REDUCED ATTRIBUTE ORIENTED HANDLING OF INCONSISTENCY IN DECISION GENERATION

Yucai Feng, Wenhai Li and Zehua Lv
Department of Computer Science,
Huazhong University of Science and Technology,
Wuhan 430074, Hubei, China,
lwhaymail@21cn.com

Abstract Due to the discarded attributes, the effectual condition classes of the decision rules are highly different. To provide a unified evaluative measure, the derivation of each rule is depicted by the reduced attributes with a layered manner. Therefore, the inconsistency is divided into two primary categories in terms of the reduced attributes. We introduce the notion of joint membership function wrt. the effectual joint attributes, and a classification method extended from the default decision generation framework is proposed to handle the inconsistency.

Keywords: reduced attributes, reduced layer, joint membership function, rough set

1. Introduction

Classification in rough set theory [1] is mainly composed of two components: *feature extraction* and *decision synthesis*. Many researches focus on the construction of classification algorithm, such as probabilistic method [2], decision trees[3] and parameterized rule inducing method [4]. The purpose of these methods is to generate rules with high precision and simple expression. In view of the comprehensiveness and conciseness of the training rules, many discernibility matrices based rule extracting methods [5] concerning both approximate inducing and accurate decision are proposed to classify the objects previously unseen. We would like to point out the dynamic reduct [6], variable thresholds based hierarchical classifier [7]. The synthesis methods place emphasis on how to efficiently resolve the conflicts of training rules for the test objects, such as the stable coverings based synthesis [6], hierarchical classifier [7] and lower frequency first synthesis [8].

This paper, based on the default rule extracting framework [5], analyzes the conflicts [9] with two categories of inconsistent rules, and a synthesis stratagem with the notion of joint membership function is proposed to resolve the inconsistency [10]. In the sequel, a report from our experiments with the medical data sets is given to indicate the availability of our classification method.

2. Rough set preliminaries

The starting point of rough set based data analysis is an *information system* denoted by IS, which is a pair $\mathcal{A}(U, A)$ [1]. An IS is a *decision system* when the attributes A can be further classified into disjoint sets of condition attributes C and decision attributes D. With every subset of attributes $B \subseteq A$ in $\mathcal{A}$, the *indiscernibility relation* denoted by $IND(B)$ is defined as follows:

$$IND(B) = \{(x, y) \in U \times U | \forall_{a \in B}, (a(x) = a(y))\}. \tag{1}$$

By $U/IND(B)$ we indicate the set of all equivalence classes in $IND(B)$. Two objects $x, y \in U$ with equation (1) held are indistinguishable from each other. In other words, each object in the universe can be expressed by its own equivalence class $E_i \in U/IND(B)$. For a set of objects $X \subseteq U$, based on $U/IND(B)$, the lower and upper approximations denoted by $\underline{B}X$ and $\overline{B}X$ are $\cup\{E \in U/IND(B)|E \subseteq X\}$ and $\{E \subseteq U/IND(B)|E \bigcap X \neq \emptyset\}$ respectively. For an information system $\mathcal{A}(U, A)$, the *discernibility matrix* denoted by $M_D(\mathcal{A})$ is expressed as an $n \times n$ matrix $\{m_D(i, j)\}$, where $n = |U/IND(A)|$ and

$$m_D(i, j) = \{a \in A | \forall_{i,j=1,2\dots n}, (a(E_i) \neq a(E_j))\}, \tag{2}$$

which implies the set of attributes of A which can distinguish between the two classes $E_i, E_j \in U/IND(A)$. For a decision system $\mathcal{A}(U, C \cup \{d\})$, the *relative discernibility matrix* $M'_D(\mathcal{A})$ is composed of $m'_D(i, j) = \emptyset$ if $d(E_i) = d(E_j)$ and $m'_D(i, j) = m_D(i, j)\backslash\{d\}$, otherwise.

Following this, a unique boolean variable $\overline{a}$ is associated with each attribute a, and $\overline{m}_D(i, j)$ is transformed from $m_D(i, j)$ in terms of $\overline{a}$. Therefore, the *discernibility function* of the attribute set A in an information system $\mathcal{A}(U, A)$ is defined by:

$$f(A) = \underset{i,j \in \{1\dots n\}}{\wedge} \vee \overline{m}_D(E_i, E_j), \tag{3}$$

where $n = |U/IND(A)|$, and the *relative discernibility function* $f'(C)$ in $\mathcal{A}(U, C \cup \{d\})$ is constructed from $\overline{M}'_D(\mathcal{A})$ like equation (3). Similarly, for $n = |U/IND(C)|$, the *local discernibility function* of any $E_i \in U/IND(C)$ is given as:

$$f'(E_i, C) = \underset{j \in \{1\dots n\}}{\wedge} \vee \overline{m}'_D(E_i, E_j). \tag{4}$$

For $\mathcal{A}(U, A)$, a *dispensable* attribute a of A implies $IND(A) = IND(A\backslash\{a\})$, and its counterpart called the *indispensable* has an opposite implication. A *reduct* of A denoted by $RED(A)$ is *a minimal set* of attributes $A' \subseteq A$ so that all attributes $a \in A\backslash A'$ are dispensable, namely $IND(A') = IND(A)$. For $\mathcal{A}(U, C \cup \{d\})$, the *relative reducts* $RED(C, d)$ of C to d are judged by $f'(C)$ similarly with the determination of $f(A)$ on $RED(A)$ [6]. Accordingly, we entitle an attribute (set) $C_{Cut} \subseteq C$ *relatively indispensable* to d iff $\forall_{c \in C_{Cut}} \vee c$ can construct a conjunct of $f'(C)$, and the *prime implicants* of $f'(E_i, C)$ is utilized to determine the *local reduct* of a condition class E_i in $\mathcal{A}$. For $X \subseteq U$ and $B \subseteq A$, the *rough membership* function of X with respect to any class $E_i \in U/IND(B)$ is

$$\mu_B(E_i, X) = \frac{|E_i \cap X|}{|E_i|}, \;\; 0 \leq \mu_B(E_i, X) \leq 1. \tag{5}$$

3. Rule extracting from training tables

Though not entirely correct wrt. the classical relative reducts oriented rule extracting methods [1, 5, 7], the *default rule extracting* framework proposed in [5] provides at lest two advantages, namely *simplicity and generalization*. Therefore, we will use this framework as a basis to validate our research under a restriction of vast rules generation.

For a given *training* table $\mathcal{A}(U, C \cup \{d\})$, taking the prime implicants of $f'(E_i, C)$ of each class $E_i \in U/IND(C)$ for the *predecessor* while regarding the prime implicants of d of each $\{X_j \in U/IND(\{d\}) \mid E_i \cap X_j \neq \emptyset\}$ as the *successor*, all the simpler rules can be expressed as $R : Des(E_i, C) \rightarrow Des(X_j, \{d\})$ with $\mu_C(E_i, X_j)$ no less than a filtering threshold μ_{tr}. By introducing an iterative reduct stratagem, thereby, new training rules by deserting the relatively indispensable attributes are generated as much as possible to handle *test* objects. Accepting $\mathcal{A}$ and a given threshold μ_{tr} as the input, the primary extracting framework can be described as the following four steps:

Step 1 $INIT(\Psi)$. Calculate $U/IND(C)$, $U/IND(\{d\})$ and $M'_D(\mathcal{A})$. For each $E_i \in U/IND(C)$, calculate $f'(E_i, C)$ and generate the rule $R : Des(E_i, C) \rightarrow Des(X_j, \{ d\}) \,|\mu_C(E_i, X_j)$ with each $X_j \in U/IND(d)$ if $\mu_C(E_i, X_j) \geq \mu_{tr}$. Let $C_{Pr} = C$ and goto Step 4.

Step 2 Exit if $ISEND(\Psi)$; let $\mathcal{A}'(U, C' \bigcup\{d\})$ equal to $NEXT(\Psi)$ and let $C_{Pr} = C'$. Calculate $U/IND(C_{Pr})$ and $M'_D(\mathcal{A}')$.

Step 3 For any $E_{(k,C_{Pr})} \in U/IND(C_{Pr})$, calculate $f'(E_i, C_{Pr})$ and generate a rule $\Delta : Des(E_{(k,C_{Pr})}, C_{Pr}) \rightarrow Des(X_j, \{d\}) \,|\mu_{C_{Pr}}(E_{(k,C_{Pr})}, X_j))$ for each $X_j \in U/IND$ (d) if $\mu_C(E_{(k,C_{Pr})}, X_j) \geq \mu_{tr}$, while the blocks to this rule $\mathcal{F} : Des(E_i, C_{Pr}) \rightarrow \neg Des(X_j, \{d\})$ are made if $\forall_{E_i \in U/IND(C)}, E_i \subseteq E_{(k,C_{Pr})} \wedge E_i \cap X_j = \emptyset$.

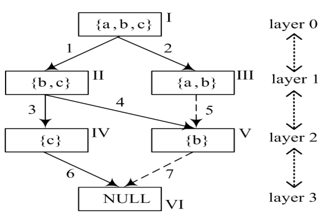
Step 4 Calculate $f'(C_{Pr})$. For each attribute set C_{Cut} emerging in the conjuncts of $f'(C_{Pr})$, select the projections $C'_{Pr} = C_{Pr} \backslash C_{Cut}$, then $INSERT(\Psi)$ with $\mathcal{A}'(U, C'_{Pr} \cup \{d\})$. Goto step2.

Where the *cursor queue* Ψ composed of all the *subtable* $\mathcal{A}'$ has four main operations $\{INIT; INSERT; ISEND; NEXT\}$. Different from the *classical queue*, $ISEND$ judges if the *cursor* is pointing to a $NULL$ subtable, and $NEXT$ is utilized to get the subtable pointed by cursor and move the cursor to the next subtable. To elucidate the generation of the *rule set*(denoted by $RUL(\mathcal{A})$), an illustrative sample displayed in figure 1 results from having observed a total of one hundred objects that were classified according to the condition attributes $C = \{a, b, c\}$ and decision attributes $\{d\}$. Furthermore, the decision classification followed with the cardinality of each $U/IND(C \cup \{d\})$ is represented as $D = \{d\}$.

Figure 1. An illustrative example

V	a	b	c	d
E_1	1	2	3	1 (50×)
E_2	1	2	1	2 (5×)
E_3	2	2	3	2 (30×)
E_4	2	3	3	2 (10×)
$E_{5,1}$	3	5	1	3 (4×)
$E_{5,2}$	3	3	1	4 (1×)

Figure 2. Flow graph of reduct



The real line with the executing sequence number in figure 2 illustrates the projection order of the default algorithm on figure 1, and the dashed denotes the duplicate projection prevented by the cursor queue. The node represents condition attribute set derived from the corresponding projection. Furthermore, the partial relation exists in the nodes which are in different layers and connected by the bidirectional line.

4. Inconsistency classifying based on Reducted Layer

The *default decision generation* method [5] extracts the rules measure up to a membership threshold as much as possible, also, it employs the membership as the interface to resolve the synthesis of the training rules for the test objects. Unfortunately, the conflict of the decision generation can not be resolved completely under this framework. Wang developed a rule-choosing stratagem named *lower frequency first* [8] to quantificationally dispose the *inconsistency* in view of a standpoint that a decision derived from the class with few test objects can represent some special cases, and the precondition of this stratagem is the rules obtained from training set with poor relativity to the test objects, but this stratagem can not work well under the situation in which the training table provides enough reliability for the universe. To parse the causation of the conflict, a notion of *reduced layer* is defined recursively as follows:

DEFINITION 1 *For a given training decision table $\mathcal{A}(U, C \cup \{d\})$, the reduced layer L of each subtable $\mathcal{A}'(U, C' \cup \{d\}) \in \Psi$ denoted by $L(\mathcal{A}')$ is*
- 0 iff $IND(C) = IND(C')$;
- k+1 iff $\exists_{\mathcal{A}''(U,C''\cup\{d\})\in\Psi}, L(\mathcal{A}'') = k \wedge C''\backslash C' \in CON(f'(C''))$.

Where $CON(f'(C''))$ accepts the attribute sets emerging in all the conjuncts of $f'(C'')$ as its elements, and each element corresponding to a *conjunct* in $f'(C'')$ includes all the attributes emerging in this conjunct. We call $\mathcal{A}''$ the *parent* of $\mathcal{A}'$ (i.e $\mathcal{A}''\mathcal{P}\mathcal{A}'$) iff $C''\backslash C' \in CON(f'(C''))$. Simultaneously, P is used to depict the partial relation between C'' and C'. If $\mathcal{A}_1\mathcal{P}\mathcal{A}_2$ and $\mathcal{A}_2\mathcal{P}\mathcal{A}_3$, due to the transitivity of $\subseteq$, subtable $\mathcal{A}_1$ is called the *forefathers* of $\mathcal{A}_3$ (i.e. $\mathcal{A}_1\mathcal{F}\mathcal{A}_3$ or C_1FC_3). From the above, obviously, the original table $\mathcal{A}(U, C \cup \{d\})$ is with the reduced layer 0. Any subtable $\mathcal{A}'(U, C' \cup \{d\})$ in Ψ with reduced layer larger than 0 is homogenous with $\mathcal{A}$ except for $C' \subseteq C$, where C' is called *reduced attributes*. Let us now assume that the considered original table had no condition attributes with the same equivalence classes, i.e. $\forall_{c_1,c_2\in C}, IND/\{c_1\} \neq IND/\{c_2\}$, and it is commonly satisfied in the large-scale environments.

PROPOSITION 2 *For two reduced attributes C'' and C' which belong to $\mathcal{A}''$ and $\mathcal{A}'$ respectively, $U/IND(C') \subseteq U/IND(C'')$ exists iff $C''FC'$, namely $\mathcal{A}''\mathcal{F}\mathcal{A}'$.*

When considering the necessity, due to the transitivity of relation $\mathcal{P}$ among all the middle subtables between $\mathcal{A}'$ and $\mathcal{A}''$, $U/IND(C') \subseteq U/IND(C'')$ can be easily proven. When considering the sufficiency, we suppose there exists another subtable $\mathcal{B}(U, B \cup \{d\})$ with $L(\mathcal{B}) = L(\mathcal{A}'') \wedge \mathcal{B}\mathcal{F}\mathcal{A}'$ held, and due to the greedy manner of the default rule extracting framework discussed in [5], we assert $U/IND(B) = U/IND(C'')$; also because both $\mathcal{B}$ and $\mathcal{A}''$ root in the original table $\mathcal{A}$ with several indispensable attributes deserted, $B = C''$ can be obtained. And thus $C''FC'$ and $\mathcal{A}''\mathcal{F}\mathcal{A}'$ are proven.

As discussed in section 3, a set of rules with the form of $r_k : Pred(r_k) \rightarrow Succ(r_k)|\mu(r_k)$ can be generated by applying the four steps to a given training table

$\mathcal{A}(U, C \cap \{d\})$. For the universe W, each object $u \in W$ can be classified to a decision class $CLS(Succ(r_k))$ iff any attribute $a \in A$ emerging in $Pred(r_k)$ is supported by u, and it's denoted by $Mat(r_k, u) : \forall_{a \in A}, a(Pred(r_k)) \neq \emptyset \rightarrow a(u) = a(Pred(r_k))$. Therefore, the inconsistency consists in $RUL(\mathcal{A})$ iff

$$\exists_{r_i, r_j \in RUL(\mathcal{A})}, Mat(r_i, u) \wedge Mat(r_j, u) \wedge CLS(Succ(r_i)) \neq CLS(Succ(r_j)), \quad (6)$$

where $Mat(r_i, u)$ denotes $Pred(r_i)$ is supported by u, and $CLS(Succ(r_i))$ denotes the decision class determined by $Succ(r_i)$. Therefore, $RUL(\mathcal{A})$ is inconsistent due to the existence of any $r_i, r_j \in RUL(A)$ with both $\forall_{a \in A}, a(Pred(r_i)) \neq \emptyset \wedge a(Pred(r_j)) \neq \emptyset \rightarrow a(Pred(r_i) = a(Pred(r_j)))$ and $CLS(Succ(r_i)) \neq CLS(Succ(r_j))$ held. To distinguish the rules derived from different subtables, each $r \in RUL(\mathcal{A})$ is expressed by $Des(E_i^r, C^r) \rightarrow Des(X_j, \{d\})$, where $Des(E_i^r, C^r)$ implies $Pred(r)$ comprising the local reduct of E_i^r in subtable $\mathcal{A}^r(U, C^r \cup \{d\})$. Based on the correlative notions of reduced layer, the inconsistency among the rules can be divided into two cases according to their condition class.

COROLLARY 3 *For two inconsistent rule r_1 and r_2 derived respectively from $\mathcal{A}^{r_1}$ and $\mathcal{A}^{r_2}$, suppose $L(\mathcal{A}^{r_2}) \geq L(\mathcal{A}^{r_1})$, we shall say that this inconsistency is:*

$$\begin{cases} inherited & \textit{iff } C^{r_2} \subseteq C^{r_1}, \quad (7b) \\ varietal & \textit{iff } C^{r_2} \not\subseteq C^{r_1}. \quad (7c) \end{cases}$$

The *inherited* inconsistency can be ulteriorly divided into two cases, i.e. $L(\mathcal{A}^{r_2}) = L(\mathcal{A}^{r_1}) \rightarrow C^{r_2} = C^{r_1}$ and $L(\mathcal{A}^{r_2}) > L(\mathcal{A}^{r_1}) \rightarrow C^{r_2} \subset C^{r_1}$, and the *varietal* inconsistency has two similar cases. In figure 2, the consistency between the rules from node II and the rules from node IV belongs to the inherited, and the consistency arising from node III and node IV is varietal. With little consideration of the difference among the subtables, many researches focus on the inconsistency of the rules derived from the same subtable, hence the rule certainty is converted into the cardinality-based evaluation measures for the sake of achieving high-frequency rule.

5. Methods of inconsistency handling

In this paper, to complement the default decision generation method, we mainly discuss the inconsistency from different layers and suppose $L(\mathcal{A}^{r_2}) > L(\mathcal{A}^{r_1})$. For two inconsistent rules $r_1 : Des(E_{i_1}^{r_1}, C^{r_1}) \rightarrow Des(X_{j_1}, \{d\})$ and $r_2 : Des(E_{i_2}^{r_2}, C^{r_2}) \rightarrow Des(X_{j_2}, \{d\})$, if $C^{r_2} \subset C^{r_1}$ exists, it's obvious that the condition classes could hold either $E_{i_1}^{r_1} \subset E_{i_2}^{r_2}$ or $E_{i_1}^{r_1} \cap E_{i_2}^{r_2} = \emptyset$. Being comparable with the condition class determined by r_1, the *effectual set* covered by r_2 is only composed of the classes which leads to $Succ(r_2)$ while belonging to $U/IND(C^{r_1})$, namely:

$$ES(E_{i_2}^{r_2}, C_{r_1}) = \{E_i^{r_1} \in IND/C^{r_1} | E_i^{r_1} \cap X_{j_2} \neq \emptyset \wedge E_i^{r_1} \subseteq E_{i_2}^{r_2}\}. \quad (8)$$

When measuring the rules r_1 and r_2 with the relation $C^{r_2} \subset C^{r_1}$ held, due to the desertion of the relatively indispensable attributes $C^{r_1} \backslash C^{r_2}$, the condition classes in $U/IND(C^{r_1})$ which could not lead to the decision $Succ(r_2)$ are taken into account, and it may depress the rule r_2. Hence, for disposing the inherited inconsistency, the notion of *joint membership function* can be determined by the cardinality-based evaluation measure of the effectual set.

DEFINITION 4 *For two inconsistent rules r_1, r_2 with $C^{r_2} \subset C^{r_1}$ held, the joint membership function of r_2 with respect to C^{r_1} is defined as:*

$$\mu_{C^{r_1}}(E_{i_2}^{r_2}, X_{j_2}) = \frac{\sum_{E_k \in ES(E_{i_2}^{r_2}, C_{r_1})} |E_k \cap X_{j_2}|}{\sum_{E_k \in ES(E_{i_2}^{r_2}, C_{r_1})} |E_k|}, \ \ 0 \leq \mu_{C^{r_1}}(E_{r_2}, X_{j_2}) \leq 1. \quad (9)$$

Where the denominator denotes the cardinality of the effectual set for r_2 under the condition attributes C^{r_1}, and the numerator denotes the cardinality of the objects which support r_2. Clearly, one can perceive that the rough membership function is a special case of the joint membership function, i.e. $\mu_{C^{r_1}}(E_{i_2}^{r_2}, X_{j_2}) = \mu_{C^{r_2}}(E_{i_2}^{r_2}, X_{j_2}))$ iff $C^{r_1} = C^{r_2}$. As shown in equation (9), on the assumption that both r_1 and r_2 are under the same condition restriction, $ES(E_{i_2}^{r_2}, C_{r_1})$ depicts the object sets contained by r_2 comparable with the ones determined by r_1, and $\mu_{C^{r_1}}(E_{i_2}^{r_2}, X_{j_2})$ is more equitable than $\mu_{C^{r_2}}(E_{i_2}^{r_2}, X_{j_2})$ for inherited inconsistency.

When considering the varietal inconsistency, for the above two rules r_1 and r_2, $C^{r_2} \not\subseteq C^{r_1}$ comes into existence as discussed in corollary 3. Similarly with the analysis of the inherited case, it can be divided into two subcases, i.e. $L(\mathcal{A}^{r_2}) = L(\mathcal{A}^{r_1}) \rightarrow C^{r_2} \neq C^{r_1}$ and $L(\mathcal{A}^{r_2}) > L(\mathcal{A}^{r_1}) \rightarrow C^{r_2} \not\subset C^{r_1}$. In figure 2, one may conclude the inconsistent rules from node II and node III to be the former and the ones from III and IV the latter. Due to the necessity of proposition 2, the condition attribute set $C^{r_1} \cup C^{r_2}$ is the forefather of the both subset, denoted by $(C^{r_1} \cup C^{r_2})FC^{r_1}$ and $(C^{r_1} \cup C^{r_2})FC^{r_2}$. Therefore, $C^{r_1} \cup C^{r_2}$ can be utilized to evaluate the rule certainty, and called by the *effectual joint attributes*.

PROPOSITION 5 *For two inconsistent rules r_1, r_2 with $L(\mathcal{A}^{r_2}) = L(\mathcal{A}^{r_1}) \rightarrow C^{r_2} \neq C^{r_1}$ held, we shall say that the rule certainty can be evaluated by the joint membership function $\mu_{C^{r_1} \cup C^{r_2}}(E_{i_1}^{r_1}, X_{j_1})$ and $\mu_{C^{r_1} \cup C^{r_2}}(E_{i_2}^{r_2}, X_{j_2})$.*

It's obvious that $C^{r_1} = C^{r_1} \cup C^{r_2}$ iff $C^{r_2} \subseteq C^{r_1}$, thus proposition 5 provides a unified evaluative condition attributes for the both rules, and the both categories of inconsistency can be disposed by choosing the rules with higher joint membership function. All the above accounts for the inconsistency between two rules, but when two rules r_1, r_2 are consistent with both the predecessor and the successor (denoted by $r_1 Cst\ r_2$), i.e. $\forall_{a \in A}, a(Pred(r_1)) \neq \emptyset \wedge a(Pred(r_2)) \neq \emptyset \rightarrow a(Pred(r_1) = a(Pred(r_2)) \wedge CLS(Succ(r_1)) = CLS(Succ(r_2))$, to compete with any $r_3 \in RUL\mathcal{A}$ which is inconsistent with (denoted by $r_3 Inc\, r_1$) the both rules, all the consistent pairs of each rule must be treated like the inconsistent pairs for obtaining the most credible rule. To achieve the forementioned, the rule is constructed by a header followed with an array of consistent rule descriptions and an array of inconsistent rule descriptions, and the header include six members:

$$Idt : Rule : Block : Strength : Layer : Pds : CstArray : IncArray. \quad (10)$$

For any $r_a \in RUL(\mathcal{A})$, the symbol Idt denotes the identifier of r_a and $Strength(r_a) = |E_{i_a}^{r_a} \cap X_{j_a}|$ denotes the cardinality of the r_a supported objects. $Layer$ denotes the reduced layer of $\mathcal{A}_a^r$. Pts points to the r_a related decision subtable in the cursor queue Ψ, and r_a is also pointed by its related subtable. During the rule extracting phase, as discussed in definition 1, these four members are obtained from Step 2 of the extracting framework with a layer marker in $\mathcal{A}'$. Each element in the last two arrays is composed of an identifier $Idt(r_b)$ and a pair of joint membership function value $(\mu_{C^{r_a}\cup C^{r_b}}(E^{r_a}_{i_a}, X_{j_a}), \mu_{C^{r_a}\cup C^{r_b}}(E^{r_b}_{i_b}, X_{j_b}))$, in which $CstArray$ records all the consistent rules to r_a and $IncArray$ includes all the inconsistent ones.

Following Step 3, according to the established subtables in Ψ, each *generated rule* is fetched to compare with r_a. And then, as discussed in definition 4, join $Idt(r_b)\ :\ (\mu_{C^{r_a}\cup C^{r_b}}(E^{r_a}_{i_a}, X_{j_a}), \mu_{C^{r_a}\cup C^{r_b}}(E^{r_b}_{i_b}, X_{j_b}))$ into $CstArray$ if $r_aCst\ r_b$ and join $Idt(r_b)\ :\ (\mu_{C^{r_a}\cup C^{r_b}}(E^{r_a}_{i_a}, X_{j_a}), \mu_{C^{r_a}\cup C^{r_b}}(E^{r_b}_{i_b}, X_{j_b}))$ into $IncArray$ if $r_aInc\ r_b$. The generated rule implies that the both arrays only record the corresponding rules which are generated from the subtable with the reduced layer smaller than $\mathcal{A}^{r_a}$, due to the reflexivity of both Cst and Inc, this can reduce the complexity of extracting and synthesis. From all above, we assert the time and space complexity of reduced attributes oriented rule extracting algorithm are of order $O(n^4 \cdot m^2)$ and $O(n \cdot m + m^2)$, respectively.

According to the above structures, suppose several rules $M = (r_1...r_k)$ are supported by a test object u, then the most credible rule can be obtained by:

1. Classify $\forall r_a \in M$ into several consistent subsets (Suppose the number is K) according to X_{j_a}.

2. For each consistent subset, with a dimidiate manner, chose the rule with the maximal joint membership value by $CstArray$; if the result is not unique, chose the rule with the largest $Strength$, and $M' = (r^1...r^K)$ is obtained.

3. For $\forall r^a, r^b \in M'$, with a dimidiate manner, chose the one with the maximal joint membership value by $IncArray$; if the result is not unique, chose the rule with the largest $Strength$, and the most credible rule for u is found.

In Step 2 and 3, for comparing rules pairs, fetch the $CstArray$ or $IncArray$ of the rule with the larger $Layer$. The random selection is applied if both the joint membership value and $Strength$ of any pair are the same.

6. Computational experiments

To indicate the validity of our method, three medical data from the UCI Machine Learning Repository is used in our experiments. Let us notice that the data sets used in our experiments are assumed to be complete. To achieve this, the data were slightly modified by removing a few attributes which result in vast incompleteness, and the other missing values with a ratio of 8% were made out by a statistic method. To insure the comparability, 10 fold cross-validation reclassification technique was performed.

In order to indicate the availability of our method, three synthesis methods based on vast rules generation algorithm [5] are given for comparison. In which, Std is the standard discernibility applying a random rule selection to the rules with equal membership, HFF uses the high frequency first strategy of inconsistenct rule-choosing and LFF is it's opposite. The Reducted Attributed oriented Rule Generation is denoted by RARG. Moreover, we consult two popular rough sets based rule induction systems, i.e. new version of LERS (New LERS) and the classification coefficient oriented synthesis system based on the object-oriented programming library (RSES-lib). For the purpose of comparison, the membership value threshold for Std, HFF, LFF and RARG are all 0.55 and the coefficient threshold for RSES-lib is 0.75, which are quoted by the corresponding authors.

Table 1. Computational result with the medical datasets

Algorithm	Lymphography			Breast cancer			Primary tumor		
	Rule number	Error rate Train	Test	Rule number	Error rate Train	Test	Rule number	Error rate Train	Test
New LERS	984	0.000	0.233	1163	0.063	0.342	8576	0.245	0.671
RSES-lib	427	0.000	0.195	756	0.152	0.277	6352	0.136	0.687
Std	1321	0.000	0.320	2357	0.060	0.361	7045	0.175	0.764
HFF	1321	0.000	0.267	2357	0.042	0.338	7045	0.147	0.742
LFF	1321	0.000	0.341	2357	0.245	0.470	7045	0.360	0.720
RARG	1321	0.000	0.207	2357	0.051	0.292	7045	0.125	0.598

As shown in table 1, since HFF refined the default decision generation framework, its performance exceeds the later in all the three datasets. Due to the different granularity distribution of both classes, LFF works well in the first and the third datasets while falling across a sharply decrease in the breast cancer dataset. Because RARG provides a unified evaluation criterion for conflicts, with the irrelevant condition classes filtered, it guarantees the decision with the largest ratio of the sustaining decision objects to the effectual condition objects. For the tested objects, it refers to the most accordant rule with respect to other conflict ones. Therefore, RARG is particularly outstanding in the applications with voluminous inconsistency, such as the Primary tumor dataset displayed in the result. In conclusion, RARG takes on a comparatively high performance in the above four methods. The results also show that RARG is comparable with the other two systems, and especially, it exceeds them in the Primary tumor dataset.

References

1. Pawlak Z., Skowron A. (1993) A rough set approach to decision rules generation, In *Research Report 23/93, Warsaw University of Technology*. 1-19.
2. Michie D., Spiegelhalter D. J., C. Taylor C. (1994) Machine learning, neural and statistical classification. Ellis Horwood, New York.
3. Quinlan J. R. (1986) Induction of decision trees. *Machine Learning*. 1:81-106.
4. Grzymala-Busse, J. W. (1997) A new version of the rule induction system LERS. *Fundamenta Informaticae*. 31:27-39.
5. Mollesta T., Skowron A. (1996) A rough set framework for data mining of propositional default rules, In *9th Int. Sym. Proc. On Found. of Intelligent Systems*. 448-457.
6. Bazan J., Nguyen H. S., Synak P., Wrĺőblewski J. (2000) Rough set algorithms in classification problems. In *Rough set methods and applications. New Developments in Knowledge Discovery in Information Systems, Studies in Fuzziness and Soft Computing*. 49-88.
7. Skowron A., Wang H., Wojna A., Bazan J. G. (2005) A hierarchical approach to multimodal classification. In *RSFDGrC*. 2:119-127.
8. Wang G. Y., Wu Y., Liu F. (2000) Generating rules and reasoning under inconsistencies, In *2000 IEEE Int. Conf. on Industrial Electronics, Control and Instrumentation*. 2536-2541.
9. Jerzy W., Grzymalao-Busse, Ming L. (2001) A comparison of several approaches to missing attribute values in data mining. In Ziarko W. and Yao Y. (ed.), *RSCTC2000*. 378-385.
10. Dzeroski S. (2000) Normalized decision functions and measures for inconsistent decision tables analysis. In *Fundamenta Informaticae*. 44(3):291-319.