Knowledge Reduction and Probability Rule Induction Based on Rough Sets

Hui Yang [1, a], Zhengsheng Ding [2, b]

[1]College of Science, Xi'an University of Science and Technology, Xi'an 710054, China;

[2]College of Geology and Environment, Xi'an University of Science and Technology, Xi'an 710054, China

[a]yangh816@aliyun.com, [b]dingzs@xust.edu.cn

Abstract

Rough sets theory is a traditional method for information reduction and rules acquisition under the circumstance of incomplete information. In this paper, the new version of reduction algorithm based on discernibility matrix was proposed through revising the common algorithm. The new algorithm is more efficient in finding out the criminal attribute. In the decision information system, using the definition of accuracy rate and coverage rate, the positive rule, negative rule and exclusive rule were given according to the definition of probability rules.

Keywords

Rough Sets, Knowledge Reduction, Discernibility Matrix, Probability Rule.

1. Introduction

Rough sets theory was proposed by Z. Pawlak to solve the problems of uncertain and inaccuracy [1]. The fuzziness and uncertainty of classification was defined in the rough sets theory. According to this theory, knowledge was the partition of domain and can be analyzed by equivalence relation in the field of algebra. Nowadays, rough sets theory has been wildly used to many fields such as data mining, machine learning and pattern recognition [2,3,4,5] etc.

In the paper, a revised reduction algorithm based on the discernibility matrix of rough sets theory was proposed. The revised algorithm is more efficient in finding out the criminal attributes. In the decision information system, using the definition of accuracy rate and coverage rate in rough sets theory, the positive rule, negative rule and elimination rule were given according to the definition of probability rules.

The paper is organized as follows. Section 2 gives the basic definition of rough sets theory. Section 3 introduces the revised algorithm based on the traditional attributes reduction. Section 4 gives three kinds of probability rule. Section5 is the conclusion of the paper.

2. Basic Definition of Rough Sets Theory

Information System (IS) was used to describe the knowledge in rough sets theory and its definition was given as follows,

Definition 1[5] (U, A, F) a Information System, where

U is a non-empty finite set of objects, $U = \{x_1, x_2, \ldots, x_n\}$

A is a non-empty finite set of attributes, $A = \{a_1, a_j, \ldots, a_p\}$

F is the relation set between U and A, $F = \{ f_j \mid f_j : U \to V_j,\, j \le p\}$ and V_j is the domain of attribute a_j

Definition 2[5] (U, A, F, D, G) is a Decision Information System (DIS), where

(U, A, F) is a Information System,

D is a non-empty finite set of decision attributes, $D = \{d_{1,} d_{2,} ... d_q\}$,

G is the relation set of U and D, $G = \{g_j \mid g_j : U \to V_j^{'}, j \le q\}$ and $V_j^{'}$ is the domain of decision attribute d_j.

Using the Definition 2, a new decision attribute can be defined as d, and its domain V_d is the subsets of $V_{d_1} \times V_{d_2}, \ldots, \times V_{d_q}$, and the following formula holds up,

$$g(x, d) = (g(x, d_1), g(x, d_2), \ldots, g(x, d_q))$$

thus a multi-objective decision information system was converted into a single objective decision information system. In this reason, only the single-object problem where $q = 1$ was discussed in the paper, correspondingly the decision information system was written by (U, A, F, d, G).

Suppose that $B = A \bigcup \{d\}$, its atomic formulas is in the form of $[a = v]$ which was the smallest set of all atomic formulas and closed for the disjunction and conjunction operator.

$\forall f \in F(B, V)$, f_A is the explanation of the formula f, which means f_A is the set of all samples (in the domain U) that were characterized by f. In other words, f_A satisfies the following conditions, suppose that f is in the form of $[a = v]$, then $f_A = \{s \in U \mid a_s = v\}$

$$(f \wedge g)_A = f_A \bigcap g_A \quad (f \vee g)_A = f_A \bigcup g_A \quad (\neg f)_A = U \setminus f_A$$

The accuracy rate and coverage rate of classification were introduced to evaluate the decision rules as follows.

Definition3 Let R be a formula of $F(B, V)$ and D be a sample set belong to decision attribute d, then the accuracy rate $\alpha_R(D)$ and coverage rate $\kappa_R(D)$ of the rule $R \to D$ can be defined as:

$$\alpha_R(D) = \frac{|R_A \bigcap D|}{|R_A|} (= P(D \mid R))$$

$$\kappa_R(D) = \frac{|R_A \bigcap D|}{|D|} (= P(R \mid D))$$

Given (U, R), where R is the equivalence relation in domain U ,then $\forall X \subseteq U$,

$$\underline{R}(X) = \{x \in U; [x]_R \subseteq X\} = \bigcup \{[x]_R; [x]_R \subseteq X\}$$

$$\overline{R}(X) = \{x \in U; [x]_R \bigcap X \ne \varnothing\} = \bigcup \{[x]_R; [x]_R \bigcap X \ne \varnothing\}$$

are the lower-approximate and upper-approximate respectively[5].

3. Attributes Reduction Algorithm Based on Discernibility Matrix

The essential of attributes reduction based on rough sets theory is the reduction of decision table, this process of reduction will not change the dependency relationship between the condition attributes and decision attributes.

Existing reduction algorithms can be classified into two categories, heuristic algorithms based on information entropy and algorithms based discernibility matrix and discernibility function .The former could not find out all the reduction of the system, so the latter is the mainstream of research. The idea of the algorithm is computing disjunctive normal form of the discernibility function which was derived from the discernibility matrix. Every conjunctive clause of the above disjunctive normal form is just a reduction of the information system.

The exact definition of the discernibility matrix and the core are defined as, Definition 4 Given decision table system $S=<U,R,V,f>$, where $R=P\bigcup D$ is the attributes set, its subset $D=\{d\}$ and $P=\{a_i|i=1,2,...,m\}$ are the decision attributes set and condition attributes set respectively. $U=\{x_1,x_2,...,x_n\}$ is the domain set, $a_i(x_j)$ is the value of attribute a_i corresponding to sample x_j, $C_D(i,j)$ is the elements of the *i*th row and *j*th column in the discernibility matrix, then the discernibility matrix C_D is

$$C_D(i,j)=\begin{cases}\{a_k|a_k\in P\wedge a_k(x_i)\neq a_k(x_j)\} & d(x_i)\neq d(x_j)\\ 0 & d(x_i)=d(x_j)\end{cases}$$

Core is a set composed of all the singleton of discernibility matrix, expressed by

$$core(A)=\{a\in A|a(x,y)=\{a\},x,y\in U\}$$

The attributes reduction algorithm based on discernibility matrix was described as follows,

Attributes Reduction Algorithm Based on Discernibility Matrix(ARAD)

Step1: computing the discernibility matrix C_D of decision table,

Step 2: for every element C_{ij} of the C_D, and $C_{ij}\neq\varnothing$, disjunctive normal form corresponding to C_{ij} can be got as $L_{ij}=\underset{a_i\in C_{ij}}{\vee}a_i$,

Step 3: According to all the disjunctive normal form , conjunction normal form L can be computed through $L=\underset{C_{ij}\neq 0,C_{ij}\neq\varnothing}{\wedge}L_{ij}$

Step 4: transforming the conjunction normal form L into the disjunctive normal form $L^{'}=\underset{i}{\vee}L_i$

Then every conjunctive clause of the above disjunctive normal form is a reduction of the information system.

The above reduction algorithm is essentially the processing of logic formulas. Then the attribute reduction algorithm can be revised to be more efficient in the following way:

Checking the discernibility matrix, if there is an element that including the Core, then this attribute is the unique crucial to distinguish two attribute-related samples that relative to this element, moreover it is unique . We can take out these crucial attributes meanwhile revised the elements that including the core to be 0. Consequently, we got a new matrix and can execute the step 2,3and 4 in the revised matrix. The improved algorithm will make the reduction more efficient than ever.

4. Three Kinds of Probability Rules

The accuracy rate and coverage rate [6] of classification are useful in application [7, 8], nevertheless they are two main index to evaluate the rules. Combining the idea of rule acquisition in rough sets theory, we have the following theorem,

Theorem1 The accuracy rate $\alpha_R(D)$ evaluates the sufficiency of a rule $R\to d$ and the coverage rate $\kappa_R(D)$ evaluates the necessity of the rule, if $\alpha_R(D)=\kappa_R(D)=1$, then $R\leftrightarrow d$ and now $R_A=D$

Proof:

$$\alpha_R(D)=1.0\Rightarrow|R_A\bigcap D|=|R_A|\Rightarrow R_A\subseteq D$$

This means R_A is a subset of D and $R\to d$ holds, similarly,

$$\kappa_R(D)=1.0\Rightarrow|R_A\bigcap D|=|D|\Rightarrow D\subseteq R_A$$

This means D is a subset of R_A and $d\to R$ holds.

Above all, $R_A=D$

The theorem shows that the accuracy rate and coverage rate of classification characterize the sufficiency and necessity of a rule.

Using the above theorem and the of definition $\alpha_R(D)$ and $\kappa_R(D)$, we deduce the probability rule, given δ_α, $\delta_\kappa \in R^+$ as threshold values of $\alpha_R(D)$ and $\kappa_R(D)$,

$$R \xrightarrow{\alpha,\kappa} d \text{, satisfies } \alpha_{R_i}(D) \ge \delta_\alpha \text{, } \kappa_{R_i}(D) \ge \delta_\kappa$$

is a probability rule, where $R = \vee_i R_i = \vee \wedge_j [a_j = v_k]$.

Especially when $\alpha_R(D) = 1.0$, the rule was only supported by positive samples, so it was named as positive rule and was defined as,

$$R \to d \text{, satisfies } \alpha_R(D) = 1.0 \text{,}$$

where $R = \wedge_j [a_j = v_k]$

Similarly, when $\kappa_R(D) = 1.0$, we have the exclusive rule that was supported by any one of the positive samples, presented by,

$$R \to d \text{, satisfies } \kappa_R(D) = 1.0 \text{,}$$

where $R = \vee_j [a_j = v_k]$

The inverse negative proposition can be written as,

$$\wedge_j \neg [a_j = v_k] \to \neg d \text{, satisfies } \forall [a_j = v_k] \text{, } \kappa_{[a_j = v_k]}(D) = 1.0$$

Which was named as negative rule that means if there exists a sample that doesn't satisfy any attribute of the negative rule, then delete its corresponding decision attribute d.

5. Conclusion

In this paper, the traditional reduction algorithm based on discernibility matrix was revised. The new algorithm is much more efficient than the traditional one. Furthermore, under the circumstance of decision information system, based on the definition of accuracy rate and coverage rate of classification, the Positive rule, Exclusive rule and Negative rule were defined.

The paper researched the field of information system through combining the probability knowledge to the rough sets theory, however, only the reduction problem and rule induction algorithm were studied. Thus in the future, there are many works can be done deeply in the research of rough sets theory which was integrated with the probability.

Acknowledgements

Project supported by the Training Program Funded by Xi'an University of Science and Technology. (2010043) and Online Open project of Probability and Statistic Funded by Higher Education Press. (6000150246)

References

[1] Pawlak Z: Rough Sets Theoretical Aspects of Reasoning about Data (Klawer Academic publishers, Bosten, 1991), p.12-15.

[2] Shusaku Tsumoto: Rule Induction System based on Characterization of Medical Diagnosis using Rough Sets, IFSA World Congress and 20th NAFIPS International Conference, 2001. Joint 9th (25-28 July 2001) , vol. 4 (2001), p.2364 – 2369.

[3] D.Molodtsov: Soft Set Theory-First Results, Computer and Mathematics with Applications,vol. 37(1999), p.19-31.

[4] P.K.Madj and A.R.Roy: An Application of Soft Sets in A Decision Making Problem, Computers and Mathematics with Applications vol. 44(2002), p.1077-1083.

[5] Zhang Wenxiu, Wu Weizhi and Liang Jiye: Rough sets theory and methods (Science publishers, China, July 2001), p.26-27.

[6] S.Tsumoto: Knowledge Discovery in Clinical Databases and Evaluation of Discovered Know-ledge in Outpatient Clinic, Information Science vol.124 (2000), p.125-137.

[7] Zhai Junhai, Zhai Mengyao and Li Shengjie: Induction of Decision Tree for Continuous -Valued Attributes Based on Tolerance Rough Sets Technique, Computer Science,vol.39(2012) No.11, p.183-186.

[8] P.Maji and P. Garai: On Fuzzy-Rough Attribute Seletion: Criteria of Max-Dependency, Max-Relevance, Min-Redundancy, and Max-Significance, Applied Soft Computing, vol.13 (2013) No.9, p.1166-1177.