Chapter 3 Variable Precision Rough Sets

The variable precision rough sets model is an extension of the original rough set model, which was proposed to analyze and identify data patterns that represent statistical trends rather than being functional. Compared to the original rough set model, VPRS introduces 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 (1993) defined the β value as a classification error and the range in the domain $[0.0, 0.5)$. However, An et al. (1996) and Beynon (2001) 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's notion.

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.

3.1 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 $R^* = \{E_1, E_{2,} ..., E_n\}$. We will assume that all sets under consideration are finite and non-empty (Ziarko, 2002). 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 $P \subseteq C$:

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

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

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

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

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

Quality of classification

Based on Ziarko (1993), the measure of quality of classification for the VPRS model is defined as:

$$\gamma(P, D, \beta) = \frac{card(\bigcup_{1-P_r(Z|x_i)\le\beta} \{x_i \in E(P)\})}{card(U)}, \tag{3.1}$$

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.

3.2 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 (Lin, et al. 1997):

(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 (Ziarko, 1993):

(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 $U=\{x_1, x_x, ..., x_n\}$. 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:

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

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 (Pawlak, 1991), i.e.

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

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 (Walczak, et al. 1999).

3.3 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 (1993) 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.