A Heuristic Algorithm for Generating Decision Rules in Variable Precision Rough Set Models

A Heuristic Algorithm for Generating Decision Rules in Variable Precision Rough Set Models

著者	KUDO Yasuo, MURAI Tetsuya
journal or publication title	Proceedings of Joint 5th International Conference on Soft Computing and Intelligent Systems and 11th International Symposium on Advanced Intelligent Systems (SCIS & ISIS 2010)
page range	1568-1573
year	2010-12
URL	http://hdl.handle.net/10258/2198

doi: info:doi/10.14864/softscis.2010.0.1568.0

A Heuristic Algorithm for Generating Decision Rules in Variable Precision Rough Set Models

Yasuo Kudo
College of Information and Systems
Muroran Institute of Technology
27-1 Mizumoto, Muroran 050-8585, Japan
Email: kudo@csse.muroran-it.ac.jp

Tetsuya Murai
Graduate School of Information Science and Technology
Hokkaido University
Kita 14, Nishi 9, Kita-ku, Sapporo 060-0814, Japan
Email: murahiko@main.ist.hokudai.ac.jp

***Abstract*—In this paper, we propose a heuristic algorithm to extract decision rules based on variable precision rough set models (VPRS models). The VPRS models provides a theoretical basis of regarding probabilistic / inconsistent information in the framework of rough set theory. The main idea of our algorithm is based on construction of suitable β-lower approximations by giving up to discern some discernible objects that belong to different decision classes each other. All decision rules extracted by our algorithm are guaranteed that the certainty of all extracted decision rules are equal to or higher than the predefined threshold of certainty.**

I. Introduction

Extraction of decision rules is an important application of rough set theory [4], [5] from a viewpoint of data analysis. Variable precision rough set models (for short, VPRS models) proposed by Ziarko [8] provides a theoretical basis of regarding probabilistic / inconsistent information in the framework of rough set theory.

In this paper, we propose a heuristic algorithm to extract decision rules based on the VPRS models. The main idea of our algorithm is based on construction of suitable β-lower approximations by giving up to discern some discernible objects that belong to different decision classes each other. All decision rules extracted by our algorithm are guaranteed that the certainty of all extracted decision rules are equal to or higher than the predefined threshold of certainty.

The rest of this paper is organized as follows. In Section II, we review Pawlak's rough set theory and the VPRS models as the background of this paper. In Section III, we introduce a heuristic algorithm to extract decision rules based on the VPRS models, and describe small examples to explain how the proposed algorithm works. We discuss a few properties of the proposed algorithm in Section IV and finally conclude this paper in Section V.

II. Rough Sets

In this section, we review the foundations of rough set theory as background for this paper. The contents of this section are based on [5], [6], [8].

A. Lower and Upper Approximations in Decision Tables

In rough set data analysis, objects as targets of analysis are illustrated by combination of multiple attributes and those values, and represented by the following decision table:

$$DT = (U, C, d), \tag{1}$$

where U is the set of objects, C is the set of condition attributes such that each attribute $a \in C$ is a function $a : U \to V_a$ from U to the value set V_a of a, and d is a function $d : U \to V_d$ called the decision attribute.

The indiscernibility relation R_B on U with respect to a subset $B \subseteq C$ is defined by

$$xR_By \Longleftrightarrow a(x) = a(y),\ \forall a \in B. \tag{2}$$

The equivalent class $[x]_B$ of $x \in U$ by R_B is the set of objects that are not discernible with x even though using all attributes in B. Any indiscernibility relation provides a partition of U. We denote the partition of U by R_B, i.e., the quotient set by R_B, by U/R_B. In particular, the partition $\mathcal{D} = \{D_1, \cdots, D_m\}$ provided by the indiscernibility relation R_d with respect to the decision attribute d is called the set of decision classes.

For any decision class D_i $(1 \le i \le m)$, the lower approximation $\underline{B}(D_i)$ and the upper approximation $\overline{B}(D_i)$ of D_i with respect to the indiscernibility relation R_B are defined as follows, respectively:

$$\underline{B}(D_i) = \{x \in U \mid [x]_B \subseteq D_i\}, \tag{3}$$
$$\overline{B}(D_i) = \{x \in U \mid [x]_B \cap D_i \neq \emptyset\}. \tag{4}$$

Note that a decision table is called consistent if and only if $\underline{C}(D_i) = D_i = \overline{C}(D_i)$ holds for all decision classes $D_i \in \mathcal{D}$.

Table I is an example of a decision table used in [2] and this decision table consists of the set of objects $U = \{x_1, \cdots, x_6\}$, the set of condition attributes $C = \{c_1, \cdots, c_6\}$ and the decision attribute d. For example, an attribute c_1 is a function $c_1 : U \to \{0, 1\}$, and the value of an object $x_1 \in U$ at c_1 is 1, that is, $c_1(x_1) = 1$. The decision attributed d provides the following two decision classes; $D_1 = \{x_1, x_2, x_3\}$, and $D_2 = \{x_4, x_5, x_6, x_7\}$. Note that this table is not consistent because $\underline{C}(D_1) = \{x_1, x_3\}$ and $\overline{C}(D_1) = \{x_1, x_2, x_3, x_5, x_7\}$, and $\underline{C}(D_1) = D_1 = \overline{C}(D_1)$ does not hold.

B. Decision Rules

We denote a decision rule constructed from a subset $B \subseteq C$ of condition attribute, the decision attribute d and an object

TABLE I
AN EXAMPLE OF DECISION TABLE

U	c_1	c_2	c_3	c_4	c_5	c_6	d
x_1	1	1	1	1	1	1	M
x_2	1	0	1	0	1	1	M
x_3	0	0	1	1	0	0	M
x_4	1	1	1	0	0	1	F
x_5	1	0	1	0	1	1	F
x_6	0	0	0	1	1	0	F
x_7	1	0	1	0	1	1	F

TABLE II
THE DISCERNIBILITY MATRIX OF TABLE I

	x_1	x_2	x_3	x_4	$\cdots$
x_1	$\emptyset$				
x_2	$\emptyset$	$\emptyset$			
x_3	$\emptyset$	$\emptyset$	$\emptyset$		
x_4	$\{c_4, c_5\}$	$\{c_2, c_5\}$	$\{c_1, c_2, c_4, c_6\}$	$\emptyset$	
x_5	$\{c_2, c_4\}$	$\emptyset$	$\{c_1, c_4, c_5, c_6\}$	$\emptyset$	
x_6	$\{c_1, c_2, c_3, c_6\}$	$\{c_1, c_3, c_4, c_6\}$	$\{c_3, c_5\}$	$\emptyset$	
x_7	$\{c_2, c_4\}$	$\emptyset$	$\{c_1, c_4, c_5, c_6\}$	$\emptyset$	

$x \in U$ by $(B, x) \to (d, x)$. The concepts of certainty and coverage are well-known criteria for evaluating decision rules, however, we only use the certainty in this paper. For any decision rule $(B, x) \to (d, x)$, the score $Cer(\cdot)$ of certainty of the decision rule is defined by

$$Cer((B, x) \to (d, x)) = \frac{|[x]_B \cap D_i|}{|[x]_B|}, \tag{5}$$

where $|X|$ is the cardinality of the set X and D_i is the decision class such that $x \in D_i$.

For example, a decision rule $(B, x_7) \to (d, x_7)$ constructed from a set $B = \{c_1, c_2\}$, the decision attribute d and an object $x_7 \in U$ has actually the following form:

$$(c_1 = 1) \wedge (c_2 = 0) \to (d = \mathrm{F}),$$

and its certainty is $\frac{2}{3}$.

C. Discernibility Matrices

The discernibility matrix [7] is generally used for computing all relative reducts in the given decision table. Let DT be a decision table with $|U|$ objects, where $|U|$ is the cardinality of U. The discernibility matrix DM of DT is a symmetric $|U| \times |U|$ matrix whose element at i-th row and j-th column is the following set of condition attributes to discern between two objects x_i and x_j. Each element $a \in \delta_{ij}$ represents that x_i and x_j are discernible by checking the value of a:

$$\delta_{ij} = \begin{cases} \{a \in C \mid a(x_i) \neq a(x_j)\}, & \text{if } d(x_i) \neq d(x_j) \text{ and} \\ & \{x_i, x_j\} \cap POS_C(\mathcal{D}) \neq \emptyset, \\ \emptyset, & \text{otherwise}, \end{cases} \tag{6}$$

where $POS_B(\mathcal{D})$ is the positive region of $\mathcal{D}$ by $B \subseteq C$ and defined by

$$POS_B(\mathcal{D}) = \bigcup_{D_i \in \mathcal{D}} \underline{B}(D_i). \tag{7}$$

Table II is the discernibility matrix of the decision table presented by Tab. I. Note that we omit upper triangular components of the discernibility matrix and the columns of x_5, x_6, and x_7 in Table II.

D. Variable Precision Rough Set Models

VPRS models [8] generalize Pawlak's rough set models by generalizing the notion of the standard set inclusion, and provide a theoretical basis for dealing with inconsistent information in the framework of rough sets. Suppose that a decision table (U, C, d) is given. For any sets $X, Y \subseteq U$ of objects, the measure $c(X, Y)$ of the relative degree of misclassification of the set X with respect to the set Y is defined by

$$c(X, Y) = \begin{cases} 1 - \dfrac{|X \cap Y|}{|X|}, & \text{if } |X| > 0, \\ 0, & \text{if } |X| = 0. \end{cases} \tag{8}$$

The relative degree $c(X, Y)$ represents that if we were to classify all objects of X into Y, then the misclassification error ratio would be $c(X, Y) \times 100\%$. It is easy to confirm that the following property holds for any sets $X, Y \subseteq U$:

$$X \subseteq Y \Longleftrightarrow c(X, Y) = 0. \tag{9}$$

Thus, by setting an admissible classification error ratio, called a precision β $(0 \leq \beta < 0.5)$, the set inclusion is generalized by

$$X \overset{\beta}{\subseteq} Y \Longleftrightarrow c(X, Y) \leq \beta. \tag{10}$$

Let R_B be an indiscernibility relation with respect to $B \subseteq C$, and U/R_B be the quotient set based on R_B. For each decision class D_i, the β-lower approximation $\underline{B}_\beta(D_i)$ and the β-upper approximation $\overline{B}_\beta(D_i)$ with respect to R_B are introduced by

$$\underline{B}_\beta(D_i) = \bigcup \{[x]_B \in U/R_B \mid [x]_B \overset{\beta}{\subseteq} D_i\} \tag{11}$$

$$= \{x \in U \mid c([x]_B, D_i) \leq \beta\}, \tag{12}$$

$$\overline{B}_\beta(D_i) = \{x \in U \mid c([x]_B, D_i) < 1 - \beta\}. \tag{13}$$

It is easy to confirm that $\underline{B}_0(D_i) = \underline{B}(D_i)$ and $\overline{B}_0(D_i) = \overline{B}(D_i)$ hold, i.e., the β-lower (upper) approximation is identical to Pawlak's lower (upper) approximation in the case of $\beta = 0$.

Similar to the case of Pawlak's rough sets, the β-positive region of D_i by $B \subseteq C$ is defined by

$$POS_B^\beta(\mathcal{D}) = \bigcup_{D_i \in \mathcal{D}} \underline{B}_\beta(D_i). \tag{14}$$

III. A HEURISTIC ALGORITHM FOR GENERATING DECISION RULES IN VPRS MODELS

In this section, we propose a heuristic algorithm for generating decision rules in VPRS models and show a small example to explain how the proposed algorithm works.

A. Main Idea and Motivation

For generating decision rules with some exceptions in VPRS models, the main idea of our algorithm that we propose later is to give up discerning some discernible objects that belong to different decision classes each other This idea is based on the following motivations:

- In some cases, by giving up discerning some discernible objects, we can generate decision rules with shorter antecedents.
- Even though the given decision table is consistent, giving up discerning some discernible objects may enable us to present concise decision rules that ignore small differences between objects as exceptions.

Thus, in the algorithm we propose later, we intend to generate decision rules that satisfy the following constraints:

1) The certainty of all generated decision rules is at least equal to or higher than the given precision $\beta \in [0, 0.5)$ in the VPRS model.
2) The length of antecedents of generated decision rules are as short as possible.

B. Algorithm for Generating Decision Rules

Based on the idea and motivations described in the previous subsection, we introduce a heuristic algorithm for generating decision rules in VPRS models. For any decision table $DT = (U, C, d)$ and any precision $\beta \in [0, 0.5)$, our algorithm guarantees that the certainty of each generated rule is at least equal to or higher than $1 - \beta$.

Algorithm 1 consists of mainly the following three components:

1) Calculation of certainty for selecting objects to stop discerning (Steps. 4–15).
2) Elimination of condition attributes that are not using for generating decision rules (Steps. 18–22).
3) Construction of decision rules (Steps. 24–30).

Algorithm 1 Decision rules generation algorithm

Input: decision table $DT = (U, C, d)$, precision $\beta \in [0, 0.5)$
Output: set of decision rules $Rules$
1: $Rules = \emptyset, Cond = C$
2: Compute the discernibility matrix DM of DT
3: Compute the complement of β-positive region of $\mathcal{D}$ by C, i.e., $U - POS_C^{\beta}(\mathcal{D})$
4: **for all** $\delta_{ij} \in DM$ such that $i > j$ **do**
5:   **if** $\delta_{ij} \neq Cond$ **and** $\delta_{ij} \neq \emptyset$ **then**
6:     $Ign_{ij} = \{x_k \in U | \delta_{kl} \in DM, \delta_{kl} \subseteq \delta_{ij}\} \cup (U - POS_C^{\beta}(\mathcal{D}))$
7:     $\mathcal{E}_{ij} = \{[x_k]_{Cond - \delta_{ij}} \mid x_k \in Ign_{ij}\}$
8:     **for all** $[x_k] \in \mathcal{E}_{ij}$ **do**
9:       $Cer_{ij}^{k} = \max_{D_m \in \mathcal{D}} \frac{|[x_k] \cap D_m|}{|[x_k]|}$
10:     **end for**
11:     $Cer_{ij} = \min_k Cer_{ij}^{k}$
12:   **else**
13:     $Cer_{ij} = 0$
14:   **end if**
15: **end for**
16: $Cer^* = \max Cer_{ij}$
17: **if** $Cer^* \geq 1 - \beta$ **then**
18:   $F = \{\delta_{ij} \in DM \mid Cer^* = Cer_{ij}\}$
19:   Select one $\delta^* \in F$ such that $|\delta^*| \geq |\delta|, \forall \delta \in F$
20:   $Cond := Cond - \delta^*$
21:   Remove all attributes in δ^* from DM
22:   Go back to Step. 4
23: **else**
24:   Construct a quotient set U/R_{Cond}
25:   **for all** $D_m \in \mathcal{D}$ **do**
26:     Construct the β-lower approximation $\underline{Cond}_{\beta}(D_m)$
27:     **for all** $[x] \in U/R_{Cond}$ such that $[x] \cap \underline{Cond}_{\beta}(D_m) \cap D_m \neq \emptyset$ **do**
28:       $Rules := Rules \cup \{(Cond, x) \rightarrow (d, x)\}$
29:     **end for**
30:   **end for**
31: **end if**
32: **return** $Rules$

To guarantee the minimum certainty of generated decision rules, we need to carefully select objects to stop discerning. Thus, for each pair of discernible objects in different decision classes, we need to check the influence of stopping these two objects. In Steps. 4–15, the score Cer_{ij} of each element δ_{ij} that corresponds to the pair of objects x_i and x_j means that, if we would stop discerning x_i and x_j, the minimum certainty of generated decision rules would be Cer_{ij}. Consequently, we have to select a pair x_i and x_j that satisfies the condition $Cer_{ij} \geq 1 - \beta$ in Steps. 18–22, and we have to stop this selection if all scores of certainty are less than $1 - \beta$ because further selection causes decrease of the minimum certainty of decision rules below to the threshold $1 - \beta$.

Moreover, to generate decision rules with as short antecedents as possible, we also need to decrease the number of condition attributes using for generating decision rules in Steps. 18–22. Thus, for eliminating as many condition attribute as possible without decreasing the certainty of decision rules below to the threshold $1 - \beta$, we should select the longest element, i.e., the number of contained condition attributes is the biggest, among the elements of the discernibility matrix with the highest certainty score.

Finally, in Steps. 24–30, we need to use objects that are not regarded as exceptions to generate decision rules that satisfy the condition $Cer(\cdot) \geq 1 - \beta$. As such the suitable objects, we select objects that belong to both a decision class D_m and its β-lower approximation $\underline{Cond}_{\beta}(D_m)$ by the set of selected condition attributes $Cond$. It is easily confirmed that all the generated decision rules by our algorithm satisfy the condition $Cer(\cdot) \geq 1 - \beta$.

C. Small Examples

In this subsection, we show two small examples to explain how the proposed algorithm works for consistent decision table and inconsistent decision table.

TABLE III
THE SCORE Cer_{ij} OF EACH ELEMENT δ_{ij} IN TAB. II

	x_1	x_2	x_3
x_4	0.5	0.75	0.5
x_5	0.5	0	0.5
x_6	0.67	0.75	0.67
x_7	0.5	0	0.5

1) Case of Inconsistent Decision Tables: Let DT be the decision table presented by Tab. I, DM be the discernibility matrix of DT presented by Tab. II, and $\beta = 0.3$. First, we construct the β-unpredictable region by the set of all condition attributes C. For both $D_1 = \{x_1, x_2, x_3\}$ and $D_2 = \{x_4, x_5, x_6, x_7\}$, only an equivalence class $[x_2]_C = \{x_2, x_5, x_7\} \in U/R_C$ does not satisfy the condition of β-lower approximation (11). Then, the complement of the β-positive region of $\mathcal{D}$ by C is

$$U - POS_C^{\beta}(\mathcal{D}) = \{x_2, x_5, x_7\}.$$

Table III presents the score Cer_{ij} of each non-empty element δ_{ij} in Tab. II. Note that all the omitted scores in Tab III are 0. For example, the value 0.5 that corresponds to objects x_7 and x_3 means that Cer_{73}, i.e., the score of $\delta_{73} = \{c_1, c_4, c_5, c_6\}$, is 0.5 and is calculated as follows; First, we construct the set $Ign_{73} = \{x_2, x_4, x_5, x_7\}$ because δ_{41}, δ_{53}, and δ_{73} itself are included in δ_{73} and $x_2 \in U - POS_C^{\beta}(\mathcal{D})$ holds.

Then, using the set $Cond - \delta_{73} = \{c_2, c_3\}$, we have a set $\mathcal{E}_{73}$ of equivalence classes that consists of

$$[x_2] = \{x_2, x_3, x_5, x_7\},\ [x_4] = \{x_1, x_4\}.$$

This meant that we could not discern any objects in $[x_2]$ (similarly, any objects in $[x_4]$) if we stopped discerning the two objects x_7 and x_3 by rejecting all condition attributes in δ_{73}. Finally, we calculate the following certainty scores of $[x_2]$ and $[x_4]$:

$$\begin{aligned} Cer_{73}^2 &= \frac{|[x_2] \cap D_2|}{|[x_2]|} = \frac{1}{2}, \\ Cer_{73}^4 &= \frac{|[x_4] \cap D_1|}{|[x_4]|} = \frac{1}{2}, \end{aligned}$$

and consequently we have the score $Cer_{73} = 0.5$ of δ_{73}. This score indicates that the minimum certainty of decision rules would be 0.5 if we construct decision rules by the set $Cond - \delta_{73} = \{c_2, c_3\}$.

Here, the highest score in Tab. III is 0.75 of $\delta_{42} = \{c_2, c_5\}$ and $\delta_{62} = \{c_1, c_3, c_4, c_6\}$ and we select δ_{62} because $|\delta_{62}| = 4 > 2 = |\delta_{42}|$ holds. This selection corresponds to give up discerning the objects x_6 and x_2. Then, we revise the set $Cond$ as follows:

$$Cond := C - \delta_{62} = \{c_2, c_5\}.$$

We also remove all condition attributes in δ_{62} from Tab. II.

Table IV and Tab. V present the revised discernibility matrix and the scores of non-empty elements in Tab IV, respectively. From the scores in Tab. V, further selection of objects to stop discerning can not provide decision rules with $Cer(\cdot) \geq 1 - \beta$, and therefore we finish the selection of condition attributes and fix the set $Cond = \{c_2, c_5\}$.

TABLE IV
THE REVISED DISCERNIBILITY MATRIX

	x_1	x_2	x_3	x_4	$\cdots$
x_1	$\emptyset$				
x_2	$\emptyset$	$\emptyset$			
x_3	$\emptyset$	$\emptyset$	$\emptyset$		
x_4	$\{c_5\}$	$\{c_2, c_5\}$	$\{c_2\}$	$\emptyset$	
x_5	$\{c_2\}$	$\emptyset$	$\{c_5\}$	$\emptyset$	
x_6	$\{c_2\}$	$\emptyset$	$\{c_5\}$	$\emptyset$	
x_7	$\{c_2\}$	$\emptyset$	$\{c_5\}$	$\emptyset$	

TABLE V
THE SCORE Cer_{ij} OF EACH ELEMENT δ_{ij} IN TAB. IV

	x_1	x_2	x_3
x_4	0.5	0	0.5
x_5	0.5	0	0.5
x_6	0.5	0	0.5
x_7	0.5	0	0.5

Finally, for generating decision rules from $Cond = \{c2, c5\}$, we construct the quotient set U/R_{Cond} and the β-lower approximation of each decision class. Equivalence classes in U/R_{Cond} are

$$\begin{aligned} &[x_1]_{Cond} = \{x_1\},\ [x_2]_{Cond} = \{x_2, x_5, x_6, x_7\}, \\ &[x_3]_{Cond} = \{x_3\},\ [x_4]_{Cond} = \{x_4\}. \end{aligned}$$

The β-lower approximations of decision classes are

$$\begin{aligned} \underline{Cond}_{\beta}(D_1) &= \{x_1, x_3\}, \\ \underline{Cond}_{\beta}(D_2) &= \{x_2, x_4, x_5, x_6, x_7\}. \end{aligned}$$

Consequently, we get the set of decision rules $Rules$ that consists of the following four decision rules such that $Cer(\cdot) \geq 1 - \beta = 0.7$:

- $(c2 = 1) \wedge (c5 = 1) \rightarrow (d = \mathrm{M})$, Certainty = 1.
- $(c2 = 0) \wedge (c5 = 0) \rightarrow (d = \mathrm{M})$, Certainty = 1.
- $(c2 = 0) \wedge (c5 = 1) \rightarrow (d = \mathrm{F})$, Certainty = 0.75.
- $(c2 = 1) \wedge (c5 = 0) \rightarrow (d = \mathrm{F})$, Certainty = 1.

Note that these rules are based on giving up discerning two discernible elements x_6 and x_2 and $x_2 \in D_1$ is regarded as an exception of $\underline{Cond}_{\beta}(D_2)$.

2) Case of Consistent Decision Tables: Let $DT2$ be a consistent decision table presented by Tab. VI. There is just one difference between Tab. I and Tab. VI, i. e., the value of the object x_2 at the attribute c_1 and this difference enable us to discern objects in $D_1 = \{x_1, x_2, x_3\}$ and $D_2 = \{x_4, x_5, x_6, x_7\}$ completely. Tab. VII presents the discernibility matrix, denoted by $DM2$, of $DT2$. Differences between Tab. II and Tab. VII appear in elements δ_{42}, δ_{52}, δ_{62}, and δ_{72}. Similar to the case of inconsistent decision table, let the precision be $\beta = 0.3$.

Because Tab. VI is consistent, it is clear that all equivalence classes in U/R_C satisfy the condition of β-lower approximation (11) and the complement of the β-positive region of $\mathcal{D}$ by C is empty, i.e., $U - POS_C^{\beta}(\mathcal{D}) = \emptyset$.

TABLE VI
AN EXAMPLE OF CONSISTENT DECISION TABLE

U	c_1	c_2	c_3	c_4	c_5	c_6	d
x_1	1	1	1	1	1	1	M
x_2	0	0	1	0	1	1	M
x_3	0	0	1	1	0	0	M
x_4	1	1	1	0	0	1	F
x_5	1	0	1	0	1	1	F
x_6	0	0	0	1	1	0	F
x_7	1	0	1	0	1	1	F

TABLE VII
THE DISCERNIBILITY MATRIX OF TABLE VI

	x_1	x_2	x_3	x_4	$\cdots$
x_1	$\emptyset$				
x_2	$\emptyset$	$\emptyset$			
x_3	$\emptyset$	$\emptyset$	$\emptyset$		
x_4	$\{c_4, c_5\}$	$\{c_1, c_2, c_5\}$	$\{c_1, c_2, c_4, c_6\}$	$\emptyset$	
x_5	$\{c_2, c_4\}$	$\{c_1\}$	$\{c_1, c_4, c_5, c_6\}$	$\emptyset$	
x_6	$\{c_1, c_2, c_3, c_6\}$	$\{c_3, c_4, c_6\}$	$\{c_3, c_5\}$	$\emptyset$	
x_7	$\{c_2, c_4\}$	$\{c_1\}$	$\{c_1, c_4, c_5, c_6\}$	$\emptyset$	

Table VIII presents the score Cer_{ij} of each non-empty element δ_{ij} in Tab. VII by computing the score Cer_{ij} of each element δ_{ij} in $DM2$ with the same procedure of the case in Tab. III. Note that all the omitted scores in Tab VIII are 0.

Here, the highest score in Tab. VIII is 0.75 of $\delta_{42} = \{c_2, c_5\}$ and we select δ_{42}, which corresponds to give up discerning the objects x_4 and x_2. Then, we revise the set $Cond$ as follows:

$$Cond := C - \delta_{42} = \{c_3, c_4, c_6\}.$$

We also remove all condition attributes in δ_{42} from Tab. VII.

Table IX and Tab. X present the revised discernibility matrix and the scores of non-empty elements in Tab IX, respectively. After this revision, however, further selection of objects to stop discerning can not provide decision rules with $Cer(\cdot) \geq 1-\beta$, and therefore we finish the selection of condition attributes and fix the set $Cond = \{c_3, c_4, c_6\}$.

Finally, for generating decision rules from $Cond = \{c4\}$, we construct the quotient set U/R_{Cond} and the β-lower approximation of each decision class. Equivalence classes in U/R_{Cond} are

$$[x_1]_{Cond} = \{x_1\},\ [x_2]_{Cond} = \{x_2, x_4, x_5, x_7\},$$
$$[x_3]_{Cond} = \{x_3\},\ [x_6]_{Cond} = \{x_6\}.$$

The β-lower approximations of decision classes are

$$\underline{Cond}_\beta(D_1) = \{x_1, x_3\},$$
$$\underline{Cond}_\beta(D_2) = \{x_2, x_4, x_5, x_6, x_7\}.$$

Consequently, we get the set of decision rules $Rules$ that consists of the following four decision rules such that $Cer(\cdot) \geq 1 - \beta = 0.7$:

- $(c3 = 1) \wedge (c4 = 1) \wedge (c6 = 1) \to (d = \mathrm{M})$, Certainty = 1.
- $(c3 = 1) \wedge (c4 = 1) \wedge (c6 = 0) \to (d = \mathrm{M})$, Certainty = 1.
- $(c3 = 1) \wedge (c4 = 0) \wedge (c6 = 1) \to (d = \mathrm{F})$, Certainty = 0.75.

TABLE VIII
THE SCORE Cer_{ij} OF EACH ELEMENT δ_{ij} IN TAB. VII

	x_1	x_2	x_3
x_4	0.5	0.75	0.5
x_5	0.67	0.67	0.5
x_6	0.67	0.5	0.5
x_7	0.67	0.67	0.5

TABLE IX
THE DISCERNIBILITY MATRIX OF TABLE VI

	x_1	x_2	x_3	x_4	$\cdots$
x_1	$\emptyset$				
x_2	$\emptyset$	$\emptyset$			
x_3	$\emptyset$	$\emptyset$	$\emptyset$		
x_4	$\{c_4\}$	$\emptyset$	$\{c_4, c_6\}$	$\emptyset$	
x_5	$\{c_4\}$	$\emptyset$	$\{c_4, c_6\}$	$\emptyset$	
x_6	$\{c_3, c_6\}$	$\{c_3, c_4, c_6\}$	$\{c_3\}$	$\emptyset$	
x_7	$\{c_4\}$	$\emptyset$	$\{c_4, c_6\}$	$\emptyset$	

- $(c3 = 0) \wedge (c4 = 1) \wedge (c6 = 0) \to (d = \mathrm{F})$, Certainty = 1.

Note that these rules are based on giving up discerning two discernible elements x_4 and x_2 and $x_2 \in D_1$ is regarded as an exception of $\underline{Cond}_\beta(D_2)$.

IV. DISCUSSION

As we described in Sec. III-A, the main idea of this paper is to give up discerning some discernible objects that belong to different decision classes each other. The examples presented in the previous section indicate that this idea may enable us to generated decision rules such that the certainty of each generated rule is at least equal to or higher than $1 - \beta$. In particular, the example of the case of consistent decision table indicate the possibility of generating decision rules with some exceptions from consistent decision tables. Thus, combining the proposed algorithm to a heuristic attribute reduction algorithm based on generating reduced decision tables [3], it is possible to generated decision rules with some exceptions from decision tables with numerous condition attributes.

Here, we consider the relationship between the set $Cond$ used for constructing decision rules at Steps. 24–30 in Algorithm 1 and β-reducts proposed by Beynon [2]. β-reducts are based on the quality of classification defined by

$$\gamma_B^\beta(\mathcal{D}) = \frac{\sum_{D_i \in \mathcal{D}} |\underline{B}_\beta(D_i)|}{|U|}. \tag{15}$$

Formally, a β-reduct is a set of condition attributes $A \subseteq C$ that satisfies the following two conditions:

1) $\gamma_A^\beta(\mathcal{D}) = \gamma_C^\beta(\mathcal{D})$.
2) $\gamma_B^\beta(\mathcal{D}) \neq \gamma_C^\beta(\mathcal{D})$ for any proper subset $B \subset A$.

Thus, the β-reduct A is a minimal set of condition attributes that preserves the quality of classification by the set of all condition attributes C.

The set $Cond = \{c_2, c_5\}$ used in Section III-C1 is, however, not a β-reduct because it does not preserve the quality of classification in the case of $\beta = 0.3$. The 0.3-lower approximations of decision classes in Tab. I by C are

$$\underline{C}_{0.3}(D_1) = \{x_1, x_3\},\ \underline{C}_{0.3}(D_2) = \{x_4, x_6\},$$

TABLE X
THE SCORE Cer_{ij} OF EACH ELEMENT δ_{ij} IN TAB. IX

	x_1	x_2	x_3
x_4	0.6	0	0.5
x_5	0.6	0	0.5
x_6	0.67	0	0.5
x_7	0.6	0	0.5

and therefore the quality of classification is

$$\gamma_C^{0.3}(\mathcal{D}) = \frac{2+2}{7} = \frac{4}{7}.$$

On the other hand, the quality of classification by $Cond$ is

$$\begin{aligned} \gamma_{Cond}^{0.3}(\mathcal{D}) &= \frac{|\underline{Cond}_{0.3}(D_1)| + |\underline{Cond}_{0.3}(D_2)|}{7} \\ &= \frac{2+5}{7} = 1, \end{aligned}$$

and therefore $\gamma_{Cond}^{0.3}(\mathcal{D}) \neq \gamma_C^{0.3}(\mathcal{D})$.

However, from the viewpoint of the quality of classification, the set $Cond$ used for generating decision rules satisfies the following good property. The proof is obvious from the construction method of $Cond$ in Algorithm 1.

Proposition 1: For any precision $\beta \in [0, 0.5)$, the following inequality holds:

$$\gamma_{Cond}^{\beta}(\mathcal{D}) \geq \gamma_C^{\beta}(\mathcal{D}). \tag{16}$$

Thus, the quality of classification by $Cond \subseteq C$ is at least equal to or higher than the quality of classification by C. This property indicates that, by not using some condition attributes for suitably ignoring some exceptions, we may be able to construct better classification of objects rather than the case of using all condition attributes.

V. CONCLUSION

In this paper, we proposed a heuristic algorithm to extract decision rules based on the VPRS models. The main idea of our algorithm is based on construction of suitable β-lower approximations by giving up to discern some discernible objects that belong to different decision classes each other. All decision rules extracted by our algorithm are guaranteed that the certainty of all extracted decision rules are equal to or higher than the predefined threshold of certainty.

There are many future issues. First, we need to refine the proposed algorithm and compare other algorithms to generated decision rules that guarantee the minimum accuracy of generated rules, for example, the Apriori algorithm proposed by Agrawal and Srikant [1] by applying our algorithm and other methods to larger datasets. Moreover, by improving our algorithm, proposal of a heuristic algorithm to compute β-reducts is also an interesting issue.

REFERENCES

[1] R. Agrawal and R. Srikant, Fast algorithms for mining association rules, *Proc. of the 20th International Conference on VLDB*, pp.487–499, 1994.

[2] M. Beynon, Reducts within the variable precision rough sets model: A further investigation, *European Journal of Operational Research*, Vol. 134, pp.592–605, 2001.

[3] Y. Kudo and T. Murai, An Attribute Reduction Algorithm by Switching Exhaustive and Heuristic Computation of Relative Reducts, *Proc. of 2010 IEEE International Conference on Granular Computing*, IEEE, pp.265–270, 2010.

[4] Z. Pawlak, Rough Sets. *International Journal of Computer and Information Science*, Vol. 11, pp.341–356, 1982.

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

[6] L. Polkowski, *Rough sets: Mathematical Foundations*, Advances in Soft Computing, Physica-Verlag, 2002.

[7] A. Skowron and C. M. Rauszer, The discernibility matrix and functions in information systems, *Intelligent Decision Support: Handbook of Application and Advance of the Rough Set Theory*, Słowiński, R. (ed.), Kluwer Academic Publishers, pp.331–362, 1992.

[8] W. Ziarko, Variable Precision Rough Set Model, *Journal of Computer and System Science*, Vol. 46, pp.39–59, 1993.