Available online at www.sciencedirect.com
SCIENCE DIRECT®
Omega 34 (2006) 149–157
omega
The International Journal of Management Science
www.elsevier.com/locate/omega

Precision parameter in the variable precision rough sets model: an application

Chao-Ton Su[a,*], Jyh-Hwa Hsu[b]

[a]*Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu, Taiwan*
[b]*Department of Industrial Engineering and Management, National Chiao Tung University, 1001 Ta Hsueh Road, Hsinchu, Taiwan*

Received 5 April 2003; accepted 30 August 2004
Available online 7 October 2004

Abstract

Despite their diverse applications in many domains, the variable precision rough sets (VPRS) model lacks a feasible method to determine a precision parameter (β) value to control the choice of β-reducts. In this study we propose an effective method to find the β-reducts. First, we calculate a precision parameter value to find the subsets of information system that are based on the least upper bound of the data misclassification error. Next, we measure the quality of classification and remove redundant attributes from each subset. We use a simple example to explain this method and even a real-world example is analyzed. Comparing the implementation results from the proposed method with the neural network approach, our proposed method demonstrates a better performance.

Keywords: VPRS model; β-reduct; Precision parameter; Neural networks

1. Introduction

Knowledge bases have been successfully applied in many real-world applications, where intelligent decisions have to be made. Knowledge bases can usually be represented as a set of decision rules that generally follow the form of: "if..., then...". These rules can be extracted from human experts or collected data. Most of the time the collected data is so huge that it is beyond the ability of a human expert to analyze it without using feasible analysis techniques. The analysis and extraction of useful information from collected data has been the subject of active research in data mining [1].

Rough sets as a mathematical methodology for data analysis were introduced by Pawlak [2]. They provide a powerful tool for data analysis and knowledge discovery from imprecise and ambiguous data. The theory of rough sets has been successfully applied to diverse areas, such as knowledge acquisition, forecasting and predictive modeling, knowledge base systems, and data mining [3,4].

The rough sets methodology is based on the premise that lowering the degree of precision in the data makes the data pattern more visible. The central premise of the rough sets philosophy is that knowledge exists in the ability to classify. In other words, the rough sets approach can be considered as a formal framework for discovering patterns from imperfect data. The results of the rough sets approach are presented in the form of classification or decision rules derived from given data sets.

The rough sets approach is inspired by the notion of inadequacy of the available information to perform a complete

* Corresponding author. Fax: +886 3 5722932.
E-mail address: ctsu@mx.nthu.edu.tw (C.-T. Su).

0305-0483/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.omega.2004.08.005

classification of objects; that is, performing a complete classification requires that the collected data must be fully correct or certain. The classification with a controlled degree of uncertainty, or a misclassification error, is outside the realm of this approach [5].

The variable precision rough sets (VPRS) model was introduced by Ziarko [5] and is an extension of the original Rough Set Theory (RST) as a tool for classification of objects. This is an important extension, since as noted by Kattan et al. [6], 'In real-world decision making, the patterns of classes often overlap, suggesting that predictor information may be incomplete... This lack of information results in probabilistic decision making, where perfect prediction accuracy is not expected'.

VPRS deals with partial classification by introducing a precision parameter β (in the rough set the β value is zero). The β value represents a bound on the conditional probability of a proportion of objects in a condition class, which are classified to the same decision class. Ziarko [5] considered β as a classification error and it is defined to be in the domain [0.0, 0.5).

Because the VPRS model has no formal historical background of having empirical evidence to support any particular method of β-reduct selection [11], VPRS-related research studies do not focus in detail on the choice of the precision parameter (β) value. Ziarko [5] proposed the β value to be specified by the decision maker. Beynon [8] proposed two methods of selecting a β-reduct without such a known β value. Beynon [9] proposed the allowable β value range to be an interval, where the quality of classification may be known prior to determining the β value range.

The extended VPRS was introduced by Katzberg et al. [10], which allowed asymmetric bounds l and u to be used. The VPRS models the restriction $l < 0.5$ and $u = 1 - l$ must hold. Beynon [7] introduced the (l, u)-quality graph, which elucidates the associated level of quality of classification, based on the selected l and u values. The results in this paper, within a criteria for the effective choice of l and u values is still required. However, without a β value to control the choice of β-reducts, this may lead to the full set of β-reducts becoming too large, such that an addition to a search scope is needed to find a suitable β-reduct.

In this study we propose a method to determine the precision parameter value based on the least upper bound of data misclassification error. In addition we will use a simple example to explain our proposed method and a medical examination example to demonstrate the method.

2. Variable precision rough sets model

The VPRS model is an extension of the original rough set model [12], which was proposed to analyze and identify data patterns that represent statistical trends rather than being functional.

VPRS deals with a partial classification by introducing a precision parameter β. The β value represents a bound on the conditional probability of a proportion of objects in a condition class that are classified to the same decision class. Ziarko [5] defined the β value as a classification error and the range in the domain $[0.0, 0.5)$. However, An et al. [13] and Beynon [9] considered β to denote the proportion of correct classifications, in which case the appropriate range is $(0.5, 1.0]$. In this study we use the Ziarko notion. The procedure of the VPRS model has five steps and they are as follows:

Step 1: Choosing the precision parameter (β) value.
Step 2: Find the full set of β-reduct.
Step 3: Elimination of duplicate objects.
Step 4: Elimination of superfluous values of attributes.
Step 5: Rules extraction.

VPRS operates on what may be described as a knowledge representation system or information system. An information system (S) consisting of four parts is shown as

$$S = (U, A, V, f),$$

where U is a non-empty set of objects; A is the collection of objects; we have $A = C \cup D$ and $C \cap D = \phi$, where C is a non-empty set of condition attributes, and D is a non-empty set of decision attributes; V is the union of attribute domains, i.e., $V = \bigcup_{a \in A} V_a$, where V_a is a finite attribute domain and the elements of V_a are called values of attribute a; F is an information function such that $f(u_i, a) \in V_a$ for every $a \in A$ and $u_i \in U$.

Every object that belongs to U is associated with a set of values corresponding to the condition attributes C and decision attributes D.

2.1. β-lower and β-upper approximations

Suppose that information system $S = (U, A, V, f)$, with each subset $Z \subseteq U$ and an equivalence relation R, referred to as an indiscernibility relation, corresponds to a partitioning of U into a collection of equivalence classes $R^* = \{E_1, E_2, \ldots, E_n\}$. We will assume that all sets under consideration are finite and non-empty [14]. The variable precision rough sets approach to data analysis hinges on two basic concepts, namely, the β-lower and the β-upper approximations of a set. The β-lower and the β-upper approximations can also be presented in an equivalent form as shown below. The β-lower approximation of the set $Z \subseteq U$ and $P \subseteq C$:

$$\underline{C}_\beta(D) = \bigcup_{1 - P_r(Z|x_i) \leqslant \beta} \{x_i \in E(P)\},$$

the β-upper approximation of the set $Z \subseteq U$ and $P \subseteq C$:

$$\bar{C}_\beta(D) = \bigcup_{1 - P_r(Z \mid x_i) < 1 - \beta} \{x_i \in E(P)\},$$

where $E(\bullet)$ denotes a set of equivalence classes (in the above definitions, they are condition classes based on a subset of attributes P).

$$Z \subset E(D), \quad P_r(Z \mid x_i) = \frac{\text{Card}(Z \cap x_i)}{\text{Card}(x_i)}.$$

2.2. Quality of classification

Based on Ziarko [5], the measure of quality of classification for the VPRS model is defined as

$$\gamma(P, D, \beta) = \frac{\text{card}\left(\bigcup_{1-P_r(Z|x_i) \leqslant \beta} \{x_i \in E(P)\}\right)}{\text{card}(U)}, \quad (2.1)$$

where $Z \subset E(D)$ and $P \subseteq C$, for a specified value of β. The value $\gamma(P, D, \beta)$ measures the proportion of objects in the universe (U) for which a classification (based on decision attributes D) is possible at the specified value of β.

2.3. Core and β-reducts

If the set of attributes is dependent, then we are interested in finding all possible minimal subsets of the attribute, which leads to the same number of elementary sets as the whole attributes (β-reduct), and in finding the set of all indispensable attributes (core). The concepts of core and β-reduct are two fundamental concepts of the VPRS. The β-reduct is the essential part of the information system, which can discern all discernable objects by the original information system. The core is the common part of all β-reducts.

We will call a β-reduct for an information system any subset B, $B \subseteq C$ such that [1]

(1) $\forall D_I \in D^*$, $B(\underline{C}_\beta D_I) = \underline{C}_\beta D_I$,
(2) $\forall A \subset B, \exists D_I \in D^*$, $\underline{A}(\underline{C}_\beta D_I) \neq \underline{C}_\beta D_I$,

where D^* denotes a set of equivalence classes; D_I denotes the ith category of D^*.

A β-reduct of the set of condition attributes P ($P \subseteq C$) with respect to a set of decision attributes D is a subset $RED(P, D, \beta)$ of P which satisfies the following two criteria [5]:

(1) $\gamma(P, D, \beta) = \gamma(RED(P, D, \beta), D, \beta)$;
(2) no attributes can be eliminated from $RED(P, D, \beta)$ without affecting the requirement (1).

To compute reducts and core, the discernibility matrix is used. Let the information system $S = (U, A)$ with $U = \{x_1, x_x, \ldots, x_n\}$. We use a discernibility matrix of S, denoted as $M(S)$, which has the dimension $n \times n$, where n denotes the number of elementary sets, defined as

$$(c_{ij}) = \{a \in A \mid a(x_i) \neq a(x_j) \text{ for } i, j = 1, 2, \ldots, n\}.$$

Thus, entry c_{ij} is the set of all attributes which discern objects x_i and x_j.

The core can be defined as the set of all single element entries of the discernibility matrix [2], i.e.,

$$\text{core}(A) = \{a \in A \mid c_{ij} = (a) \text{ for some } i, j\}.$$

The discernibility matrix can be used to find the minimal subset(s) of attributes, which leads to the same partition of the data as the whole set of attributes A. To do this, we have to construct the discernibility function $f(A)$. This is a Boolean function constructed in the following way: to each attribute from the set of attributes, which discern two elementary sets, (e.g., $\{a_1, a_2, a_3, a_4\}$), we assign a Boolean variable 'a', and the resulting Boolean function attains the form $(a_1 + a_2 + a_3 + a_4)$, or it can be presented as $(a_1 \wedge a_2 \wedge a_3 \wedge a_4)$. If the set of attributes is empty, then we assign to it the Boolean constant 1 [15].

2.4. Rules extraction

Procedures for generating decision rules from an information system has two main steps as follows:

Step 1: Selection of the best minimal set of attributes (i.e., β-reduct selection).

Step 2: Simplification of the information system can be achieved by dropping certain values of attributes that are unnecessary for the information system.

Ziarko [5] indicated that every minimal set of attributes may be perceived as an alternative group of attributes, which could be used instead of all the available attributes in the decision making based on cases. The main difficulty is how to select an optimal β-reduct. Two approaches can be used in this case. In the first one, the β-reduct with the minimal number of attributes is selected. In the second approach, the β-reduct that has the least number of combinations of values of its attributes is selected.

3. Determination of the precision parameter

When performing a VPRS analysis, how the β-reducts are selected is a key point of the process. The precision parameter value can control the choice of β-reducts. Previous related research studies lacked an effective method to determine a precision parameter value. Ziarko [5] defined the measure of the relative degree of misclassification of the set X with respect to Y as

$$c(X, Y) = \begin{cases} 1 - \dfrac{\text{card}(X \cap Y)}{\text{card}(X)} & \text{if card}(X) > 0, \\ 0 & \text{if card}(X) = 0, \end{cases}$$

where card denotes set cardinality.

Let X and Y be non-empty subsets of U. We say that X is included in Y, if for every element that belongs to X, then that also belongs to Y. The measure of relative misclassification can define the inclusion relationship between X and Y without explicitly using a general quantifier

$$Y \supseteq X \Leftrightarrow c(X, Y) = 0.$$

According to the specified majority requirement, the admissible β must be within the range $0 \leqslant \beta < 0.5$. Based on this assumption the majority inclusion relation is defined as

$$Y \overset{\beta}{\supseteq} X \Leftrightarrow c(X, Y) \leqslant \beta.$$

The above definition covers the entire family of β-majority relations.

In this study we propose an effective method to find the β-reducts, which involves two steps:

Step 1: Find the candidates of β-reducts using precision parameter (β).

In this study we propose an effective method to determine the β value in the VPRS model, which is based on the least upper bound $\xi(C, D)$ of the data set, where C is the condition attributes set, D is the decision attributes set, and $C^* = \{E_1, E_2, \ldots, E_n\}$ is the equivalence classes. The following equation is used for calculating the least upper bound of data set.

$$\xi(C, D) = \max(m_1, m_2), \tag{3.1}$$

where

$$m_1 = 1 - \min\{c(E, D) | E \in C^* \text{ and } \theta < c(E, D)\},$$
$$m_2 = \max\{c(E, D) | E \in C^* \text{ and } c(E, D) < \theta\},$$
$$c(E, D) = 1 - \frac{\text{card}(E \cap D)}{\text{card}(E)}$$

denotes the relative degree of misclassification of the set E with respect to D.

θ denotes the threshold, which is determined by the decision maker based on the relative degree. Usually, θ is set at 0.5.

Step 2: Find the β-reducts:

(1) For each candidate of β-reducts (subset P), calculate the quality of classification based on (2.1).
(2) Remove redundant attributes.
Let subset $X \subseteq P$. For each subset ($\gamma(C, D, \beta) = \gamma(P, D, \beta)$), if $\gamma(X, D, \beta) = \gamma(C, D, \beta)$, then remove the attributes $P \backslash X$. Otherwise, do not remove any attribute from subset P.
(3) Find the β-reducts.
Any subset X, which has $\gamma(X, D, \beta) = \gamma(C, D, \beta)$ is a β-reduct.

4. Rules extraction in the neural networks

Neural networks possess the unique capability of learning arbitrary non-linear mappings between noisy sets of input and output patterns. A neural network approach can usually be constructed without requiring any information concerning the functional form of the relationship between the predictors and the response [16]. It learns and extracts the process behavior from the past operating information. Once trained, a neural network can be evaluated very quickly, and the knowledge pertaining to the relationships between the input and output is stored in the network weights.

The extraction of symbolic rules from trained neural networks can alleviate the knowledge acquisition problem and refine the initial domain knowledge. Using extracted rules, neural network users can understand what the neural networks have learned and how the neural network makes predictions.

Su et al. [17] presented that the rules for the extraction procedure in the neural networks contains three steps as follows:

Step 1: Select important input features.
Step 2: Delete unnecessary connections.
Step 3: Extract rules.

For rules extraction, the back-propagation network is the most popular neural network model. Funahashi [18] proved that any continuous mapping can be approximately realized by multilayer neural networks with one hidden layer whose output functions are sigmoid functions. Thus, a multilayer neural network with one hidden layer using a back-propagation algorithm network is used in this study.

5. A simple example

The data sets taken from literature [9] are given in Table 1. There exists a set of objects $U(1, 2, \ldots, 7)$ contained in the rows of the table, with the columns denoting the condition attributes $C(a, b, c, d, e, f)$ of these objects, and a related decision attribute D.

In this information system the objects have been classified into one of two categories, M and F. In the information system, objects 1, 3, 4, and 6 are unambiguously classified, in the sense that all objects with a given set of attribute values are assigned to the same category. Objects 2, 5, and 7 are ambiguously classified since they have the same combination of condition attributes, but they are not all classified to the same decision category. Subsequently, the condition classes of objects as groupings of indiscernible objects are

$$C^* = \{C_1, C_2, C_3, C_4, C_5\}, \quad \text{where } C_1 = \{1\},$$
$$C_2 = \{2, 5, 7\},\ C_3 = \{3\},\ C_4 = \{4\},\ C_5 = \{6\}.$$

Similarly, the decision classes of the categories are

$$D^* = \{M, F\},$$
$$\text{where } D_M = \{1, 2, 3\} \quad \text{and} \quad D_F = \{4, 5, 6, 7\}.$$

Table 1
Information system

Objects	a	b	c	d	e	f	D
1	1	1	1	1	1	1	M
2	1	0	1	0	1	1	M
3	0	0	1	1	0	0	M
4	1	1	1	0	0	1	F
5	1	0	1	0	1	1	F
6	0	0	0	1	1	0	F
7	1	0	1	0	1	1	F

5.1. Choose the precision parameter value of the information system

Based on (3.1), $\xi(C, D) = \max(m_1, m_2)$, where

$$m_1 = 1 - \min[0.5 < c(C, D)], \quad m_2 = \max[c(C, D) < 0.5],$$
$$c(C_1, D_M) = 1 - \tfrac{1}{1} = 0, \quad c(C_1, D_F) = 1 - \tfrac{0}{1} = 1,$$
$$c(C_2, D_M) = 1 - \tfrac{1}{3} = \tfrac{2}{3}, \quad c(C_2, D_F) = 1 - \tfrac{2}{3} = \tfrac{1}{3},$$
$$c(C_3, D_M) = 1 - \tfrac{1}{1} = 0, \quad c(C_3, D_F) = 1 - \tfrac{0}{1} = 1,$$
$$c(C_4, D_M) = 1 - \tfrac{0}{1} = 1, \quad c(C_4, D_F) = 1 - \tfrac{1}{1} = 0,$$
$$c(C_5, D_M) = 1 - \tfrac{0}{1} = 1, \quad c(C_5, D_F) = 1 - \tfrac{1}{1} = 0,$$
$$m_1 = 1 - \min\left(\tfrac{2}{3}, 1\right) = \tfrac{1}{3},$$
$$m_2 = \max\left(0, \tfrac{1}{3}\right) = \tfrac{1}{3}.$$

Thus, $\xi(C, D) = \max\left(\tfrac{1}{3}, \tfrac{1}{3}\right) = \tfrac{1}{3}$.

Therefore, the precision parameter value is equal to $\tfrac{1}{3}$.

5.2. Find the full set of β-reducts

Since the β value is equal to $\tfrac{1}{3}$, then $\underline{C}_\beta(D_M) = C_1 \cup C_3 = \{1, 3\}$ and $\underline{C}_\beta(D_F) = C_2 \cup C_4 \cup C_5 = \{2, 4, 5, 6, 7\}$. The discernibility matrix, $M(S)$, for the seven elementary sets presented, is shown in Table 2. The relative discernibility functions are

$$\begin{aligned}\Delta_{\underline{\beta}}(1) &= (d+e)(b+d)(a+b+c+f)\\ &= be + ad + bd + cd + df,\end{aligned}$$

$$\begin{aligned}\Delta_{\underline{\beta}}(2) &= (b+e)(a+c+d+f)\\ &= ab + bc + bd + bf + ae + ce + de + ef,\end{aligned}$$

$$\begin{aligned}\Delta_{\underline{\beta}}(3) &= (a+b+d+f)(c+e)(a+d+e+f)\\ &= ac + ae + cd + de + be + cf + ef,\end{aligned}$$

$$\begin{aligned}\Delta_{\underline{\beta}}(4) &= (d+e)(b+e)(a+b+d+f)\\ &= ae + bd + be + ed + ef,\end{aligned}$$

$$\begin{aligned}\Delta_{\underline{\beta}}(5) &= (b+d)(a+d+e+f)\\ &= d + ab + be + bf,\end{aligned}$$

$$\begin{aligned}\Delta_{\underline{\beta}}(6) &= (c+e)(a+b+c+f)(a+c+d+f)\\ &= c + ae + ef + bed,\end{aligned}$$

$$\begin{aligned}\Delta_{\underline{\beta}}(7) &= (b+d)(a+d+e+f)\\ &= \Delta_{\underline{\beta}}(5),\end{aligned}$$

$$\begin{aligned}\Delta_{\underline{\beta}}(D_M) &= \Delta_{\underline{\beta}}(1)^*\Delta_{\underline{\beta}}(3)\\ &= (be + ad + bd + cd + df)\\ &\quad \times (ac + ae + cd + de + be + cf + ef)\\ &= cd + be + ade + def,\end{aligned}$$

$$\begin{aligned}\Delta_{\underline{\beta}}(D_F) &= \Delta_{\underline{\beta}}(2)^*\Delta_{\underline{\beta}}(4)^*\Delta_{\underline{\beta}}(5)^*\Delta_{\underline{\beta}}(6)^*\Delta_{\underline{\beta}}(7)\\ &= (ab + bc + bd + bf + ae + cd + de + ef)\\ &\quad \times (ae + bd + be + ed + ef)\\ &\quad \times (d + ab + be + bf)(c + ae + ef + bde)\\ &\quad \times (d + ab + be + bf)\\ &= abe + bce + cde + abe + def + bde\\ &\quad + bcd + bcf + bef,\end{aligned}$$

$$\begin{aligned}\Delta_{\underline{\beta}}(D) &= \Delta_{\underline{\beta}}(D_M)^*\Delta_{\underline{\beta}}(D_F)\\ &= (cd + be + ade + def)(abe + bce + cde + abe\\ &\quad + def + bde + bcd + bcf + bef)\\ &= abe+bce+cde+ade+def+bde+bcd+bef.\end{aligned}$$

By step 1, we have eight subsets, which are

$$\{a, b, e\}, \{b, c, e\}, \{c, d, e\}, \{a, d, e\}, \{d, e, f\}, \{b, d, e\},$$
$$\{b, c, d\}, \text{ and } \{b, e, f\}.$$

Since $\{c, d\}$ has the least number of attributes, it is selected. In Table 3, the $M(S)$-information system for β-reduct $\{c, d\}$ is presented. By step 2, we have four β-reducts, which are: $\{c, d\}$, $\{b, e\}$, $\{d, e, f\}$, and $\{a, d, e\}$. Since β-reducts $\{c, d\}$ has the least number of combinations of values of its attributes, it is selected for further study. In Table 3, the $M(S)$-information system for β-reducts $\{c, d\}$ is presented.

We also are interested in the elimination of superfluous values of condition attributes in the $M(S)$-information system. To do this, we must compute the relative attributes values of subset $\{c, d\}$, based on the $M(S)$-discernibility matrix constructed for the $M(S)$-information system. Table 4 presents the $M(S)$-discernibility matrix for the β-reducts $\{c, d\}$. The relative discernibility functions are

$$f_1(A) = cd,$$
$$f_2(A) = cd,$$
$$f_3(A) = cd,$$
$$f_4(A) = d,$$
$$f_5(A) = d,$$
$$f_6(A) = c,$$
$$f_7(A) = d.$$

Table 5 shows the information system's final version in the subspace $\{c, d\}$.

Table 2
Discernibility matrix

	1	2	3	4	5	6	7
1		—	—	d, e	b, d	a, b, c, f	b, d
2	—		—	b, e	—	a, c, d, f	—
3	—	—		a, b, d, f	a, d, e, f	c, e	a, d, e, f
4	d, e	b, e	a, b, d, f		—	—	—
5	b, d	—	a, d, e, f	—		—	—
6	a, b, c, f	a, c, d, f	c, e	—	—		—
7	b, d	—	a, d, e, f	—	—	—	

Table 3
$M(S)$-information system

Objects	c	d	D
1	1	1	M
2	1	0	M
3	1	1	M
4	1	0	F
5	1	0	F
6	0	1	F
7	1	0	F

Table 4
$M(S)$-discernibility matrix

	1	2	3	4	5	6	7
1		—	—	d	d	c	d
2	—		—	—	—	c, d	—
3	—	—		d	d	c	d
4	d	—	d		—	—	—
5	d	—	d	—		—	—
6	c	c, d	c	—	—		—
7	d	—	d	—	—	—	

Table 5
Final version of information system (literature data)

Objects	c	d	D
1	1	1	M
2	1	0	M
3	1	1	M
4	*	0	F
5	*	0	F
6	0	*	F
7	*	0	F

Note: "*" indicates: do not care.

Table 6
Decision rules (by the proposed approach)

Rules	Accuracy
1. If $c = 1$ and $d = 1$ then $D = M$	100% (2/2)
2. If $d = 0$ then $D = F$	75% (3/4)
3. If $c = 0$ then $D = F$	100% (1/1)

Note: (/) indicates (number of correct instances/number of total instances).

Table 7
Decision rules (by the Beynon approach)

Rules	Accuracy
1. If $b = 1$ and $e = 1$ then $D = M$	100% (1/1)
2. If $b = 0$ and $e = 1$ then $D = F$	75% (3/4)
3. If $b = 0$ and $e = 0$ then $D = M$	100% (1/1)
4. If $b = 1$ and $e = 0$ then $D = F$	100% (1/1)

Note: (/) indicates (number of correct instances/number of total instances).

5.3. Rules extraction

According to Table 5, the generalized rules are listed in Table 6. Comparing the implementation results from the proposed method with the literature approach (the extracted rules and results are listed in Table 7), the extracted rules' numbers generated by our method are less than those of the literature's approach, though the classification accuracy in the literature's approach is as good as our proposed method.

6. A case study

6.1. The problem

This case utilizes medical data to diagnose liver malfunctions. The data are from the general medical examination

Table 8
Condition attributes ranges for MES discretization

Examination items	Range '1'	Range '2'	Range '3'
Age (a_1)	23–34	35–55	56–63
Neutrophil (a_2)	0.0–36.9	37.0–75.0	75.1–
Lymphovyte (a_3)	0.0–19.9	20.0–55.0	55.1–
Moncyte (a_4)	0.0–2.4	2.5–10.0	10.1–
Basophil (a_5)	—	0.0–2.0	2.1–
GLUAC (a_6)	0–69	70–110	111–
ALK-P (a_7)	0–59	60–205	206–
GOT (a_8)	0–7	8–35	36–
GPT (a_9)	—	0–35	36–
γ-GT (a_{10})	—	0–45	46–
D-Bil. (a_{11})	0	0.1–0.5	0.6–
T-Protein (a_{12})	0.0–6.2	6.3–8.5	8.6–
TG (a_{13})	0–59	60–105	106–
BUN (a_{14})	0–7	8–25	26–
Uric Acid (a_{15})	0.0–2.4	2.5–8.0	8.1–

Table 9
Final version of the information system (liver data)

Objects	a_7	a_8	a_9	a_{10}	a_{13}	D
1	2	2	2	2	2	0
2	3	*	2	*	3	0
3	*	*	3	*	*	1
4	2	*	*	*	3	1
5	*	3	*	*	*	1
6	*	*	3	*	*	1
7	*	*	3	*	*	1
8	*	3	*	*	*	1
9	*	*	*	3	*	1
10	*	*	*	3	*	1
11	*	*	*	3	*	1
12	3	*	*	*	2	1
13	*	3	*	*	*	1
14	*	*	*	3	*	1
15	*	*	3	*	*	1
16	*	*	3	*	*	1
17	1	*	*	*	*	1

Note: "*" indicates: do not care.

items at a hospital located in Taipei, Taiwan. The examination data has fifteen items. They are Age, Neutrophil, Lymphocyte, Moncyte, Basophil, GLUAC, ALK-P, GOT, GPT, γ-GT, D-Bil., T-Protein, TG, BUN, and Uric Acid. These items are characterized by multi-dimensional information about the current health status of patients, which makes it difficult to diagnose other diseases based on such a large amount of information. Until now, the relationship between the medical examination data and liver malfunction symptom is still ambiguous.

In this case 168 instances are collected. These instances are separated into a training set that includes 101 instances (54 instances that are normal; 47 instances of liver malfunctions) and a test set that includes 67 instances (35 instances that are normal; 32 instances of liver malfunctions). Labeling "liver malfunction patients" is based on the medical history of the patients as judged by medical doctors.

6.2. Using the proposed approach

Since VPRS needs the data in a categorical form, the continuous attributes must be discretized before the VPRS analysis is performed. In this case the items of the medical examination standard (MES) are utilized to discretize the continuous attributes. The results are listed in Table 8. From Table 8, we know that each condition attribute is classified into two or three ranges.

In this information system the objects have been classified into one of two categories, 0 (normal) and 1 (malfunction). The condition classes of objects can distinguish 40 groups, and the precision parameter value is equal to $\frac{1}{6}$. Following the method of analysis given previously, four subsets and a β-reducts can be obtained. The subsets are $\{a_2, a_7, a_8, a_9, a_{10}, a_{13}\}$, $\{a_3, a_7, a_8, a_9, a_{10}, a_{13}\}$, $\{a_7, a_8, a_9, a_{10}, a_{13}, a_{15}\}$, and $\{a_7, a_8, a_9, a_{10}, a_{12}, a_{13}\}$; the β-reduct is $\{a_7, a_8, a_9, a_{10}, a_{13}\}$. The final version of the information system in the subspace $\{a_7, a_8, a_9, a_{10}, a_{13}\}$ is shown in Table 9.

According to Table 9, the extracted rules are listed in Table 10. From Table 10, we know that the instances of the test set at rule 2 and rule 7 are null, while rule 8 shows only one instance in the test set. Since these rules are not a matter for the judgment of liver diseases, they are deleted. The final extraction rules are listed in Table 11.

6.3. Using the neural network approach

In this section the *Professional II Plus* software package (Neural Ware, Inc., 1992) is used to perform the computation in order to obtain the structure with a maximum classification rate. After trial and error, we choose 0.25 and 0.30 as the learning rates in the hidden layer and the output layer, respectively. The momentum is set at 0.95, and the number of iterations is set at 20,000. Structure 15-10-1 is the optimal structure through the trained back-propagation neural network.

After the features selection, eight features are deleted and seven features are retained. They are Lymphocyte, Monocyte, GOT, GPT, γ-GT, D-Bil, and TG. These seven features are used to retrain a new network. Structure 7-5-1 is chosen for further analysis.

After pruning the unnecessary connections from network 7-5-1, only three attributes, GOT, GPT, and γ-GT could affect the result. The simplified 7-5-1 structure is used to extract the rules, and the results are listed in Table 12.

Table 10
Results of rule extraction (VPRS)

Rules	Accuracy (%)	
	Training set	Test set
1. If $60 \leqslant$ ALK-P $\leqslant 205$, $8 \leqslant$ GOT $\leqslant 35$, $0 \leqslant$ GPT $\leqslant 35$, $0 \leqslant \gamma$-GT $\leqslant 45$ and $60 \leqslant$ TG $\leqslant 105$, then one is normal.	100% (52/52)	94.59% (35/37)
2. If $206 \leqslant$ ALK-P, $0 \leqslant$ GPT $\leqslant 35$ and $106 \leqslant$ TG, then one is normal.	100% (1/1)	—
3. If $36 \leqslant$ GPT, then one has a malfunction.	96.43% (27/28)	100% (13/13)
4. If $60 \leqslant$ ALK-P $\leqslant 205$ and $106 \leqslant$ TG, then one has a malfunction.	100% (4/4)	100% (2/2)
5. If $36 \leqslant$ GOT, then one has a malfunction.	100% (8/8)	100% (8/8)
6. If $46 \leqslant \gamma$-GT, then one has a malfuncion.	100% (6/6)	100% (6/6)
7. If $206 \leqslant$ ALK-P and $60 \leqslant$ TG $\leqslant 105$, then one has a malfunction.	100% (1/1)	—
8. If ALK-P $\leqslant 59$, then one has a malfunction.	100% (1/1)	100% (1/1)

Note: "—" indicates the instance in the set is null. (/) indicates (number of correct instances/number of total instances).

Table 11
Final results of rule extraction (VPRS)

Rules	Accuracy (%)	
	Training set	Test set
1. If $60 \leqslant$ ALK-P $\leqslant 205$, $8 \leqslant$ GOT $\leqslant 35$, $0 \leqslant$ GPT $\leqslant 35$, $0 \leqslant \gamma$-GT $\leqslant 45$ and $60 \leqslant$ TG $\leqslant 105$, then one is normal.	100% (52/52)	94.59% (35/37)
2. If $36 \leqslant$ GPT, then one has a malfunction.	96.43% (27/28)	100% (13/13)
3. If $60 \leqslant$ ALK-P $\leqslant 205$ and $106 \leqslant$ TG, then one has a malfunction.	100% (4/4)	100% (2/2)
4. If $36 \leqslant$ GOT, then one has a malfunction.	100% (8/8)	100% (8/8)
5. If $46 \leqslant \gamma$-GT, then one has a malfunction.	100% (6/6)	100% (6/6)

Note: (/) indicates (number of correct instances/number of total instances).

Table 12
Results of rule extraction (neural networks)

Rules	Accuracy (%)	
	Training set	Test set
1. If GOT $\leqslant 35$, GPT $\leqslant 35$ and γ-GT $\leqslant 45$, then one is normal.	88.14% (52/57)	90% (36/40)
2. If $36 \leqslant$ GOP, then one has a malfunction.	100% (12/12)	100% (9/9)
3. If $36 \leqslant$ GPT, then one has a malfunction.	97.56% (40/41)	100% (23/23)
4. If $46 \leqslant \gamma$-GT, then one has a malfunction.	100% (6/6)	100% (11/11)

Note: (/) indicates (number of correct instances/number of total instances).

7. Conclusion

Despite its diverse applications in many domains, the VPRS model lacks a feasible method to determine a precision parameter value to control the choice of β-reducts. The primary motivation of this study is to develop a method to select a precision parameter to control the choice of β-reducts.

The proposed method is based on the least upper bound of the data misclassification error. A simple example has been provided to illustrate the effectiveness and speed in obtaining a suitable β-reduct with our proposed method. Furthermore, a medical examination case has also been analyzed.

In this study we propose an effective method to find the β-reducts. The notion of the approach being effective is a subjective opinion. First, we calculate a precision parameter value to find the subsets of an information system that is based on the least upper bound of the data misclassification error. Next, we measure the quality of classification and remove redundant attributes from each subset. Two numerical examples have been conducted to demonstrate the feasibility of the proposed method.

Future work should consider if the information system has any missing data and should develop a continuous attributes' discretization method without the domain knowledge of experts.

References

[1] Lin TY, Cercone N. Rough sets and data mining analysis of imprecise data. Dordrecht: Kluwer Academic Publishers; 1997.

[2] Pawlak Z. Rough sets theoretical aspects of reasoning about data. Dordrecht: Kluwer Academic Publishers; 1991.

[3] Slowinski R. Intelligent decision support handbook of application and advances of the rough sets theory. Dordrecht: Kluwer Academic Publishers; 1992.

[4] Beynon M, Peel MJ. Variable precision rough set theory and data discretisation: an application to corporate failure prediction. Omega-International Journal of Management Science 2001;29(6):561–76.

[5] Ziarko W. Variable precision rough set model. Journal of Computer and System Science 1993;46:39–59.

[6] Kattan MW, Cooper RB. The predictive accuracy of computer-based classification decision techniques. A review and research directions. Omega-International Journal of Management Science 1998;26(4):467–82.

[7] Beynon M. Investigating the choice of l and u values in the extended variable precision rough sets model. In: The Third International Conference on Rough Sets and Current Trend in Computing. Lecture notes in artificial intelligence series. Berlin: Springer; 2002. p. 61–8.

[8] Beynon M. An investigating of β-reduct selection within the variable precision rough sets model. The Second International Conference on Rough Sets and Current Trend in Computing. Lecture notes in artificial intelligence series. Berlin: Springer; 2000. p. 114–22.

[9] Beynon M. Reducts within the variable precision rough sets model: a further investigation. European Journal of Operational Research 2001;134:592–605.

[10] Katzberg JD, Ziarko W. Variable precision extension of rough sets. Fundamenta Informaticae 1996;27:155–68.

[11] Beynon M. The identification of low-paying workplaces: an analysis using the variable precision rough sets model. The Third International Conference on Rough Sets and Current Trend in Computing. Lecture notes in artificial intelligence Series. Berlin: Springer; 2000. p. 530–7.

[12] Ziarko W. Probabilistic decision tables in the variable precision rough set model. Computational Intelligence 2001;17(3):593–603.

[13] An A, Shan N, Chan C, Cercone N, Ziarko W. Discovering rules for water demand prediction: an enhanced rough-set approach. Engineering Applications in Artificial Intelligence 1996;9(6):645–53.

[14] Ziarko W, Fei X. VPRSM approach to WEB searching. The Third International Conference on Rough Sets and Current Trend in Computing. Lecture notes in artificial intelligence series. Berlin: Springer; 2002. p. 514–21.

[15] Walczak B, Massart. Rough sets theory. Chemometrics and Intelligent Laboratory Systems 1999;47:1–16.

[16] Stern HS. Neural networks in applied statistics (with discussion). Technometrics 1996;38:205–20.

[17] Su CT, Hsu JH, Tsai CH. Knowledge mining from trained neural networks. Journal of Computer Information Systems 2002;42(4):61–70.

[18] Funahashi KI. On the approximate realization of continuous mappings by neural networks. Neural Networks 1989;2: 183–92.