Information Sciences 290 (2015) 72–85

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Automatic determination about precision parameter value based on inclusion degree with variable precision rough set model

Yusheng Cheng [a,*], Wenfa Zhan [b,1], Xindong Wu [c,d,2], Yuzhou Zhang [a,3]

[a] School of Computers and Information, Anqing Normal University, Anqing 246011, China
[b] Department of Science and Technology, Anqing Normal University, Anqing 246011, China
[c] School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230009, China
[d] Department of Computer Science, University of Vermont, Burlington, VT 05405, USA

ARTICLE INFO

Article history:
Received 25 April 2013
Received in revised form 3 August 2014
Accepted 13 August 2014
Available online 23 August 2014

Keywords:
Precision parameter value
Decision-theoretic rough set
Symmetric variable precision rough set
Decision table

ABSTRACT

The rough set theory provides a powerful approach for attributes reduction and data analysis. The variable precision rough set (VPRS) model, an extension of the original rough set approach, tolerates misclassifications of the training data to some degree, which promotes the applications of rough set theory in inconsistent information systems. However, in most existing algorithms of feature reduction based on VPRS, the precision parameter (β) is introduced as prior knowledge, which restricts their applications because it is not clear how to set the β value. By studying β-consistency in the measurement of a decision table and the threshold value of the β-consistent decision table, this paper presents an algorithm for automatic determination of the precision parameter value from a decision table based on VPRS. At the same time, the precision parameter value from our proposed method is compared with the thresholds from the decision-theoretic rough set (DTRS). The influences of the precision parameter are also discussed on attribute reduction, which shows the necessity of the estimated precision parameter from a decision table. The simulation results including VPRS and other classification methods in real data further indicate that different precision parameter values make a great difference on rules and setting a precise parameter near the threshold value of the β-consistent decision table can precisely reflect the decision distribution of the decision table.

1. Introduction

As a new mathematical tool for dealing with inexact, uncertain knowledge, the rough set theory (RST) has been successfully employed in machine learning, data mining and other fields since it was put forward by Pawlak [24]. RST can be used to model classification but the classification must be fully correct or certain, which limits the practical applicability of RST in real world applications. As a probabilistic rough set model (PRS), the decision-theoretic rough set model

* Corresponding author. Tel.: +86 18055665796.
E-mail addresses: chengyshaq@163.com (Y. Cheng), zhanwf@aqtc.edu.cn (W. Zhan), xwu@uvm.edu (X. Wu), zhangyzh@aqtc.edu.cn (Y. Zhang).
[1] Tel.: +86 13955612021.
[2] Tel.: +1 802 656 7839.
[3] Tel.: +86 15385562628.

http://dx.doi.org/10.1016/j.ins.2014.08.034
0020-0255/

(DTRS) [2,7,8,10–22,26,27,32–36,38] extended RST by considering probabilistic information of objects into a set of positive, negative or boundary regions, by introducing a pair of thresholds (α, β). When proper thresholds are used, we can derive several existing probabilistic rough set models, such as the variable precision rough set (VPRS) [40,41], the game-theoretic rough set (GTRS) [2,12], and the Bayesian rough set (BRS) [23] from the decision-theoretic rough set model. However, how to determine and interpret the thresholds is fundamental issues in probabilistic rough sets. At present, many methods about the thresholds have been proposed by utilizing different cost functions to minimize the overall classification cost [2,10,16]. Deng and Yao [10]proposed an approach that can obtain the required threshold parameters directly from the data by utilizing the measure of Shannon entropy, which is known as the information-theoretic method. How to find an effective threshold pair was formulated as mining the uncertainty of Shannon entropy of the three regions. The information-theoretic method only provides a symmetric threshold to search the optimal threshold in the matrix. That is, the thresholds (α, β) were converted into $(\alpha, 1 - \alpha)$. In order to determine an optimal threshold in the space of all possible threshold pairs, Azam and Yao [2,12] proposed a GTRS method to obtain the required threshold parameters. The GTRS method reformulates the conditional entropy in [10] as the overall uncertainty $\Delta(\alpha, \beta)$. A repetitive game was formulated where the players cooperatively determine suitable thresholds. Jia [15,16] also proposed an optimization method by minimizing the decision cost through six cost functions. Based on the suggested minimization problem, the thresholds and cost functions can be automatically obtained from data directly. These PRS based efforts have provided methods that can learn the threshold parameters automatically by searching the minimization of uncertainty of the classification of the three regions from the data.

As one of the most important branches in the rough set theory,VPRS was proposed by Ziarko in 1993. The majority inclusion relation and the weak dependence were the core of VPRS [40,41]. So our proposed method based on VPRS with the inclusion degree can also obtain the optimal threshold automatically by only searching the maximization of the inclusion degree of the positive or negative region [37].

VPRS deals with partial classification by introducing a probability value i.e. the precision parameter β. This parameter represents a bound on the conditional probability of a proportion of objects in a condition class that are classified into the same decision class. An et al. [1] used the symbol β to denote the proportion of correct classifications, the appropriate value range of which is (0.5 1.0]. However, Ziarko [1,40] considered β as a classification error, defined in the domain of [0.0 0.5). In this paper, we use the original meaning of precision parameter β [1,40], which is equivalent to the β parameter of the threshold pair (α, β) in DTRS. Therefore, we also provide a symmetric threshold $(1 - \beta, \beta)$ method from the viewpoint of DTRS [2,7,8,10–22,26,27,32–36,38].

On the other hand, the VPRS model is parametric and it is difficult to set the precision parameter value (β) in some applications. Therefore, relevant β values are often given subjectively and they need to be adjusted in order to improve the quality of classification. At present, in most VPRS-related studies, β is introduced as prior knowledge, which runs against the prominent advantage of RST—'Let data speak by itself', i.e. avoid any other information outside the underlying information system. If β could be generated from the data to be processed during the reduction, it will surely play an important role in promoting the development and application of the VPRS model. Su et al. [30]provided a method to calculate the precision parameter value based on the least upper bound of the data misclassification error. Ziarko [40] proposed that the precision parameter value be specified by the decision maker. Beynon [3] proposed two methods of selecting a β-reduct without such a known β value. Slezak [28,29] put forward a Bayesian rough set model, in which the parameter was defined by a prior probability. Beynon [4,5] proposed an allowable β value range to be an interval, where the quality of classification may be known prior to determining the β value range. Katzberg et al. [17] allowed asymmetric bounds l and u to be used, with which the restrictions $l < 0.5$ and $u = 1 - l$ must hold. Zhou et al. [39] proposed a method to determine the precision parameter value based on the least upper bound of the data misclassification error. Beynon [6] utilized the $(l - u)$-graphs to choose the values of l and u based on the associated levels of the quality of classification and the degree of dependency. Despite that all these methods aim at searching for an optimal precision parameter objectively, they still focus on a special β value and attribute reduction anomalies will occur inevitably under the classic reduct definition [39]. Three-region decision about DTRS [2,7,8,10–22,26,27,32–36,38] is currently a hot topic. Yao [10,34–36], Jia [15,16] and others [2,12,19,20] have provided several methods for finding the parameters. In our previous study [9], we also discussed the influences of the precision parameter on attribute reduction.

As a flexible tool for analyzing the optimal thresholds, the information-theoretic method requires tremendous computations and is difficult to implement. Inspired by the discernibility matrices proposed by Skowron [25], the β-consistency about an information system is proposed in this paper, which develops the β-discernibility about a set X. The threshold about a β-consistent decision table, can be obtained by computing the threshold about the set from different decision classes. In order to facilitate the realization about our method, we further propose a combined probability matrix of a decision information system. The time complexity of our method is significantly smaller than DTRS methods.

Moreover, the paper is further concerned with the relations between β and the β-consistent decision table and influences of β on attribute reduction such as β lower distribution reduction, β upper distribution reduction and β lower approximate reduction.

The rest of the paper is organized as follows. Section 2 gives some basic notions related to VPRS and the relative discernibility of a set. Section 3 introduces the measurement approach of the β-consistent decision table and how to compute the value of β from a decision table. Our proposed method is compared with the thresholds from DTRS in Section 4. In Section 7, we prove that for some special threshold values, β lower distribution reduction is equivalent to β upper distribution reduction, whereas β lower approximate reduction can obtain classification decision rules with certainty under a given classification error. How to set the precision parameter is discussed in Section 6. In the last section, we sum up what has been discussed and point out further research.

2. Basic notations and properties related to VPRS

An information system is usually denoted as a triplet $S = (U, C \cup D, f)$, called a decision table, where U is the universe which consists of a finite set of objects, C is the set of condition attributes and D is the set of decision attributes. With every attribute $a \in C \cup D$, a set of its values V_a is associated. Each attribute a determines an information function f: $U \to V_a$ such that for any $a \in C \cup D$, and $x \in U$, $f(x) \in V_a$. Each non-empty subset $B \subseteq C$ determines an indiscernible relation

$$R_B = \{(x,y) : \forall a \in B, f_a(x) = f_a(y), x, y \in U\}$$

R_B is called an equivalence relation and partitions U into a family of disjoint subsets U/R_B called a quotient set of U:

$$U/R_B = \{[x]_B : x \in U\}$$

where $[x]_B$ denotes the equivalence class determined by x with respect to B, i.e., $[x]_B = \{y \in U: (x, y) \in R_B\}$

If X, $Y \subseteq U$ are subsets of U, then the ratio of classification error, denoted as $c(X, Y)$ [1,40], is defined as follows,

$$c(X,Y) = \begin{cases} 1 - \frac{|X \cap Y|}{|X|} & |X| > 0 \\ 0 & |X| = 0 \end{cases}$$

where $|X|$ is the cardinality of set X.

Definition 2.1 [40]. Let X, $Y \subseteq U$. The β-majority inclusion relation is defined as $Y \overset{\beta}{\supseteq} X \iff c(X,Y) \leqslant \beta$, where $0 \leqslant \beta < 0.5$.

The β-majority inclusion relation can be explained as the degree of Y containing X, called the inclusion degree [37]. Based on a threshold, the variable precision rough set is called a symmetric variable precision rough set. In this paper, our proposed method is based on the symmetry of VPRS.

Definition 2.2 [40]. Let $X \subseteq U$, $B \subseteq C$, $U/R_B = \{X_1, X_2, \ldots, X_n\}$. The β lower and β upper approximations of the set X are defined as follows respectively:

$$\underline{R_B^\beta}(X) = \cup\{X_j | c(X_j, X) \leqslant \beta\}$$
$$\overline{R_B^\beta}(X) = \cup\{X_j | c(X_j, X) < 1 - \beta\}$$

The set $\underline{R_B^\beta}(X)$ is also called the β positive region of X, denoted by $pos_\beta(X)$. Correspondingly, the β negative region and β boundary region of X can be defined as follows respectively:

$$negr_\beta(X) = \cup\{X_j | c(X_j, X) \geqslant 1 - \beta\}$$
$$bnr_\beta(X) = \cup\{X_j | \beta < c(X_j, X) < 1 - \beta\}$$

When $\beta = 0$, $pos_0(X)$, is defined as the positive region of X in the Pawlak rough set.

Definition 2.3 [40]. If $bnr_\beta(X) = \phi$ then set X is called β discernibility, or else it is called β indiscernibility.

Proposition 2.1 [40]. *If set X is discernible on the classification error level* $0 \leqslant \beta < 0.5$, *then X is also discernible at any level* $\beta < \beta_1 < 0.5$.

Proposition 2.2 [40]. *If* $\overline{R_{0.5}}(X) \neq \underline{R_{0.5}}(X)$, *then X is discernible at any level* $0 \leqslant \beta < 0.5$.

This proposition states that a set with a non-empty absolute boundary region is indiscernible.

Proposition 2.3 [40]. *If set X is not discernible on the classification error level* $0 \leqslant \beta < 0.5$, *then X is also not discernible at any level* $\beta_2 < \beta$.

Ziarko [40] also stated that a set which is not given a classification for every β is called absolutely rough, while one only given a classification for a range of β is called relatively rough. These statements of Ziarko indicate some move to the exposition of the role of ranges of β rather than specific β values, which was studied in [4,40]. The minimal threshold value, denoted as $\xi(X)$ that can discern set X, is defined as follows.

Proposition 2.4 [40].

$$\xi(X) = \max(m1, m2) \tag{1}$$

where

$$m1 = 1 - \min\{c(X_i, X) | \forall X_i \in U/R, c(X_i, X) > 0.5\}$$
$$m2 = \max\{c(X_i, X) | \forall X_i \in U/R, c(X_i, X) < 0.5\}$$

Proposition 2.4 provides a simple property that can be used to find the discernibility threshold of a weakly discernible set X.

Example 1 [40]. Supposing the basic classes in an approximate space $K = (U, R)$ are as follows,

$$X_1 = \{x_1, x_2, x_3, x_4, x_5\}, X_2 = \{x_6, x_7, x_8\}, X_3 = \{x_9, x_{10}, x_{11}, x_{12}\}, X_4 = \{x_{13}, x_{14}, x_{15}, x_{16}\}$$

$X_5 = \{x_{17}, x_{18}\}$ and give a set $X = \{x_4, x_5, x_8, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}\}$, then

$$c(X_1, X) = 0.6, c(X_2, X) = 0.66, c(X_3, X) = 1.0, c(X_4, X) = 0.25, c(X_5, X) = 0$$

Therefore, the minimal threshold of the set X is $\xi(X) = 0.4$, which makes this set β discernible under the minimal classification error β.

3. β-Consistent decision table and precision parameter

Given a decision table $S = (U, C \cup D, f)$, a partition of the condition attributes $U/R_C = \{X_1, X_2, \ldots, X_n\}$ and a partition of the decision attributes $U/R_D = \{D_1, D_2, \ldots, D_m\}$, it is easy to prove that any $D_i\underline{R_B^\beta}$ and $\overline{R_B^\beta}$ satisfy the following properties [40]:

$$\underline{R_B^\beta}(D_i) \cap \underline{R_B^\beta}(D_j) = \phi \quad i \neq j \tag{2}$$

$$\bigcup_{i=1}^{m} \underline{R_B^\beta}(D_i) \subseteq \bigcup_{i=1}^{m} \overline{R_B^\beta}(D_i) \subseteq U \tag{3}$$

3.1. β-Consistent decision table

In this section, we discuss the value β related to the β-consistent decision table based on VPRS and introduce a method to get the minimal threshold of β from a decision table.

Definition 3.1. Given a decision table $S = (U, C \cup D, f)$, for $0 \leqslant \beta < 0.5$, we define the β boundary region and the β positive region of S respectively as:

$$bnr_\beta(S) = \bigcup_{i=1}^{m} bnr_\beta(D_i)$$

$$pos_\beta(S) = \bigcup_{i=1}^{m} \underline{R_C^\beta}(D_i)$$

So, the β negative region of S can be computed as follows:

$$negr_\beta(S) = U - bnr_\beta(S) - pos_\beta(S) \tag{4}$$

Definition 3.2. A decision table $S = (U, C \cup D, f)$ is absolutely rough iff $bnr_{0.5}(S) \neq \phi$.

Definition 3.3. For a decision table $S = (U, C \cup D, f)$, if $bnr_\beta(S) = \phi$ then S is β-consistent; otherwise it is β-inconsistent.

Proposition 3.1. *If a decision table S is β consistent on the classification error level $0 \leqslant \beta < 0.5$, then it is also β consistent at any level $\beta < \beta_1 < 0.5$.*

For any $D_i \in U/R_D$, we can obtain the value of $\xi(D_i)$ according to Proposition 2.4. Therefore, for all decision classes $D_1, D_2, \cdots, D_m$, we can obtain a series of $\xi(D_1), \xi(D_2), \ldots, \xi(D_m)$. Thus the minimal threshold, denoted by $\xi(S)$, can be found according to Proposition 3.1:

$$\xi(S) = \max(\xi(D_1), \xi(D_2), \ldots, \xi(D_m))$$

Lemma 3.1. *For $S = (U, C \cup D, f)$ and $D_i \in U/R_D$, we have the relation*

$$negr_\beta(D_i) = pos_\beta(\sim D_i)$$

where $\sim D_i = U - D_i$.

Proof. For $X_j \in U/R_C$, supposing $X_j \subseteq negr_\beta(D_i)$, according to Definition 2.2, we have

$$c(X_j, D_i) \geqslant 1-\beta \Longleftrightarrow 1-\frac{|X_j \cap D_i|}{|X_j|} \geqslant 1-\beta \Longleftrightarrow \frac{|X_j \cap D_i)}{|X_j|} \leqslant \beta$$

In addition, due to

$$|X_j \cap (U-D_i)| = |X_j \cap U - X_j \cap D_i| = |X_j - X_j \cap D_i| \geqslant |X_j| - |X_j \cap D_i|, \text{ we have}$$
$$\frac{|X_j \cap \sim D_i|}{|X_j|} = \frac{|X_j \cap (U-D_i)|}{|X_j|} \geqslant \frac{|X_j| - |X_j \cap D_i|}{|X_j|} = 1 - \frac{|X_j \cap D_i|}{|X_j|} \geqslant 1-\beta$$
$$\Longleftrightarrow 1 - \frac{|X_j \cap \sim D_i|}{|X_j|} \leqslant \beta$$

Therefore, $c(X_j \sim D_i) \leqslant \beta \Leftrightarrow X_j \subseteq pos_\beta(\sim D_i) \Leftrightarrow negr_\beta(D_i) = pos_\beta(\sim D_i)$ □

Proposition 3.2. *If $\beta \in [\xi(S), 0.5)$ then a decision table $S = (U, C \cup D, f)$* is *β-consistent and $pos_\beta(S) = U$.*

Proof. According to Proposition 3.1 and Definition 3.3, when $\beta \in [\xi(S), 0.5)$, the equation of $bnr_\beta(S) = \phi$ must be satisfied. Therefore, the decision table is β-consistent.

Next we show by contradiction the proposition of $pos_\beta(S) = U$.

Because $bnr_\beta(S) = \phi$, $pos_\beta(S) \cup negr_\beta(s) = U$ according to (4)

If $pos_\beta(S) \neq U$, then $pos_\beta(S) \cup negr_\beta(s) = U$, and therefore there is at least one element $x \in U$ and $x \notin pos_\beta(S)$, denoted as $x \in X_k$

$\Rightarrow X_k \not\subseteq pos_\beta(S) \Rightarrow X_k \subseteq negr_\beta(S)$, $\exists D_i \in U/R_D$

$\Rightarrow c(X_k, D_i) \geqslant 1-\beta$, according to Lemma 3.1

$\Rightarrow c(X_k \sim D_i) \leqslant \beta$, i.e. $X_k \subseteq pos_\beta(\sim D_i) = pos_\beta(U - D_i)$

In addition, $U - D_i = D_1 \cup D_2 \cup \cdots \cup D_{i-1} \cup D_{i+1} \cup \cdots \cup D_m$

$\Rightarrow X_k \subseteq pos_\beta(D_1 \cup D_2 \cup \cdots \cup D_{i-1} \cup D_{i+1} \cup \cdots \cup D_m)$, and according to (2): any basic class can be classified into a decision class according to the majority inclusion

$\Rightarrow \exists D_j, j \neq i, X_k \subseteq pos_\beta(D_j) \subseteq pos_\beta(S)$

$\Rightarrow X_k \subseteq pos_\beta(S)$ is in contradiction with the equation $X_k \not\subseteq pos_\beta(S)$. Therefore, it has been proved. □

Proposition 3.2 implies that every condition attribute supports a decision rule when β is assigned to be in the domain $[\xi(S), 0.5)$, i.e., $C \Rightarrow D$ when the decision table is β-consistent according to the majority inclusion relation.

3.2. Getting β from a decision table

The β value plays an important role in attribute reduction based on VPRS. An algorithm of self-determining the precision parameter can be realized if we get the minimal threshold from a decision table. In order to obtain $\xi(S)$, several steps are designed as follows:

Step 1. Obtain the matrix of combined probabilities of the decision table, denoted as CoD:

$$CoD = \begin{bmatrix} D(D_1/X_1) & D(D_1/X_2) & \cdots & D(D_1/X_n) \\ D(D_2/X_1) & D(D_2/X_2) & \cdots & D(D_2/X_n) \\ \vdots & \vdots & \vdots & \vdots \\ D(D_m/X_1) & D(D_m/X_2) & \cdots & D(D_m/X_n) \end{bmatrix}_{m\times n}$$

where $D(Y/X) = \frac{|X \cap Y|}{|X|}$ if $|X| > 0$, and $D(Y/X) = 0$ otherwise.

Step 2. Due to $c(X, Y) + D(Y/X) = 1$ based on the definitions of $c(X, Y)$ and $D(Y/X)$, the matrix CoD is transformed into $\overline{CoD}$ which represents the probability distribution of classification errors in the decision table;

Step 3. Compute $\xi(D_1), \xi(D_2), \ldots, \xi(D_m)$ according to $\overline{CoD}$;

Step4. Compute the minimal threshold value of the decision table, $\xi(S) = \max(\xi(D_1), \ldots, \xi(D_m))$.

Example 2. Given a decision table $S = (U, C \cup D, f)$, $U/R_C = \{X_1, X_2, X_3, X_4\}$and $U/R_D = \{D_1, D_2, D_3\}$, where $X_1 = \{1, 2, 19, 20, 21\}$, $X_2 = \{3\}$, $X_3 = \{4–13\}$, $X_4 = \{14–18\}$,

$D_1 = \{1–12\}$, and $D_2 = \{13–17\}$, $D_3 = \{18–21\}$, it is easy to get $CoD = \begin{bmatrix} 0.4 & 1 & 0.9 & 0 \\ 0 & 0 & 0.1 & 1 \\ 0.6 & 0 & 0 & 0 \end{bmatrix}$ and $\overline{CoD} = \begin{bmatrix} 0.6 & 0 & 0.1 & 1 \\ 1 & 1 & 0.9 & 0 \\ 0.4 & 1 & 1 & 1 \end{bmatrix}$.

Then from $\overline{CoD}$, we have $(\xi(D_1), \xi(D_2), \xi(D_3)) = (0.4, 0.1, 0.4)$. Therefore, $\xi(S) = \max((0.4, 0.1, 0.4)) = 0.4$, i.e., when $\beta \in [0.4, 0.5)$, the inconsistency of Table 1 can be treated as a β-consistent system based on the VPRS approach.

Table 1
Decision table.

U	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
a	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0
b	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0
c'	0	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
e	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
d	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	2	2	2	2

4. $\xi(S)$ Comparison with thresholds of DTRS

4.1. DTRS based on the equivalent classes

Suppose a decision table $S = (U, C \cup D, f)$, $U = \{u_1, u_2, \ldots, u_n\}$ and a partition of decision attributes $U/R_D = \{D_1, D_2, \ldots, D_m\}$. With respect to the three regions of $pos_0(D_j)$, $bnd_0(D_j)$ and $negr_0(D_j)$, the set of actions related to a state is given by $A = (a_{P_j}, a_{B_j}, a_{N_j})$, where a_{P_j}, a_{B_j} *and* a_{N_j} represent the three actions in classifying an object μ_i, namely, deciding $\mu_i \in pos_0(D_j)$, $\mu_i \in bnd_0(D_j)$ *or* $\mu_i \in negr_0(D_j)$, respectively. Let $\lambda_{P_jD_k}, \lambda_{B_jD_k}$ and $\lambda_{\mathrm{N_jD_k}}$ denote the costs incurred for taking actions a_{P_j}, a_{B_j} *and* a_{N_j} in classifying an object D_k is classified into $pos_0(D_j)$, $bnd_0(D_j)$ and $negr_0$ (D_j), respectively. Similarly, $\lambda_{P_j\sim D_k}, \lambda_{B_j\sim D_k}$ and $\lambda_{\mathrm{N_j\sim D_k}}$ denote the costs incurred for taking the same actions when the object D_k does not belong to $pos_0(D_j)$, $bnd_0(D_j)$ and $negr_0(D_j)$, respectively.

Consider a special kind of cost functions with $\lambda_{P_jD_j} \leqslant \lambda_{B_jD_j} < \lambda_{\mathrm{N_jD_j}}, \lambda_{N_jD_k} \leqslant \lambda_{B_jD_k} < \lambda_{\mathrm{P_jD_k}}$ $(k \neq j)$, and $\lambda_{P_jD_k} = \lambda_{P_j\sim D_j}, \lambda_{B_jD_k} = \lambda_{B_j\sim D_j}, \lambda_{N_jD_k} = \lambda_{N_j\sim D_j}$ $(k \neq j)$. That is, the cost of classifying an object μ in D_j into the positive region $pos_0(D_j)$ is less than or equal to the cost of classifying μ into the boundary region $bnd_0(D_j)$, and both of these costs are strictly less than the cost of classifying μ into the negative region bnd_0 (D_j). The reverse order of these costs is used for classifying an object not in the region.

Definition 4.1. Suppose a decision table $S = (U, C \cup D, f)$, $U = \{u_1, u_2, \ldots, u_n\}$ and a partition of decision attributes $U/R_D = \{D_1, D_2, \ldots, D_m\}$. The matrix $\lambda_{\mathrm{j}}(= P_j, B_j, N_j)$ of the cost functions by taking different actions can be defined as:

$$\lambda_j = \begin{bmatrix} \lambda_{1D_1} & \lambda_{1D_2} & \ldots & \lambda_{1D_n} \\ \lambda_{2D_1} & \lambda_{2D_2} & \ldots & \lambda_{2D_n} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_{jD_1} & \lambda_{jD_2} & \ldots & \lambda_{jD_n} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_{mD_1} & \lambda_{mD_2} & \ldots & \lambda_{mD_n} \end{bmatrix}$$

Therefore, the risk matrix $COST_j(= P_j, B_j, N_j)$ can be defined as:

$$COST_j = \begin{bmatrix} R(a_{1|[u_1]_B}) & R(a_{1|[u_2]_B}) & \ldots & R(a_{1|[u_n]_B}) \\ R(a_{2|[u_1]_B}) & R(a_{2|[u_2]_B}) & \ldots & R(a_{2|[u_n]_B}) \\ \vdots & \vdots & \vdots & \vdots \\ R(a_{j|[u_1]_B}) & R(a_{j|[u_2]_B}) & \ldots & R(a_{j|[u_n]_B}) \\ \vdots & \vdots & \vdots & \vdots \\ R(a_{m|[u_1]_B}) & R(a_{m|[u_2]_B}) & \ldots & R(a_{m|[u_n]_B}) \end{bmatrix}$$

The values of the risk matrix can be obtained through matrixes *CoD and* λ_j, that is:

$$R(a_{j|[u_i]_B}) = \sum_{k=1}^{m} \lambda_{jD_k} P(D_k|[u_i]_B)$$

The Bayesian decision procedure suggests the following minimum cost decision rules:

$$(\mathrm{PR}_1) \quad \textit{if } R(a_{P_j|[u_i]_B}) \leqslant R(a_{B_j|[u_i]_B}) \textit{ and } R(a_{P_j|[u_i]_B}) \leqslant R(a_{N_j|[u_i]_B}), \textit{ then } u_i \in pos_0(D_j)$$
$$(\mathrm{BR}_1) \quad \textit{if } R(a_{B_j|[u_i]_B}) \leqslant R(a_{P_j|[u_i]_B}) \textit{ and } R(a_{B_j|[u_i]_B}) \leqslant R(a_{N_j|[u_i]_B}), \textit{ then } u_i \in bnd_0(D_j)$$
$$(\mathrm{NR}_1) \quad \textit{if } R(a_{N_j|[u_i]_B}) \leqslant R(a_{P_j|[u_i]_B}) \textit{ and } R(a_{N_j|[u_i]_B}) \leqslant R(a_{B_j|[u_i]_B}), \textit{ then } u_i \in negr_0(D_j)$$

For the first condition of PR_1:

$$
\begin{aligned}
&R(a_{P_j|[u_i]_B}) \leqslant R(a_{B_j|[u_i]_B}) \\
&\iff \sum_{k=1}^{m} \lambda_{P_jD_k} P(D_k|[u_i]_B) \leqslant \sum_{k=1}^{m} \lambda_{B_jD_k} P(D_k|[u_i]_B) \\
&\iff \lambda_{P_jD_j} P(D_j|[u_i]_B) + \sum_{k=1,k\neq j}^{m} \lambda_{P_jD_k} P(D_k|[u_i]_B) \\
&\leqslant \lambda_{B_jD_j} P(D_j|[u_i]_B) + \sum_{k=1,k\neq j}^{m} \lambda_{B_jD_k} P(D_k|[u_i]_B) \\
&\iff \lambda_{P_jD_j} P(D_j|[u_i]_B) + \lambda_{P_jD_{\sim j}} \sum_{k=1,k\neq j}^{m} P(D_k|[u_i]_B) \\
&\leqslant \lambda_{B_jD_j} P(D_j|[u_i]_B) + \lambda_{B_jD_{\sim j}} \sum_{k=1,k\neq j}^{m} P(D_k|[u_i]_B) \\
&\iff \lambda_{P_jD_j} P(D_j|[u_i]_B) + \lambda_{P_jD_{\sim j}} (1 - P(D_j|[u_i]_B)) \\
&\leqslant \lambda_{B_jD_j} P(D_j|[u_i]_B) + \lambda_{B_jD_{\sim j}} (1 - P(D_j|[u_i]_B) \\
&\iff P(D_j|[u_i]_B) \geqslant \frac{\lambda_{P_jD_{\sim j}} - \lambda_{B_jD_{\sim j}}}{(\lambda_{P_jD_{\sim j}} - \lambda_{B_jD_{\sim j}}) + (\lambda_{B_jD_j} - \lambda_{P_jD_j})}
\end{aligned}
$$

For the second condition of, similarly, the result can be obtained as:

$$P(D_j|[u_i]_B) \geqslant \frac{\lambda_{P_jD_{\sim j}} - \lambda_{N_jD_{\sim j}}}{(\lambda_{P_jD_{\sim j}} - \lambda_{N_jD_{\sim j}}) + (\lambda_{N_jD_j} - \lambda_{P_jD_j})}$$

Let α_j, γ_j be expressed as:

$$\alpha_j = \frac{\lambda_{P_jD_{\sim j}} - \lambda_{B_jD_{\sim j}}}{(\lambda_{P_jD_{\sim j}} - \lambda_{B_jD_{\sim j}}) + (\lambda_{B_jD_j} - \lambda_{P_jD_j})}$$

$$\gamma_j = P(D_j|[u_i]_B) \geqslant \frac{\lambda_{P_jD_{\sim j}} - \lambda_{N_jD_{\sim j}}}{(\lambda_{P_jD_{\sim j}} - \lambda_{N_jD_{\sim j}}) + (\lambda_{N_jD_j} - \lambda_{P_jD_j})}$$

Then the decision rule of PR_1 can be re-expressed as:

$$(\mathrm{NewPR}_1) \quad \textit{if } P(D_j|[u_i]_B) \geqslant \alpha_j \textit{ and } P(D_j|[u_i]_B) \geqslant \gamma_j, \textit{ then } x \in pos_0(D_j)$$

Similarly, we omit the derivations of BR_1 and NR_1.

Let $\beta_j = \frac{\lambda_{B_jD_{\sim j}} - \lambda_{N_jD_{\sim j}}}{(\lambda_{B_jD_{\sim j}} - \lambda_{N_jD_{\sim j}}) + (\lambda_{N_jD_j} - \lambda_{B_jD_j})}$

The decision rules can be modified to:

$$
\begin{aligned}
&(\mathrm{NewPR}_1) \quad \textit{if } P(D_j|[u_i]_B) \geqslant \alpha_j \textit{ and } P\big(D_j|[u_i]_B\big) \geqslant \gamma_j, \textit{ then } u_i \in pos_0(D_j) \\
&(\mathrm{NewBR}_1) \quad \textit{if } P(D_j|[u_i]_B) < \alpha_j \textit{ and } P(D_j|[u_i]_B) \geqslant \beta_j, \textit{ then } u_i \in bnd_0(D_j) \\
&(\mathrm{NewNR}_1) \quad \textit{if } P(D_j|[u_i]_B) \leqslant \beta_j \textit{ and } P(D_j|[u_i]_B) \leqslant \gamma_i, \textit{ then } u_i \in negr_0(D_j)
\end{aligned}
$$

Each rule is defined by two out of the three parameters. The conditions of rule (NewBR_1) suggest that $\alpha_j > \beta_j$ may be a reasonable constraint. So, the following condition on the cost functions can be obtained:

$$
\begin{aligned}
&\frac{\lambda_{P_jD_{\sim j}} - \lambda_{B_jD_{\sim j}}}{(\lambda_{P_jD_{\sim j}} - \lambda_{B_jD_{\sim j}}) + (\lambda_{B_jD_j} - \lambda_{P_jD_j})} > \frac{\lambda_{B_jD_{\sim j}} - \lambda_{N_jD_{\sim j}}}{(\lambda_{B_jD_{\sim j}} - \lambda_{N_jD_{\sim j}}) + (\lambda_{N_jD_j} - \lambda_{B_jD_j})} \\
&\iff \frac{1}{1 + \frac{\lambda_{B_jD_j} - \lambda_{P_jD_j}}{\lambda_{P_jD_{\sim j}} - \lambda_{B_jD_{\sim j}}}} > \frac{1}{1 + \frac{\lambda_{N_jD_j} - \lambda_{B_jD_j}}{\lambda_{B_jD_{\sim j}} - \lambda_{N_jD_{\sim j}}}} \\
&\iff \frac{\lambda_{B_jD_j} - \lambda_{P_jD_j}}{\lambda_{P_jD_{\sim j}} - \lambda_{B_jD_{\sim j}}} < \frac{\lambda_{N_jD_j} - \lambda_{B_jD_j}}{\lambda_{B_jD_{\sim j}} - \lambda_{N_jD_{\sim j}}}
\end{aligned}
$$

Therefore, $1 \geqslant \alpha_j > \gamma_j > \beta_j \geqslant 0$. In this case, after tie-breaking, the following simplified rules are obtained:

$$
\begin{aligned}
&(\mathrm{NewPR}'_1) \; \textit{if } P(D_j|[u_i]_B) \geqslant \alpha_j, \textit{ then } u_i \in pos_0(D_j) \\
&(\mathrm{NewBR}'_1) \; \textit{if } \beta_j < P(D_j|[u_i]_B) < \alpha_j, \textit{ then } u_i \in bnd_0(D_j) \\
&(\mathrm{NewNR}'_1) \; \textit{if } P(D_j|[u_i]_B) \leqslant \beta_j, \textit{ then } u_i \in negr_0(D_j)
\end{aligned}
$$

By using the thresholds, we can divide the universe into three regions of a decision partition U/R_D based on (α_j, β_j) in [2,10,12,15,16,19,20,34–36]. When $\psi(S) = min(\alpha_1, \alpha_2, \ldots, \alpha_m)$, for any $u_i \in U$, the following proposition can be obtained.

Proposition 4.1. *If* $\delta \in [\psi(S), 1)$ *then* $pos_0(S) = U$.

The classification problem of each object was discussed above. It has been shown that the VPRS model is only a special case of decision-theoretic rough sets, when the "majority inclusion relation" is interpreted using the conditional probability.

4.2. DTRS based on majority inclusion relation of VPRS

Suppose a decision table $S = (U, C \cup D, f)$, a partition of condition attributes $U/R_C = \{X_1, X_2, \ldots, X_t$ and a partition of decision attributes $U/R_D = \{D_1, D_2, \ldots, D_m\}$. Similar to the above discussion, the following minimum cost decision rules can be expressed as:

(VPRSNR$_1$) $\quad$ *if* $R(a_{P_j|X_i}) \leqslant R(a_{B_j|X_i})$ *and* $R(a_{P_j|X_i}) \leqslant R(a_{N_j|X_i})$, *then* $X_i \subseteq pos_\beta(D_j)$

(VPRSBR$_1$) $\quad$ *if* $R(a_{B_j|X_i}) \leqslant R(a_{P_j|X_i})$ *and* $R(a_{B_j|X_i}) \leqslant R(a_{N_j|X_i})$, *then* $X_i \subseteq bnd_\beta(D_j)$

(VPRSPR$_1$ $\quad$ *if* $R(a_{N_j|X_i}) \leqslant R(a_{P_j|X_i})$ *and* $R(a_{N_j|X_i}) \leqslant R(a_{B_j|X_i})$, *then* $X_i \subseteq negr_\beta(D_j)$

For the first condition of:

$$\begin{aligned}
& R(a_{P_j|X_i}) \leqslant R(a_{B_j|X_i}) \\
& \iff \sum_{k=1}^{t} \lambda_{P_jD_k} P(D_k|X_i) \leqslant \sum_{k=1}^{t} \lambda_{B_jD_k} P(D_k|X_i) \\
& \iff \lambda_{P_jD_j} P(D_j|X_i) + \sum_{k=1,k\neq j}^{t} \lambda_{P_jD_k} P(D_k|X_i) \leqslant \lambda_{B_jD_j} P(D_j|X_i) + \sum_{k=1,k\neq j}^{t} \lambda_{B_jD_k} P(D_k|X_i) \\
& \iff \lambda_{P_jD_j} P(D_j|X_i) + \lambda_{P_jD_{\sim j}} \sum_{k=1,k\neq j}^{t} P(D_k|X_i) \leqslant \lambda_{B_jD_j} P(D_j|X_i) + \lambda_{B_jD_{\sim j}} \sum_{k=1,k\neq j}^{t} P(D_k|X_i) \\
& \iff \lambda_{P_jD_j} P(D_j|X_i) + \lambda_{P_jD_{\sim j}} (1 - P(D_j|X_i)) \leqslant \lambda_{B_jD_j} P(D_j|X_i) + \lambda_{B_jD_{\sim j}} (1 - P(D_j|X_i) \\
& \iff P(D_j|X_i) \geqslant \frac{\lambda_{P_jD_{\sim j}} - \lambda_{B_jD_{\sim j}}}{(\lambda_{P_jD_{\sim j}} - \lambda_{B_jD_{\sim j}}) + (\lambda_{B_jD_j} - \lambda_{P_jD_j})}
\end{aligned}$$

We can also derive the corresponding results and define three parameters α_j, γ_j and β_j, and the rules can be simplified as:

$\left(\text{VPRSNewPR}_1'\right)$ *if* $P(D_j|X_i) \geqslant \alpha_j$, *then* $X_i \subseteq pos_\beta(D_j)$

$\left(\text{VPRSNewBR}_1'\right)$ *if* $\beta_j < P(D_j|X_i) < \alpha_j$, *then* $X_i \subseteq bnd_\beta(D_j)$

$\left(\text{VPRSNewNR}_1'\right)$ *if* $P(D_j|X_i) \leqslant \beta_j$, *then* $X_i \subseteq negr_\beta(D_j)$

By using the thresholds, one equivalence class can be classified into the three regions of a decision partition $pos_\beta(D_j)$, $bnd_\beta(D_j)$ *and* $negr_\beta(D_j)$ based on (α_j, β_j) in some classification error like VPRS. When $\psi'(S) = min(\alpha_1, \alpha_2, \ldots, \alpha_m)$, for any $X_i \in U/R_C$, the following proposition can be obtained.

Proposition 4.2. *If* $\delta' \in [\psi'(S), 1)$ *then* $pos_\beta(S) = U$.

It is not difficult to find that the parameters of the decision-theoretic rough set theory can be obtained by giving the cost functions and then calculating the parameters $\psi'(S)$,which can be used as the variable precision parameter. The proposed method is automatic parameter acquisition from a decision system without any prior knowledge and the conclusion $pos_\beta(S) = U$ is also the same with the decision-theoretic RST. Our method does not consider the risk factors, and therefore, $\xi(S) \geqslant 1 - \psi'(S)$.

Example 3. With the decision table in Table 1, the cost functions are defined in Table 2.

Then we have:

Table 2
The cost functions of Table 1.

	$\lambda_{P_jD_j}$	$\lambda_{B_jD_j}$	$\lambda_{N_jD_j}$	$\lambda_{N_j\sim D_k}$	$\lambda_{B_j\sim D_k}$	$\lambda_{P_j\sim D_k}$
D_1	0	2	9	0	4	12
D_2	0	3	12	0	6	13
D_3	0	4	15	0	8	14

$$\alpha_1 = \frac{12-4}{12-4+2-0} = 0.8$$
$$\alpha_2 = \frac{13-6}{13-6+3-0} = 0.7$$
$$\alpha_3 = \frac{14-8}{14-8+4-0} = 0.6$$

So:

$$\psi'(S) = min(\alpha_1, \alpha_2, \ldots, \alpha_m) = 0.6$$

Therefore, $\xi(S) \geqslant 1 - \psi'(S) = 0.4$.

Moreover, when we take action a_{P_1}, the equivalence classes $X_2 \subseteq D_1$, $X_3 \subseteq D_1$ according to the parameter $\alpha_1 = 0.8$, which is similar to VPRS when the majority inclusion relation is interpreted.

5. Influences of β on attribute reduction

In this section, we introduce several types of attribute reduction, such as β lower distribution reduction, β upper distribution reduction and β lower approximate reduction. After giving their notations, we discuss the influences of the threshold value obtained from a data set on attribute reduction.

5.1. β Lower (upper) distribution reduct [23]

Definition 5.1. Given a decision table $S = (U, C \cup D, f)$ and $B \subseteq C$, we denote

$$L_B^\beta = \left(\underline{R_B^\beta}(D_1), \ldots \underline{R_B^\beta}(D_m)\right),\ H_B^\beta = (\overline{R_B^\beta}(D_1), \ldots \left(\overline{R_B^\beta}(D_m)\right)$$

1) If $L_B^\beta = L_C^\beta$, we say that B is a β lower distribution consistent set of S. If B is a β lower distribution consistent set, and no proper subset of B is β lower distribution consistent, then B is referred to as a β lower distribution reduct of S.
2) If $H_B^\beta = H_C^\beta$, we say that B is a β upper distribution consistent set of S. If B is a β upper distribution consistent set, and no proper subset of B is β upper distribution consistent, then B is referred to as a β upper distribution reduct of S.

Proposition 5.1. *Given a decision table $S = (U, C \cup D, f)$ and $B \subseteq C$, with*

$$G_B^\beta(x) = \{D_j | x \in \underline{R_B^\beta}(D_j), x \in U\}, M_B^\beta(x) = \{D_j | x \in \overline{R_B^\beta}(D_j), x \in U\}$$

then (1) B is a β lower distribution consistent set iff $\forall x \in U, G_B^\beta(x) = G_C^\beta(x)$

(2) B is a β upper distribution consistent set iff $\forall x \in U, M_B^\beta(x) = M_C^\beta(x)$

Definition 5.2. For a decision table $S = (U, C \cup D, f)$ and $U/R_C = \{X_1, X_2, \ldots, X_n\}$, with $D_1^{*\beta} = \left\{([x]_C, [y]_C) | M_C^\beta(x) \neq M_C^\beta(y)\right\}, D_2^{*\beta} = \left\{([x]_C, [y]_C) | G_C^\beta(x) \neq G_C^\beta(y)\right\}$, we have:

$$D_k^\beta(X_i, X_j) = \begin{cases} a_l \in C | f(X_i, a_l) \neq f(X_j, a_l), & (X_i, X_j) \in D_k^{*\beta} \\ C, & (X_i, X_j) \notin D_k^{*\beta} \end{cases}, \quad k = 1, 2$$

where $f(X_i, a_l)$ stands for the value of a_l with respect to the objects in X_i. Then $D_k^\beta(X_i, X_j)(k = 1, 2)$ are referred to as β upper (lower) distribution and β lower distribution consistent discernible attribute sets respectively and $D_k^\beta(k = 1, 2)$ is referred to as β upper (lower) distribution and β lower distribution matrices respectively.

5.2. Influences of β on lower (upper) distribution reduct

In general, if β is not in the interval $[\xi(S), 0.5)$, then a β upper distribution consistent set may not be β lower distribution consistent. However, the following proposition will imply that β lower distribution reduction is equivalent to β upper distribution reduction for some special threshold values.

Proposition 5.2. *Given a decision table $S = (U, C \cup D, f)$ and $B \subseteq C$, if $\beta \in [\xi(S), 0.5)$ then a β upper distribution consistent set is equal to a β lower distribute consistent set.*

Table 3
Decision table.

U	a_1	a_2	a_3	a_4	d
x_1	1	0	0	0	1
x_2	0	1	1	1	2
x_3	0	1	0	0	2
x_4	0	1	0	1	2
x_5	0	1	0	0	1
x_6	0	1	0	0	1

Table 4
Influences of precision parameter on knowledge acquisition.

Credit-g	Selected conditional attributes (SCA1)	{a3,a6,a7,a8,a9,a12,a14,a19}+{d}							
	β	0	0.17	0.2	0.25	0.33	0.5		
	Number of rules	763	764	765	773	785	46		
	Coverage	0.862	0.868	0.873	0.909	0.951	0.049		
	Selected conditional attribute (SCA2)	{a3,a6,a7,a9,a10,a11,a12,a15,a16,a19,a20}+{d}							
	β	0	0.25	0.33	0.4	0.43	0.5		
	Number of rules	836	838	844	845	846	46		
	Coverage	0.913	0.921	0.939	0.944	0.951	0.049		
	Selected conditional attributes (SCA3)	{a3,a6,a12,a14,a19}+{d}							
	β	0	0.2	0.3	0.38	0.42	0.44	0.47	0.5
	Number of rules	153	166	181	196	205	207	208	48
	Coverage	0.275	0.449	0.660	0.728	0.886	0.903	0.929	0.071
KR-VS-KP	Selected conditional attributes	{a3,a6,a12,a14,a19}+{d}							
	β	0	0.2	0.35	0.4	0.42	0.44	0.5	
	Number of rules	1	3	4	5	6	8	NULL	
	Coverage	0.001	0.015	0.038	0.044	0.135	1		

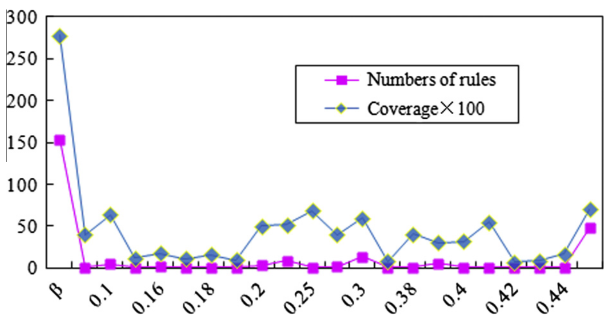

Fig. 1. Changes about numbers of rules and coverage in credit-g with β.

Proof. According to Proposition 3.2, for any $D_i \in U/R_D$, if $\beta \in [\xi(S), 0.5)$, then $bnr_\beta(D_i) = \phi$, i.e., $\overline{R_B^\beta}(D_i) = \underline{R_B^\beta}(D_i)$. As a result we have $L_B^\beta = H_B^\beta$. Therefore, a β upper distribution consistent set is equal to aβlower distribution consistent set when β is in the domain $[\xi\ (S), 0.5)$. □

Example 4. Consider a decision table $S = (U, C \cup D, f)$ in Table 3 [23].

The decision classes of objects are:

$$D_1 = \{x_1, x_5, x_6\},\ D_2 = \{x_2, x_3, x_4\}$$

The condition attributes of objects are:

$$X_1 = \{x_1\}, X_2 = \{x_2\},\ X_3 = \{x_3, x_5, x_6\},\ X_4 = \{x_4\}$$

As $M_A^{0.4}(x_1) = \{D_1\}, M_A^{0.4}(x_2) = \{D_2\}, M_A^{0.4}(x_3) = \{D_1\}, M_A^{0.4}(x_4) = \{D_2\}$, we have $D_1^{*0.4} = \{(X_1, X_2), (X_1, X_4), (X_2, X_3), (X_3, X_4)\}$.

Due to the symmetry of the matrix, we can denote it as an upper triangular matrix:

$$\begin{bmatrix} C & a_1a_2a_3a_4 & C & a_1a_2a_4 \\ & C & a_3a_4 & C \\ & & C & a_4 \\ & & & C \end{bmatrix}$$

Therefore, when β = 0.4, the upper distribution reduct set of S is

$(a_1 \vee a_2 \vee a_3 \vee a_4) \wedge (a_1 \vee a_2 \vee a_4) \wedge \ (a_3 \vee a_4) \wedge (a_4)$ = $\{a_4\}$ which is the same as in [10] for the same decision table. The reason that leads to the same result is as follows.

Since

$$CoD = \begin{bmatrix} 1 & 0 & \frac{2}{3} & 0 \\ 0 & 1 & \frac{1}{3} & 1 \end{bmatrix} \overline{CoD} = \begin{bmatrix} 0 & 1 & \frac{1}{3} & 1 \\ 1 & 0 & \frac{2}{3} & 0 \end{bmatrix}$$

we have $\xi(S) = \frac{1}{3}$.

According to Proposition 5.1, because $\beta = 0.4 \in [\frac{1}{3}, 0.5)$, the β lower distribution reduct set of $S$is equal to the upper distribution reduct set and then we can conclude that the same result is relative to the β value rather than the decision table itself.

5.3. Influence of β value on lower approximate reduction

For a decision table $S = (U, C \cup D, f)$and $B \subseteq C$, we denote $\gamma_B^\beta = pos_\beta(S)/card(U)$. When $\gamma_B^\beta = \gamma_A^\beta$, we say that B is a β lower approximate consistent set of S.

If B is a β lower approximate consistent set, but no proper subset of B is β lower approximate consistent, then B is referred to as a β lower approximate reduction of S.

In general, the value of γ_A^β is different for a different β value, which leads to a different reduction set. With the method we proposed above, a proper β value and the right reduction set can be obtained from the decision table.

According to Proposition 3.2, if $\beta \in [\xi(S), 0.5)$ then $\gamma_A^\beta = 1$. Therefore the verification of lower approximate reduction can be changed into the question: whether γ_B^β is equal to 1. Besides, the value γ_A^β measures the proportion of objects in the universe for which classification (based on decision attributes D) is possible at the specified value of β. In other words, it involves combining all β positive regions. Then we can conclude that the rules are $C \Rightarrow D$ which preserve the classification errors when $\gamma_A^\beta = 1$. It can be easily calculated that the 0.4 lower approximate reduction set is $\{a_4\}$.

Table 5
Simulation results with the VPRS model.

β		0.2		0.3		0.38		0.42		0.44		0.47		0.48	0.49	0.5
Category of rules		NSR	NWR	NSR	NWR	NSR	NWR	NSR	NWR	NSR	NWR	NSR	NWR	NSR	NSR	NCR
Credit-g	Numbers of rules	166	42	181	27	196	12	205	3	207	1	208	0	208	208	48
	Coverage	0.449	0.47	0.66	0.259	0.728	0.19	0.886	0.033	0.903	0.016	0.929	0	0.929	0.929	0.07
KR-VS-KP	Numbers of rules	2	6	2	6	4	4	6	2	6	2	7	1	8	8	NULL
	Coverage	0.03	0.137	0.03	0.986	0.038	0.978	0.137	0.879	0.137	0.879	0.955	0.063	1	1	

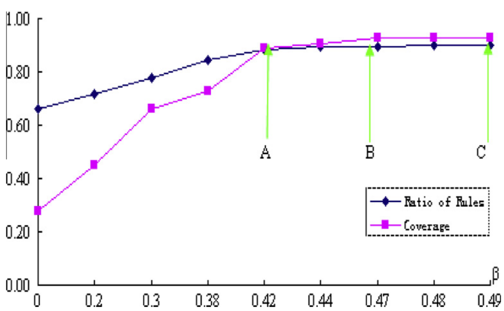

Fig. 2. Influences of precision parameter on knowledge acquisition in credit-g.

6. Simulations in real data

6.1. Simulation method of VPRS

To evaluate the influences of β on knowledge acquisition, some of the following measures need to be considered:

(1) **N**umbers of **S**trong **R**ules, usually abbreviated as **NSR**, and **N**umbers of **W**eak **R**ules, abbreviated as **NWR**, and **N**umbers of **C**onflict **R**ules, abbreviated as **NCR**

(2) Classification Accuracy

In a case of a classification system, the quality of a complete set of rules is evaluated through using the classification accuracy rate (or conversely the misclassification error rate). It is defined by the ratio, denoted by Coverage: $Coverage = n_c/n$

where n is the number of U, and n_c is the number of the β positive region, i.e., $n_c = |pos_\beta(S)|$.

The experiments were performed on several different real-life data sets, which were obtained from http://www.mkp.com/datamining. In order to simplify the evaluation problem, we first select the part of condition attributes, and then observe the influences on the Number of Rules and the Coverage with Precision Parameter. The simulation results are shown in Table 4.

Fig. 1 shows the Number of Rules and Coverage changes with β values.

Please notice that the data sets used in the experiments were not assumed to be absolutely inconsistent, i.e. we did not consider $\xi(S) = 0.5$. Due to this fact, some of the data were slightly modified by removing a few examples or attributes. For example, the minimal threshold values were obtained from real data sets such as Credit-g and KR-VS-KP, i.e. $\xi(Credit-g) = 0.47$, $\xi(KR-VS-KP) = 0.44$. Simulation results (see Table 5) show that the smaller the precision parameter is, the less the coverage of the rules is. When the precision parameter reaches $\xi(S)$, the coverage arrives at the maximum, being able to cover all the records of the decision table.

Table 6
The simulation results with other classification models.

Tools	Coverage					
	VPRS($\beta = 0$)	C4.5	AdaBoostM1	Logistic	PART	Naive bayes
Credit-g						
SCA1	0.862	0.709	0.701	0.722	0.691	0.728
SCA2	0.913	0.685	0.706	0.713	0.712	0.723
SCA3	0.275	0.716	0.706	0.718	0.704	0.718

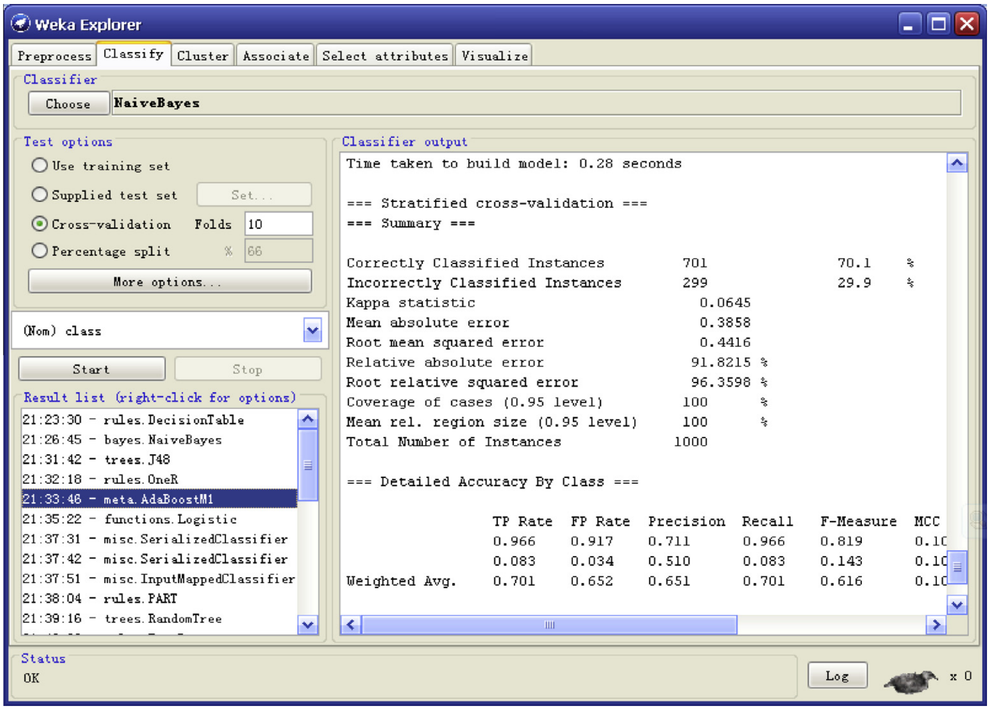

Fig. 3. A screenshot of other mining tools in WEKA.

Taking the Credit-g data set for example, Fig. 2 reflects the change curve of strong rule coverage and the ratio of rules with the precision parameter after normalization. Point B is the $\xi(S)$. It is shown from the figure that when the precision parameter is set between points B and C, i.e.,when $\beta \in [\xi(S), 0.5)$, the coverage reaches the maximum, having no effect on knowledge acquisition. Meanwhile, when point B is near the curve, there is a clear transitional point A, which can roughly reflect the decision distribution of the decision table. From the angle of knowledge acquisition, point A as a precision parameter is sufficient to reflect a rough decision distribution of the original decision table. As a result, setting the precision parameter close to $\xi(S)$ can relatively precisely reflect the decision distribution of the decision table.

6.2. Comparison with other classification models

In order to compare our method of VPRS, we used the WEKA software (Waitato Environment for Knowledge Analysis). Within Weka, we use C4.5, AdaBoostM1, Logistic and other classification methods to derive useful knowledge from Credit-g. As these classifiers in WEKA do not consider partially incorrect classification, the precision parameter is set to zero.The comparison was made between VPRS when $\beta = 0$ and other mining tools (see Table 6). Table 6 shows that the VPRS method is better than others when some parts of the conditional attributes were selected such as SCA1 and SCA2. As the choices of condition attributes are too few; the database sample conflicts increase. The results of SCA3 are relatively worse than others because partially incorrect classification should not be taken into account of the VPRS method when $\beta = 0$.

Fig. 3 shows a screenshot of the AdaBoostM1 algorithm.

7. Conclusion

The decision-theoretic rough set (DTRS) uses conditional probability to measure the uncertainty of the three regions by introducing a pair of thresholds. By applying an uncertainty degree or decision cost in the DTRS theory, some efficient methods to systematically compute and interpret the required thresholds have been suggested in the literature. As a special case of DTRS, the variable precision rough set (VPRS) promotes the applications of rough set theory in inconsistent information systems. However, in most existing algorithms of attribute reduction based on VPRS, the value of β is introduced as prior knowledge, which restricts their applications. The paper proposed the concept of a β-consistent decision table and an algorithm for determining the discernible threshold value. It has been proven that a value of β chosen from the interval of threshold values can get an attribute reduction set at a certain level of classification errors. Thus, the proposed method can automatically acquire the parameter value from the decision system without any prior knowledge like existing DTRS approaches.However, our method uses the maximal inclusion degree and is different from applying an uncertain degree or decision cost function in DTRS.Also, the time complexity of our method is significantly smaller than DTRS methods. Meanwhile, from the viewpoint of DTRS, we should point out that our method only provides a symmetric threshold $(1-\beta, \beta)$ method to obtain an optimal or close to optimal threshold. In this case, how to estimate an asymmetric threshold pair (α, β) from the data to be processed is our future work.

Acknowledgements

The authors sincerely thank the anonymous reviewers for their valuable and constructive comments.This research is supported by the National Natural Science Foundation of China (NSFC Nos. 61306046 and 61229301), the Program for Changjiang Scholars and Innovative Research Team in University (PCSIRT) of the Ministry of Education, China (under Grant IRT13059), the National 973 Program of China under Grant 2013CB329604, the Natural Science Foundation of Higher Education of Anhui Province (No. KJ2013A177), and the Natural Science Funds for Young Scholars's Research of Anhui Province (No. 10040606Q42).

References

[1] A. An, N. Shan, C. Chan, N. Ziarko, Discovering rules for water demand prediction: an enhanced rough set approach, Eng. Appl. Artif. Intell. 9 (6) (1996) 645–653.
[2] N. Azam, J.T. Yao, Analyzing uncertainties of probabilistic rough set regions with game-theoretic rough sets, Int. J. Approx. Reason. 55 (2014) 142–155.
[3] M. Beynon, An investigating of β-reduct selection within the variable precision rough sets model, in: The Second International Conference on Roughsets and Current Trend in Computing, Lecture Notes in Artificial Intelligence Series, Springer, Berlin, 2000, pp. 114–122.
[4] M. Beynon, Reducts within the variable precision rough sets model: a further investigation, Eur. J. Oper. Res. 134 (2001) 592–605.
[5] M. Beynon, K.L. Buchanan, An illustration of variable precision rough set theory: the gender classification of the European barn swallow (Hirundorustica), Bull. Math. Biol. 65 (5) (2003) 835–858.
[6] M. Beynon, The introduction and utilization of (l-u)-graphs inthe extended variable precision rough sets model, Int. J. Intell. Syst. 18 (2003) 1035–1055.
[7] H.M. Chen, T.R. Li, D. Ruan, Dynamic maintenance of approximations under a rough-set based variable precision limited tolerance relation, J. Multiple-Valued Logic Soft Comput. 18 (5–6) (2012) 577–598.
[8] Y.S. Chen, C.H. Cheng, Application of rough set classifiers for determining hemodialysis adequacy in ESRD patients, Knowl. Inform. Syst. 34 (2) (2013) 453–482.
[9] Y.S. Cheng, Y.S. Zhang, X.G. Hu, The Relationships Between Variable Precision value and Knowledge Reduction Based on Variable Precision Rough Sets Model, RSKT2006, Springer-Verlag, Berlin, 2006. LNAI 4062, pp. 122–128.
[10] X.F. Deng, Y.Y. Yao, An information-theoretic interpretation of thresholds in probabilistic rough sets, RSKT 2012, LNCS(LNAI), vol. 7414, 2012, pp. 369–378.
[11] X.F. Deng, Y.Y. Yao, Decision-theoretic three-way approximations of fuzzy sets, Inform. Sci. 279 (2014) 702–715.

[12] J.P. Herbert, J.T. Yao, Game-theoretic rough sets, Fundamenta Informaticae 108 (3–4) (2011) 267–286.
[13] B.Q. Hu, Three-way decisions space and three-way decisions, Inform. Sci. 281 (2014) 21–52.
[14] M. Inuiguchi, Y. Yoshioka, Y. Kusunoki, Variable-precision dominance-based rough set approach and attribute reduction, Int. J. Approx. Reason. 50 (8) (2009) 1199–1214.
[15] X.Y. Jia, W.H. Liao, Z.M. Tang, L. Shang, Minimum cost attribute reduction in decision-theoretic rough set models, Inform. Sci. 219 (2013) 151–167.
[16] X.Y. Jia, Z.M. Tang, W.H. Liao, L. Shang, On an optimization representation of decision-theoretic rough set model, Int. J. Approx. Reason. 55 (2014) 156–166.
[17] J.K. Katzberg, W. Ziarko, Variable precision rough sets with asymmetric bounds, in: RSKD1993, Springer, New York, 1993, pp. 167–177.
[18] G.M. Lang, Q.G. Li, L.K. Guo, Discernibility matrix simplification with new attribute dependency functions for incomplete information systems, Knowl. Inform. Syst. 37 (3) (2013) 611–638.
[19] H.X. Li, X.Z. Zhou, Risk decision making based on decision-theoretic rough set: a multi-view decision model, Int. J. Comput. Intell. Syst. 4 (1) (2011) 1–11.
[20] D. Liu, T.R. Li, D. Ruan, Probabilistic model criteria with decision-theoretic rough sets, Inform. Sci. 181 (17) (2011) 3709–3722.
[21] G.L. Liu, K. Zhu, The relationship among three types of rough approximation pairs, Knowl.-Based Syst. 60 (2014) 28–34.
[22] C. Luo, T.R. Li, H.M. Chen, D. Liu, Incremental approaches for updating approximations in set-valued ordered information systems, Knowl.-Based Syst. 50 (2013) 218–233.
[23] J.S. Mi, W.Z. Wu, W.X. Zhang, Approaches to knowledge reduction based on variable precision rough sets model, Inform. Sci. 159 (2004) 255–272.
[24] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning About Data, Kluwer Academic Publishers, Boston, 1991.
[25] W. Pedrycz, Granular Computing: Analysis and Design of Intelligent Systems, CRC Press/Francis Taylor, Boca Raton, 2013.
[26] Y.H. Qian, J.Y. Liang, W. Pedrycz, C.Y. Dang, Positive approximation: an accelerator for attribute reduction in rough set theory, Artif. Intell. 174 (9–10) (2010) 597–618.
[27] A. Skowron, C. Rauszer, The discernibility matrices and functions in information systems, in: R. Slowinski (Ed.), Intelligent Decision Support, Handbook of Applications and Advances of the Rough Sets Theory, Kluwer Academic Publishers., Dordrecht, 1992, pp. 331–362.
[28] D. Slezak, Attribute reduction in the Bayesian version of variable precision rough set model, Electron. Theor. Comput. Sci. 82 (2003) 1–11.
[29] D. Slezak, W. Ziarko, The investigation of the Bayesian rough set model, Int. Approx. Reason. 40 (2005) 81–91.
[30] C.T. Su, J.H. Hsu, Precision parameter in the variable precision rough sets model: an application, Omega-Int. J. Manage. Sci. 34 (2) (2006) 149–157.
[32] Z. Xiao, W.J. Chen, L.L. Li, A method based on interval-valued fuzzy soft set for multi-attribute group decision-making problems under uncertain environment, Knowl. Inform. Syst. 34 (3) (2013) 653–669.
[33] X.B. Yang, Y.S. Qi, X.N. Song, J.Y. Yang, Test cost sensitive multigranulation rough set: model and minimal cost selection, Inform. Sci. 250 (2013) 184–199.
[34] Y.Y. Yao, S.K.M. Wong, A decision theoretic framework for approximating concepts, Int. J. Man–Machine Stud. 37 (6) (1992) 793–809.
[35] Y.Y. Yao, Y. Zhao, Attribute reduction in decision-theoretic rough set models, Inform. Sci. 178 (17) (2008) 3356–3373.
[36] Y.Y. Yao, The superiority of three-way decisions in probabilistic rough set models, Inform. Sci. 181 (6) (2011) 1080–1096.
[37] W.X. Zhang, Y. Liang, P. Xu, Uncertainty Reasoning Based on Inclusion Degree, Tsinghua University Press, Beijing, 2007. in Chinese.
[38] W.X. Zhang, W.Z. Wu, H.Y. Zhang, Y. Leung, L. Zhou, Variable-precision-dominance-based rough set approach to interval-valued information systems, Inform. Sci. 244 (2013) 75–91.
[39] J. Zhou, D.Q. Miao, β-Interval attribute reduction in variable precision rough set model, Soft Comput. 15 (2011) 1643–1656.
[40] W. Ziarko, Variable precision rough set model, J. Comput. Syst. Sci. 46 (1) (1993) 39–59.
[41] W. Ziarko, Set approximation quality measures in the variable precision rough set model, in: Soft Computing Systems Management Applications, IOS Press, 2001, pp. 442–452.

Further Reading

[31] S.B. Sun, M.C. Zhang, C.X. Hu, VPRS-based attribute reduction approaches in incomplete information systems, J. Inform. Comput. Sci. 9 (8) (2011) 1515–1523.