the universe constitute one class, the average granularity get the biggest value $|U|$;This is consistent with the idea of the references[10].

From the perspective of the consistency for the rule we can see that the average granularity has great influence on the degree of uncertainty of the rules. If the granularity of the condition class $U/IND(C)$ is too large, then it's easy to compute a lot of uncertainty rules, and the classification rate is also lower. On the contrary, if the granularity is too small, then the classification rate will be higher. But it makes every rule support less object, and the known data to the rule's support degree is also not enough, lacking of representativeness of the data, then the classification predicted reliability of new object is greatly decreased. From the decision classification aspect, the situation is just on the opposite.

3.2 The Average Support

Let $S= (U, Q=C \cup D, V, F)$ be a decision table, $X_i \in U/IND(C)$, $Y_j \in U/IND(D)$. The average support is defined as follows:

$$D\,\text{supp}(C \to D) = \sum_{i=1}^{n} \sum_{j=1}^{m} \left[\frac{|X_i \cap Y_j|}{|U|} \bullet \frac{|X_i \cap Y_j|}{|U|} \right] = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{|X_i \cap Y_j|^2}{|U|^2}$$

The average support describes the average strength of the rule set. From the decision table we can see that the average support is weighted sum of the multiply of the probability of each rule and the rule's support .The probability of each rule is the ratio of the number of objects in the universe, which is the same with the support from the size of the form and value.

From the reduction aspect, for example, suppose an attribute reduction, made the condition equivalence class X_{n-1} and X_n combining, and get a new condition equivalence class $X_{n-1}+ X_n$. Then we can see the average support's change:

$$\Delta D\,\text{supp}(C \to D) = \sum_{j=1}^{m} \frac{\left|(X_{n-1} \cup X_n) \cap Y_j\right|^2 - \left|X_{n-1} \cap Y_j\right|^2 - \left|X_n \cap Y_j\right|^2}{|U|^2} = \sum_{j=1}^{m} \frac{T}{|U|^2}$$

$$\text{Let } \left| X_n \cap Y_j \right| = a_1, \left| X_{n-1} \cap Y_j \right| = a_2$$

Thus, we have $T=(a_1+a_2)^2-a_1^2-a_2^2=2a_1a_2 \geq 0$

Obviously, $\Delta D \operatorname{supp}(C) \geq 0$. So, we can conclude that reduction makes the rule set average support increasing.

3.3 The Average Confidence

Let $S= (U, Q=C \cup D, V, F)$ be a decision table, $X_i \in U/IND(C)$, $Y_j \in U/IND(D)$. We define the confidence measure as follows:

$$\text{D } cer(C \to D) = \sum_{i=1}^{n} \sum_{j=1}^{m} \operatorname{supp}(X_i, Y_j) cer(X_i, Y_j)$$

That is, for rule set, its confidence is weighted sum of the strength and confidence of individual rules in the set. We can see some special cases of the average confidence from Table 1.

Table 1. Analysis of the rule set confidence in the special case

Condition classification	Decision classification	supp(X_i, Y_j)	cer(X_i, Y_j)	Dcer(C→D)	Conclusion
Random	|U/IND(D)|=1	$\pi(X_i)$	1	1	Decision constant The classification ratio constant 1
|U/IND(C)|=1	|U/IND(D)|=|U|	1/|U|	1/|U|	1/|U|	Complete uncertainty Minimum confidence
|U/IND(C)|=|U|		1/|U|	1	1	Certainty decision Single object rule

From the table we can see that complete uncertainty of the rule set calculated by formula, is entirely consistent with the usual understanding of the maximum uncertainty. This shows the Dcer(C→D) objectivity.

Thus, for the confidence of the rule set, its confidence presents a non-monotonically decreasing trend with the conduct of the reduction, its also reflects the uncertainty of the decision table increased, they have the consistency.

3.4 The Average Coverage

Let $S= (U, Q=C \cup D, V, F)$ be a decision information system, $X_i \in U/IND(C)$, $Y_j \in U/IND(D)$. We define the average coverage as following:

$$\text{D } cov(C \to D) = \sum_{j=1}^{m} \sum_{i=1}^{n} \operatorname{supp}(X_i, Y_j) cov(X_i, Y_j)$$

For rule set, its coverage is the weighted sum of the strength and coverage of individual rules in a set. We can see some extreme cases from Table 2.

Table 2. Analysis of rule set coverage in the special cases

Condition Classification	Decision Classification	Supp($X_i,$ Y_j)	cov($X_i,$ Y_j)	Dcov(C→D)	Conclusion
|U/IND(C)|=1	Random	$\pi(Y_i)$	1	1	Condition constant
|U/IND(C)|=|U|	|U/IND(D)|=1	1/|U|	1/|U|	1/|U|	Complete uncertainty
	|U/IND(D)|=|U|	1/|U|	1	1	certainty decision

From the table we can see that the complete uncertainty of the rule set calculated by formula, which is entirely consistent with the usual understanding of the maximum uncertainty. This also shows the Dcov($C{\to}D$) objectivity.

Thus, for the coverage of the set of rules, its presents a non-monotonically increasing trend with the conduct of the reduction, while it had an opposite trend to the confidence. This can be intuitively considered that, with the conduct of the reduction, the condition granule increased. So when a decision class happened, judge the credibility of the conditions of its occurrence increases.

4. Multiple Knowledge Model Integration

A rule set can be generated by the information system with a reduction, it can be used as redundant knowledge to classifying and discerning the new target. Rule set for an important objective is to obtain a minimum set of rules and the minimum length of the descriptor set of rules. However, in many cases, using a minimum set of rules classify new objects is not sufficient. Because a minimum set of rules only represent an individual knowledge base classification ability, and it's only generated by information system with a reduction which is sensitive to noise.

Naturally, the redundant knowledge can be generated by the information system with multiple reduction, each knowledge base can be viewed as a single decision model. On this basis, we can select model and implementation model integration. We can show the redundant knowledge model integration framework as follow:
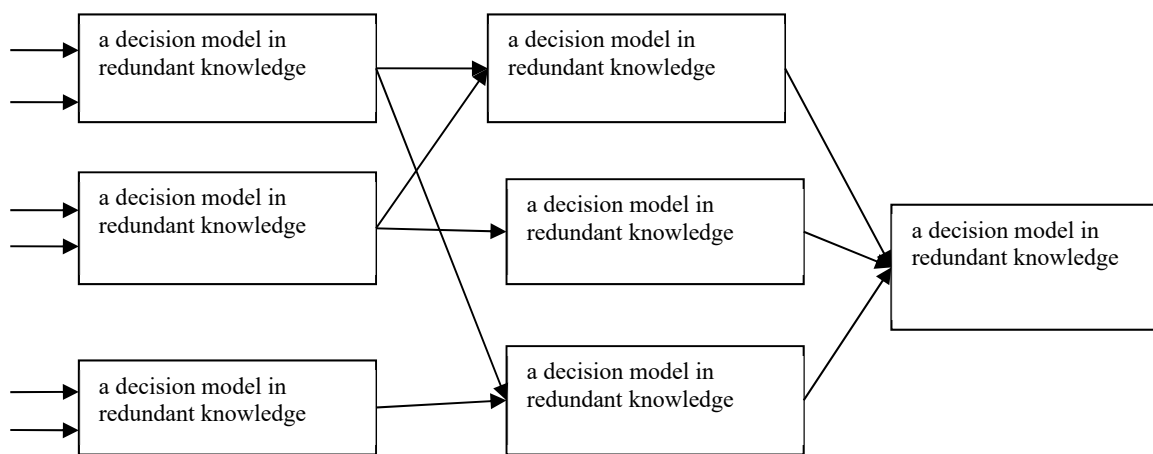

Fig 1. model integration based on redundant knowledge

Figure 1 shows a composite model and each model is a knowledge base decision model from information systems, the selection criteria is according to the information systems metrics to weigh, from left to right reflects the model integration. Model selection is a combination of dynamic process,

complex model ultimately represents a comprehensive decision on an issue, to achieving the modeling of complex problems.

In our analysis of the project power load, we can analysis the week load from the day load and use the week load to analysis the month or year load. Model integration is a process of progressive, which is the concrete realization of the system we propose a new method. Redundant knowledge generated by the reduction and according to the selected metrics through integration to achieve integrated its biggest feature is that it can meet the various needs of the comprehensive structural analysis model.

5. Summary

For the decision rules generated from the information system, there have been many methods. The algorithms often use one of the following strategies:1. generating minimum rule set which cover all objects from the information systems. 2. generating rule set which have the maximum generalization ability from the information systems. 3. generating approximate rule set which has good noise immunity.

Usually we want the data in knowledge representation system to be consistently, or covering more condition class, or covering more decision class. This is a general goal to pursue and it is particularly effective in the information systems which contains many classes. For the class in information system which has a small support, as less than a given threshold of support, etc., in certain circumstances they can be regarded as noise data, helping to get more simple and applicable rules.

We have argued the evaluation metrics of the decision rules; the single rule evaluation metrics is a measure of a individual rule. Evaluation metrics of the rule set reflects the property of rule set in general, and reduction analysis is a manifestation of the changes in the measure of rule set. These comprehensive descriptions of the property of the rule set from multiple perspectives providing an integrity basis of the decision-making model selection for the redundant knowledge decision fusion.

Acknowledgments

References

[1]. ZHANG Wen-xiu, WU wei-zhi, LIANG Jie-ye, et al. Theory and method of rough sets[M]. Beijing: Science and Technology Press, 2001.

[2]. Pawlak Z. Rough sets [J]. International Journal of Information and Computer Science, 1982 ,11 (5) :341-356.

[3]. Pawlak Z. Decisions rules and flow networks. European Journal of Operational Research, 2004,154(1):184 ~190.

[4]. Hu X, Cercone N. Mining knowledge rules from databases: A rough set approach[C]. Data Engineering, 1996. Proceedings of the Twelfth International Conference on. IEEE, 1996: 96-105.

[5]. Pawlak Z. Theorize with Data using Rough Sets, Proceedings of the26th Annual International Computer Software and Applications Conference, 2002.

[6]. D' Eer L, Cornelis C, Yao Y. A semantically sound approach to Pawlak rough sets and covering-based rough sets[J]. International Journal of Approximate Reasoning, 2016, 78(C):62-72.

[7]. Narli S. An alternative evaluation method for Likert type attitude scales: rough set data analysis[J]. Sci Res Essays, 2010, 5.

[8]. Ju H, Li H, Yang X, et al. Cost-sensitive rough set: A multi-granulation approach[J]. Knowledge-Based Systems, 2017.

[9]. Pawlak Z. In pursuit of patterns in data reasoning from data-the rough set way[C]: Rough Sets and Current Trends in Computing. Springer Berlin Heidelberg, 2002: 1-9.

[10].   MIAO Duo-qian, FAN Shi-dong. The Calculation of Knowledge Granulation and Its Application[J], Systems Engineering-theory & Practice, 2002,22(1):48~56.