

A Heuristic Approach to Acquisition of Minimum Decision Rule Sets in Decision Systems

Zuqiang Meng, Liang Jiang, Hongyan Chang, and Yuansheng Zhang

College of Computer, Electronics and Information, Guangxi University Nanning 530004, China
mengzuqiang@163.com

Abstract. In rough set theory, not too much work pays attention to the acquisition of decision rules and to the uses of the obtained rule set as classifier to predict data. In fact, rough set theory also can be applied to train data and create classifiers and then complete data prediction. This paper systematically studies the problem of acquisition of decision rules in decision systems. The main outcomes of this research are as follows: (1) the specific definition of minimum rule set is given, and such a minimum rule set can be used as a classifier to predict new data; (2) a new approach to finding out all minimum rule sets for a decision system, Algorithm 1, is proposed based on discrimination function, but with relatively low execution efficiency; (3) By improving Algorithm 1, a heuristic approach to computing a special minimum rule set, Algorithm 3, is proposed, which works far more efficiently than Algorithm 1. The outcomes can form the foundation for applying rough set theory to data classification and offer a new resolution to data classification.

Keywords: Heuristic approach, minimum decision rule set, rule acquisition, classifier, data classification.

1 Introduction

Rough set theory[1] is a powerful mathematical tool to deal with insufficient, incomplete or vague information. Nowadays the majority of work in rough set theory focuses on attribute reduction[2]. Attribute reduction is in fact feature selection and has been applied to address many practical problems[3-6]. After attribute reduction, data sets can be used for a variety of applications and data classification is an important application. However, not too much work pays attention to data classification with rough set theory. In rough set theory, most methods of data classification are to use rough set theory to select features and then use other tools to train data and create classifier[7]. In fact, rough set theory also can be used to extract decision rules from decision systems. Although some scholars have made contribution to this work[8-10], there is still a lack of systematic foundation for rule acquisition in rough set theory. This paper systematically studies the problem of acquisition of decision rules in decision systems. First, the specific definition of minimum rule sets is given, and then a new approach to finding out all minimum rule sets for a decision system is proposed based on discrimination function, and finally a heuristic approach

to computing a special minimum rule set is proposed. Each of the two proposed approaches has its own advantages and disadvantages. The work provides a relatively complete solution for extracting decision rules from decision systems and using rough set theory to classify data.

The rest of the paper is organized as follows. Section 2 reviews some basic concepts linked to Information system and decision logic. Section 3 gives the definition of minimum rule set. Section 4 proposes a general approach to the acquisition of all minimum rule sets in a given decision system. Section 5 then proposes a heuristic approach to the acquisition of a minimum rule set. Section 6 gives an example to show how to use the heuristic approach. Section 7 finally concludes the paper.

2 Information System and Decision Logic

An **information system** is usually expressed in the following form: $IS = (U, A, \{V_a\}, f_a)_{a \in A}$, where U is a nonempty finite set of objects, standing for a given universe; A is a nonempty finite set of attributes; V_a is a value set(domain) of attribute a ; f_a is called information function, i.e., $f_a: U \rightarrow V_a$, which denotes the value of function f on attribute a ; If V_a and f_a are obvious, then $(U, A, \{V_a\}, f_a)_{a \in A}$ can be denoted as (U, A) for short.

For any $B \subseteq A$, subset B determines a binary relation, denoted as $TR(B)$, which is defined as follows: $TR(B) = \{(x, y) \mid f_a(x) = f_a(y) \text{ for any } a \in B \text{ and any } x, y \in U\}$.

It is easy to prove that $TR(B)$ is reflexive, symmetric, transitive, and thereby is an equivalence relation on U . Thus, equivalence relation $TR(B)$ can divide the universe U into several disjoint subsets, which are known as equivalence classes. Suppose that X_1, X_2, \dots, X_n are all equivalence classes induced by attribute set B , then $\{X_1, X_2, \dots, X_n\}$ is a partition $U/TR(B)$ of the universe U , denoted by $U/TR(B) = \{X_1, X_2, \dots, X_n\}$.

Extraction of decision rules is in fact the problem of extracting description of granules (equivalence classes), which are expressed with decision logic.

Let $IS(B) = \langle U, B, \{V_a\}, f_a \rangle_{a \in B}$, $B \subseteq A$. Then decision logic language $DL(B)$ with respect to B is defined as following:

- (1) (a, v) is an atomic formula, where $a \in B, v \in V_a$,
- (2) an atomic formula is a formula in $DL(B)$;
- (3) if φ is a formula, then $\sim\varphi$ is also a formula in $DL(B)$;
- (4) if both φ and ψ are formulae, then $\varphi \vee \psi, \varphi \wedge \psi, \varphi \rightarrow \psi, \varphi \equiv \psi$ are all formulae;
- (5) only the formulae obtained according to the above Steps (1) to (4) are formulae in $DL(B)$.

If φ is a simple conjunction, which consists only of atomic formulae and connectives \wedge , then φ is called a **basic formula**.

The following definition gives the relationship between formulae in $DL(B)$ and granules $IS(B)$:

For any $s \in U$, the relationship between s and formulae in $DL(B)$ is defined as following:

- (1) $s \models (a, v)$ iff $f_a(s) = v$
- (2) $s \models \sim\varphi$ iff not $s \models \varphi$
- (3) $s \models \varphi \wedge \psi$ iff $s \models \varphi$ and $s \models \psi$
- (4) $s \models \varphi \vee \psi$ iff $s \models \varphi$ or $s \models \psi$
- (5) $s \models \varphi \rightarrow \psi$ iff $s \models \sim\varphi \vee \psi$
- (6) $s \models \varphi \equiv \psi$ iff $s \models \varphi \rightarrow \psi$ and $s \models \psi \rightarrow \varphi$.

For formula φ , if $s \models \varphi$, then we say that the object s satisfies formula φ . Let $m(\varphi) = \{ s \mid s \models \varphi \}$, that is, $m(\varphi)$ is the set of all those objects that satisfy formula φ ; For subset $g \subseteq U$, if $m(\varphi) = g$, then g is said to be **descriptive** and φ is a **description** of g , denoted by $DES(g)$. It is easy to find that for the same subset $g \subseteq U$, there are possibly many different descriptions: $\varphi_1, \varphi_2, \dots, \varphi_m$, such that $m(\varphi_1) = m(\varphi_2) = \dots = m(\varphi_m) = g$. This is one of the reasons why extracting all decision rules from a decision system is a NP-hard problem.

3 Minimum Rule Set in Decision System

We note that some principles of attribute reduction can also be applied to extracting decision rules from decision systems and then to acquiring minimum rule set, which can be used as a classifier.

Decision system is a special case of information system, which can be regarded to be generated by partitioning attribute set A into two disjoint subsets, or by adding some attributes to A .

A **decision system** is usually denoted by $DS = (U, C \cup D, \{V_a\}, f_a)_{a \in A = C \cup D}$, where U, V_a and f_a have the same meanings as that in the above section; C is a nonempty finite set of attributes, called condition attribute set; D is a nonempty finite set of attributes, called decision attribute set, and $C \cap D = \emptyset$; The 4-triple $(U, C \cup D, \{V_a\}, f_a)_{a \in A = C \cup D}$ is usually denoted as $(U, C \cup D)$ for short, namely, $DS = (U, C \cup D)$.

As mentioned above, we can consider that decision system $(U, C \cup D)$ consists of two information systems: (U, C) and (U, D) .

Suppose that $\phi \in DL(C)$ and $\psi \in DL(D)$. Implication form $\phi \rightarrow \psi$ is said to be a **decision rule** in decision system $(U, C \cup D)$. If both ϕ and ψ are basic formula, then $\phi \rightarrow \psi$ is called **basic decision rule**.

Decision rule has two important measuring indices, confidence and support, which are defined as following:

$$conf(\phi \rightarrow \psi) = \frac{|m(\phi) \cap m(\psi)|}{|m(\phi)|} \quad [11] \tag{1}$$

$$sup(\phi \rightarrow \psi) = \frac{|m(\phi) \cap m(\psi)|}{|U|} \tag{2}$$

where $conf(\phi \rightarrow \psi)$ and $sup(\phi \rightarrow \psi)$ are confidence and support of decision rule $\phi \rightarrow \psi$, respectively.

In general, we usually find those decision rules whose confidence is equal to 1. Obviously, if $m(\phi) \subseteq m(\psi)$, then $conf(\phi \rightarrow \psi) = 1$. Of course, it is not enough to guarantee that confidence is equal to 1. In fact, the larger the confidence, the better it is, although it is small than 1 with certainty. That is, extracting decision rules is the procedure of finding those rules whose confidence is equal to 1 and whose support is as large as possible. In order to illustrate this procedure, we further introduce some related concepts.

Definition 1. For decision system $(U, C \cup D)$, $B \subseteq C$ and $x \in U$, let $[x]_B = \{y \in U \mid (x, y) \in TR(B)\}$, then $[x]_B$ is said to be an equivalence class with respect to attribute subset B .

Property 1. For decision system $(U, C \cup D)$, $B_1 \subseteq B_2 \subseteq C$ and any $x \in U$, $[x]_{B_2} \subseteq [x]_{B_1}$.

Proof. It is straightforward.

Property 1 shows that with the decrease of attribute subset, corresponding equivalence classes become larger and larger, thus increase corresponding rule's support. That is, by deleting some attributes from C , we can increase rule's support according to formula (2), which is the foundation of finding decision rules in this paper.

In fact, by deleting some attributes from condition attribute set, every object in the universe U can induce at least one decision rule. Thus, all objects from U can induce a lot of decision rules and then form a very large decision rule set. Obviously, many rules are redundant and should be removed from rule set. Therefore, one basic problem is that what is the criterion of a minimum rule set?

Definition 2. For decision system $DS = (U, C \cup D)$, if rule $\phi \rightarrow \psi$ is true in $DL(C \cup D)$, i.e., for any $x \in U \mid x \models \phi \rightarrow \psi$, then rule $\phi \rightarrow \psi$ is said to be **consistent** in DS , denoted by $\models_{DS} \phi \rightarrow \psi$; if there exists at least object $x \in U$ such that $x \models \phi \wedge \neg \psi$, then rule $\phi \rightarrow \psi$ is said to be **satisfiable** in DS .

Consistency and satisfiability are the basic properties that must be satisfied by decision rules. In addition, an efficient and effective decision rule set should be as small as possible, that is, it should be minimized.

Definition 3. For decision system $(U, C \cup D)$, object $x \in U$ and decision rule $r: \phi \rightarrow \psi$, if $x \models r$, then it is said that rule r **covers** object x , or object x is **covered** by rule r ; let $coverage(r)$ denote the set of all objects that are **covered** by rule r and $coverage^{-1}(x)$ the set of all rules that cover object x , that is:

$$coverage(r) = \{x \in U \mid x \models r\},$$

$$coverage^{-1}(x) = \{r' \mid r' \text{ cover } x, \text{ i.e., } x \in coverage(r')\}.$$

Definition 4. For decision system $(U, C \cup D)$, decision rules r_1 and r_2 , if $\text{coverage}(r_1) \subseteq \text{coverage}(r_2)$, then it is said that r_2 **functionally cover** r_1 , denoted by $r_1 \leq r_2$.

Obviously, those rules that are functionally covered by other rules should be removed out from rule set.

Definition 5. For decision system $(U, C \cup D)$ and $x \in U$, $[x]_B$ is said to be **maximized** if $[x]_B \subseteq [x]_C$ and for any $B' \subset B$ $[x]_{B'} \not\subseteq [x]_D$, where $B \subseteq C$; rule $DES([s]_B) \rightarrow DES([s]_D)$ is said to be **reduced** if $[x]_B$ is maximized.

A decision rule set \wp is said to be minimal if it satisfies the following properties:

- (1) any rule from \wp should be consistent;
- (2) any rule from \wp should be satisfiable;
- (3) any rule from \wp should be reduced;
- (4) for any two rules $r_1, r_2 \in \wp$, neither $r_1 \leq r_2$ nor $r_2 \leq r_1$.

When a decision rule set satisfies all the four properties, each rule is effective and the whole rule set is minimal and then matching efficiency can be improved greatly.

4 A General Approach to Acquisition of All Minimum Rule Sets

Suppose that φ be a formula and $B = \{a_1, a_2, \dots, a_m\}$ be a condition attribute subset. Let $\text{set}(\varphi)$ be the set of all attributes that appear in formula φ , and $\wedge B$ and $\vee B$ respectively be the simple conjunction and the simple disjunction that consists of all the attributes appear in B with \wedge as connective, i.e., $\wedge B = a_1 \wedge a_2 \wedge \dots \wedge a_m$. For example, let $\varphi = a \wedge b \wedge c$ then $\text{set}(\varphi) = \{a, b, c\}$; let $B = \{a, b, c\}$ then $\vee B = a \vee b \vee c$ and $\wedge B = a \wedge b \wedge c$.

Definition 6. For decision system $(U, C \cup D)$, let $\alpha([x]_C, x') = \vee \{a \in C \mid x' \notin [x]_{\{a\}}\}$, where $x, x' \in U$; suppose that $U - [x]_D = \{x'_1, x'_2, \dots, x'_m\}$, then the discrimination function of $[x]_C$ is defined as following:

$$f([x]_C) = \alpha([x]_C, x'_1) \wedge \alpha([x]_C, x'_2) \wedge \dots \wedge \alpha([x]_C, x'_m).$$

The discrimination function is a conjunctive normal form. By using absorption law and distribution law, it can be converted to be a disjunctive normal form, which consists of several simple conjunctions. Each simple conjunction corresponds to a decision rule. Without loss of generality, suppose that $f([x]_C)$ is converted to disjunctive normal form $\rho_1 \vee \rho_2 \vee \dots \vee \rho_{m'}$, where ρ_i is a simple conjunction, $i=1, 2, \dots, m'$. Then all the rules that induced by equivalence class $[x]_C$ are as follows:

$$\begin{aligned} r_1: & DES([x]_{\text{set}(\rho_1)}) \rightarrow DES([x]_D), \\ r_2: & DES([x]_{\text{set}(\rho_2)}) \rightarrow DES([x]_D), \\ r_{m'}: & DES([x]_{\text{set}(\rho_{m'})}) \rightarrow DES([x]_D). \end{aligned}$$

Let $R([x]_C)$ be the set of all rules that induced by $[x]_C$, i.e., $\mathbf{R}(\mathbf{x})_C = \{r_1, r_2, \dots, r_m\}$. Suppose that decision class $[x]_D$ is an union of several equivalence classes: $[x_1]_C, [x_2]_C, \dots, [x_p]_C$, i.e., $[x]_D = [x_1]_C \cup [x_2]_C \cup \dots \cup [x_p]_C$. Let $\mathbf{R}(\mathbf{x})_D = R([x_1]_C) \cup R([x_2]_C) \cup \dots \cup R([x_p]_C)$. Obviously, some induced rules in $R([x]_D)$ are redundant and should be removed from rule set. Or in other words, $R([x]_D)$ needs to be further reduced.

Definition 7. For decision system $(U, C \cup D)$ and decision class $[x]_D$, suppose \wp is a subset of rule set $R([x]_D)$, i.e., $\wp \subseteq R([x]_D)$, then \wp is a **reduct** of $R([x]_D)$ if any rule in \wp is consistent, satisfiable, reduced and for any two rules $r_1, r_2 \in \wp$, neither $r_1 \leq r_2$ nor $r_2 \leq r_1$.

Similar to attribute reduction, $R([x]_D)$ possibly has more than one reduct. We can also use discrimination function to find all reducts of $R([x]_D)$.

Suppose $[x]_D = \{x_1, x_2, \dots, x_h\}$. Let $f([x]_D) = [\vee \text{coverage}^{-1}(x_1)] \wedge [\vee \text{coverage}^{-1}(x_2)] \wedge \dots \wedge [\vee \text{coverage}^{-1}(x_h)]$. Obviously, $f([x]_D)$ is a conjunctive normal form. Similarly, it can be converted to a disjunctive normal form by using absorption law and distribution law, in which each simple conjunction corresponds to a reduct of $R([x]_D)$. Assume that $f([x]_D)$ is converted to disjunctive normal form $\beta_1 \vee \beta_2 \vee \dots \vee \beta_{h'}$. Then all reducts of $R([x]_D)$ are as follows: $set(\beta_1), set(\beta_2), \dots, set(\beta_{h'})$.

It is not difficult to prove that for any $r \in set(\beta_i)$, r is consistent, satisfiable, reduced and for any two rules $r_1, r_2 \in set(\beta_i)$, neither $r_1 \leq r_2$ nor $r_2 \leq r_1$, $i \in \{1, 2, \dots, h'\}$. Therefore, $set(\beta_i)$ is a reduct of $R([x]_D)$.

Again, we know that that if $[x_1]_D \neq [x_2]_D$ then $[x_1]_D \cap [x_2]_D = \emptyset$ and then $R([x_1]_D) \cap R([x_2]_D) = \emptyset$. Hence for any two different decision classes $[x_1]_D$ and $[x_2]_D$, we may concurrently calculate all reducts of $[x_1]_D$ and $[x_2]_D$.

According to analysis above, we can give a complete solution to acquisition of all minimum rule sets in decision system $DS = (U, C \cup D)$, which is called Algorithm 1.

Algorithm 1. For calculating all minimum rule sets.

- (1) compute all different equivalence classes: $[x_1]_D, [x_2]_D, \dots, [x_n]_D$, and let $DC(DS) = \{[x_1]_D, [x_2]_D, \dots, [x_n]_D\}$;
- (2) for decision class $[x_i]_D \in DC(DS)$, find out all equivalence classes whose union is equal to $[x_i]_D$, and suppose the set of all such equivalence classes is $EC([x_i]_D)$;
- (3) for each $[x']_C \in EC([x_i]_D)$, compute $R([x']_C)$ using discrimination function and then obtain $R([x_i]_D) = \cup \{R([x']_C) \mid [x']_C \in EC([x_i]_D)\}$;
- (4) find all reducts of $R([x_i]_D)$ using discrimination function, and suppose the set of all reducts of $R([x_i]_D)$ is $RED[R([x_i]_D)]$;
- (5) compute the set of all minimum rule sets: $\{\bigcup_{i=1}^n \wp_i \mid \wp_i \in RED[R([x_i]_D)], \text{ where } i \in \{1, 2, \dots, n\}\}$.

Steps (2)-(4) can be performed concurrently, with which we can improve the efficiency of the approach. The biggest advantage of this approach lies in that it can find out all minimum rule sets for a decision system. However, we know that in steps

(3) and (4) discrimination function is used, where a conjunctive normal form is converted to be a disjunctive normal form using absorption law and distribution law. This conversion involves problem of combination explosion and is a NP-hard problem. Therefore, the approach is suitable for large data set. Furthermore, it is unnecessary to find all minimum rule sets for a decision system in real applications. In fact, a special minimum rule set is more desired in most cases. Therefore, the time-consuming steps, conversion of conjunctive normal form to disjunctive normal form, should be abandoned in real applications.

5 A Heuristic Approach to Acquisition of a Minimum Rule Set

Suppose that $f(t_1, t_2, \dots, t_r)$ is a conjunctive normal form, where t_1, t_2, \dots, t_r are r different items that occur in the formula. They can stand for attributes or rules. Further, suppose that after using absorption law: $f(t_1, t_2, \dots, t_r) = \gamma_1 \wedge \gamma_2 \wedge \dots \wedge \gamma_k$, where γ_i is a simple disjunction, and for any different simple disjunctions γ_i and γ_j , neither $set(\gamma_i) \subseteq set(\gamma_j)$ nor $set(\gamma_j) \subseteq set(\gamma_i)$, $i, j \in \{1, 2, \dots, k\}$. Now we introduce how to obtain a special simple conjunction of $f(t_1, t_2, \dots, t_r)$, which corresponds to a reduct of attribute set or rule set.

Let us sort items t_1, t_2, \dots, t_r in an ascending order by their occurring frequencies in formula $f(t_1, t_2, \dots, t_r)$ and suppose the ascending order is $\langle t'_1, t'_2, \dots, t'_r \rangle$, where t'_r occurs most frequently in formula $f(t_1, t_2, \dots, t_r)$ and therefore it is viewed as the most important item. That is, the finally obtained simple conjunction should contain some important items like \dots, t_{r-1}, t_r and remove some unimportant items like t_1, t_2, \dots as possible. The approach to obtaining a special simple conjunction of $f(t_1, t_2, \dots, t_r)$ is called Algorithm 2, which is described as follows.

Algorithm 2. For calculating a special simple conjunction (called a reduct of $f(t_1, t_2, \dots, t_r)$ for simplicity).

- (1) let $f' = f(t_1, t_2, \dots, t_r)$
- (2) for $i = 1$ to r do
 - (3) for each γ_j in f' , if there is no such γ_j in f' that $set(\gamma_j) = \{t'_i\}$ then remove item t_i from γ_j for each γ_j , denoted by $\gamma_j = \gamma_j - \{t_i\}$;
 - (4) apply absorption law to f' ;
- (5) $set(f')$ is reduct.

Take $f(a_1, a_2, a_3, a_4, a_5) = (a_2 \vee a_3) \wedge (a_3 \vee a_4 \vee a_5) \wedge (a_1 \vee a_3) \wedge (a_2 \vee a_4) \wedge (a_1 \vee a_4 \vee a_5)$ for example. It can be observed that the occurring frequencies of a_1, a_2, a_5, a_3, a_4 are 2, 2, 2, 3, 3, respectively. According to Algorithm 2, the reduction steps are as follows:

- (1) let $f' = (a_2 \vee a_3) \wedge (a_3 \vee a_4 \vee a_5) \wedge (a_1 \vee a_3) \wedge (a_2 \vee a_4) \wedge (a_1 \vee a_4 \vee a_5)$;
- (2) remove a_1 from f' : $f' = (a_2 \vee a_3) \wedge (a_3 \vee a_4 \vee a_5) \wedge (a_3) \wedge (a_2 \vee a_4) \wedge (a_4 \vee a_5)$;
- (3) after applying absorption law: $f' = (a_3) \wedge (a_2 \vee a_4) \wedge (a_4 \vee a_5)$;

- (4) remove a_2 from f' : $f' = (a_3) \wedge (a_4) \wedge (a_4 \vee a_5)$;
- (5) after applying absorption law: $f' = (a_3) \wedge (a_4)$;
- (6) as a result $set(f') = \{a_3, a_4\}$.

By using Algorithm 2, we can modify Algorithm 1 to a heuristic approach, which is used to acquire a special minimum rule set and is described as follows:

Algorithm 3. For acquiring a special minimum rule set.

- (1) compute all different equivalence classes: $[x_1]_D, [x_2]_D, \dots, [x_n]_D$, and let $DC(DS) = \{[x_1]_D, [x_2]_D, \dots, [x_n]_D\}$;
- (2) for decision class $[x_i]_D \in DC(DS)$, find out all equivalence classes whose union is equal to $[x_i]_D$, and suppose the set of all such equivalence classes is $EC([x_i]_D)$;
- (3) for each $[x'_c] \in EC([x_i]_D)$, sort attributes occurring $f([x'_c])$ in an ascending order by their occurring frequencies, and then use Algorithm 2 to obtain a reduct of $f([x'_c])$, with which we create a rule set, denoted as $R''([x'_c])$, and finally let $R''([x_i]_D) = \cup \{R''([x'_c]) \mid [x'_c] \in EC([x_i]_D)\}$;
- (4) sort rules occurring in $f([x_i]_D)$ in an ascending order by their occurring frequencies, and then use Algorithm 2 to obtain a reduct of $f([x_i]_D)$, with which we obtain a rule set, denoted as $RED''[R''([x_i]_D)]$;
- (5) return a minimum rule set: $\cup \{RED''[R''([x_i]_D)] \mid [x_i]_D \in DC(DS)\}$.

6 Example Analysis

Consider decision system $DS = (U, C \cup D)$, where $U = \{x_1, x_2, \dots, x_6\}$, $C = \{a_1, a_2, a_3\}$ and $D = \{d\}$, presented in Table 1.

Table 1. A decision system

U	a_1	a_2	a_3	d
x_1	2	3	2	1
x_2	1	1	1	2
x_3	3	3	3	3
x_4	2	2	2	3
x_5	2	2	1	2
x_6	2	3	3	1

According to Algorithm 3, steps to obtain a special minimum rule set are as follows:

- (1) by computing, $DC(DS) = \{\{x_1, x_6\}, \{x_2, x_5\}, \{x_3, x_4\}\} = \{[x_1]_D, [x_2]_D, [x_3]_D\}$;
- (2) for decision class $[x_3]_D \in DC(DS)$, $EC([x_3]_D) = \{\{x_3\}, \{x_4\}\} = \{[x_3]_C, [x_4]_C\}$; $f([x_3]_C) = (a_1 \vee a_3) \wedge (a_1 \vee a_2 \vee a_3) \wedge (a_1 \vee a_2 \vee a_3) \wedge a_1$, and through the statistics the ascending order is $\langle a_2, a_3, a_1 \rangle$, and then by using Algorithm 2, the obtained reduct of $f([x_3]_C)$ is $\{a_1\}$; so we have $R''([x_3]_C) = \{(a_1, 3) \rightarrow (d, 3)\}$, and similarly, we have $R''([x_4]_C) = \{(a_2, 2) \wedge (a_3, 2) \rightarrow (d, 3)\}$; finally, $R''([x_3]_D) = R''([x_3]_C) \cup R''([x_4]_C) = \{(a_1, 3) \rightarrow (d, 3), (a_1, 2) \wedge (a_1, 2) \rightarrow (d, 3)\}$;

(3) by the similar way, we can compute $R''([x_1]_D)$ and $R''([x_2]_D)$. All the results are as follows:

$$\begin{aligned} R''([x_1]_D) &= \{r_1: (a_2, 3) \wedge (a_3, 2) \rightarrow (d, 1), & r_2: (a_1, 2) \wedge (a_3, 3) \rightarrow (d, 1)\} \\ R''([x_2]_D) &= \{r_3: (a_3, 1) \rightarrow (d, 2), & r_4: (a_3, 1) \rightarrow (d, 2)\} \\ R''([x_3]_D) &= \{r_5: (a_1, 3) \rightarrow (d, 3), & r_6: (a_2, 2) \wedge (a_3, 2) \rightarrow (d, 3)\} \end{aligned}$$

(4) by computing, we have: $\text{coverage}^{-1}(x_1) = \{r_1\}$, $\text{coverage}^{-1}(x_2) = \{r_3, r_4\}$, $\text{coverage}^{-1}(x_3) = \{r_5\}$, $\text{coverage}^{-1}(x_4) = \{r_6\}$, $\text{coverage}^{-1}(x_5) = \{r_3, r_4\}$, $\text{coverage}^{-1}(x_6) = \{r_2\}$. Take $RED''[R''([x_2]_D)]$ for example. Because $f([x_2]_D) = (r_3 \vee r_4) \wedge (r_3 \vee r_4)$ and the ascending order is $\langle r_3, r_4 \rangle$, so it can be found that $\{r_4\}$ is a reduct of $f([x_2]_D)$ and therefore $RED''[R''([x_2]_D)] = \{r_4\}$; by the same way, $RED''[R''([x_1]_D)] = \{r_1, r_2\}$, $RED''[R''([x_3]_D)] = \{r_5, r_6\}$;

(5) the finally obtained minimum rule set is $RED''[R''([x_1]_D)] \cup RED''[R''([x_2]_D)] \cup RED''[R''([x_3]_D)] = \{r_1, r_2\} \cup \{r_4\} \cup \{r_5, r_6\} = \{r_1, r_2, r_4, r_5, r_6\}$.

That is, $\{r_1, r_2, r_4, r_5, r_6\}$ is a special minimum rule set that obtained by Algorithm 3. We can prove that any rule in the rule set is consistent, satisfiable, reduced and for any two rules r_1, r_2 in the rule set, neither $r_1 \leq r_2$ nor $r_2 \leq r_1$. So it is a minimum rule set.

7 Conclusion

In rough set theory, many scholars pay more attention to attribute reduction, and not too much work pays attention to the acquisition of decision rules with rough set theory. Most methods use rough set theory to reduce data in the first place, and then use other algorithms, such as CART, SVM^[7], to create classifier based on the reduced data and use it to classify or predict new data. In fact, rough set theory can also be used to extract decision rules from data sets. By creating minimum rule sets, which is used as classifiers, we can also achieve better data prediction. In this paper, we propose two important algorithms, Algorithm 1 and Algorithm 3. Algorithm 1 can find out all minimum rule sets for a given data set theoretically. But it applies discrimination function to complete its computing, which is a time-consuming operation, so its disadvantage is low efficiency, especially for large data sets. By improving Algorithm 1, a heuristic approach to computing a special minimum rule set, Algorithm 3, is proposed. Using occurring frequency of attributes or rules as heuristic information, Algorithm 3 can rapidly obtain a special minimum rule set. It works far more efficiently than Algorithm 1 and has high application value.

Acknowledgement. This work is supported by the National Natural Science Foundation of China (No. 61363027) and the Natural Science Foundation of Guangxi Province, China (No. 2012GXNSFAA053225).

References

- [1] Pawlak, Z.: Rough Sets. International Journal of Computer and Information Sciences 11(5), 341–356 (1982)
- [2] Li, M., Shang, C.X., Feng, S.Z., Fan, J.P.: Quick Attribute Reduction in Inconsistent Decision Tables. Information Sciences 254, 155–180 (2014)

- [3] Liu, J., Hu, Q., Yu, D.: A Weighted Rough Set Based Method Developed for Class Imbalance Learning. *Information Sciences* 4, 1235–1256 (2008)
- [4] Meng, Z.Q., Shi, Z.Z.: A Fast Approach to Attribute Reduction in Incomplete Decision Systems with Tolerance Relation-based Rough Sets. *Information Sciences* 179, 2774–2793 (2009)
- [5] Meng, Z.Q., Shi, Z.Z.: Extended Rough Set-based Attribute Reduction in Inconsistent Incomplete Decision Systems. *Information Sciences* 204, 44–66 (2012)
- [6] Yang, H.H., Liu, T.C., Lin, Y.T.: Applying rough sets to prevent customer complaints for IC packaging foundry. *Expert Systems with Applications* 32, 151–156 (2007)
- [7] Hu, Q.H., Liu, J.F., Yu, D.: Mixed Feature Selection Based on Granulation and Approximation. *Knowledge-Based Systems* 21(4), 294–304 (2008)
- [8] Chang, L.Y., Wang, G.Y., Wu, Y.: An Approach for Attribute Reduction and Rule Generation Based on Rough Set Theory. *Journal of Software* 10(11), 1206–1211 (1999)
- [9] Shao, E.X., Zhang, L.S., Yang, Y.Z., Li, F., Yang, H., J.X.: Method of Rule Extraction Based on Rough Set Theory. *Computer Science* 38(1), 232–235 (2011)
- [10] Du, Y., Hu, Q., Zhu, P., Ma, P.: Rule Learning for Classification Based on Neighborhood Covering Reduction. *Information Sciences* 181 (2011)
- [11] Pawlak, Z., Skowron, A.: Rough Membership Functions. In: Yager, R.R., Fedrizzi, M., Kacprzyk, J. (eds.) *Advances in the Dempster–Shafer Theory of Evidence*, pp. 251–271. John Wiley and Sons (1994)