Maximum Condition Entropy Based Attribute Reduction in Variable Precision Rough Set Model

Can Gao, Duoqian Miao, Jie Zhou
Department of Computer Science and Technology, Tongji University
Shanghai, 201804, P.R.China
2005gaocan@163.com, miaoduoqian@163.com, jie_jpu@163.com

Abstract

Variable precision rough set model, as a probabilistic extension of original rough set model, is a very useful approach to inducing probabilistic rules from datasets. In this paper, some anomalies in present definition of attribute reduction based on variable precision rough set model are discussed. Maximum condition entropy is introduced to analyze the mergers of condition classes in the process of reduction and construct a monotonic measure for attribute reduction. Then, based on core attributes under maximum condition entropy, a new heuristic algorithm is proposed to compute reduct, which eradicates all anomalies in variable precision rough set model. Finally, an example is used to show the validity of proposed algorithm.

1. Introduction

Since the initial work of Z. Pawlak [1,2], rough set theory, as a new mathematic approach for dealing with imprecise, uncertain and incomplete information, has been used in many research fields successfully, such as pattern recognition, artificial intelligent, machine learning, knowledge acquisition, data mining and so on [3,4,5]. In recent years, probabilistic approaches are incorporated into rough set theory by many researchers [6]. Several probabilistic extensions model such as decision-theoretic rough set model [7,8], variable precision rough set model [9], Bayesian rough set model [10] and others [11], have been proposed. These models develop the theory of rough sets and its domain of applications in some extent.

Variable precision rough set model (VPRSM), as an important development of original rough set model, was firstly introduced by Ziarko [9] in 1993. With softening standard inclusion relation to majority inclusion relation, VPRSM can be able to allow for some degree of misclassification in classification analysis. So it can discover the non-functional or non-deterministic dependencies between attributes and induce probabilistic rules. Beynon [12] analyzed the defects of Ziarko's attribute reduction model and founded the connection between the range of β (viz. the proportion of correct classification) and quality of classification. A β-reduct should not only satisfy Ziarko's criteria but also be unrestricted in the range of β. In other words, a β-reduct is the minimal set of condition attributes which holds the quality of classification invariably and its β interval under the quality of classification should include original β interval in whole condition attributes. In VPRSM, however, the same quality of classification may not always mean same positive region because of non-monotonicity of measure [13]. Wang [14] showed some anomalies with an example, such as classification error, dynamic changes of β interval. They employed positive region of decision classes and limitative condition on β interval as attribute reduction criteria, not the quality of classification, to avoid classification error and dynamic changes of β interval. But the definition of attribute reduction becomes more complex, and most seriously, the confidence of objects may be changed after attribute reduction in above definition. That means the confidence of rules may increase or decrease in the process of reduction, which violates the principle of attribute reduction. Mi [15] proposed the concept of β lower distribution reduct and β upper distribution reduct. The relations between distribution reduct, β lower distribution reduct, maximum distribution reduct, β upper distribution reduct and possible reduct are also analyzed. But these reducts except distribution reduct may also have the problems of confidence anomaly and dynamic changes of interval. Distribution reduct can eliminate all anomalies, but it is so rigorous that there may have some superfluous attributes in a sense. Section 5 will give an example to show these anomalies.

In some extent, knowledge reduction [16,17] is used to simplify rules and acquire useful knowledge. In rough set theory, the confidence of rules derived from positive region is equal to 1, so it can be hold with same positive region in the process of reduction. But the confidence of positive rules in VPRSM varies from classification precision β to 1, it may be different with the same positive region. In this paper, maximum condition entropy is used as a measure for attribute reduction, which takes positive region and the confidence of rules into account, to conduct the combinations of condition classes. Furthermore, the monotonicity of new measure has been proved, and an algorithm based on maximum condition entropy is also proposed. New algorithm eliminates all anomalies effectively and achieves the objective of attribute reduction in VPRSM.

2. Preliminaries

This section will present some basic concepts of VPRSM and new notions used in the paper. Detail description of the theory can be found in [2, 3, 4, 5, 9].

Definition 1 A decision information system is defined as $S=<U, A=C\cup D, V, \rho>$, where U is a non-empty finite set of objects, called universe; C is a non-empty finite set of condition attributes or features, D is a finite set of decision attributes, $C\cap D=\varnothing$; V is the union of attribute domains, i.e., $V=\cup V_a$, where V_a denotes the domain for each attribute $a\in A$ and ρ is an information function which associates a unique value of each attribute with every object belonging to U.

Definition 2 Let $S=<U, A=C\cup D, V, \rho>$ be a decision information system, the partition of indiscernibility relation of attribute set C and D over U are denoted as $U/C=\{C_1, C_2, ..., C_{|U/C|}\}$ and $U/D=\{D_1, D_2, ..., D_{|U/D|}\}$, the symbol |C| means the cardinality of set C. Each element $C_i\in U/C$ and $D_i\in U/D$ are called as condition class and decision class respectively.

Definition 3 Let $S=<U, A=C\cup D, V, \rho>$ be a decision information system, for any classification precision $\beta\in(0.5, 1]$, the β lower and β upper approximation of decision class $D_j\in U/D$ with respect to indiscernibility relation C are denoted as follows:

$$\underline{C}_\beta(D_j)=\cup\{C_i\in U/C\ |P(D_j|C_i)\geq\beta\}$$
$$\overline{C}_\beta(D_j)=\cup\{C_i\in U/C\ |P(D_j|C_i)>1-\beta\}$$

Where $P(D_j|C_i)=|C_i\cap D_j|/|C_i|$. Then, the positive, boundary and negative regions of decision class D_j can be presented as follows:

$$POS(C, D_j, \beta)=\cup\{C_i\in U/C\ |P(D_j|C_i)\geq\beta\}$$
$$BND(C, D_j, \beta)=\cup\{C_i\in U/C\ |1-\beta<P(D_j|C_i)<\beta\}$$
$$NEG(C, D_j, \beta)=\cup\{C_i\in U/C\ |P(D_j|\ C_i)\leq 1-\beta\}$$

Definition 4 Let $S=<U, A=C\cup D, V, \rho>$ be a decision information system, for any condition class $C_i\in U/C$, its maximum inclusion degree is denoted as $MP(C_i)=Max\{|C_i\cap D_1|/|C_i|, |C_i\cap D_2|/|C_i|, \ldots, |C_i\cap D_{|U/D|}|/|C_i|\}$, the decision value of objects with maximum inclusion degree in C_i is a set $MD(C_i)=\cup\{\rho(o_k, D)|\ o_k\in D_j\wedge P(D_j|C_i)=MP(C_i)\}$, the distribution of maximum inclusion degree in indiscernibility relation C can be denoted as $MS(C)=(MP(C_1), MP(C_2), \ldots, MP(C_{|U/C|}))$.

Definition 5 Let $S=<U, A=C\cup D, V, \rho>$ be a decision information system, the distribution of maximum inclusion degree under C is $MS(C)=(MP(C_1), MP(C_2), \ldots, MP(C_{|U/C|}))$, then the objects with maximum inclusion degree more than β can be denoted as $MO_\beta=\cup\{C_i\ |\ C_i\in U/C\wedge MP(C_i)>\beta\}$

Definition 6 Let $S=<U, A=C\cup D, V, \rho>$ be a decision information system and classification precision $\beta\in(0.5, 1]$, for any $B\subseteq C$, the distribution of maximum inclusion degree $MS(B)=(MP(B_1), MP(B_2), \ldots, MP(B_{|U/B|}))$, and maximum condition entropy of condition attributes set B with respect to decision attributes set D is denoted as follows:

$$ME_\beta(B) = -\sum_{B_i\cap MO_\beta\neq\varnothing}\frac{|B_i|}{|U|}MP(B_i)logMP(B_i)$$

In practice, we often use the rules with the maximum confidence higher than 0.5 to classify new objects. So, maximum condition entropy only takes the objects with the maximum inclusion degree more than 0.5 into account, which may be beneficial for classification.

3. The mergers of condition classes

In VPRSM, as removing condition attributes from decision system, some condition classes will be merged. So, if the model of attribute reduction doesn't educe any anomaly at every step of combining condition classes, its outcome will be an expected reduct.

Without loss of generality, we assume that condition class C_i and C_j will only be merged as excluding attribute a from condition attributes set C. The maximum inclusion degree and decision value of classes C_i, C_j are denoted as $MP(C_i)$, $MD(C_i)$, $MP(C_j)$, $MD(C_j)$. The united condition class $C_i\cup C_j$ has maximum inclusion degree $MP\ (C_i\cup C_j)$ and decision value $MD\ (C_i\cup C_j)$. For simplicity, we replace maximum inclusion degrees $MP(C_i)$, $MP(C_j)$, $MP(C_i\cup C_j)$ by symbols β_i, β_j, β_{ij}. Denote the difference between $ME_\beta(C\text{-}\{a\})$ and $ME_\beta(C)$ as MH. With different values of β_i, β_j, β_{ij}, the values of MH are discussed as follows.

I. $Min\ \{\beta_i, \beta_j\}>0.5$

(1) $\beta_i=\beta_j=\beta_{ij}=\beta>0.5$

This situation will only happen with the condition $MD(C_i)\cap MD(C_j)\cap MD(C_i\cup C_j)\neq\varnothing$. The decision value of united condition class $C_i\cup C_j$ with maximum inclusion degree is in common with that of C_i and C_j. That means the maximum confidence of objects in C_i and C_j and their decision values doesn't change after removing condition attribute *a*. So, condition attribute *a* is a superfluous attribute.

When $\beta_i=\beta_j=\beta_{ij}=\beta$, then $MH= ME_\beta(C\text{-}\{a\}) - ME_\beta(C) = -1/|U|[(|C_i|+|C_j|)\beta_{ij}log\beta_{ij} - |C_i|\beta_i log\beta_i - |C_j|\beta_j log\beta_j]=0$. The maximum condition entropy also doesn't change after removing attribute *a*.

(2) $0.5<\beta_j\leq\beta_{ij}<\beta_i$

In this case, the confidences of objects in condition class C_i and C_j without attribute *a* are vary from β_i and β_j to β_{ij} respectively. The classification information of decision information system after removing attribute *a* will be changed undoubtedly. So, condition attribute *a* can't be excluded from condition attributes set *C*.

With different decision value of $MD(C_i)$, $MD(C_j)$, $MD(C_i\cup C_j)$, the values of *MH* are analyzed as follows.

a. $MD(C_i)\cap MD(C_j)\cap MD(C_i\cup C_j)=k$

The decision value of united condition class $C_i\cup C_j$ with maximum inclusion degree is the same to that of condition class C_i and C_j. So, $|C_i\cap D_k|+|C_j\cap D_k|= |(C_i\cup C_j)\cap D_k|$. Then, $(|C_i|+|C_j|)\beta_{ij} = |C_i|\beta_i + |C_j|\beta_j$ will be hold.

$MH=ME_\beta(C\text{-}\{a\})- ME_\beta(C)$
$= 1/|U|\{|C_i|\beta_i log\beta_i + |C_j|\beta_j log\beta_j$
$- (|C_i|\beta_i + |C_j|\beta_j)\, log[(|C_i|\beta_i + |C_j|\beta_j)/(|C_i|+|C_j|)]\}$
$= 1/|U|\ \{|C_i|\beta_i(log\beta_i - log[(|C_i|\beta_i+|C_j|\beta_j)/(|C_i|+|C_j|)])$
$+|C_j|\beta_j(log\beta_j - log[(|C_i|\beta_i+|C_j|\beta_j)/(|C_i|+|C_j|)])\}$
$= 1/|U|\ \{|C_i|\beta_i log[\beta_i(|C_i|+|C_j|)/(|C_i|\beta_i+|C_j|\beta_j]$
$+|C_j|\beta_j log[\beta_j(|C_i|+|C_j|)/(|C_i|\beta_i+|C_j|\beta_j]\}$

Assume $|C_i|\beta_i =\varphi$, $|C_j|\beta_j =\psi$, $\lambda=\beta_i/\beta_j$, then

$$MH= \varphi log\frac{\varphi+\lambda\psi}{\varphi+\psi}+\psi log\frac{\psi+\frac{1}{\lambda}\varphi}{\varphi+\psi}$$

$$\frac{d(MH)}{d(\lambda)}=\frac{\varphi\psi(\lambda-1)}{\lambda(\varphi+\lambda\psi)}=\begin{cases}<0, & 0<\lambda<1\\ =0, & \lambda=1\\ >0, & \lambda>1\end{cases}$$

But $\beta_i/\beta_j>1$, so $MH>0$.

b. $D(C_i)\cap MD(C_i\cup C_j)=p\wedge MD(C_j)\cap MD(C_i\cup C_j)=\varnothing$

$MH=1/|U|\{|C_i|\beta_i log\beta_i +|C_j|\beta_j log\beta_j$
$- (|C_i|\beta_i + |C_j\cap D_p|)log[(\beta_i|C_i|+|C_j\cap D_p|)/(|C_i|+|C_j|)]\}$

But $0.5<\beta_j$, then $0<|C_j\cap D_p|<|C_j|\beta_j$

$MH>1/|U|\{|C_i|\beta_i log\beta_i +|C_j|\beta_j log\beta_j$
$- (|C_i|\beta_i + |C_j|\beta_j)log[(\beta_i|C_i|+\beta_j|C_j|)/(|C_i|+|C_j|)]\}$

The right part of inequation is the same to maximum condition entropy of case a. So, *MH* is larger than zero.

c. $D(C_j)\cap MD(C_i\cup C_j)=q\wedge MD(C_i)\cap MD(C_i\cup C_j)=\varnothing$

It is similar to case b.

In a word, when $0.5<\beta_j\leq\beta_{ij}<\beta_i$, maximum condition entropy without attribute *a* will be increased.

(3) $0.5<\beta_i\leq\beta_{ij}<\beta_j$

It is similar to case (2).

(4) $\beta_{ij}\leq\beta_j<\beta_i$

This case will appear with the condition $MD(C_i)\cap MD(C_j)=\varnothing$. The confidence of objects in C_i or C_j decreases to β_{ij}, and their decision value may also be changed. Therefore, condition attribute *a* can't be removed in this case.

When $0.5<\beta_j<\beta_i$, then $0.25<\beta_{ij}$, $\beta_j/2\leq\beta_{ij}$ and $\beta_i/2<\beta_{ij}$ will be hold.

$MH=|C_i|/|U|\beta_i log\beta_i+|C_j|/|U|\beta_j log\beta_j$
$- (|C_i|+|C_j|)/|U|\beta_{ij}log\beta_{ij}$
$=|C_i|/|U|\ (\beta_i log\beta_i -\beta_{ij}log\beta_{ij})$
$+|C_j|/|U|\ (\beta_j log\beta_j -\beta_{ij}log\beta_{ij})$
$>|C_i|/|U|\ (\beta_i log\beta_i - \beta_i/2log\beta_i/2)$
$+|C_j|/|U|\ (\beta_j log\beta_j -\beta_j/2log\beta_j/2)$
$=|C_i|/|U|\beta_i log(2\beta_i)^{1/2}+ |C_j|/|U|\beta_j log(2\beta_j)^{1/2}$
$>0\ (0.5<\beta_j<\beta_i)$

When $\beta_{ij}\leq\beta_j<\beta_i$, the maximum condition entropy without attribute *a* will also be increased.

(5) $\beta_{ij}\leq\beta_i<\beta_j$

It is similar to case (4).

(6) $\beta_{ij}<\beta_i=\beta_j$

$MH=(|C_i|+|C_j|)/|U|(\beta_i log\beta_i - \beta_{ij}log\beta_{ij})$
$>(|C_i|+|C_j|)/|U|\ (\beta_i log\beta_i - \beta_i/2log\beta_i/2)$
$=(|C_i|+|C_j|)/|U|\beta_i log(2\beta_i)^{1/2}$
$>0\ (0.5<\beta_i)$

(7) $\beta_j\leq\beta_i<\beta_{ij}$

Intuitively, If the values of β_i, β_i are maximum inclusion degree in classes C_i, C_j, the value β_{ij} of united class $C_i\cup C_j$ won't exceed the maximum of β_i, β_i. From the view of information entropy, the uncertainty information in two independent set will be no more than that of its united set. So, this case will not happen.

II. *Min* $\{\beta_i, \beta_j\}<0.5$ and *Max* $\{\beta_i, \beta_j\}>0.5$

According to definition 6, maximum condition entropy will not consider the uncertainty information of condition classes whose maximum inclusion degree is below to 0.5.

Without loss of generality, we suppose that *Max* $\{\beta_i, \beta_j\}=\beta_i$, then $\beta_{ij}\geq\beta_i/2$

$MH=ME_\beta(C_i\cup C_j) - ME_\beta(C_i)$
$= - (|C_i|+|C_j|)/|U|\beta_{ij}log\beta_{ij} +|C_i|/|U|\beta_i log\beta_i$
$\geq|C_i|/|U|\ (\beta_i log\beta_i -\beta_i/2log(\beta_i/2))- |C_j|/|U|\beta_{ij}log\beta_{ij}$
$>0(\beta_i>0.5)$

III. *Max* $\{\beta_i, \beta_j\}<0.5$

In this case, maximum inclusion degrees of C_i and C_j are both below to 0.5. The merger of these condition classes will not affect the objects in MO_β, which just include the objects with confidence higher than 0.5 in original decision information system. Therefore, condition attribute a is a superfluous in this case. The maximum condition entropy also doesn't change after removing attribute a.

On all accounts, the confidence of objects with maximum value more than 0.5 will not be affected only with the same maximum inclusion degree in the process of reduction. In other cases, it will fluctuate, which is the main reason for some anomalies. But, with the measure of maximum condition entropy, we can discover superfluous attributes and eliminate all anomalies.

4. Maximum condition entropy based attribute reduction in VPRS

4.1 Related definition for attribute reduction

Theorem 1 (Monotonicity) Let $S=<U, A=C\cup D, V, \rho>$ be a decision information system and classification precision $\beta\in(0.5, 1]$, for any $a\in C$, then $ME_\beta(C\text{-}\{a\}) \geq ME_\beta(C)$.

Proof: From the section 3, maximum condition entropy will not vary only with case I-(1) and III. In other cases, the maximum condition entropy will increase. So, the measure with maximum condition entropy is monotonic.

Definition 7 Let $S=<U, A=C\cup D, V, \rho>$ be a decision information system and classification precision $\beta\in(0.5, 1]$, for any subset $P\subseteq C$, if P satisfies the following two conditions:

(1) $ME_\beta(P)=ME_\beta(C)$;

(2) no attribute can be eliminated from P without affecting the requirement (1);

then, P is a reduct of decision information system S.

The definition firstly maintains maximum condition entropy invariably in the process of reduction. The confidence of objects with maximum inclusion degree higher than β will not be changed after removing some superfluous attributes from C. Therefore, the β positive region and quality of classification will retain undoubtedly. The requirement (2) guarantees that there is not superfluous attribute in P, so, the reduct P is a minimum subset of C that preserves the ability of classification in original decision information system.

Definition 8 Let $S=<U, A=C\cup D, V, \rho>$ be a decision information system, and classification precision $\beta\in(0.5, 1]$, for any $a\in C$, if $ME_\beta(C\text{-}\{a\}) > ME_\beta(C)$, then a is a core attribute.

Definition 9 Let $S=<U, A=C\cup D, V, \rho>$ be a decision information system and classification precision $\beta\in(0.5, 1]$, $P\subseteq C$, for any $a\in C\text{-}P$, the significance of attribute a is denoted as $Sig(a, P, D)= ME_\beta(P) - ME_\beta(P\cup\{a\})$

4.2 Algorithm

Based on above reduct definition, the bottom-up algorithm for attribute reduction with maximum condition entropy can be constructed as follows.

Algorithm 1: maximum condition entropy based attribute reduction algorithm (MCEARA)

Input: an decision information system $S=<U, A=C\cup D, V, \rho>$ and classification precision β

Output: a reduct Red

Step1: compute the maximum inclusion degree for each condition class in U/C and MO_β;

Step2: compute $ME_\beta(C)$ and core attributes $Core$, $Red=Core$;

Step3: while $ME_\beta(Red)\neq ME_\beta(C)$;

(1)choose a condition attribute a_i with maximum significance from $C\text{-}Red$;

(2)$Red=Red\cup\{a_i\}$;

Step4: check superfluous attributes in Red again;

Step5: return reduct Red.

4.3 Complexity Analysis

Let $S=<U, A=C\cup D, V, \rho>$ be a decision information system, suppose that $|C| = m$ and $|U| = n$. For algorithm 1 above, the majority computation lies in step 3. In [18], the authors proved that the complexity of computing a partition is $O(mn)$. Thus, the time complexity of our algorithm is $O(m^2n)$.

5. Example

In order to illustrate some anomalies in present attribute reduction model and the validity of our algorithm, we give a decision information system S_1 in following table.

Table 1. Decision information system S_1

U	a_1	a_2	a_3	a_4	d
o_1	0	0	0	0	1
o_2	0	0	0	1	2
o_3	0	0	1	1	2
o_4	0	0	1	1	2
o_5	0	0	1	1	2
o_6	0	0	1	1	3
o_7	0	1	1	1	1
o_8	0	1	1	1	2
o_9	0	1	1	1	2

o_{10}	0	1	1	1	2
o_{11}	1	1	1	1	2
o_{12}	1	1	1	1	3

In decision information system S_1, there are twelve objects, four condition attributes and one decision attribute. U/C={C_1:{o_1}, C_2:{o_2}, C_3:{o_3, o_4, o_5, o_6}, C_4:{o_7, o_8, o_9, o_{10}}, C_5:{o_{11}, o_{12}}}, $MP(C_1)$=1, $MP(C_2)$=1, $MP(C_3)$=3/4, $MP(C_4)$=3/4, $MP(C_5)$=1/2, $MD(C_1)$={1}, $MD(C_2)$={2}, $MD(C_3)$={2}, $MD(C_4)$={2}, $MD(C_5)$= {2,3}.

Given classification precision β=0.7, the positive region and $MO_{0.7}$ in S_1 is {o_1, o_2, o_3, o_4, o_5, o_6, o_7, o_8, o_9, o_{10}}. The set {a_1}is a β reduct in Ziarko [9] and Beynon [12]. But after removing condition attribute a_3, the maximum confidence of object o_2 in positive region decreases from 1 to 4/5 and that of objects o_3, o_4, o_5, o_6 increases from 3/4 to 4/5 reciprocally. When a_4 is excluded from condition attribute set C, the decision value of object o_2 is turned from 1 to 2, which is the anomaly of classification error. {a_1, a_4} is a β reduct in Wang [14] and MI [15]. There are also some anomalies of confidence and dynamic interval.

Condition attribute set {a_1, a_2, a_3, a_4} is a distribute reduct in MI [15], which maintains the maximum confidence of objects in S_1, but condition attribute set {a_1, a_3, a_4 } can also hold that requirement. That means distribute reduct may have superfluous attributes. But, in our algorithm, {a_1, a_3, a_4 } will be extracted from the whole condition attributes C because of equation $ME_{0.7}(C\text{-}\{a_2\})=ME_{0.7}(C)$.

6. Conclusion

Quite a few datasets of practical problems may contain useful probabilistic patterns (viz. probabilistic rules), which can be used to classification analysis with least misclassification rate. VPRSM is just an approach to induce probabilistic rules with minimum condition attributes, but there are some serious problems in existed model for attribute reduction. This paper illustrates the anomalies in VPRSM with example, and a monotonic measure with maximum condition entropy is adopted to develop algorithm for attribute reduction, which has some properties to eradicate anomalies. Further research is planned to introduce our approach into decision-theoretic rough set.

Acknowledgements

This work is partially supported by National Natural Science Foundation of China (Serial No. 60475019, 60775036) and the Research Fund for the Doctoral Program of Higher Education (Serial No. 200602470-39).

References

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

[2] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, Dordrecht: Kluwer Academic Publishers, 1991

[3] W.X. Zhang, W.Z. Wu, J.Y. Liang, et al. Rough set theory and methods, Beijing: Science Press, 2003

[4] G.Y. Wang, Rough set theory and knowledge acquisition, Xi'an: Xi'an Jiaotong university Press, 2001

[5] D.Q. Miao, D.G. Li, Rough set theory, algorithms and applications, Beijing: Tsinghua university press, 2008

[6] Z. Pawlak, S.K.M Wong, W. Ziarko, Rough sets: probabilistic versus deterministic approach, International Journal of Man-machine Studies, 29 (1988) 81-95

[7] Y.Y. Yao, S.K.M. Wong, P. Lingras, A decision-theoretic rough set model, in: Z.W. Ras, M. Zemankova, M.L. Emrich (Eds.), Methodologies for Intelligent Systems, North-Holland, New York, 5 (1990) 17-24

[8] Y.Y. Yao, S.K.M Wong, A decision theoretic framework for approximating concepts, International Journal of Man-machine Studies, 37 (1992) 793-809

[9] W. Ziarko, Variable precision rough set model, Journal of Computer and System Science, 46 (1993) 39-59

[10] D. Slezak, W. Ziarko, The investigation of the Bayesian rough set model, International Journal of Approximate Reasoning, 40 (2005) 81-91

[1 1] Y.Y. Yao, Probabilistic rough set approximations, International Journal of Approximate Reasoning, 49 (2008) 255-271

[12] M. Beynon, Reducts within the variable precision rough sets model: An further investigation. European Journal of Operational Research, 134 (2001) 592-605

[1 3] Y.Y. Yao, Y Zhao, Attribute reduction in decision-theoretic rough set models, Information Sciences, 178 (2008) 3356-3373

[14] J.Y. Wang, J. Zhou, Research of reduct features in the variable precision rough set model, Neurocomputing, 72(2009) 2643-2648

[15] J.S. Mi, W.Z. Wu, W.X. Zhang, Approaches to knowledge reduction based on variable precision rough set model, Information Sciences, 159 (2004) 255-272

[16] D.Q. Miao, J. Wang, Information-Based Algorithm for Reduction of knowledge, In: Proceedings of the IEEE International Conference on Intelligent Processing Systems, Beijing, China,1997: 1155-1158

[17] R.Z. Wang, D.Q. Miao, G.R. Hu, Discernibility matrix based algorithm for reduction of attributes, 2006 IEEE/WIC/ ACM International Conference, 2006:477-480

[18] Z.Y. Xu, Z.P. Liu, B.R. Yang, et al, A quick attribute reduction algorithm with complexity of max(O(|C||U|), O(|C|^2|U/C|), Chinese Journal of Computers 29(2006) 391-399