Soft Comput (2011) 15:1643–1656
DOI 10.1007/s00500-011-0693-4

ORIGINAL PAPER

β-Interval attribute reduction in variable precision rough set model

Jie Zhou · Duoqian Miao

Published online: 24 February 2011

Abstract The differences of attribute reduction and attribute core between Pawlak's rough set model (RSM) and variable precision rough set model (VPRSM) are analyzed in detail. According to the interval properties of precision parameter β with respect to the quality of classification, the definition of attribute reduction is extended from a specific β value to a specific β interval in order to overcome the limitations of traditional reduct definition in VPRSM. The concept of β-interval core is put forward which will enrich the methodology of VPRSM. With proposed ordered discernibility matrix and relevant interval characteristic sets, a heuristic algorithm can be constructed to get β-interval reducts. Furthermore, a novel method, with which the optimal interval of precision parameter can be determined objectively, is introduced based on shadowed sets and an evaluation function is also given for selecting final optimal β-interval reduct. All the proposed notions in this paper will promote the development of VPRSM both in theory and practice.

Keywords Variable precision rough set model (VPRSM) · β-Interval reduct · β-Interval core · Interval characteristic sets · Shadowed sets

1 Introduction

Rough set theory (Pawlak 1982; Pawlak and Skowron 2007) has developed dramatically since it was introduced by Pawlak in 1982. It is a new mathematic methodology aimed at data analysis problems involving uncertain, imprecise or incomplete information, and has had widespread success in many research fields, such as knowledge acquisition, pattern recognition, economic forecast, data mining and others (Jelonek et al. 1995; Shen and Chouchoulas 2002; Dimitras et al. 1999; Anantaram et al. 1998; Mushrif and Ray 2008). However, Pawlak's rough set model (RSM) is founded on classical set theory and only the information gathered from positive region will be utilized to acquire decision rules. So, original rough set model cannot deal with data sets which have noisy data effectively, and some latent useful knowledge in boundary region may be fully abandoned. One of the major challenges in rough set model for data analysis is to minimize the size of the boundary region (Herbert and Yao 2009), and some extended rough set models have been put forward (Pawlak et al. 1988; Dubois and Prade 1999; Zhu and Wang 2003; Wojciech et al. 2008; Greco et al. 2008).

Variable precision rough set model (VPRSM) (Ziarko 1993), as one of the most important branches, was proposed by Ziarko in 1993. Standard inclusion relation is extended to majority inclusion relation and data patterns can be analyzed from the perspective of statistics. Strict functional dependence between attributes is relaxed to weak dependence, and more general association rules can be mined in VPRSM. Subsequently, Ziarko et al. put forward asymmetric variable precision rough set model (AVPRSM) (Katzberg and Ziarko 1993), thus the model becomes more general. Variable precision rough set models, symmetric or asymmetric, have also been applied in many domains (An et al. 1996; Beynon and Peel 2001). However, they have some parameters involved, β or l and u. Different parameter values will result in distinct rule sets.

J. Zhou (✉) · D. Miao
Department of Computer Science and Technology, Tongji University, Shanghai 201804, People's Republic of China
e-mail: jie_jpu@163.com

D. Miao
e-mail: miaoduoqian@163.com

Attribute reduction (Pawlak and Skowron 2007; Thangavel and Pethalakshmi 2009) is one of the most fundamental notions in RSM, as well as in VPRSM. Ziarko's reduct definition is widely used both in theory and practice, but some anomalies may be produced in the procedures of attribute reduction under this definition. Because some monotonic properties in RSM are not satisfied in VPRSM. Attribute reduction anomalies are first analyzed by Beynon (2001) and the concept of unrestricted β-reduct is proposed which, yet, is too strict to discover latent useful knowledge from data sets. Attribute reduction anomalies under Ziarko's definition are further analyzed (Wang and Zhou 2009) and reduct hierarchy is established to eliminate the anomalies gradually.

In practical applications, it is difficult to choose optimal parameters in VPRSM. Relevant parameter values are often given subjectively. Ziarko (1999) considers that β value should be chosen not only according to the external knowledge that will improve the predictions, but also depending on the intrinsical quality of information that included in the data sets. Su and Hsu (2006) gives a method to calculate precision parameter value based on the least upper bound of the data misclassification error. Beynon (2003) utilizes the proposed $(l-u)$-graphs to choose the values of l and u based on the associated levels of the quality of classification and the degree of dependency. Slezak and Ziarko (2002, 2005) and Slezak (2003) put forward Bayesian rough set model, in which parameter in VPRSM is defined by the prior probability of occurrence of the event under consideration. Despite that all of these methods are aimed at searching for an optimal precision parameter objectively, they still focus on a special β value and attribute reduction anomalies will occur inevitably under classical reduct definition.

Based on Bayesian decision procedure with minimum risk, Yao (Yao et al. 1990; Yao and Wong 1992; Yao 2007) puts forward decision-theoretic rough set model (DTRSM) which brings new insights into the probabilistic approaches to rough set model (Yao 2008). DTRSM not only has good semantic interpretation, but also is beneficial for rule acquisition in the applications involving costs and risks (Yao 2003; Yao and Zhao 2008). VPRSM and AV-PRSM can be directly derived from DTRSM under appropriate loss functions, namely, parameters in VPRSM can be determined by the loss functions. Therefore, both VPRSM and VPRSM can be considered as the special cases of DTRSM. The loss functions which play an important role in Bayesian decision procedure are often determined by experts' prior knowledge or domain knowledge combining with concrete problems.

Beynon (2000) has discussed the relationship between precision parameter and classification quality for a given decision table in VPRSM. Two kinds of descriptive principles for selecting β-reduct are proposed in light of β interval properties. However, different classification qualities correspond to different β intervals; therefore, how to choose an appropriate β interval based on data set itself needs to be resolved. In addition, the reduct definition can be extended from a specific β value to a specific β interval according to the interval properties of precision parameter, and constructive algorithms for obtaining β-interval reducts are still on the way.

Attribute core is an important notion in RSM and plays significant role in heuristic algorithms for attribute reduction (Wang and Miao 1998; Yao et al. 2006). Since it is included in all reducts, it is often selected as the starting point in heuristic algorithms. Up to now, attribute core in VPRSM is rarely discussed in literatures. This sounds imperfect for VPRSM compared with RSM methodology.

In this paper, we only concentrate on symmetric variable precision rough set model. The differences of attribute reduction and attribute core between RSM and VPRSM are analyzed critically. The concept of β-interval core, which enriches the methodology of VPRSM, is put forward in light of β interval properties. Three kinds of interval characteristic sets are defined based on proposed ordered discernibility matrix, and their relationships with β-interval core are discussed. Since these characteristic sets can be easily and automatically obtained according to β-separate lines, a heuristic algorithm to get β-interval reducts can be constructed based on them. Furthermore, a novel method to determine optimal β interval is proposed on the basis of shadowed sets which will be more objective. Finally, an evaluation function for choosing β-interval reducts is also introduced.

2 Preliminaries

For convenience in presenting our notions, some basic concepts in VPRSM will be briefly introduced in this section. More detailed description can be found in (Pawlak 1982; Pawlak and Skowron 2007; Yao et al. 1990; Ziarko 1993).

Definition 1 (Pawlak and Skowron 2007). A decision information system DT can be described as the tuple: $DT = (U, A = C \cup D, V, \rho)$. U is a universe that will be interested in(usually, a finite and nonempty set of objects). A is a set of attributes, where C and D are disjoint, C is the set of condition attributes and D is the set of decision attributes. $V = \cup_{a \in (C \cup D)} V_a$ is the union of attribute value domains, V_a is a nonempty set of values for attribute $a \in A$. $\rho : U \times A \rightarrow V$ is an information mapping function, $\rho\ (o, a)$ means that object o has the value on attribute a.

Decision information system is often called as decision table briefly. With equivalence relation C, referred to as an indiscernibility relation, universe U can be partitioned into a collection of equivalence classes $U/C = \{C_1, C_2, \ldots, C_{|U/C|}\}$. Similarly, with equivalence relation D, $U/D = \{D_1, D_2, \ldots, D_{|U/D|}\}$. Each element of U/C is called condition class and each element of U/D is called decision class. The notation $|X|$ denotes the cardinality of set X.

Definition 2 (Ziarko 1993). Given a decision table $DT = (U, C \cup D, V, \rho)$ and precision parameter $\beta \in (0.5, 1]$, $\forall D_j \in U/D$ $(j = 1, 2, \ldots, |U/D|)$, its β-upper and β-lower approximation with respect to C are denoted as $C^{\beta}D_j$ and $C_{\beta}D_j$, respectively, and defined as

$$\begin{aligned} C^{\beta}D_j &= \cup\left\{ C_i \middle| \frac{|C_i \cap D_j|}{|C_i|} > 1 - \beta, C_i \in U/C \right\} \\ C_{\beta}D_j &= \cup\left\{ C_i \middle| \frac{|C_i \cap D_j|}{|C_i|} \geq \beta, C_i \in U/C \right\} \end{aligned} \tag{1}$$

where $i = 1, 2, \ldots, |U/C|$. $BN(C, D_j, \beta) = C^{\beta}D_j - C_{\beta}D_j$ is called as β-boundary region of decision class D_j with respect to C. For each decision class $D_j \in U/D$, *its* β-lower approximation is the collection of all those condition classes that can be classified into D_j with the certainty degree not lower than β; *its* β-upper approximation is composed of all those condition classes that can be classified into D_j with the certainty degree higher than $1 - \beta$; and its β-boundary region is composed of all those condition classes which can be classified into D_j with the certainty degree between $1 - \beta$ and β.

Obviously, the β-boundary region will be contracted and the β-lower approximation will be extended for each decision class $D_j \in U/D$ in VPRSM compared with its boundary region and lower approximation under the same attribute set in RSM, respectively. More objects, which are in the boundary region that cannot be classified in RSM, can be classified definitely under a reasonable misclassification rate in VPRSM.

Definition 3 (Ziarko 1993). Given a decision table $DT = (U, C \cup D, V, \rho)$ and precision parameter $\beta \in (0.5, 1]$, the quality of classification (or called as β-approximation degree of dependence) of decision attribute set D with respect to condition attribute set C is defined as

$$\gamma(C, D, \beta) = \frac{|POS(C, D, \beta)|}{|U|} \tag{2}$$

where $POS(C, D, \beta) = \cup_{D_j \in U/D} C_{\beta}D_j$ is called as β-positive region of decision attribute set D with respect to condition attribute set C. In the sequel, $POS(C, D, \beta)$ is referred to β-positive region briefly.

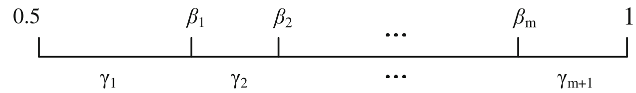

Fig. 1 The relationship between classification qualities and β intervals

With a given decision table, we can establish the relationship between classification qualities and β intervals under the entire attribute set. The procedure for establishing this intrinsic relationship does not need any additional information, only the data set itself. The details can be referenced from Wang and Zhou (2009). Different qualities of classification will be one-to-one correspondence with different β intervals (Beynon 2001), as shown in Fig. 1, where $1 \geq \gamma_1 > \gamma_2 > \cdots > \gamma_{m+1} \geq 0, 0.5 < \beta_1 < \beta_2 < \cdots < \beta_m < 1$.

As discussed in the introduction, it is not easy to choose an optimal precision value in VPRSM. However, the intrinsic relationship between classification qualities and β intervals under entire set of attributes can be established directly from data set itself. Given a specific classification quality, we can get its associated β interval. The lower the classification quality is, the higher the associated β interval will be. In this case, objects can be classified certainly with higher accuracy rate. It indicates that fewer objects may be misclassified. On the other side, the higher the classification quality is, the lower the associated β interval will be. Thus, more objects can be classified with lower correct rate. It indicates that more objects may be misclassified.

Definition 4 (Wang and Zhou 2009). Given a decision table $DT = (U, C \cup D, V, \rho)$, for each condition class $C_i \in U/C(i = 1, 2, \ldots, |U/C|)$, its threshold of inclusion degree over all decision classes is defined as

$$TS(C_i) = \max_{j=1}^{|U/D|} \frac{|C_i \cap D_j|}{|C_i|} \tag{3}$$

$TS(C_i)$ means the highest inclusion degree of C_i over all decision classes. If $TS(C_i) > 0.5$, then the decision value of C_i is the value which corresponds to the decision class D_k, viz. $\rho(D_k, D)$, where D_k satisfies $TS(C_i) = \frac{|C_i \cap D_k|}{|C_i|}$. If $TS(C_i) < 0.5$, it is considered that C_i has no decision value. In this case, C_i is absolutely unclassified with respect to D. Directly, if $TS(C_i) > 0.5$ and $C_i \subseteq POS(C, D, TS(C_i))$, then $\forall \beta \in (0.5, TS(C_i)]$; it has $C_i \subseteq POS(C, D, \beta)$.

3 Attribute reduction in VPRSM

The quality of classification γ is often used to measure the classification ability for a decision table. Nevertheless, the final rule inference is based on the concrete objects in positive region. In RSM, the monotonicity of classification quality and positive region are uniform during the procedures of attribute reduction. The same classification quality means the same positive region. In VPRSM, these monotonic properties will not be satisfied and the same classification quality may result in different positive regions (Yao 2008; Wang and Zhou 2009), namely, the final rule sets maybe distinct under the same classification quality. So the measurement of classification ability with classification quality is no longer reasonable in VPRSM. In addition, Ziarko's attribute reduct definition is under a special value, but this definition will bring some anomalies into the procedures of attribute reduction (Beynon 2001). According to β interval properties in VPRSM, the reduct definition can be extended from a specific β value to a specific β interval.

Definition 5 Given a decision table $DT = (U, C \cup D, V, \rho)$ and classification quality γ, if subset $RED \subseteq C$ satisfies

(1) $\beta_C^{\gamma} \cap \beta_{RED}^{\gamma} \neq \emptyset$;
(2) $POS(C, D, \beta_C^{\gamma}) = POS(RED, D, \beta_{RED}^{\gamma})$;
(3) $\forall RED' \subset RED$, at least one of conditions (1) and (2) is not satisfied.

Then RED is called as a β_C^{γ} interval reduct of condition attribute set C with respect to decision attribute set D. In the sequel, it is called as a β_C^{γ} interval reduct briefly.

Where β_C^{γ} and β_{RED}^{γ} denote the β intervals in which classification qualities are equal to γ under condition attribute set C and reduct RED, respectively. $\beta_C^{\gamma} \cap \beta_{RED}^{\gamma}$ is the final precision parameter interval for reduct RED.

In Definition 5, condition (1) indicates there are some common β values for reduct RED and the entire condition attribute set C under the same classification quality. It presents the interval properties of attribute reduction. Condition (2) preserves the same β-positive region after reducing some attributes. Condition (3) indicates that there are no redundant attributes in the final reduct; in other words, the classification information of original decision table can be retained with minimal attribute set. Compared with RSM, a β-interval reduct is not only focused on material attributes, but also associated with β interval characteristics. We can use the tuple *<reduct*, *reduct interval>* to describe a β-interval reduct.

The classification quality γ in definition 5 should be predetermined. It can be chosen based on the relationship between classification qualities and β intervals which is embedded in data set (as shown in Fig. 1). In fact, with a given decision table, the alternative classification qualities are finite discrete values. Generally, the number of these discrete values are very small under a given decision table. According to different γ values, attribute reduction model can be established, respectively.

According to Definition 5, if $B \subset C$ is not a β_C^{γ} interval reduct, it indicates (1) $\beta_C^{\gamma} \cap \beta_B^{\gamma} = \emptyset$, namely, after reducing some attributes, no common β values will exist for attribute set C and B under the given classification quality; (2) $\beta_C^{\gamma} \cap \beta_B^{\gamma} \neq \emptyset$, but $POS(C, D, \beta_C^{\gamma}) \neq POS(B, D, \beta_B^{\gamma})$, it means β-positive region has been changed, despite that there are some common β values under the same classification quality after reducing some attributes.

Due to the monotonic properties of classification quality and positive region with respect to attribute set inclusion in RSM, whether an attribute can be reduced or not, it can be judged only through the change of classification quality or positive region after removing this attribute. However, Attribute reduction is more complicated in VPRSM. The threshold of inclusion degree of each condition class will be dynamically changed during the procedures of attribute reduction, and the monotonic properties of classification quality and positive region with respect to attribute set inclusion would not be satisfied. After reducing some attributes, positive region may be extended and classification quality may be elevated in the same β interval. These phenomena can be shown by the following example.

Example 1 Suppose a decision table $DT1$, as shown in Table 1. $C = \{a_1, a_2, a_3, a_4, a_5\}$ is condition attribute set, $D = \{d\}$ is decision attribute set. The relationship between classification qualities and β intervals under entire attribute set C in $DT1$ is presented as in Fig. 2.

According to β-interval reduct definition 5, we can get a $\beta_C^{7/9}$ interval reduct $\{a_1, a_2, a_3\}$ and its associated reduct interval is (0.5, 0.6]. The procedures of attribute reduction, in which attribute a_4 and a_5 will be removed one by one, can be shown in Table 2, where $U/C = \{C_1, C_2, \ldots, C_6\}$, $U/B = \{B_1, B_2, \ldots, B_5\}$ and $U/B' = \{B'_1, B'_2, \ldots, B'_4\}$.

After removing attribute a_4, the quality of classification is decreased from 7/9 to 3/9 in interval (0.5,0.667]. In this

Table 1 Decision table $DT1$

U	a_1	a_2	a_3	a_4	a_5	d
o_1	1	1	1	2	1	Y
o_2	1	1	1	0	1	Y
o_3	0	1	0	1	0	Y
o_4	0	1	1	2	1	Y
o_5	1	1	1	0	0	N
o_6	1	1	1	0	1	N
o_7	0	0	0	1	1	N
o_8	1	1	1	0	1	N
o_9	0	1	1	2	1	N

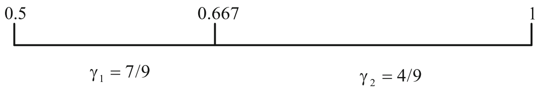

Fig. 2 The relationship between classification qualities and β intervals under attribute set C in $DT1$

case, attribute set $C - \{a_4\}$ is not a $\beta_C^{7/9}$ interval reduct. Subsequently, we remove attribute a_5 right along, and the quality of classification is elevated from 3/9 to 7/9 in interval (0.5,0.6] and the β-positive region in this interval is not changed compared with entire condition attribute set. Consequently, we get the $\beta_C^{7/9}$ interval reduct $\{a_1, a_2, a_3\}$. AS shown in Table 2, the monotonic property of classification quality is broken in interval (0.5,0.6]. Attribute a_4 and a_5 can be reduced at the same time. However, a_4 will be judged as an irreducible attribute in RSM.

Definition 6 Given a decision table $DT = (U, C \cup D, V, \rho)$ and precision parameter $\beta \in (0.5, 1], \forall a \in C$, if $\exists B \subseteq C - \{a\}$ satisfies $POS(C, D, \beta) = POS(B, D, \beta)$, *then* a is called as a β-reducible attribute. Otherwise, if $\forall B \subseteq C - \{a\}$, it has $POS(C, D, \beta) \neq POS(B, D, \beta)$, a is called as a β-irreducible attribute.

An attribute, irrespective of whether it can be reduced or not is distinct between RSM and VPRSM. In RSM, we can judge an attribute by only one step in the procedures of attribute reduction. There are two important propositions as follows:

(1) $a \in C$ is reducible $\Leftrightarrow POS(C, D) = POS(C - \{a\}, D)$;
(2) $a \in C$ is irreducible $\Leftrightarrow POS(C, D) \neq POS(C - \{a\}, D)$.

In VPRSM, the above two propositions would not be satisfied, and the equivalence propositions just become to implications as follows:

(1) $POS(C, D, \beta) = POS(C - \{a\}, D, \beta) \Rightarrow a \in C$ is reducible;
(2) $a \in C$ is irreducible $\Rightarrow POS(C, D, \beta) \neq POS(C - \{a\}, D, \beta)$.

The dissatisfaction of monotonic properties of some approximate measurements will lead to more complex procedures of attribute reduction in VPRSM. We cannot qualitatively judge an attribute whether it can be reduced or not just by one step in the procedures of attribute reduction in VPRSM. Such as attribute a_4 in Table 2.

4 β-Interval core

The concept of attribute core is an important notion in RSM. However, in VPRSM, the essence of attribute reduction is to preserve the same classification information as the entire attribute set C in the same β interval. It is unable to define attribute core over all β intervals, but we can define attribute core for a special β interval. Here, it is called as β-interval core.

Definition 7 Given a decision table $DT = (U, C \cup D, V, \rho)$, classification quality γ and its associated interval $(\beta_l, \beta_u]$. The β_C^{γ}-interval core, denoted as $core(\beta_C^{\gamma})$, is defined as the intersection over all interval reducts in interval $(\beta_l, \beta_u]$. Namely, $core(\beta_C^{\gamma}) = \cap\, RED$, where RED denotes the β_C^{γ} interval reducts.

The notion of β-interval core corresponds to the notion of attribute core in RSM. Since there may be more than one β interval in a given decision table and different intervals will have their relevant β-interval reducts, the notion of attribute core in VPRSM must be restricted in a special β interval. It can be considered as 'local' attribute core, not the 'global' attribute core over all β intervals.

Theorem 1 *Given a decision table* $DT = (U, C \cup D, V, \rho)$, *classification quality* γ. $\forall a \in C, a \in core(\beta_C^{\gamma})$ *if and only if* $\forall B \subseteq C - \{a\}$ *satisfies one condition as follows:*

(1) $\beta_C^{\gamma} \cap \beta_B^{\gamma} = \emptyset$;
(2) *if* $\beta_C^{\gamma} \cap \beta_B^{\gamma} \neq \emptyset$,

Table 2 The procedures of attribute reduction in $DT1$

Attribute set	Condition classes and associated thresholds of inclusion degree	Qualities of classification
C	$1.0/C_1 = \{o_1\}; 1.0/C_2 = \{o_3\}; 1.0/C_3 = \{o_5\}; 1.0/C_4 = \{o_7\}$;	$\gamma = 7/9, \beta \in (0.5, 0.667]$
	$0.667/C_5 = \{o_2, o_6, o_8\}$;	$\gamma = 4/9, \beta \in (0.667, 1]$
	$0.5/C_6 = \{o_4, o_9\}$.	
$B = C - \{a_4\}$	$1.0/B_1 = \{o_3\}; 1.0/B_2 = \{o_5\}; 1.0/B_3 = \{o_7\}$;	$\gamma = 3/9, \beta \in (0.5, 1]$
	$0.5/B_4 = \{o_1, o_2, o_6, o_8\}; 0.5/B_5 = \{o_4, o_9\}$.	
$B' = C - \{a_4, a_5\}$	$1.0/B'_1 = \{o_3\}; 1.0/B'_2 = \{o_7\}$;	$\gamma = 7/9, \beta \in (0.5, 0.6]$
	$0.6/B'_3 = \{o_1, o_2, o_5, o_6, o_8\}$;	$\gamma = 2/9, \beta \in (0.6, 1]$
	$0.5/B'_4 = \{o_4, o_9\}$.	

then $POS(C,D,\beta_C^\gamma) \neq POS(B,D,\beta_B^\gamma)$.

Proof suppose $\{RED_1, RED_2, \ldots, RED_n\}$ is the collection of all β_C^γ interval reducts.

Sufficiency: suppose $a \notin core(\beta_C^\gamma)$, then $\exists RED_k \in \{RED_1, RED_2, \ldots, RED_n\}$, such that $a \notin RED_k$. Then it has $RED_k \subseteq C - \{a\}$. Since RED_k is a β_C^γ interval reduct, it has $\beta_C^\gamma \cap \beta_{RED_k}^\gamma \neq \emptyset$ and $POS(C,D,\beta_C^\gamma) = POS(RED_k, D, \beta_{RED_k}^\gamma)$. It is a contradiction with conditions (1) and (2). Consequently, $a \in core(\beta_C^\gamma)$.

Necessity: since $a \in core(\beta_C^\gamma)$, it has $a \in RED_1$, $a \in RED_2$, ..., $a \in RED_n$. So $\forall B \subseteq C - \{a\}$, B is not a β_C^γ interval reduct. According to the discussion of definition 5, it can be proved directly. □

For Table 1, $\beta_C^{7/9} = (0.5, 0.667]$ and its relevant β-interval core is $core(\beta_C^{7/9}) = \{a_1\}$. Similarly, $\beta_C^{4/9} = (0.667, 1.0]$ and its relevant interval core is $core(\beta_C^{4/9}) = \{a_1, a_4, a_5\}$.

β-interval core embodies the essential classification information in a special β interval. The classification ability in this interval will be changed no matter whichever attribute in β-interval core is removed. $core(\beta_C^\gamma)$ is the intersection of all β_C^γ interval reducts, namely, if an attribute belongs to β_C^γ interval core, then it must not be reduced during the procedures of β_C^γ attribute reduction. As discussed in Sect. 3, it need to check all the subsets of full attribute set C in order to judge whether an attribute can be reduced or not. So it is complicated to get the whole interval core attributes in a specific β interval.

5 β-Interval characteristic sets

5.1 Ordered discernibility matrix

Discernibility matrices and discernibility functions are often used to describe discrepant attribute information in rough set theory. Skowron and Rauszer (1991) have proved that all reducts are in one-to-one correspondence with the prime implicants of the discernibility function in a given decision table. The threshold of inclusion degree and the decision value of each condition class will be dynamically changed during the procedures of attribute reduction in VPRSM. Each attribute has affection on whether any other attributes can be reduced or not (for instance, a_4 is affected by a_5 as shown in Table 2). So we should consider not only the discrepant attributes between condition classes which have different decision values, but also the discrepant attributes between condition classes which have the same decision value.

Definition 8 Given a decision table $DT = (U, C \cup D, V, \rho)$, its ordered discernibility matrix in VPRSM can be defined as

$$DM[i,j] = \begin{cases} \{a \mid a \in C \wedge \rho(C_i, a) \neq \rho(C_j, a)\} & \\ \qquad j < i \wedge TS(C_i) \geq TS(C_{i+1}) & \\ \emptyset \qquad \text{others} & \end{cases} \tag{4}$$

where $C_i, C_j \in U/C (i,j = 1, 2, \ldots, |U/C|)$. Ordered discernibility matrix is a lower triangle matrix which elements reflect the discernible attributes between condition classes. The rows in ordered discernibility matrix are arranged according to the descending thresholds of inclusion degree of condition classes. The sequence of columns corresponds to the sequence of rows. For Table 1, its ordered discernibility matrix is shown as in Table 3.

5.2 β-Interval characteristic sets

With a given quality of classification γ and its associated interval $\beta_C^\gamma = (\beta_l, \beta_u]$, the ordered discernibility matrix DM can be partitioned into four disjoint blocks as follows:

$$\begin{aligned} &DM1[i,j] : TS(C_i) \geq \beta_u \wedge TS(C_j) \geq \beta_u; \\ &DM2[i,j] : TS(C_i) \leq \beta_l \wedge TS(C_j) \geq \beta_u; \\ &DM3[i,j] : TS(C_i) \leq \beta_l \wedge TS(C_j) \leq \beta_l; \\ &DM4[i,j] : TS(C_i) \geq \beta_u \wedge TS(C_j) \leq \beta_l. \end{aligned} \tag{5}$$

where $1 \leq i,j \leq |U/C|$. $DM1$, $DM2$, $DM3$ and $DM4$ corresponds to the top left corner, bottom left corner, bottom right corner and top right corner, respectively, in an ordered discernibility matrix.

Two separation lines, which partition DM into above four blocks, are called β-horizontal line and β-vertical line respectively. β-horizontal line and β-vertical line appear in

Table 3 The ordered discernibility matrix for Table 1

$TS(C_i)$		$\{o_1\}$	$\{o_3\}$	$\{o_5\}$	$\{o_7\}$	$\{o_2, o_6, o_8\}$	$\{o_4, o_9\}$
1	$\{o_1\}$						
1	$\{o_3\}$	$a_1a_3a_4a_5$					
1	$\{o_5\}$	a_4a_5	$a_1a_3a_4$				
1	$\{o_7\}$	$a_1a_2a_3a_4$	a_2a_5	$a_1a_2a_3a_4a_5$			
0.667	$\{o_2, o_6, o_8\}$	a_4	$a_1a_3a_4a_5$	a_5	$a_1a_2a_3a_4$		
0.5	$\{o_4, o_9\}$	a_1	$a_3a_4a_5$	$a_1a_4a_5$	$a_2a_3a_4$	a_1a_4	

Fig. 3 *β-horizontal line* and *β-vertical line* for $\beta_C^{4/9}$ interval in table 1

decision value	$TS\ (C_i)$		$\{o_1\}$	$\{o_3\}$	$\{o_5\}$	$\{o_7\}$	$\{o_2,o_6,o_8\}$	$\{o_4,o_9\}$
Y	1	$\{o_1\}$						
Y	1	$\{o_3\}$	$a_1a_3a_4a_5$					
N	1	$\{o_5\}$	a_4a_5	$a_1a_3a_4$				
N	1	$\{o_7\}$	$a_1a_2a_3a_4$	a_2a_5	$a_1a_2a_3a_4a_5$			
N	0.667	$\{o_2,o_6,o_8\}$	a_4	$a_1a_3a_4a_5$	a_5	$a_1a_2a_3a_4$		
–	0.5	$\{o_4,o_9\}$	a_1	$a_3a_4a_5$	$a_1a_4a_5$	$a_2a_3a_4$	a_1a_4	

pairs and in the corresponding place in an ordered discernibility matrix. The cross point of these two lines will move from bottom right to top left as increasing the value of precision parameter. For Table 1, $\beta_C^{4/9} = (0.667,1]$ and its β-horizontal line and β-vertical line are shown in Fig. 3.

According to the definition of interval reduct, in order to get β_C^{γ} interval reduct (without lose of generality, suppose $\beta_C^{\gamma} = (\beta_l, \beta_u]$), the β-positive region cannot be changed during the procedures of attribute reduction. Consequently, condition classes which $TS(C_i) \geq \beta_u$ (included in β-positive region) can be merged possibly or condition classes which $TS(C_i) \leq \beta_l$ (not included in β-positive region) can be merged possibly after some attributes reduced. The mergence between two condition classes that one is with $TS(C_i) \geq \beta_u$ and the other is with $TS(C_i) \leq \beta_l$ is not allowed. From the view of separation lines, condition classes which are above the β-horizontal line (or beneath the β-horizontal line) can be merged possibly. But condition classes which are on the different side of β-horizontal line cannot be merged. Similarly, it can be analyzed on the basis of β-vertical line.

All the elements in $DM4$ are equal to empty and all the nonempty elements in $DM3$ can only result in the mergence between condition classes which $TS(C_i) \leq \beta_l$, so we can only consider the nonempty elements in $DM1$ and $DM2$ for attribute reduction.

Theorem 2 *Given a decision table* $DT = (U, C \cup D, V, \rho)$, *classification quality* γ *and its associated interval* $(\beta_l, \beta_u]$. *In this* β *interval, we have*

(1) *if* $c_{ij} \in DM1$ *with a singleton attribute, and the decision values of condition class* C_i *and* C_j *are the same, then this singleton attribute is β-reducible;*

(2) *if* $c_{ij} \in DM2$ *with a singleton attribute, then this singleton attribute is β-irreducible.*

Where $1 \leq i, j \leq |U/C|$.

Proof (1) suppose $c_{ij} = \{a\}$. Since $c_{ij} \in DM1$ and condition class C_i and C_j have the same decision value, then $\exists D_k \in U/D$, such that $C_i \subseteq C_\beta D_k$ and $C_j \subseteq C_\beta D_k (\forall \beta \in (\beta_l, \beta_u])$. After removing attribute a, condition class C_i and C_j will be merged.

Due to $\frac{|C_i \cap D_k|}{|C_i|} \geq \beta_u, \frac{|C_j \cap D_k|}{|C_j|} \geq \beta_u$; it has

$$\begin{aligned} \frac{|(C_i \cup C_j) \cap D_k|}{|C_i \cup C_j|} &= \frac{|(C_i \cap D_k) \cup (C_j \cap D_k)|}{|C_i| + |C_j|} \\ &= \frac{|(C_i \cap D_k)| + |(C_j \cap D_k)|}{|C_i| + |C_j|} \\ &\geq \beta_u. \end{aligned}$$

It indicates that $C_i \cup C_j \subseteq POS(C - \{a\}, D, \beta_{C-\{a\}}^{\gamma})$, namely, $POS(C, D, \beta_C^{\gamma}) = POS(C - \{a\}, D, \beta_{C-\{a\}}^{\gamma})$. Thus, attribute a is β-reducible.

(2) suppose $c_{ij} = \{a'\}$. $\forall \beta \in (\beta_l, \beta_u]$, since $c_{ij} \in DM2$, for condition class C_i, $\forall D_k \in U/D$, it has $C_i \not\subseteq C_\beta D_k$; for condition class C_j, $\exists D_p \in U/D$ such that $C_j \subseteq C_\beta D_p$. Namely, $C_i \not\subseteq POS(C, D, \beta_C^{\gamma})$ and $C_j \subseteq POS(C, D, \beta_C^{\gamma})$.

After removing attribute a', C_i and C_j will be merged. There are two cases:

(I) if $C_i \cup C_j \subseteq POS(C - \{a'\}, D, \beta_{C-\{a'\}}^{\gamma})$, it has $POS(C, D, \beta_C^{\gamma}) \neq POS(C - \{a'\}, D, \beta_{C-\{a'\}}^{\gamma})$. Because $C_i \not\subseteq POS(C, D, \beta_C^{\gamma})$;

(II) if $C_i \cup C_j \not\subseteq POS(C - \{a'\}, D, \beta_{C-\{a'\}}^{\gamma})$, it also has $POS(C, D, \beta_C^{\gamma}) \neq POS(C - \{a'\}, D, \beta_{C-\{a'\}}^{\gamma})$. Because $C_j \subseteq POS(C, D, \beta_C^{\gamma})$.

Thus $POS(C, D, \beta_C^{\gamma}) \neq POS(C - \{a'\}, D, \beta_{C-\{a'\}}^{\gamma})$. In the succedent attribute reduction procedures, the granularity of condition classes will only become coarser. So $C_i \cup C_j$ cannot be fined under any new attribute set $B \subseteq C - \{a'\}$; it always has $POS(C, D, \beta_C^{\gamma}) \neq POS(B, D, \beta_B^{\gamma})$. It means that $\forall B \subseteq C - \{a'\}$, B is not a β_C^{γ} interval reduct. Consequently, attribute a' is β-irreducible. □

If $c_{ij} \in DM1$ with a singleton attribute, and the decision values of condition class C_i and C_j are different, then we cannot simply determine whether this singleton attribute can be reduced. For example, attribute a_4 in the procedures of $\beta_C^{7/9}$ interval attribute reduction in Table 1.

Definition 9 Given a decision table $DT = (U, C \cup D, V, \rho)$, classification quality γ and its associated interval $(\beta_l, \beta_u]$. The ordered discernibility matrix can be blocked with β separation lines caused by interval $(\beta_l, \beta_u]$. Three kinds of interval characteristic sets can be defined as follows:

(1) *if* $c_{ij} \in DM1$ with a singleton attribute, and the decision values of condition class C_i and C_j are the same, then the collection of all these singleton attributes is called as interval reducible set, denoted as $IR(\beta_C^\gamma)$;
(2) *if* $c_{ij} \in DM1$ with a singleton attribute, and the decision values of condition class C_i and C_j are different, then the collection of all these singleton attributes is called as interval 'maybe' reducible set, denoted as $IMR(\beta_C^\gamma)$;
(3) if $c_{ij} \in DM2$ with a singleton attribute, then the collection of all these singleton attributes is called as interval irreducible set, denoted as $IIR(\beta_C^\gamma)$.

Where $1 \le i,j \le |U/C|$.

If an attribute belongs to more than one interval characteristic set, the preference of which characteristic set it will belong to is agreed to $IIR(\beta_C^\gamma) > IMR(\beta_C^\gamma) > IR(\beta_C^\gamma)$. Notation '>' means 'precede'.

Theorem 3 *Given a decision table* $DT = (U, C \cup D, V, \rho)$, *classification quality* γ *and its associated interval* $(\beta_l, \beta_u]$, *it has* $IIR(\beta_C^\gamma) \subseteq core(\beta_C^\gamma) \subseteq IMR(\beta_C^\gamma) \cup IIR(\beta_C^\gamma)$.

Proof According to Theorems 1 and 2, if $\forall a \in C$, such that $a \in IIR(\beta_C^\gamma)$, it indicates that a is β-irreducible, so $a \in core(\beta_C^\gamma)$. $core(\beta_C^\gamma) \subseteq IMR(\beta_C^\gamma) \cup IIR(\beta_C^\gamma)$ can be proved from the discussion in Sect. 3 directly.

Theorem 3 indicates that interval irreducible set is a subset of interval core and can be obtained automatically according to β separation lines. On the other hand, interval core is not equal to interval irreducible set, because interval core may include some attributes in interval 'maybe' reducible set.

Theorem 4 *Given a decision table* $DT = (U, C \cup D, V, \rho)$, *if the quality of classification is* γ (C, D) *in RSM, then it has* $core(\beta_C^{\gamma(C,D)}) = IMR(\beta_C^{\gamma(C,D)}) \cup IIR(\beta_C^{\gamma(C,D)})$.

Proof There is only one β interval in RSM, namely, $\beta = 1$. The thresholds of inclusion degree of condition classes above β-horizontal line are equal to one. In this case, the condition classes, which are included in the same lower approximation of a decision class, can be allowed to merge during the procedures of attribute reduction. Consequently, attributes belonging to *IMR* will be definitely irreducible. So attribute core in RSM can be divided into two parts, $IMR(\beta_C^{\gamma\,(C,D)})$ and $IIR(\beta_C^{\gamma\,(C,D)})$, respectively. □

Theorem 4 indicates that attribute core in RSM is a special case of β interval core in VPRSM. Attribute core is composed of interval irreducible set and interval 'maybe' reducible set in RSM.

6 *IIR* based heuristic algorithm for β-interval attribute reduction

We have discussed that it is not easy to get the whole β-interval core in Sect. 4. However, β-interval irreducible set is a subset of β-interval core and can be obtained easily in an ordered discernibility matrix. So $core(\beta_C^\gamma)$ can be replaced with $IIR(\beta_C^\gamma)$ as the starting points of heuristic algorithms for attribute reduction in VPRSM. Given a classification quality γ and its associated interval $(\beta_l, \beta_u]$, $IIR(\beta_C^\gamma)$ can be obtained dynamically in different β intervals only by adjusting the β separation lines.

Algorithm 1 Heuristic algorithm based on *IIR* for β-interval attribute reduction

Input: decision table $DT = (U, C \cup D, V, \rho)$;
Output: interval reduct *RED* which satisfies a given classification quality, final interval β_{RED} for *RED*.
Step1: establish the relationship between classification qualities and β intervals;
Step2: select a classification quality γ and its associated interval $(\beta_l, \beta_u]$;
Step3: construct ordered discernibility matrix *DM* and partition it into four blocks based on interval $(\beta_l, \beta_u]$. Set $RED = \emptyset$ and $IIR = \emptyset$;
Step4: scan *DM*2 and add its elements with singleton attribute to *IIR*, then set $RED = RED \cup IIR$;
Step5: $\forall c_{ij} \in DM$, if $c_{ij} \cap RED \neq \emptyset$, then set $c_{ij} = \emptyset$. If all elements in *DM* are equal to empty or $(\beta_l, \beta_u] \cap \beta_{RED}^\gamma \neq \emptyset$ and $POS(C, D, \beta_C^\gamma) = POS(RED, D, \beta_{RED}^\gamma)$, then go to step 7;
Step6: choose an attribute in *DM* which has the highest significance, set $RED = RED \cup \{a\}$ and go to step 5. If there are more than one attribute with the highest significance at the same time, then choose one from them randomly and add it to *RED*;
Step7: return *RED* and its interval $\beta_{RED} = (\beta_l, \beta_u] \cap \beta_{RED}^\gamma$.

Two kinds of attribute significance definition are proposed in paper (Wang et al. 2006). Only one uses the occurrence frequencies of attributes in the discernibility matrix. The improved definition combines with the occurrence frequencies of attributes and the length of elements in the discernibility matrix, and the elements which have the shortest length will be considered preferentially. We utilize the latter in Algorithm 1.

For Table 1, $\beta_C^{4/9} = (0.667,1]$. It can be found from Fig. 3 that $IIR(\beta_C^{4/9}) = \{a_1, a_4, a_5\}$. After removing the elements which include a_1, a_4 or a_5 in *DM*, all the elements in *DM* are equal to empty; thus $\{a_1, a_4, a_5\}$ is a $\beta_C^{4/9}$ interval reduct, and its associated interval is $\beta_C^{4/9} \cap \beta_{\{a_1,a_4,a_5\}}^{4/9} = (0.667, 1] \cap (0.667, 1] = (0.667, 1]$.

In order to get $\beta_C^{7/9} = (0.5,0.667]$ interval reducts in Table 1, the β-separation lines can be adjusted automatically, viz., the cross point will be moved to the bottom right as shown in Fig. 4.

Since $IIR(\beta_C^{7/9}) = \{a_1\}$ and $\beta_{\{a_1\}}^{7/9} = \emptyset$, after removing attribute a_1, the ordered discernibility matrix is shown in Fig. 5.

It needs to compute the significance of attribute a_2, a_3, a_4 and a_5. Based on the length of elements in Fig. 5, we have

$$L_1 = \{a_4, a_5\},$$
$$L_2 = \{a_2a_5, a_4a_5\},$$
$$L_3 = \{a_2a_3a_4, a_3a_4a_5\}.$$

where $L_k(k = 1, 2, 3)$ means the collection of elements whose lengths are equal to k. $SIG(a)/L_k$ means the occurrence frequency of attribute a in L_k. If more than one attribute has the highest frequency in L_k, then their frequencies in L_{k+1} need to be compared.

Because $SIG(a_4)/L_1 = SIG(a_5)/L_1 = 1$, $SIG(a_5)/L_2 = 2 > SIG(a_4)/L_2 = 1$, so attribute a_5 is selected preferentially. However, $\beta_C^{7/9} \cap \beta_{\{a_1,a_5\}}^{7/9} = \emptyset$, $\{a_1, a_5\}$ is not a $\beta_C^{7/9}$ interval reduct. Subsequently, after removing attribute a_5 in Fig. 5 and computing attribute significance again, attribute a_4 is selected. At this time, $\beta_C^{7/9} \cap \beta_{\{a_1,a_4,a_5\}}^{7/9} \neq \emptyset$ and $POS(C, D, \beta_C^{7/9}) = POS(\{a_1, a_4, a_5\}, D, \beta_{\{a_1,a_4,a_5\}}^{7/9})$, so $\{a_1, a_4, a_5\}$ is a $\beta_C^{7/9}$ interval reduct, its associated interval is $\beta_C^{7/9} \cap \beta_{\{a_1,a_4,a_5\}}^{7/9} = (0.5, 0.667] \cap (0.5, 0.667] = (0.5, 0.667]$. It can be found that attribute set $\{a_1, a_4, a_5\}$ preserves the classification information in both intervals $\beta_C^{4/9}$ and $\beta_C^{7/9}$ as the entire attribute set C.

7 Determination of β interval based on shadowed sets

Different classification qualities will correspond to different precision parameter intervals and there are often more than one β-interval reduct in each interval. Up to now, it does not have a method to determine optimal intervals objectively. Though we can choose a β interval according to the given classification quality, it is still a problem as to which classification quality is appropriate.

Shadowed sets (Pedrycz 1998) are one among several key contributors to the area of granular computing. They could be considered as new and stand-alone constructs, yet they often induced by some fuzzy sets. They are simpler and more practical than fuzzy sets and can be sought as a symbolic representation of numeric fuzzy sets.

Shadowed sets utilize three logical vales {0, 1, (0, 1)} to simplify the relevant fuzzy sets. Obviously, they not only simplify the interpretation but also avoid a number of computations of numeric membership grades compared with the methodology of fuzzy sets. Conceptually, shadowed sets are close to rough sets even though their mathematical foundations are very different. The notions of negative region, positive region and boundary region in rough set theory are related to three logical vales in shadowed sets, namely, excluded, included and uncertain(shadow), respectively. In this sense, shadowed sets can be considered as the bridge between fuzzy and rough sets (Pedrycz 1999).

The principle of constructing shadowed sets is based on balancing the uncertainty that is inherently associated with fuzzy sets; in other words, the total uncertainty will be relocated and concentrated. As elevating membership values (high enough) of some regions of universe to 1 and at the same time, reducing membership values (low enough) of some regions of universe to 0, the uncertainty in these

Fig. 4 *β-horizontal line* and *β-vertical line* for $\beta_C^{7/9}$ interval

decision value	$TS\ (C_i)$		$\{o_1\}$	$\{o_3\}$	$\{o_5\}$	$\{o_7\}$	$\{o_2,o_6,o_8\}$	$\{o_4,o_9\}$
Y	1	$\{o_1\}$						
Y	1	$\{o_3\}$	$a_1a_3a_4a_5$					
N	1	$\{o_5\}$	a_4a_5	$a_1a_3a_4$				
N	1	$\{o_7\}$	$a_1a_2a_3a_4$	a_2a_5	$a_1a_2a_3a_4a_5$			
N	0.667	$\{o_2,o_6,o_8\}$	a_4	$a_1a_3a_4a_5$	a_5	$a_1a_2a_3a_4$		
—	0.5	$\{o_4,o_9\}$	a_1	$a_3a_4a_5$	$a_1a_4a_5$	$a_2a_3a_4$	a_1a_4	

Fig. 5 The procedure of $\beta_C^{7/9}$ interval attribute reduction

decision value	$TS\ (C_i)$		$\{o_1\}$	$\{o_3\}$	$\{o_5\}$	$\{o_7\}$	$\{o_2,o_6,o_8\}$	$\{o_4,o_9\}$
Y	1	$\{o_1\}$						
Y	1	$\{o_3\}$	Φ					
N	1	$\{o_5\}$	a_4a_5	Φ				
N	1	$\{o_7\}$	Φ	a_2a_5	Φ			
N	0.667	$\{o_2,o_6,o_8\}$	a_4	Φ	a_5	Φ		
—	0.5	$\{o_4,o_9\}$	Φ	$a_3a_4a_5$	Φ	$a_2a_3a_4$	Φ	

regions can be eliminated. In order to balance the total uncertainty, it needs to compensate the eliminated parts by allowing for the emergence of uncertainty regions, namely, it results in shadowed region.

The optimal separate threshold in shadowed sets can be determined by the principle of balance of uncertainty. The detailed descriptions for this determination optimization task can be found in Pedrycz (2002, 2005, 2009). Due to the reverse relationship between precision parameter and the quality of classification, in the following, we will introduce a novel method to choose optimal β interval based on shadowed sets.

Suppose a decision table $DT = (U, C \cup D, V, \rho)$, $U/C = \{C_1, C_2, \ldots, C_{|U/C|}\}$. As discussed in Sect. 2, each condition class $C_i \in U/C (i = 1, 2, \ldots, |U/C|)$ has a threshold of inclusion degree $TS(C_i)$ over all decision classes. $TS(C_i)$ can be considered as the membership grade that C_i could be included in positive region. Since $\forall C_i \in U/C$, it has $0 < TS(C_i) \le 1$, a fuzzy set defined on U/C with respect to positive region is given as follows:

$$\tilde{F}(U/C) = \left(\frac{\tilde{F}(C_1)}{C_1}, \frac{\tilde{F}(C_2)}{C_2}, \ldots, \frac{\tilde{F}(C_{|U/C|})}{C_{|U/C|}}\right) = \left(\frac{TS(C_1)}{C_1}, \frac{TS(C_2)}{C_2}, \ldots, \frac{TS(C_{|U/C|})}{C_{|U/C|}}\right) \tag{6}$$

Based on shadowed sets, in order to win the total balance of uncertainty, a certain optimization process can be established as follows:

$$V(\beta_{opt}) = \min_{\beta \in (0.5, 1]} V(\beta) \tag{7}$$

$$V(\beta) = |\Re_1 + \Re_2 - \Re_3|. \tag{8}$$

where

$$\Re_1 = \sum_{i:\tilde{F}(C_i) \le 1-\beta} P(C_i)\tilde{F}(C_i)$$

$$\Re_2 = \sum_{i:\tilde{F}(C_i) \ge \beta} P(C_i)\left(1 - \tilde{F}(C_i)\right)$$

$$\Re_3 = \sum_{1-\beta < \tilde{F}(C_i) < \beta} P(C_i)$$

$P(C_i) = |C_i|/|U|$ is the estimation of probability distribution for each condition class in university U. $\Re_1$ means the elimination of uncertainty by reducing the membership values, viz. low-membership grades are reduced to 0 (excluded). $\Re_2$ means the elimination of uncertainty by elevating the membership values, viz. high-membership grades are elevated to 1 (included). $\Re_3$ means the shadow.

By this optimization process, we can determine the optimal β interval which preserves the balance of total uncertainty for a given decision table. Then the positive region will be extended and at the same time, the boundary region will be concentrated. More rules, deterministic rules or probabilistic rules (Herbert and Yao 2009; Yao 2003), can be acquired from the new positive region with a reasonable certainty degree.

For Table 1, it has

$$\tilde{F}(U/C) = \left(\frac{\tilde{F}(C_1)}{C_1}, \frac{\tilde{F}(C_2)}{C_2}, \frac{\tilde{F}(C_3)}{C_3}, \frac{\tilde{F}(C_4)}{C_4}, \frac{\tilde{F}(C_5)}{C_5}, \frac{\tilde{F}(C_6)}{C_6}\right) = \left(\frac{1}{C_1}, \frac{1}{C_2}, \frac{1}{C_3}, \frac{1}{C_4}, \frac{0.667}{C_5}, \frac{0.5}{C_6}\right).$$

According to formula (8), it has

$$V(\beta) = \begin{cases} 5/9 & 0.667 < \beta \le 1 \\ 1/9 & 0.5 < \beta \le 0.667 \end{cases}$$

So the optimal β interval is (0.5, 0.667]. In this case, condition class C_5 will be included in the positive region and relevant probabilistic rules can be gotten.

8 Evaluation for β-interval reducts

Generally, there are more than one interval reduct in a specific β interval; it is a problem how to evaluate the interval reducts in a special β interval. In VPRSM, Ziarko (1993) gives two approaches to select reduts: the reduct with the minimal number of attributes or the reduct which has the least number of combinations of values of attributes. The first criterion focuses on the reduct that most redundant information can be removed from original decision table. The second criterion focuses on the reduct which will represent the strongest patterns or data regularities. These two criteria are also often used in RSM. Nevertheless, both of them evaluate the reduct under a special β value.

A β-interval reduct in VPRSM can be depicted with a binary tuple <*reduct*, *reduct interval*>. It is not reasonable to evaluate a β-interval reduct only by means of attributes. With β interval properties, an evaluation function for β-interval reducts can be defined as follows:

$$g(RED) = \frac{1}{|RED|}\frac{|\beta^{\gamma}_{RED} \cap \beta^{\gamma}_{C}|}{|\beta^{\gamma}_{C}|} \tag{9}$$

where $|RED|$ denotes the total number of attributes in RED; $|\beta^{\gamma}_{C}|$ presents the length of interval in which classification quality is equal to γ under entire condition attribute set C; $|\beta^{\gamma}_{RED}|$ presents the length of interval in which classification quality is equal to γ under reduct RED.

Since $0 < \frac{1}{|RED|} \le 1$ and $0 < \frac{|\beta^{\gamma}_{RED} \cap \beta^{\gamma}_{C}|}{|\beta^{\gamma}_{C}|} \le 1$, so for each interval reduct RED, it has $0 < g(RED) \le 1$. $\beta^{\gamma}_{RED} \cap \beta^{\gamma}_{C}$ reflects the common β region between reduct interval and the interval of entire attribute set C under the given classification quality γ. $\frac{|\beta^{\gamma}_{RED} \cap \beta^{\gamma}_{C}|}{|\beta^{\gamma}_{C}|} = 1$ indicates that reduct RED preserves the whole β interval as the same as entire attribute set C. The smaller the value $|RED|$ and the higher the ratio $\frac{|\beta^{\gamma}_{RED} \cap \beta^{\gamma}_{C}|}{|\beta^{\gamma}_{C}|}$ are, then the higher the value of evaluation function $g(RED)$ will be. When $g(RED) = 1$, namely, $|RED| = 1$ and $\beta^{\gamma}_{RED} \cap \beta^{\gamma}_{C} = \beta^{\gamma}_{C}$, it indicates that a singleton attribute can completely preserve the information as in original β interval. In this case, the data set will be furthest simplified without losing classification information. Practically, we can choose the β-interval reduct with the highest value of $g(RED)$ as the optimal β-interval reduct.

According to Ziarko's criteria, the part $|RED|$ can be replaced with the number of combinations of values of attributes in RED. In addition, we can also assign weights to $|RED|$ and its associated interval, respectively, and then use weighted approaches to evaluate and select β-interval reducts.

For Table 1, we get three $\beta^{7/9}_{C} = (0.5, 0.667]$ interval reducts; the relevant parameters are shown in Table 4.

From the Table 4, $\beta^{7/9}_{C}$ interval reducts $\{a_1, a_4, a_5\}$ and $\{a_1, a_2, a_4\}$ have the highest value of evaluation function; both of them can be chosen as the optimal $\beta^{7/9}_{C}$ interval reduct. In addition, it can be found that $\{a_1, a_4, a_5\} \cap \{a_1, a_2, a_3\} \cap \{a_1, a_2, a_4\} = \{a_1\}$ which is equal to *core*$(\beta^{7/9}_{C})$ defined in Sect. 4.

9 Experiment and analysis

In the following, we will use the discrete data set WINE, as shown in (Beynon 2001), to illustrate the proposed notions in this paper. There are eight condition attributes and one decision attribute. The data set WINE can be partitioned into a collection of 12 condition classes with respect to $C = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8\}$ and 3 decision classes with respect to $D = \{d_1\}$. Therein, the thresholds of inclusion degree of 9 condition classes are equal to one and the others are equal to 0.87, 0.667 and 0.625, respectively. The relationship between classification qualities and β intervals is shown in Fig. 6. The associated ordered discernibility matrix is constructed as shown in Table 5.

Table 4 $\beta^{7/9}_{C}$ interval reducts in Table 1

RED	β^{γ}_{RED}	$\beta^{\gamma}_{RED} \cap \beta^{\gamma}_{C}$	$g(RED)$
$\{a_1, a_4, a_5\}$	(0.5,0.667]	(0.5,0.667]	0.3333
$\{a_1, a_2, a_3\}$	(0.5,0.6]	(0.5,0.6]	0.1996
$\{a_1, a_2, a_4\}$	(0.5,0.75]	(0.5,0.667]	0.3333

According to dynamic β-separation lines and proposed heuristic algorithm, we can get irreducible set and β-interval reduct in each β interval which are shown in Table 6.

An interesting thing is that the results of heuristic attribute reduction in each interval are unrestricted reduct which is defined in Beynon (2001), namely, the reduct interval should include original β interval, despite that this restriction is not necessary in the β-interval reduct definition. It indicates that the result in each interval, which is obtained by the proposed algorithm, can availably preserve the information as the entire condition attribute set C.

As discussed in Sect. 4, it is not easy to get the whole attribute core in a given β interval. However, β-interval irreducible set is beneficial for heuristic interval attribute reduction instead of β-interval core and then search space can be reduced effectively. Such as $RED = IIR$ in intervals (0.667,0.87] and (0.87,1]. A β-interval reduct in each of these two intervals can be obtained just by β-interval irreducible set.

Furthermore, we can establish $\tilde{F}(U/C)$ as follows:

$$\tilde{F}(U/C) = \left(\frac{1}{C_1}, \frac{1}{C_2}, \frac{1}{C_3}, \frac{1}{C_4}, \frac{1}{C_5}, \frac{1}{C_6}, \frac{1}{C_7}, \frac{1}{C_8}, \frac{1}{C_9}, \frac{0.87}{C_{10}}, \frac{0.667}{C_{11}}, \frac{0.625}{C_{12}}\right).$$

Then it has

$$V(\beta) = \begin{cases} 37/60 & 0.87 < \beta \leq 1 \\ 11/60 & 0.667 < \beta \leq 0.87 \\ 1/20 & 0.625 < \beta \leq 0.667 \\ 2/15 & 0.5 < \beta \leq 0.625 \end{cases}$$

So β interval (0.625,0.667] is chosen as the optimal β interval. In this case, $POS(C, D, \beta) = U - C_{12}$, namely, positive region is extended which will include condition class C_{10} and C_{11}.

For interval (0.625,0.667], the classification quality is equal to 52/60; all $\beta^{52/60}_{C}$ interval reducts and their relevant parameters are given in Table 7.

From Table 7, it can be found that $\beta^{52/60}_{C}$ interval reducts $\{c_1, c_4, c_5, c_8\}$ and $\{c_3, c_4, c_5, c_8\}$ have the same value of evaluation function, so both of them can be chosen as the optimal β-interval reduct in interval (0.625,0.667] for WINE data set.

10 Summary and conclusion

Traditional attribute reduction model based on Ziarko's definition may produce some anomalies in the procedures of attribute reduction. According to the interval properties of precision parameter, we extend reduct definition from a specific β value to a specific β interval in VPRSM.

Table 5 The ordered discernibility matrix of WINE data set

Decision	$TS(C_i)$		C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}
2	1	C_1												
2	1	C_2	$c_3c_4c_5$											
2	1	C_3	c_3c_4	c_5										
2	1	C_4	c_3c_5	c_4	c_4c_5									
2	1	C_5	c_1c_4	$c_1c_3c_5$	c_1c_3	$c_1c_3c_4c_5$								
3	1	C_6	$c_1c_3c_4c_5$	c_1	c_1c_5	c_1c_4	c_3c_5							
3	1	C_7	$c_1c_3c_4c_5c_8$	c_1c_8	$c_1c_5c_8$	$c_1c_4c_8$	$c_3c_5c_8$	c_8						
3	1	C_8	$c_1c_3c_5$	c_1c_4	$c_1c_4c_5$	c_1	$c_3c_4c_5$	c_4	c_4c_8					
3	1	C_9	$c_1c_3c_5c_8$	$c_1c_4c_8$	$c_1c_4c_5c_8$	c_1c_8	$c_3c_4c_5c_8$	c_4c_8	c_4	c_8				
1	0.87	C_{10}	$c_1c_3c_8$	$c_1c_4c_5c_8$	$c_1c_4c_8$	$c_1c_5c_8$	$c_3c_4c_8$	$c_4c_5c_8$	c_4c_5	c_5c_8	c_5			
2	0.667	C_{11}	$c_1c_3c_4$	c_1c_5	c_1	$c_1c_4c_5$	c_3	c_5	c_5c_8	c_4c_5	$c_4c_5c_8$	c_4c_8		
3	0.625	C_{12}	c_1c_3	$c_1c_4c_5$	c_1c_4	c_1c_5	c_3c_4	c_4c_5	$c_4c_5c_8$	c_5	c_5c_8	c_8	c_4	

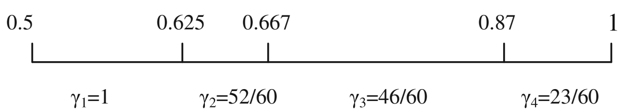

Fig. 6 The relationship between classification qualities and β intervals under attribute set C in WINE

Table 6 The result of interval reduction in WINE

β interval	(0.5,0.625]	(0.625,0.667]	(0.667,0.87]	(0.87,1]
IIR	$\emptyset$	$\{c_4, c_5, c_8\}$	$\{c_1, c_3, c_5, c_8\}$	$\{c_1, c_3, c_5\}$
RED	$\{c_4, c_5\}$	$\{c_1, c_4, c_5, c_8\}$	$\{c_1, c_3, c_5, c_8\}$	$\{c_1, c_3, c_5\}$
β^{γ}_{RED}	(0.5,0.625]	(0.625,0.714]	(0.5,0.87]	(0.541,1]
$\beta^{\gamma}_{C} \cap \beta^{\gamma}_{RED}$	(0.5,0.625]	(0.625,0.667]	(0.667,0.87]	(0.87,1]

Table 7 $\beta_C^{52/60}$ interval reducts and their relevant parameters in WINE

RED	β^{γ}_{RED}	$\beta^{\gamma}_{RED} \cap \beta^{\gamma}_{C}$	$g(RED)$
$\{c_1, c_4, c_5, c_8\}$	(0.625,0.714]	(0.625,0.667]	0.25
$\{c_3, c_4, c_5, c_8\}$	(0.625,0.8]	(0.625,0.667]	0.25

Furthermore, the concept of β interval core is proposed which will enrich the methodology of VPRSM. Three kinds of interval characteristic sets can be easily and automatically gotten by β separation lines. With these interval characteristic sets, a heuristic algorithm for interval attribute reduction can be constructed.

The value of precision parameter β is critical in VPRSM. A novel method, which can determine this parameter objectively, is introduced on the basis of shadowed sets. It does not need any additional information except data set itself. In addition, an evaluation function is also given for selecting final optimal β-interval reduct. All the notions in this paper will promote the development of VPRSM both in theory and practice. Different β-interval reducts will result in different rule sets; how to integrate these rule sets will be the aim of our future work.

Acknowledgments The authors wish to grateful acknowledge anonymous reviews for their valuable comments and recommendations. This work was supported by National Natural Science Foundation of China (Serial No. 60475019, 60775036, 60970061) and the Ph.D. programs Foundation of Ministry of Education of China (Serial No. 20060247039).

Appendix

See Table 8.

Table 8 WINE data set

WINE index	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	d_1
2	1	0	1	1	1	0	0	1	1
5	1	0	1	1	1	0	0	1	1
16	1	0	1	1	1	0	0	1	1
17	1	0	1	1	1	0	0	1	1
20	1	0	1	1	1	0	0	1	1
22	1	0	1	1	1	0	0	1	1
24	1	0	1	1	1	0	0	1	1
28	1	0	1	1	1	0	0	1	1
30	1	0	1	1	1	0	0	1	1
34	1	0	1	1	1	0	0	1	1
38	1	0	1	1	1	0	0	1	1
40	1	0	1	1	1	0	0	1	1
43	1	0	1	1	1	0	0	1	1
47	1	0	1	1	1	0	0	1	1

Table 8 continued

WINE index	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	d_1
50	1	0	1	1	1	0	0	1	1
51	1	0	1	1	1	0	0	1	1
53	1	0	1	1	1	0	0	1	1
54	1	0	1	1	1	0	0	1	1
57	1	0	1	1	1	0	0	1	1
58	1	0	1	1	1	0	0	1	1
63	1	0	1	1	1	0	0	0	2
64	1	0	1	0	1	0	0	0	2
66	1	0	1	1	1	0	0	0	2
69	1	0	1	1	1	0	0	1	2
70	1	0	1	1	1	0	0	1	2
74	1	0	1	1	1	0	0	1	2
77	1	0	0	0	1	0	0	0	2
78	0	0	1	1	0	0	0	0	2
87	0	0	1	0	0	0	0	0	2
88	0	0	1	0	1	0	0	0	2
89	0	0	1	0	1	0	0	0	2
90	0	0	1	0	1	0	0	0	2
94	1	0	1	0	1	0	0	0	2
98	1	0	1	0	1	0	0	0	2
101	0	0	0	1	1	0	0	0	2
104	0	0	1	0	1	0	0	0	2
105	1	0	1	0	1	0	0	0	2
123	1	0	1	1	1	0	0	0	2
125	0	0	1	0	1	0	0	0	2
131	1	0	1	1	0	0	0	0	3
132	1	0	1	1	0	0	0	0	3
137	1	0	1	0	0	0	0	1	3
139	1	0	1	0	0	0	0	0	3
143	1	0	1	1	0	0	0	0	3
144	1	0	1	1	1	0	0	0	3
146	1	0	1	1	0	0	0	1	3
148	1	0	1	0	0	0	0	0	3
149	1	0	1	1	1	0	0	0	3
150	1	0	1	1	0	0	0	0	3
151	1	0	1	1	0	0	0	0	3
152	1	0	1	1	0	0	0	0	3
153	1	0	1	1	1	0	0	0	3
154	1	0	1	1	0	0	0	0	3
160	1	0	1	0	1	0	0	0	3
161	1	0	1	0	1	0	0	0	3
163	1	0	1	1	0	0	0	0	3
165	1	0	1	0	0	0	0	0	3
170	1	0	1	1	1	0	0	0	3
172	1	0	1	0	0	0	0	0	3
178	1	0	1	1	1	0	0	0	3

References

An A, Shan N, Chan C (1996) Discovering rules for water demand prediction: an enhanced rough set approach. Eng Appl Artif Intell 9:645–653

Anantaram C, Nagaraja G, Nori KV (1998) Verification of accuracy of rules in a rule based system. Data Knowl Eng 27:115–138

Beynon M (2000) An investigating of β-reduct selection within the variable precision rough sets model. In: The second international conference on rough sets and current trend in computing (RSCTC 2000). LNAI 2005:114–122

Beynon M (2001) Reducts within the variable precision rough sets model: a further investigation. Eur J Oper Res 134:595–597

Beynon M (2003) The introduction and utilization of (l-u)-graphs in the extended variable precision rough sets model. Int J Intell Syst 18:1035–1055

Beynon M, Peel MJ (2001) Variable precision rough set theory and data discretization:an application to corporate failure prediction. Int J Manag Sci 29:561–576

Dimitras AI et al (1999) Business failure using rough sets. Eur J Oper Res 114:263–280

Dubois D, Prade H (1999) Rough fuzzy sets and fuzzy rough sets. Int J Gen Syst 17:191–209

Greco S, Matarazzo B, Slowinski R (2008) Parameterized rough set model using rough membership and Bayesian confirmation measures. Int J Approx Reason 49:285–300

Herbert JP, Yao JT (2009) Criteria for choosing a rough set model. Comput Math Appl 57:908–918

Jelonek J, Krawiec K, Slowinski R (1995) Rough set reduction of attributes and their domains for neural networks. Int J Comput Intell 11:339–347

Katzberg JK, Ziarko W (1993) Variable precision rough sets with asymmetric bounds. In: Ziarko W, Rijsbergen CJ (eds) Rough sets, fuzzy sets and knowledge discovery. Springer, New York, pp 167–177

Mushrif MM, Ray AK (2008) Color image segmentation: rough-set theoretic approach. Pattern Recogn Lett 29:483–493

Pawlak Z (1982) Rough sets. Int J Comput Inf Sci 11:314–356

Pawlak Z, Skowron A (2007) Rudiments of rough sets. Inf Sci 177:3–27

Pawlak Z, Wong SKM, Ziarko W (1988) Rough sets: probabilistic versus deterministic approach. Int J Man Mach Stud 29:81–95

Pedrycz W (1998) Shadowed sets: representing and processing fuzzy sets. IEEE T Syst Man Cybern B 28:103–109

Pedrycz W (1999) Shadowed sets: bridging fuzzy and rough sets. In: Pal SK, Skowron A (eds) Rough fuzzy hybridization a new trend in decision-making. Springer, Singapore, pp 179–199

Pedrycz W (2005) Interpretation of clusters in the framework of shadowed sets. Pattern Recogn Lett 26:2439–2449

Pedrycz W (2009) From fuzzy sets to shadowed sets: interpretation and computing. Int J Intell Syst 24:48–61

Pedrycz W, Bukovich G (2002) Granular computing with shadowed sets. Int J Intell Syst 17:173–197

Shen Q, Chouchoulas A (2002) A rough-fuzzy approach for generating classification rules. Pattern Recogn Lett 35:2425–2438

Skowron A, Rauszer C (1991) The discernibility matrices and functions in information systems. In: Slowinski R (eds) Intelligent decision support: handbook of applications and advances of the rough sets theory. Kluwer, Dordrecht, pp 331–362

Slezak D (2003) Attribute reduction in the Bayesian version of variable precision rough set model. Electron Notes Theor Comput Sci 82:1–11

Slezak D, Ziarko W(2002) Bayesian rough set model. In: Proceedings of the international workshop on foundation of data mining (FDM'2002). Maebashi, Japan, pp 131–135

Slezak D, Ziarko W (2005) The investigation of the Bayesian rough set model. Int J Approx Reason 40:81–91

Su C-T, Hsu J-H (2006) Precision parameter in the variable precision rough sets model: an application. Int J Manag Sci 34:149–157

Thangavel K, Pethalakshmi A (2009) Dimensionality reduction based on rough set theory: a review. Appl Soft Comput 9:1–12

Wang J, Miao DQ (1998) Analysis on attribute reduction strategies of rough set. J Comput Sci Technol 13:189–193

Wang JY, Zhou J (2009) Research of reduct features in the variable precision rough set model. Neurocomputing 72:2643–2648

Wang RZ, Miao DQ, Hu GR et al (2006) Discernibility matrix based algorithm for reduction of attributes. In: Butz CJ (eds) Web intelligence and intelligent agent technology. IEEE Computer Society, Hong Kong, pp 477–480

Wojciech K et al (2008) Stochastic dominance-based rough set model for ordinal classification. Inf Sci 178:4019–4037

Yao YY (2003) Probabilistic approaches to rough sets. Expert Syst 20:287–297

Yao Y Y (2007) Decision-theoretic rough set modes. In: Proceedings of RSKT'07. LNAI 4481:1–12

Yao YY (2008) Probabilistic rough set approximations. Int J Approx Reason 49:255–271

Yao YY, Wong SKM (1992) A decision theoretic framework for approximating concepts. Int J Man Mach Stud 37:793–809

Yao YY, Zhao Y (2008) Attribute reduction in decision-theoretic rough set model. Inf Sci 178:3356–3373

Yao YY, Wong SKM, Lingras P (1990) A decision-theoretic rough set model. In: Ras ZW et al (eds) Methodologies for intelligent systems. North-Holland, New York, pp 17–24

Yao Y Y, Zhao Y, Wang J (2006) On reduct construction algorithms. In: Proceedings of RSKT'06. LNAI 4062:297–304

Zhu W, Wang FY (2003) Reduction and axiomization of covering generalized rough sets. Inf Sci 152:217–230

Ziarko W (1993) Variable precision rough set model. J Comput Syst Sci 46:44–54

Ziarko W (1999) Decision making with probabilistic decision tables. In: Proceedings of RSFDGrC'99. LNAI 1711:463–471