A rough set approach for the discovery of classification rules in interval-valued information systems

Leung, Yee and Fischer, Manfred M. and Wu, Wei-Zhi and Mi, Ju-Sheng

The Chinese University of Hong Kong, Vienna University of Economics and Business, Zhejiang Ocean University, Hebei Normal University

2008

A rough set approach for the discovery of classification rules in interval-valued information systems

Yee Leung [a] *, **Manfred M. Fischer** [b], **Wei-Zhi Wu** [c], **Ju-Sheng Mi** [d]

[a] *Center for Environmental Policy and Resource Management and Joint Laboratory for GeoInformation Science, The Chinese University of Hong Kong, Shatin, Hong Kong*

[b] *Institute for Economic Geography and GIScience, Vienna University of Economics and Business Administration, Nordbergstr, 15/4/A, A-1090 Vienna, Austria*

[c] *Information College, Zhejiang Ocean University, Zhoushan, Zhejiang, 316004, P. R. China*

[d] *College of Mathematics and Information Science, Hebei Normal University, Shijiazhuang, Hebei, 050016, P. R. China*

Abstract

A novel rough set approach is proposed in this paper to discover classification rules through a process of knowledge induction which selects optimal decision rules with a minimal set of features necessary and sufficient for classification of real-valued data. A rough set knowledge discovery framework is formulated for the analysis of interval-valued information systems converted from real-valued raw decision tables. The optimal feature selection method for information systems with interval-valued features obtains all classification rules hidden in a system through a knowledge induction process. Numerical examples are employed to substantiate the conceptual arguments.

Keywords: Classification; Interval-valued information systems; Knowledge discovery; Knowledge reduction; Rough sets

---

1. Introduction

The discovery of non-trivial, previously unknown, and potentially useful knowledge from databases is important in the processing and utilization of

* Corresponding author. Fax: (852)26037242; Tel.: (852)26096473.

E-mail addresses: yeeleung@cuhk.edu.hk (Y. Leung), Manfred.Fischer@wu-wien.ac.at (M.M. Fischer), wuwz@zjou.net.cn (W.-Z. Wu), mijsh@263.net (J.-S. Mi).

voluminous information. A basic issue of a rule-based system is the determination of a minimal set of features (and feature values) and the optimal set of consistent rules for classification or inference. All of this has to be achieved with data available. The theory of rough sets, proposed by Pawlak [28], has recently been used to analyze data sets for such a purpose. This theory is an extension of classical set theory for the study of systems characterized by insufficient and incomplete information, and has been demonstrated to be useful in fields such as pattern recognition, machine learning, and automated knowledge acquisition [14,27,30-32,46,48]. Rough-set data analysis uses only internal knowledge, avoids external parameters, and does not rely on prior model assumptions such as probabilistic distribution in statistical methods, membership function in fuzzy sets theory, and basic probability assignment in Dempster-Shafer theory of evidence [7,33]. Its basic idea is to unravel an optimal set of decision rules from an information system (basically a feature-value table) via an objective knowledge induction process which determines the necessary and sufficient features constituting the rules for classification.

Classical definitions of lower and upper approximations, sometimes called Pawlak's rough approximations, were originally introduced with reference to an indiscernibility relation which is assumed to be an equivalence relation (reflexive, symmetric and transitive) [28,29]. This model is useful in the analysis of data presented in terms of complete information systems and complete decision tables. Pawlak's rough approximations may be generalized to nonequivalence relations [10,16,37,38,41,43,47,49]. The extensions of Pawlak's rough set model may be used in reasoning and knowledge acquisition in incomplete decision tables [5,8,11,18-22]. A more general definition of lower and upper approximations, called fuzzy lower and upper approximations, can be defined by using fuzzy relation and may be applied to fuzzy information systems [2,6,9,15,17,26,42,44,45].

When the rough set approach is used to unravel decision rules from a given information system, two types of decision rules may be derived. Based on the lower approximation of a decision class, certain information can be discovered and certain rules can be derived, whereas by using the upper approximation of a decision class, uncertain or partially certain information may be discovered and possible rules induced. Various approaches[①] using rough set theory have been proposed to discover decision rules from data sets taking the form of decision tables [3,8,11,13,15,17-25, 34-40,45,50,51].

Whilst conventional rough set models may be constructed for the analysis of categorical data, real-world problems often involve real-valued attributes characterizing objects of interest. In the classification of remotely-sensed images, for example, spectral bands whose gray values may be interpreted to be continuous are generally used to give spectral signatures to pixels constituting objects. Under such a situation, the Pawlak rough set model may generate an unacceptably large number of equivalence classes resulting in too many classification rules. Though the rules may be more deterministic with reference to the training data set, their generalization ability will most likely be rather low since perfect match of attribute values in real numbers is generally difficult. To make the identified classification rules more comprising and practical, a preprocessing step which can transform the real-numbered attribute values into a sufficiently small number of meaningful intervals is thus necessary.

Most of the current methods focus on the discretization of continuous attribute values by dividing the range of real numbers into a certain number of partitioning intervals [1,4,12]. Essentially, the methods transform an attribute with real values into an attribute with discrete real-valued intervals. It is, however, difficult or controversial to decide on the cut-off points separating the intervals. To circumvent such a problem, other conversion methods, such as more sophisticated statistical procedures, may be

[①] Greco et al. (1999), Grzymala-Busse (1991), and Kryszkiewicz (1998,1999) extended the rough set model to reason in incomplete information systems with missing values. Lingras and Yao (1998) employed two different generalizations of rough set models to generate plausibilistic rules with incomplete databases instead of probabilistic rules generated by a Pawlak's rough set model with complete decision tables, while other researchers such as Hong et al. (2000), Korvin et al. (1998), and Wu et al. (2003), used rough set models to handle fuzzy and quantitative data.

employed to preprocess a real-valued information system into an interval-valued information system. Once the conversion is done, the corresponding knowledge induction method needs to be investigated.

In this paper, a novel rough set approach for discovering classification rules from an interval- valued information system is proposed. The approach involves the transformation of real-valued information into interval-valued information, and the formulation of a knowledge induction procedure to identify optimal classification rules with a minimal set of features necessary and sufficient for classification with continuous attribute values.

To facilitate our discussion, we first present some basic notions of information systems and decision tables in the section that follows. New concepts of misclassification rate related to interval-valued information systems are then introduced in Section 3. Section 4 serves to discuss the α -tolerance relations, while in Section 5 and Section 6 we continue with the concept of α -classification reduction and induction of optimal decision rules, respectively. We then conclude the paper with a summary of the proposed approach and pointing to some conclusions.

2. Information systems and decision tables

The notion of an information system provides a convenient basis for the representation of objects in terms of their attributes. Without loss of generality, let us assume 1-ary attributes. A complete information system S may then be defined as a pair (U, A), where U is a nonempty finite set of, say n objects[②], $\{x_1, x_2, \ldots, x_n\}$, called the universe of discourse, and $A = \{a_1, a_2, \ldots, a_m\}$ is a nonempty finite set of m attributes, such that $a : U \to V_a$ for any $a \in A$, i.e., $a(x) \in V_a$. V_a is called the domain of attribute a.

② In a remote sensing context, the objects of interest are pixels and the attributes spectral bands. Note that in hyperspectral classification there are hundreds of spectral bands that might be used.

If the precise values of some of the attributes in an information system are not known, i.e., missing or known only partially, then such a system is called an incomplete information system and still can be denoted without any confusion by (U, A). Such a situation can be described by a set-valued information system in which the attribute value function a is defined as a mapping from U to the power set of V_a. For example, the missing values $a(x)$ can be represented by the set of all possible values for the attribute, i.e., $a(x) = V_a$; and if $a(x)$ is known partially, for instance, if we know that $a(x)$ is not $b, c \in V_a$, then the value $a(x)$ is specified as $V_a - \{b, c\}$.

A decision table is an information system $S = (U, A \cup \{d\})$, where d, $d \notin A$, is a complete attribute called a decision, and A is termed the conditional attribute set. If (U, A) is a complete information system, then $(U, A \cup \{d\})$ is referred to as a complete decision table. If (U, A) is an incomplete information system, then $(U, A \cup \{d\})$ is referred to as an incomplete decision table. We can treat the decision attribute as a kind of classifier on the universe of objects given by an expert or a decision-maker. In machine learning, decision tables are called sets of training examples.

Without loss of generality, we assume that $V_d = \{1, 2, \ldots, I\}$. We can observe that the decision d determines a partition of the universe of discourse,

$$U / d = \{[x]_d : x \in U\} = \{X_1^d, X_2^d, \ldots, X_I^d\},$$

where $X_i^d = \{x \in U : d(x) = i\}$, $i = 1, 2, \ldots, I$. The set X_i^d is termed the i-th decision class of decision table $S = (U, A \cup \{d\})$. Thus i may be regarded as the label of the class X_i^d.

For an information system $S = (U, A)$, one can describe relationships between objects through their attribute values. With respect to an attribute subset $B \subseteq A$, a binary equivalence relation R_B may be defined as

$$x, y \in U \ , \ (x, y) \in R_B \Leftrightarrow a(x) = a(y), \forall a \in B .$$

R_B is referred to as the relation with respect to B derived from information system S, and we call (U, R_B) the Pawlak approximation space with respect to B induced from S. With relation B, two objects are considered to be indiscernible if and only if they have the same value on each $a \in B$. Based on the approximation space (U, R_B), one can derive the lower and upper approximations of an arbitrary subset X of U. They are defined as

$\underline{B}(X) = \{x \in U : [x]_B \subseteq X\}$, and $\overline{B}(X) = \{x \in U : [x]_B \cap X \neq \varnothing\}$, respectively,

where $[x]_B = \{y \in U : (x, y) \in R_B\}$ is the B-equivalence class containing x. The pair $(\underline{B}(X), \overline{B}(X))$ is the representation of X in the Pawlak approximation space (U, R_B), or is referred to as the Pawlak rough set of X with respect to (U, R_B). Based on the lower and upper approximations of the decision classes X_i^d ($i = 1,2,\ldots, I$) with respect to (U, R_A) in the decision table $(U, A \cup \{d\})$, it is easy to unravel all of the certain and possible decision rules [29].

Given a number of facts, generalization can be performed in many different directions [14]. In order to extract interesting rules from databases, learning should be directed by background knowledge. For example, in the classification of remotely sensed images, land covers such as vegetation species can exhibit individual spectral signatures in a number of spectral bands. One can classify vegetation covers according to the gray levels in the corresponding bands. The task is to select a smaller number of spectral bands such that they have the same classification ability as the prespecified set which often contains more than sufficient bands. The mining of classification rules thus needs to be performed in real-valued databases. Real-valued data are in fact a very common information source in real-life problems. Remote sensing application is just a typical example.

If the values of each attribute in an information system are real numbers, then such a system is commonly called a real-valued information system in the rough set literature. If data are real-valued, then the conventional Pawlak rough-set model may yield a very large number of equivalence classes which will eventually unravel a very large number of classification rules in the knowledge induction process. Having too many classification rules, however, may give a more deterministic result in the training data, but their generalization capability will be substantially hampered. This is simply due to the difficulty in having a perfect match of attribute values in real numbers in the condition parts of the rules. Thus, to mine rules which are more encompassing and general, it is pertinent to first convert the real-valued information system into an interval-valued information system so that attributes in the mined rules are interval-valued in nature.

An interval-valued information system is a pair $K = (U, A)$, where $U = \{u_1, u_2, \ldots, u_I\}$ is a nonempty finite set of classes and $A = \{a_1, a_2, \ldots, a_m\}$ is a nonempty finite set of attributes, such that $a_k(u_i) = [l_i^k, u_i^k]$, $l_i^k < u_i^k$, for all $i = 1,2,\ldots,I$ and $k = 1,2,\ldots,m$. To obtain an interval-valued information system, we can employ methods such as discretization or other more sophisticated statistical procedures. Discretization may be based on experience or specification of arbitrary cut-off points. Statistical methods, on the other hand, may be based on the capturing of data variation under some probability density functions depicting the attributes. For example, it makes good statistical sense to specify an interval such as $\mu \pm 2\sigma$ (μ: mean, σ: standard deviation) under, say, normal distribution. Similar methods can be used for other probability density functions fitting the data. It should be noted that the statistical method is just employed for data conversion whenever appropriate. The rough-set knowledge induction approach suggested in this contribution has absolutely no bearing on any statistical arguments.

It should, however, be noted that, unlike the discretization methods by which the interval-valued attribute value interval set $\{a_k(u_i): u_i \in U\}$ forms a partition of the set for the same attribute a_k, the interval-valued attribute set obtained by statistical methods may have non-empty intersection for distinct classes in the universe of discourse. This is rather natural, because, for example, the gray values of different vegetations under the same band may have strong spectral affinity.

Example 1. Table 1 depicts an interval-valued information system about 10 species of vegetations. $U = \{u_1, u_2, \ldots, u_{10}\}$ is the universe of discourse which comprises 10 vegetation classes, $A = \{a_1, a_2, a_3, a_4, a_5\}$ the set of attributes (spectral bands) , the attribute value $a_k(u_j)$ an interval $[l_i^k, u_i^k]$ where

$$l_i^k = \max\{\mu_i^k - 2\sigma_i^k, 0\}, \quad u_i^k = \min\{\mu_i^k + 2\sigma_i^k, 255\},$$

obtained by including all real-valued data points that fall within $\mu \pm 2\sigma$ under the normal distribution. Hence, we can transform the raw data set into an interval-valued information system as shown by Table 1.

Table 1 to be placed about here

In the sections that follow, we propose some concepts and formulate a framework for mining classification rules in interval-valued information systems.

3. Misclassification rates

We start to introduce some concepts for the modeling of uncertainty in interval-valued information systems. Let $K=(U,A)$ be an interval-valued information system. Denote for any $i,j\le I$ with $i\ne j$ and $k\le m$

$$\alpha_{ij}^{k}=\begin{cases}0, & \text{if} \quad [l_i^k,u_i^k]\cap[l_j^k,u_j^k]=\varnothing,\\ \min\left\{\dfrac{\min\{u_i^k-l_j^k,u_j^k-l_i^k\}}{u_i^k-l_i^k},1\right\}, & \text{if} \quad [l_i^k,u_i^k]\cap[l_j^k,u_j^k]\ne\varnothing;\end{cases}$$

where α_{ij}^{k} is the probability that objects in class u_i are misclassified into class u_j according to attribute a_k. It is the length of the intersection of $[l_i^k,u_i^k]$ and $[l_j^k,u_j^k]$ divided by the length of the interval $[l_i^k,u_i^k]$. If $\alpha_{ij}^{k}=0$, then objects in class u_i will not be misclassified into the class u_j according to attribute a_k. If $\alpha_{ij}^{k}=1$ — note that $[l_i^k,u_i^k]\subseteq[l_j^k,u_j^k]$ in such a case — objects in class u_i will be completely misallocated to class u_j according to attribute a_k. It should be pointed out that $\alpha_{ij}^{k}=\alpha_{ji}^{k}$ does not hold in general.

Define

$$\alpha_{ij}=\min\{\alpha_{ij}^{k}:k\le m\},$$

where α_{ij} is the error that objects in class u_i being misclassified into class u_j in the system K. Moreover define

$$\alpha_i=\max\{\alpha_{ij}:j\le I,j\ne i\},$$

where α_i may be called the permissible misclassification rate that objects in class u_i being discerned (separated) from other classes in the system K. Finally, define the maximal mutual classification error between classes u_i and u_j, according to attribute a_k, as

$$\beta_{ij}^k = \max\{\alpha_{ij}^k, \alpha_{ji}^k\},$$

where $\beta_{ij}^k = \beta_{ji}^k$. If $\beta_{ij}^k = 0$, then we can distinguish classes u_i and u_j completely by using attribute a_k, while if $0 < \beta_{ij}^k < 1$, we can distinguish these classes up to a mutual classification error β_{ij}^k. But if $\beta_{ij}^k = 1$, we cannot distinguish the two classes at all.

Let us continue to define the permissible misclassification rate between classes u_i and u_j in the system K as

$$\beta_{ij} = \min_{1 \le k \le m} \beta_{ij}^k .$$

Then — if $\beta_{ij} \le \alpha$, there must exist an attribute a_k such that, by using a_k, within the permissible misclassification rate α — the two classes u_i and u_j can be separated. If $\beta_{ij} = 0$, they can be distinguished completely. If $\beta_{ij} = 1$, they cannot be separated in the system K. If $0 < \beta_{ij} < 1$, they can possibly be separated to a certain extent.

For a given permissible misclassification rate α, if $\beta_{ij} \le \alpha$, then there exists an attribute $a_k \in A$ such that, the classes u_i and u_j can be separated. If $\beta_{ij} > \alpha$, then there is no attribute in A such that with the permissible misclassification rate, u_i and u_j can be separated. In such a case, we claim that the two classes u_i and u_j cannot be distinguished in the system within α.

Define

$$\beta_i = \max\{\beta_{ij} : j \le m, j \ne i\}, \quad i \le I .$$

β_i is called the permissible misclassification rate such that class u_i can be separated from other classes in the system, that is, if $\beta_i \le \alpha$, within the given classification error α, the class u_i can be discerned.

Define, moreover, the minimal permissible misclassification rate such that all classes can be pairwise separated in the system as

$$\beta = \max\{\beta_{ij} : i \le I, j \le I, i \ne j\}.$$

That is, if $\beta \le \alpha$, then within the given misclassification rate α, any pair of classes in the system can be separated. In such a case, all classification rules derived from the system are consistent in the sense of α. If $\beta > \alpha$, within the given classification error α, classes cannot be pairwise distinguished.

Table 2 to be placed about here

Example 2. The classification errors for the information system described in Table 1 are displayed in Table 2. For example, $\alpha_{16} = 0.19$ means that the error of having objects in class u_1 misclassified into class u_6 is 0.19, from which we can find a spectral band (e.g., a_5) such that the rate of misallocating objects from class u_1 to class u_6 will not be more than 0.19. On the other hand, $\alpha_{61} = 0.13$ means that the error of having objects from class u_6 misallocated to class u_1 is 0.13. It is obvious that $\alpha_1 = 0.64$ which means that the permissible misclassification rate of separating objects in class u_1 from other classes in the system is 0.64; and $\alpha_3 = 0$ implies that there must exist some spectral bands such that, by using these bands, objects in class u_3 can be unmistakenly distinguished from other classes.

Table 3 to be placed about here

Table 3 shows the permissible misclassification rates for different classes in the interval-valued information system described in Example 1. For instance, $\beta_{85} = 0.38$ implies that the permissible misclassification rate between classes u_8 and u_5 is 0.38. $\beta_{25} = 0.88$ indicates that u_2 and u_5 are very similar and, thus, very hard to distinguish. β_8 shows a permissible misclassification rate of 0.49 for separating class u_8 from other classes, while $\beta = 0.88$ implies that the minimal permissible misclassification rate for pairwise separation of all classes is 0.88 in this example.

4. α - tolerance relations

This section continues to define α-tolerance relations in an interval-valued information system crucial for the search of the minimal number of features and the notion of attribute reducts to be discussed in the next section.

Let $K = (U, A)$ be an interval-valued information system. For a given permissible misclassification rate $\alpha \in [0,1]$ and an attribute subset $B \subseteq A$, we define a binary relation, denoted by R_B^α, on U as:

$$R_B^\alpha = \{(u_i, u_j) \in U \times U : \beta_{ij}^k > \alpha, \forall a_k \in B\}.$$

Two classes u_i and u_j have relation R_B^α if and only if they cannot be separated by the attribute set B under the misclassification rate a. We call R_B^α the α-tolerance relation with respect to B.

Obviously, R_B^α is reflexive and symmetric, but it may be not transitive. Thus, R_B^α is a tolerance relation which satisfies

$$R_B^\alpha = \bigcap_{b \in B} R_{\{b\}}.$$

Denote $S_B^\alpha(u) = \{v \in U : (u,v) \in R_B^\alpha\}$ which is called the α-tolerance class of u with respect to R_B^α or B. $v \in S_B^\alpha(u)$ if and only if u and v cannot be distinguished according to attributes in B within the misclassification rate α. It is easy to see that $0 \le \alpha \le \gamma \le 1$ implies $R_B^\alpha \subseteq R_B^\gamma$, $S_B^\alpha(u) \subseteq S_B^\gamma(u)$, for all $B \subseteq A$ and $u \in U$.

Example 3. It should be noted that $\beta_{ij}^k > \alpha$ for all $a_k \in A$ iff $\beta_{ij} > \alpha$. If we consider the interval-valued information system given by Table 1 and assume that a permissible misclassification rate $\alpha = 0.2$ is given, then we can obtain from Table 3 the Boolean matrix corresponding to $R_A^{0.2}$ as follows:

$$R_A^{0.2} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Consequently,

$S_A^{0.2}(u_1) = S_A^{0.2}(u_7) = \{u_1, u_7\}$, $S_A^{0.2}(u_2) = S_A^{0.2}(u_5) = S_A^{0.2}(u_8) = \{u_2, u_5, u_8\}$,

$S_A^{0.2}(u_3) = \{u_3\}$, $S_A^{0.2}(u_4) = S_A^{0.2}(u_{10}) = \{u_4, u_{10}\}$, $S_A^{0.2}(u_6) = \{u_6\}$, $S_A^{0.2}(u_9) = \{u_9\}$.

Hence under the given permissible misclassification rate $\alpha = 0.2$, classes u_3, u_6, and u_9 can be separated from other classes, and any one of the remaining classes cannot be discerned from others.

5. α-classification reduction and α-classification core

One fundamental aspect of rough set theory involves the search for particular subsets of attributes which provide the same information for classification purposes as the full set of available attributes. Such subsets are called attribute reducts. To acquire concise decision rules from systems, knowledge reduction is, thus, necessary. Many types of attribute reducts and decision results have been proposed in the rough set literature. For example, Kryszkiewicz [20] has established static relationships among conventional types of knowledge reduction in inconsistent complete decision tables. Zhang et al. [50] have introduced a new kind of knowledge reduction, called a maximum distribution reduct, which preserves maximum decision rules. Mi et al. [25] have proposed approaches to knowledge reduction based on variable precision rough set model. In this section, we study knowledge reduction in interval-valued information systems which can be used in the construction of optimal classification rules from the interval-valued information systems.

Let $K = (U, A)$ be an interval-valued information system, with $\alpha \in [0,1]$ and $B \subseteq A$. If $R_B^\alpha = R_A^\alpha$, then B is called an α-classification consistent set in K. If B is an α-classification consistent set, $B - \{b\}$ is not an α-classification consistent set in K for all $b \in B$, i.e. $R_{B-\{b\}}^\alpha \neq R_A^\alpha$, then B is termed an α-classification reduct in K. The set of all α-classification reducts in K is denoted by $re^\alpha(K)$. The intersection of all α-classification reducts is called the

α -classification core in K .

If $S_B^\alpha(u) = S_A^\alpha(u)$, then B is called an α -classification consistent set of u in K . If B is an α -classification consistent set of u in K , $B-\{b\}$ is not an α -classification consistent set of u in K for all $b \in B$, i.e. $S_{B-\{b\}}^\alpha(u) \neq S_A^\alpha(u)$, then B is called an α -classification reduct of u in K . The set of all α -classification reducts of u in K is denoted by $re^\alpha(u)$. The intersection of all α -classification reducts of u is called the α -classification core of u in K .

An α -classification consistent set in K is a subset of the attribute set that preserves the α -tolerance classes of all classes, while an α -classification reduct is a minimal α -consistent set that preserves the α -tolerance relation and, consequently, leads to the same classification in the sense of α . The remaining attributes are then redundant, and their removal does not affect the classification in the sense of α .

Let define the α -discernibility set of the two classes u_i and u_j in K as

$$D_{ij}^\alpha = \{a_k \in A : \beta_{ij}^k \leq \alpha\}, \quad i \neq j, \text{ and } \quad D_{ii}^\alpha = \varnothing \quad \text{for all} \quad i = 1,2,\ldots,I .$$

Then D_{ij}^α consists of a set of attributes separating classes u_i and u_j with a misclassification rate being not greater than α . Define, moreover, the α -discernibility matrix

$$\mathrm{M}^\alpha = \{D_{ij}^\alpha : i, j = 1,2,\ldots,I\}$$

and let

$$\mathrm{M}_0^\alpha = \{D_{ij}^\alpha : D_{ij}^\alpha \neq \varnothing\}.$$

Then we can use the following theorem to determine an α -classification consistent set according to the α -discernibility matrix.

Theorem 1. *Let* $K = (U, A)$ *be an interval-valued information system,*

$\alpha \in [0,1]$, *then* $B \subseteq A$ *is an* α*-classification consistent set in* K, *i.e.* $R_B^\alpha = R_A^\alpha$, *iff* $B \bigcap D \neq \varnothing$, $\forall D \in \mathrm{M}_0^\alpha$.

Proof. "$\Rightarrow$" Suppose that $R_B^\alpha = R_A^\alpha$. If $D \in \mathrm{M}_0^\alpha$, then by definition of M_0^α, there exist $1 \le i, j \le I$ with $i \neq j$ such that $D = D_{ij}^\alpha \neq \varnothing$. By definition of R_A^α we can see then that $(u_i, u_j) \notin R_A^\alpha$. Since $R_B^\alpha = R_A^\alpha$, we have $(u_i, u_j) \notin R_B^\alpha$, which implies that there exists an attribute $a_k \in B$ such that $\beta_{ij}^k \le \alpha$, that is, $a_k \in D_{ij}^\alpha$. Hence $a_k \in B \bigcap D \neq \varnothing$.

"$\Leftarrow$" Assume that $B \bigcap D \neq \varnothing$, $\forall D \in \mathrm{M}_0^\alpha$. If by contradiction $R_B^\alpha \neq R_A^\alpha$, then we know from $R_A^\alpha \subseteq R_B^\alpha$ that $R_A^\alpha \subset R_B^\alpha$. Thus, there exists $(u_i, u_j) \in R_B^\alpha$ such that $(u_i, u_j) \notin R_A^\alpha$. By $(u_i, u_j) \notin R_A^\alpha$ we see that $D_{ij}^\alpha \neq \varnothing$. Then there exists $a_k \in A$ such that $\beta_{ij}^k \le \alpha$. Hence $a_k \in D_{ij}^\alpha$, from which we can conclude that $D_{ij}^\alpha \in \mathrm{M}_0^\alpha$. Since by assumption $B \bigcap D_{ij}^\alpha \neq \varnothing$, there exists $a_l \in B$ such that $a_l \in D_{ij}^\alpha$. This means $\beta_{ij}^l \le \alpha$. Hence $(u_i, u_j) \notin R_B^\alpha$, which contradicts $(u_i, u_j) \in R_B^\alpha$. Therefore $R_B^\alpha = R_A^\alpha$.∎

Remark. According to Theorem 1, we can see that $B \subseteq A$ is an α-classification reduct in K iff B is the minimal set satisfying $B \bigcap D \neq \varnothing$, $\forall D \in \mathrm{M}_0^\alpha$.

Theorem 2. *Let* $K = (U, A)$ *be an interval-valued information system,* $\alpha \in [0,1]$, *then* $a_k \in A$ *is an element of* α*-classification core in* K *iff there exists* $D \in \mathrm{M}_0^\alpha$ *such that* $D = \{a_k\}$.

Proof. "$\Rightarrow$" Assume that $a_k \in A$ is an element of the α-classification core in K. Let

$$\mathrm{M}_k^\alpha = \{D \in \mathrm{M}_0^\alpha : a_k \in D\}.$$

If $\mathrm{card}(D) \geq 2$ for all $D \in \mathrm{M}_k^\alpha$, define

$$B = \bigcup_{D \in \mathrm{M}_0^\alpha} (D - \{a_k\}).$$

It is easy to see that

$$B \cap D \neq \varnothing, \quad \forall D \in \mathrm{M}_0^\alpha.$$

By Theorem 1 we know that B is an α-classification consistent set in K. Then there exists $C \subseteq B$ such that C is an α-classification reduct in K. Clearly, $a_k \notin C$, this contradicts a_k being an element of the α-classification core in K.

"$\Leftarrow$" Suppose that there exists $D \in \mathrm{M}_0^\alpha$ such that $D = \{a_k\}$. Then there exist $1 \leq i, j \leq I$ with $i \neq j$ such that $D_{ij}^\alpha = \{a_k\}$. By definition, we have $\beta_{ij}^k \leq \alpha$ and $\beta_{ij}^l > \alpha$ for all $l \neq k$ with $1 \leq l \leq m$. Consequently, $(u_i, u_j) \in R_{A-\{a_k\}}^\alpha$ and $(u_i, u_j) \notin R_A^\alpha$. It follows that

$$R_{A-\{a_k\}}^\alpha \neq R_A^\alpha.$$

Note that a_k is an element of the α-classification core in K iff $R_{A-\{a_k\}}^\alpha \neq R_A^\alpha$. Therefore, a_k is an element of the α-classification core in K.█

Table 4 to be placed about here

Example 4. In the information system described in Table 1 and under the given permissible misclassification rate of $\alpha = 0.2$, the discernibility sets are obtained as shown in Table 4. Since $D_{ij}^\alpha = D_{ji}^\alpha$, for simplicity, we only list D_{ij}^α's with

$1 \le j < i < I$. According to Theorem 1 and Theorem 2, it can easily be shown that the 0.2-classification reducts in the system are the two sets: $\{a_1, a_3, a_4, a_5\}$ and $\{a_2, a_3, a_4, a_5\}$; and the 0.2-classification core is $\{a_3, a_4, a_5\}$.

Reduct computation can also be translated into the computation of prime implicants of a Boolean function. It has been shown by Skowron and Rauszer[32] that the problem of finding reducts of a given Pawlak (complete) information system may be solved as a case in Boolean reasoning. The idea of Boolean reasoning is to represent a problem with a Boolean function and to interpret its prime implicants[③] as solutions to the problem. This approach is very useful to the calculation of reducts of classical information systems. We will generalize this approach to interval-valued information systems here. It should be pointed out that we are interested in implicants of monotone Boolean functions only, i.e. functions constructed without negation.

Let $K = (U, A)$ be an interval-valued information system. An α-discernibility function f_K^α for the interval-valued information system K is a Boolean function of m Boolean variables $\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_m$ corresponding to the attributes $a_1, a_2, \ldots, a_m$ respectively, and defined as follows:

$$f_K^\alpha(\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_m) = \wedge\{\vee D_{ij}^\alpha : D_{ij}^\alpha \in \mathrm{M}_0^\alpha\},$$

where $\vee D_{ij}^\alpha$ is the disjunction of all variables $\bar{a}$ such that $a \in D_{ij}^\alpha$, while $\wedge$ denotes conjunction.

Theorem 3. *Let $K = (U, A)$ be an interval-valued information system. Then an attribute subset $B \subseteq A$ is an α-classification reduct in K iff $\wedge_{a_k \in B} \overline{a_k}$ is a prime implicant of the α-discernibility function f_K^α.*

---

③ An implicant of a Boolean function f is any conjunction of literals such that for each valuation v of variables, the value of the function f under v is also true if these literals are true under v. A prime implicant is a minimal implicant.

Proof. "$\Rightarrow$". Assume that $B \subseteq A$ is an α-classification reduct in K. By Theorem 1 we have

$$B \bigcap D_{ij}^{\alpha} \neq \varnothing, \text{ for all } D_{ij}^{\alpha} \in \mathrm{M}_0^{\alpha}.$$

We claim that $\forall b \in B$ there must exist $D_{ij}^{\alpha} \in \mathrm{M}_0^{\alpha}$ such that $B \bigcap D_{ij}^{\alpha} = \{b\}$. In fact, if $\mathrm{card}\left(B \bigcap D_{ij}^{\alpha}\right) \geq 2$ for all $D_{ij}^{\alpha} \in \mathrm{M}_0^{\alpha}$ with $b \in D_{ij}^{\alpha}$, let $B' = B - \{b\}$. Then by Theorem 1 we can see that B' is an α-classification consistent set in K, which contradicts that B is an α- classification reduct. It follows that $\wedge B$ is a prime implicant of the α-discernibility function f_K^{α}.

"$\Leftarrow$". If $\wedge B$ is a prime implicant of the α-discernibility function f_K^{α}. Then $B \bigcap D_{ij}^{\alpha} \neq \varnothing$ for all $D_{ij}^{\alpha} \in \mathrm{M}_0^{\alpha}$, and moreover, $\forall b \in B$, there exists $D_{ij}^{\alpha} \in \mathrm{M}_0^{\alpha}$ such that $B \bigcap D_{ij}^{\alpha} = \{b\}$. Consequently, $B - \{b\}$ is not an α-classification consistent set in K. Thus we conclude that B is an α-classification reduct. ∎

From the above theorem, we know if

$$f_K^{\alpha}\left(\overline{a}_1, \overline{a}_2, \ldots, \overline{a}_m\right) = \wedge\left\{\vee D_{ij}^{\alpha} : D_{ij}^{\alpha} \in \mathrm{M}_0^{\alpha}\right\} = \bigvee_{l=1}^{t}\left(\bigwedge_{q=1}^{s_l} \overline{a_{p_q}}\right),$$

where $\bigwedge_{q=1}^{s_l} \overline{a_{p_q}}$, $l \leq t$, are all the prime implicants of the α-discernibility function f_K^{α}, then $B_l = \left\{a_{p_q} : q \leq s_l\right\}$, $l \leq t$, are all the α-classification reducts in K.

In what follows, we shall write a_k instead of $\overline{a_k}$ without any confusion.

Example 5. In the interval-valued information system given in Table 1 and, under the given classification error $\alpha = 0.2$, we obtain the Boolean function:

$$f_K^{0.2}\left(a_1, a_2, \ldots, a_5\right) = \left(a_1 \vee a_2 \vee a_3 \vee a_4 \vee a_5\right) \wedge a_3 \wedge \left(a_1 \vee a_2 \vee a_5\right) \wedge a_5 \wedge \left(a_3 \vee a_4 \vee a_5\right)$$
$$\wedge \left(a_1 \vee a_2 \vee a_4 \vee a_5\right) \wedge \left(a_1 \vee a_2\right) \wedge \left(a_3 \vee a_4\right) \wedge a_3 \wedge \left(a_4 \wedge a_5\right) \wedge a_4.$$

After simplification (using the absorption laws) we obtain the prime implicants representation of the Boolean function as:

$$f_K^{0.2}(a_1, a_2, \ldots, a_5) = (a_1 \vee a_2) \wedge a_3 \wedge a_4 \wedge a_5$$

$$= (a_1 \wedge a_3 \wedge a_4 \wedge a_5) \vee (a_2 \wedge a_3 \wedge a_4 \wedge a_5).$$

Hence there are two 0.2-classification reducts in the system: $\{a_1, a_3, a_4, a_5\}$ and $\{a_2, a_3, a_4, a_5\}$.

If we instead construct a Boolean function by restricting the conjunction to run over only column i (instead of over all columns) in the α-discernibility matrix, we then set the so-called i α-discernibility function, denoted by f_i^α. That is,

$$f_i^\alpha(\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_m) = \bigwedge_{\{j: D_{ij}^\alpha \in \mathrm{M}_0^\alpha\}} (\vee D_{ij}^\alpha), \quad i = 1, 2, \ldots, I.$$

The set of all prime implicants of function f_i^α determines the set of all α-classification reducts of u_i in K. These α-classification reducts reveal the minimum amount of information needed to discern class u_i from all other classes which are not included in the α-tolerance classes of u_i. We summarize this into the following theorem without proof.

Theorem 4. *Let $K = (U, A)$ be an interval-valued information system, $u_i \in U$. Then an attribute subset $B \subseteq A$ is an α-classification reduct of u_i in K iff $\bigwedge_{a_k \in B} \overline{a_k}$ is a prime implicant of the i α-discernibility function f_i^α.*

Example 6. Under the given permissible misclassification rate $\alpha = 0.2$, we can obtain the Boolean function with respect to u_i, for $i = 1, 2, \ldots, 10$. The α-classification reduct of each class can then be calculated as follows.

Since

$$f_1^{0.2}(a_1, a_2, \ldots, a_5) = \underset{\{j: D_{1j}^{\alpha} \in \mathrm{M}_0^{\alpha}\}}{\wedge} (\vee D_{1j}^{0.2})$$

$$= (a_1 \vee a_2 \vee a_3 \vee a_4 \vee a_5) \wedge a_3 \wedge (a_1 \vee a_2 \vee a_5) \wedge a_5 \wedge (a_3 \vee a_4 \vee a_5) = a_3 \wedge a_5,$$

$\{a_3, a_5\}$ is the unique 0.2-classification reduct of u_1.

Similarly, since

$$f_2^{0.2}(a_1, a_2, \ldots, a_5) = \underset{\{j: D_{2j}^{\alpha} \in \mathrm{M}_0^{\alpha}\}}{\wedge} (\vee D_{2j}^{0.2})$$

$$= (a_1 \vee a_2 \vee a_3 \vee a_4 \vee a_5) \wedge (a_1 \vee a_2 \vee a_4 \vee a_5) \wedge (a_1 \vee a_2) = a_1 \vee a_2,$$

there are two 0.2-classification reducts of u_2 in the system: $\{a_1\}$ and $\{a_2\}$.

Likewise,

$f_3^{0.2}(a_1, a_2, \ldots, a_5) = a_3 \wedge a_5$, $re^{0.2}(u_3) = \{\{a_3, a_5\}\}$,

$f_4^{0.2}(a_1, a_2, \ldots, a_5) = a_3$, $re^{0.2}(u_4) = \{\{a_3\}\}$,

$f_5^{0.2}(a_1, a_2, \ldots, a_5) = a_1 \vee a_2$, $re^{0.2}(u_5) = \{\{a_1\}, \{a_2\}\}$,

$f_6^{0.2}(a_1, a_2, \ldots, a_5) = (a_1 \wedge a_3 \wedge a_4 \wedge a_5) \vee (a_2 \wedge a_3 \wedge a_4 \wedge a_5)$,

$re^{0.2}(u_6) = \{\{a_1, a_3, a_4, a_5\}, \{a_2, a_3, a_4, a_5\}\}$,

$f_7^{0.2}(a_1, a_2, \ldots, a_5) = (a_1 \wedge a_3) \vee (a_2 \wedge a_3) \vee (a_3 \wedge a_5)$,

$re^{0.2}(u_7) = \{\{a_1, a_3\}, \{a_2, a_3\}, \{a_3, a_5\}\}$,

$f_8^{0.2}(a_1, a_2, \ldots, a_5) = a_4$, $re^{0.2}(u_8) = \{\{a_4\}\}$,

$f_9^{0.2}(a_1, a_2, \ldots, a_5) = a_3 \wedge a_5$, $re^{0.2}(u_9) = \{\{a_3, a_5\}\}$,

$f_{10}^{0.2}(a_1, a_2, \ldots, a_5) = a_5$, $re^{0.2}(u_{10}) = \{\{a_5\}\}$.

6. Induction of classification rules

After an α -classification reduct B^{α} of class u_i has been calculated, classification knowledge corresponding to u_i hidden in the interval-valued information system may be discovered and expressed in the form of a α -classification rule of the following kind:

If $a_k(x) \in [l_i^k, u_i^k]$ *for all* $a_k \in B^{\alpha}$ *, then object* x *should be classified into one of the classes in* $S_B^{\alpha}(u_i)$ *within a permissible misclassification rate* α *.*

If the cardinality of $S_B^{\alpha}(u_i)$ is one, that is, $S_B^{\alpha}(u_i) = \{u_i\}$, then the rule is regarded as certain in the sense of α. In such a case, the class u_i can be discerned from other classes under the permissible misclassification rate α. A certain α -classification rule can then be derived and represented as:

If $a_k(x) \in [l_i^k, u_i^k]$ *for all* $a_k \in B^{\alpha}$*, then object* x *should be classified into class* u_i *within the permissible misclassification rate* α *.*

Obviously, with the increase of permissible misclassification rate, more certain rules from the system may be derived. If the cardinality of $S_B^{\alpha}(u_i)$ is not one, then the corresponding classification rules are referred to as uncertain or possible. A certain 0-classification rule is called a completely certain classification rule. In such a case, the corresponding class can be discerned without error.

Now we define two measures that reflect the separation accuracy of classes under a permissible misclassification rate. α - separated accuracy of the interval-valued information system K is defined as follows:

$$\eta^{\alpha} = \frac{\sum_{\{i: |S_A^{\alpha}(u_i)|=1, i \le I\}} 1}{I}.$$

That is, the α-separated accuracy is the ratio of the number of classes which can be correctly classified within the permissible misclassification rate α to the total number of classes.

The pairwise separated accuracy under the permissible misclassification rate α in the system K, denoted by γ^{α}, is defined as follows:

$$\gamma^{\alpha} = \frac{\sum_{\{(i,j): 1 \le i < j \le I, D_{ij}^{\alpha} \ne \varnothing\}} 1}{\sum_{i=1}^{I-1} i}.$$

It is easy to verify

$$0 \le \alpha \le \beta \le 1 \Rightarrow \eta^{\beta} \ge \eta^{\alpha}, \quad \gamma^{\beta} \ge \gamma^{\alpha}.$$

Thus, the greater the permissible misclassification rate, the higher is the (pairwise) separated accuracy.

Example 7. In the information system given by Table 1, within the given permissible misclassification rate $\alpha = 0.2$, and based on the α-classification reduct of each class provided in Example 6, all certain and uncertain classification rules hidden in the interval-valued information system can be discovered and expressed as follows:

Within the permissible misclassification rate $\alpha = 0.2$, the ***certain decision rules*** are:

r_1 (u_3) : If $a_3(x) \in [7.23, 10.27]$ and $a_5(x) \in [2.06, 2.79]$, then x should be classified into class u_3.

r_2 (u_6) : If $a_1(x) \in [2.29, 3.43]$, $a_3(x) \in [6.71, 8.81]$, $a_4(x) \in [3.30, 4.23]$, and

$a_5(x) \in [3.01, 3.84]$, then x should be classified into class u_6.

r_2' (u_6) : If $a_2(x) \in [2.60, 3.48]$, $a_3(x) \in [6.71, 8.81]$, $a_4(x) \in [3.30, 4.23]$, and $a_5(x) \in [3.01, 3.84]$, then x should be classified into class u_6.

r_3 (u_9) : If $a_3(x) \in [3.83, 5.31]$ and $a_5(x) \in [1.72, 2.34]$, then x should be classified into class u_9.

Within the permissible misclassification rate $\alpha = 0.2$, the ***uncertain decision rule*s** are:

r_4 (u_1) : If $a_3(x) \in [5.32, 7.23]$ and $a_5(x) \in [2.54, 3.12]$, then x should be classified into u_1 or u_7.

r_5 (u_2): If $a_1(x) \in [3.37, 4.75]$, then x should be classified into u_2 or u_5 or u_8.

r_5' (u_2): If $a_2(x) \in [3.43, 4.85]$, then x should be classified into u_2 or u_5 or u_8.

r_6 (u_4): If $a_3(x) \in [2.59, 3.93]$, then x should be classified into u_4 or u_{10}.

r_7 (u_5): If $a_1(x) \in [3.46, 5.35]$, then x should be classified into u_2 or u_5 or u_8.

r_7' (u_5): If $a_2(x) \in [3.37, 5.11]$, then x should be classified into u_2 or u_5 or u_8.

r_8 (u_7) : If $a_1(x) \in [2.22, 3.07]$ and $a_3(x) \in [4.37, 7.05]$, then x should be classified into u_1 or u_7.

r_8' (u_7) :If $a_2(x) \in [2.43, 3.32]$ and $a_3(x) \in [4.37, 7.05]$, then x should be classified into u_1 or u_7.

r_8'' (u_7) : If $a_3(x) \in [4.37, 7.05]$ and $a_5(x) \in [2.39, 3.20]$, then x should be classified into u_1 or u_7.

$r_9\ (u_8)$: If $a_4(x) \in [4.44, 6.91]$, then x should be classified into u_2 or u_5 or u_8.

$r_{10}\ (u_{10})$: If $a_5(x) \in [1.10, 1.84]$, then x should be classified into u_4 or u_{10}.

The 0.2-separated accuracy $\eta^{0.2}$ in the information system described in Table 1 is 0.3 and $\gamma^{0.2} = 40/45$. It can easily be computed that $\eta^0 = 0.1$, $\gamma^0 = 34/45$, $\eta^{0.5} = 0.5$, $\gamma^{0.5} = 42/45$, $\eta^{0.9} = 1$, $\gamma^{0.9} = 1$.

We can see that although the pairwise separated accuracy under a permissible misclassification rate α is high, the α-separated accuracy may be low. In general, $\eta^\alpha \geq \gamma^\alpha$.

7. Summary and conclusions

In this paper we have developed a general framework for mining of classification rules in interval-valued information systems. In the approach, an interval-valued information system is first converted from a real-valued decision table by means of a statistical method. Useful concepts related to rough set data analysis in interval-valued information systems have been proposed subsequently. The concept of α-misclassification rate is employed to compare different classes of objects. Under a given permissible misclassification rate α, α-classification reducts and α-classification core can be calculated. This is very important in classification tasks involving a large number of features, e.g. spectral bands in hyperspectral classification problems. After an effective reduction of dimensions has been achieved, minimal feature sets determining the classification can be found and knowledge hidden in the systems can be unraveled and expressed in the form of α-classification rules. Such an extension of rough set theory enables rough set models to analyze effectively real-valued data commonly encountered in real-life applications. The implementations of such a knowledge discovery system in other application domains remain to be investigated in further studies.

Acknowledgement

This work has been supported by the earmarked grant CUHK 4126/04H of the Hong Kong Research Grants Council.

References

[1] M.J. Beynon, Stability of continuous value discretisation: an application within rough set theory, International Journal of Approximate Reasoning 35 (1) (2004) 29-53.

[2] S. Bodjanova, Approximation of fuzzy concepts in decision making, Fuzzy Sets and Systems 85 (1997) 23-29.

[3] C.C. Chan, J.W. Grzymala-Busse, On the two local inductive algorithms: PRISM and LEM2, Foundations of Computing and Decision Sciences 19 (1994) 185-204.

[4] M.R. Chmielewski, J.W. Grzymala-Busse, Global discretization of continuous attributes as preprocessing for machine learning, International Journal of Approximate Reasoning 15 (4) (1996) 319-331.

[5] K. Dembczynski, S. Greco, R. Slowinski, Methodology of rough-set-based classification and sorting with hierarchical structure of attributes and criteria, Control and Cybernetics 31 (4) (2002) 891-920.

[6] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, International Journal of General Systems 17 (1990) 191-208.

[7] I. Duntsch, G. Gediga, Uncertainty measures of rough set prediction, Artificial Intelligence 106 (1998) 109-137.

[8] S. Greco, B. Matarazzo, R. Slowinski, Handling missing values in rough set analysis of multi-attribute and multi-criteria decision problems, in: N. Zhong, A. Skowron, S. Ohsuda (eds.), New Directions in Rough Sets, Data Mining, and Granular-Soft Computing, LNAI 1711, Springer-Verlag, Berlin, 1999, pp.146-157.

[9] S. Greco, M. Inuiguchi, R. Slowinski, Fuzzy rough sets and multiple-premise gradual decision rules, International Journal of Approximate Reasoning 41 (2) (2006) 179-211.

[10] S. Greco, B. Matarazzo, R. Slowinski, Rough sets theory for multicriteria decision analysis, European Journal of Operational Research 129 (2001) 1-47.

[11] J.W. Grzymala-Busse, On the unknown attribute values in learning from examples, in: Z.W. Ras, M. Zemankova (eds.), Proceedings of the ISMIS-91, 6th International Symposium on Methodologies for Intelligent Systems, LNAI 542, Springer-Verlag, New York, 1991, pp.368-377.

[12] J.W. Grzymala-Busse, J. Stefanowski, Three discretization methods for rule induction, International Journal of Intelligent Systems 16 (1) (2001) 29-38.

[13] J.W. Grzymala-Busse, LERS--a system for learning from examples based on rough sets, in: R. Slowinski (ed), Intelligent Decision Support--Handbook of Applications and Advances of the Rough Sets Theory, Dordrecht: Kluwer Academic Publishers, 1992, pp.3-18.

[14] J. Han, Y. Cai, N. Cercone, Data-driven discovery of quantitative rules in relational databases, IEEE Transactions on Knowledge and Data Engineering 5 (1) (1993) 29-40.

[15] T.P. Hong, T.T. Wang, S.L. Wang, Knowledge acquisition from quantitative data using the rough-set theory, Intelligent Data Analysis 4 (2000) 289-304.

[16] D. Kim, Data classification based on tolerant rough set, Pattern Recognition 34 (2001) 1613-1624.

[17] A. de Korvin, C. McKeegan, R. Kleyle, Knowledge acquisition using rough sets when membership values are fuzzy sets, Journal of Intelligent and Fuzzy Systems 6 (1998) 237-244.

[18] M. Kryszkiewicz, Rough set approach to incomplete information systems, Information Sciences 112 (1998) 39-49.

[19] M. Kryszkiewicz, Rules in incomplete information systems, Information Sciences 113 (1999) 271-292.

[20] M. Kryszkiewicz, Comparative study of alternative types of knowledge reduction in inconsistent systems, International Journal of Intelligent Systems 16 (2001) 105-120.

[21] Y. Leung, D. Li, Maximal consistent block technique for rule acquisition in incomplete information systems, Information Sciences 153 (2003) 85-106.

[22] Y. Leung, W.Z. Wu, W.X. Zhang, Knowledge acquisition in incomplete information systems: A rough set approach, European Journal of Operational Research 168 (1) (2006) 164-180.

[23] P.J. Lingras, Y.Y. Yao, Data mining using extensions of the rough set model., Journal of the American Society for Information Science 49 (1998) 415-422.

[24] J.S. Mi, Y. Leung, W.Z. Wu, An uncertainty measure in partition-based fuzzy rough sets, International Journal of General Systems 34 (2005) 77-90.

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

[26] J.S. Mi, W.X. Zhang, An axiomatic characterization of a fuzzy generalization of rough sets, Information Sciences 160 (1-4) (2004) 235-249.

[27] A. Mrozek, Rough sets and dependency analysis among attributes in computer implementations of expert's inference models, International Journal of Man-Machine Studies 30 (4) (1989) 457-473.

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

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

[30] Z. Pawlak, Rough classification, International Journal of Human-Computer Studies 51 (1999) 369-383.

[31] L. Polkowski, A. Skowron, Rough Sets in Knowledge Discovery 1: Methodology and Applications, 2: Applications, Physica-Verlag, Heidelberg, 1998.

[32] L. Polkowski, S. Tsumoto, T.Y. Lin, Rough Set Methods and Applications, Physica-Verlag, Heidelberg, 2000.

[33] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, New York, 1976.

[34] A. Skowron, Boolean reasoning for decision rules generation, in: J. Komorowski, Z. Ras (eds), Methodologies for Intelligent Systems, LNAI 689, Springer-Verlag, Berlin, 1993, pp.295-305.

[35] A. Skowron, C. Rauszer, The discernibility matrices and functions in information systems, in: R. Slowinski (ed), Intelligent Decision Support-Handbook of Applications and Advances of the Rough Sets Theory, Kluwer Academic Publishers, London, 1992, pp.331-362.

[36] D. Slezak, W. Ziarko, The investigation of the Bayesian rough set model, International Journal of Approximate Reasoning 40 (1-2) (2005) 81-91.

[37] R. Slowinski, D. Vanderpooten, Similarity relation as a basis for rough approximations, in: P.P. Wang (ed), Advances in Machine Intelligence and Soft-Computing vol.6, Duke University Press, Durham, 1997, pp 17-33.

[38] R. Slowinski, D. Vanderpooten, A generalized definition of rough approximations based on similarity, IEEE Transactions on Knowledge and Data Engineering 12 (2000) 331-336.

[39] J. Stefanowski, On rough set based approaches to induction of decision rules, in: L. Polkowski, A. Skowron (eds), Rough Sets in Knowledge Discovery 1, Physica-Verlag, Heidelberg, 1998, pp. 500-529.

[40] J. Stefanowski, D. Vanderpooten, A general two stage approach to rule induction for examples, in: W. Ziarko (ed), Rough Sets, Fuzzy Sets and Knowledge Discovery, Springer-Verlag, Berlin, 1994, pp. 317-325.

[41] W.Z. Wu, Y. Leung, W.X. Zhang, Connections between rough set theory and Dempster-Shafer theory of evidence, International Journal of General Systems 31 (4) (2002) 405-430.

[42] W.Z. Wu, J.S. Mi, W.X. Zhang, Generalized fuzzy rough sets, Information Sciences 151 (2003) 263-282.

[43] W.Z. Wu, W.X. Zhang, Neighborhood operator systems and approximations, Information Sciences 144 (2002) 201-217.

[44] W.Z. Wu, W.X. Zhang, Constructive and axiomatic approaches of fuzzy approximation operators, Information Sciences 159 (3-4) (2004) 233-254.

[45] W.Z. Wu, W.X. Zhang, H.Z. Li, Knowledge acquisition in incomplete fuzzy information systems via rough set approach, Expert Systems 20 (5) (2003) 280-286.

[46] R. Yasdi, Combining rough sets learning and neural learning: method to deal with uncertain and imprecise information, Neuralcomputing 7 (1) (1996) 61-84.

[47] Y.Y. Yao, Generalized rough set model, in: L. Polkowski, A. Skowron (eds.), Rough Sets in Knowledge Discovery 1, Physica-Verlag, Heidelberg, 1998, pp.286-318.

[48] Y.Y. Yao, Two views of the theory of rough sets in finite universes, International

Journal of Approximate Reasoning 15 (4) (1996) 291-317.

[49] Y.Y. Yao, Information granulation and rough set approximation, International Journal of Intelligent Systems 16 (1) (2001) 87-104.

[50] W.X. Zhang, J.S. Mi, W.Z. Wu, Approaches to knowledge reductions in inconsistent systems, International Journal of Intelligent Systems 18 (2003) 989-1000.

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

Table 1

An interval-valued information system

	a_1	a_2	a_3	a_4	a_5
u_1	[2.17, 2.86]	[2.45, 2.96]	[5.32, 7.23]	[3.21, 3.95]	[2.54, 3.12]
u_2	[3.37, 4.75]	[3.43, 4.85]	[7.24, 10.47]	[4.00, 5.77]	[3.24, 4.70]
u_3	[1.83, 2.70]	[1.78, 2.98]	[7.23, 10.27]	[2.96, 4.07]	[2.06, 2.79]
u_4	[1.35, 2.12]	[1.42, 2.09]	[2.59, 3.93]	[1.87, 2.62]	[1.67, 2.32]
u_5	[3.46, 5.35]	[3.37, 5.11]	[6.37, 10.28]	[3.76, 5.70]	[3.41, 5.28]
u_6	[2.29, 3.43]	[2.60, 3.48]	[6.71, 8.81]	[3.30, 4.23]	[3.01, 3.84]
u_7	[2.22, 3.07]	[2.43, 3.32]	[4.37, 7.05]	[2.66, 3.68]	[2.39, 3.20]
u_8	[2.51, 4.04]	[2.52, 4.12]	[7.12, 11.26]	[4.44, 6.91]	[3.06, 4.65]
u_9	[1.24, 2.00]	[1.35, 1.91]	[3.83, 5.31]	[2.13, 3.01]	[1.72, 2.34]
u_{10}	[1.00, 1.72]	[1.10, 1.82]	[3.58, 5.65]	[1.67, 2.53]	[1.10, 1.84]

Table 2

The classification errors for information system described in Table 1

α_{ij}	1	2	3	4	5	6	**7**	8	9	10
1	1	0	0	0	0	0.19	0.64	0	0	0
2	0	1	0	0	0.88	0.035	0	0.49	0	0
3	0	0	1	0	0	0	0	0	0	0
4	0	0	0	1	0	0	0	0	0.07	0.26
5	0	0.68	0	0	1	0	0	0.31	0	0
6	0.13	0.05	0	0	0	1	0.16	0	0	0
7	0.47	0	0	0	0	0.13	1	0	0	0
8	0	0.43	0	0	0.38	0	0.08	1	0	0
9	0	0	0	0.07	0	0	0	0	1	0.19
10	0	0	0	0.17	0	0	0	0	0.16	1

Table 3

The permissible misclassification rates for the information system described in Table 1

β_{ij}	1	2	3	4	5	6	**7**	8	9	10
1	1	0	0	0	0	0.19	0.64	0	0	0
2	0	1	0	0	0.88	0.05	0	0.49	0	0
3	0	0	1	0	0	0	0	0	0	0
4	0	0	0	1	0	0	0	0	0.07	0.26
5	0	0.88	0	0	1	0	0	0.38	0	0
6	0.19	0.05	0	0	0	1	0.16	0	0	0
7	0.64	0	0	0	0	0.16	1	0	0	0
8	0	0.49	0	0	0.38	0	0.08	1	0	0
9	0	0	0	0.07	0	0	0	0	1	0.19
10	0	0	0	0.26	0	0	0	0	0.19	1

Table 4

The 0.2-discernibility set for the information system given by Table 1

	u_1	u_2	u_3	u_4	u_5	u_6	u_7	u_8	u_9	u_{10}
u_1										
u_2	$a_1\ a_2\ a_3\ a_4\ a_5$									
u_3	a_3	$a_1\ a_2\ a_4\ a_5$								
u_4	$a_1\ a_2\ a_3\ a_4\ a_5$	$a_1\ a_2\ a_3\ a_4\ a_5$	$a_3\ a_4$							
u_5	$a_1\ a_2\ a_5$		$a_1\ a_2\ a_5$	$a_1\ a_2\ a_3\ a_4\ a_5$						
u_6	a_5	$a_1\ a_2$	a_5	$a_1\ a_2\ a_3\ a_4\ a_5$	$a_1\ a_2$					
u_7		$a_1\ a_2\ a_3\ a_4\ a_5$	a_3	$a_1\ a_2\ a_3\ a_4\ a_5$	$a_1\ a_2\ a_4\ a_5$	a_3				
u_8	$a_3\ a_4\ a_5$		$a_4\ a_5$	$a_1\ a_2\ a_3\ a_4\ a_5$		a_4	$a_3\ a_4\ a_5$			
u_9	$a_1\ a_2\ a_3\ a_4\ a_5$	$a_1\ a_2\ a_3\ a_4\ a_5$	$a_3\ a_4$	a_3	$a_1\ a_2\ a_3\ a_4\ a_5$	$a_1\ a_2\ a_3\ a_4\ a_5$	$a_1\ a_2\ a_5$	$a_1\ a_2\ a_3\ a_4\ a_5$		
u_{10}	$a_1\ a_2\ a_3\ a_4\ a_5$	$a_1\ a_2\ a_3\ a_4\ a_5$	$a_1\ a_2\ a_3\ a_4\ a_5$		$a_1\ a_2\ a_3\ a_4\ a_5$	$a_1\ a_2\ a_3\ a_4\ a_5$	$a_1\ a_2\ a_4\ a_5$	$a_1\ a_2\ a_3\ a_4\ a_5$	a_5	