The Discussion of Certain Increment Operator in Rough Set Data Analysis

Pinjia Zhang[1], and Hongli Liang[2]

[1] Department of Electrical Engineering, Tsinghua University,Beijing 100084, China

[2] School of Information Science and Engineering, Northeastern University,Shenyang 110004, China

zpj02@mails.tsinghua.edu.cn
duduniao0104@sina.com.cn

Abstract

Two inequalities well known in the rough set theory have been modified to become equalities by using certain increment operator and uncertain decrement operator, which introduced in our previous papers. The union, intersection, complement operations have also been defined based on the equalities. Boolean algebra property can be satisfied based on these operations. Hence, many properties in rough set theory are improved and the applications of rough set theory are extended. In this paper, we will discuss the properties of certain increment operator farther in rough set data analysis (RSDA), and get some interesting and useful theorems. These theorems will extend the concept of reduct and the generation of reduct algorithm in RSDA. Particularly, a hierachical reduct algorithm (HRA) has been introduced to find the optimal reduct for huge information system.
Keyword: Certain Increment Operator; Rough Set; Data Analysis.

I. Introduction

In 1982, Z. Pawlak. introduced the rough set theory [1], which has emerged as another major mathematical tool for modeling the vagueness presented in human classification mechanism [8-10]. After more than 20 years' development in this field, rough set theory has been successfully applied in machine learning, pattern recognition, decision support systems, expert systems, data analysis, data mining and so on [2,7]. In reference [3], we have defined certain increment operator and uncertain decrement operator, which can modify inequalities $\underline{R}(X \cup Y) \supseteq \underline{R}X \cup \underline{R}Y$ and $\overline{R}(X \cap Y) \subseteq \overline{R}X \cap \overline{R}Y$ into equalities, based on which we can redefine the algebraic operation of rough sets, discuss the properties of rough sets, and prove the Boolean algebra of rough sets. In this paper we will discuss the properties of certain increment operator further. Some useful theorems will be established. More important, a hierachical reduct algorithm (HRA) is proposed.

II. Basic Concepts and Properties

In this section, we will review the definitions and properties of certain increment operator, uncertain decrement operator in reference [3], and some basic concepts of rough set data analysis [1,2].

Definition 1: Let *U* be the universe and *R* be an equivalence relation on *U*. For any $X \subseteq U$ and $x \in X$, the two sets

$$h_X(x) = \{y \mid y \in [x]_R \wedge y \notin X\} = [x]_R - X ,$$
$$l_X(x) = \{y \mid y \in [x]_R \wedge y \in X\} = [x]_R - h_X(x) , \tag{1}$$

are called the basic factor of inducing rough and the correlation basic factor of inducing rough of *X*, respectively.

Definition 2: Let *U* be the universe and *R* be an equivalence relation on *U*. For any $X \subseteq U$ and $x \in X$, the two sets

$$H(X) = \bigcup_{x \in Bn_R(X)} h_X(x) ,$$
$$L(X) = \bigcup_{x \in Bn_R(X)} l_X(x) ,$$

are called *R*-inducing rough region and *R*-inducing rough correlation region of *X*, respectively.

Definition 3: Let *U* be the universe and *R* be an equivalence relation on *U*. Let $X, Y \subseteq U$. When *X* is extended by *Y* (i.e. $X \cup Y$), $\underline{Z}_{(\cdot)}(\cdot): U \times U \to U$ is defined by

$$\underline{Z}_X(Y) = \bigcup \{[x]_R \mid x \in L(X) ,\, l_X(x) \not\subset Y ,\ h_X(x) \subseteq Y\} \ . \tag{2}$$

$\underline{Z}_{(\cdot)}(\cdot)$ is called the certain increment operator.

Definition 4: Let *U* be the universe and *R* be an equivalence relation on *U*. Let $X, Y \subseteq U$. When *X* is cut by *Y* (i.e. $X \cap Y$), $\overline{Z}_{(\cdot)}(\cdot): U \times U \to U$ is defined by

$$\overline{Z}_X(Y) = \bigcup \{[x]_R \mid x \in L(X),\ l_X(x) \cap Y = \phi,\ \ h_X(x) \cap Y \neq \phi\} . \tag{3}$$

$\overline{Z}_{(\cdot)}(\cdot)$ is called the uncertain decrement operator.

Definition 5: Let *U* be the universe and *R* be an equivalence relation on *U*. Let $X, Y \subseteq U$. We have

$\underline{R}(X \cup Y) = \underline{R}X \cup \underline{R}Y \cup \underline{Z}_X(Y)$

$\overline{R}(X \cap Y) = \overline{R}X \cap \overline{R}Y - \overline{Z}_X(Y)$

Definition 6: Let $S = \{U, A, V_a, f\}$ be an information system, where *U* is a non-empty finite set of objects, called universe, the elements in *U* are called records. *A* is the attribute set of *S*, $\forall a \in A$, V_a is the value set of attribute *a*, $f: U \times A \to V_a$ is called information function. If $A = C \cup D$, *C* and *D* are called the condition attribute and decision attribute respectively. When $C \cup D = A$ and $C \cup D = \phi$, information system *S* is also called decision system and is expressed by a decision table.

Definition 7: Let $S = \{U, C \cup D, V_a, f\}$ be a decision system, any subset of attribute $B \subseteq C$, we define equivalence relation $IND(B)$ as $IND(B) = \{(x, y) \in U \times U, \quad a(x) = a(y), \forall a \in B\}$, which is also called indiscernibility relation.

Definition 8: Let $S = \{U, C \cup D, V_a, f\}$ be an information system, if $X \subseteq U$, $B \subseteq A$, $\underline{B}(X) = \{x \in U \mid [x]_B \subseteq X\}$, $\overline{B}(X) = \{x \in U \mid [x]_B \cap X \neq \phi\}$ will denote the lower and upper approximation of *X* with respect to attribute set *B*, respectively. $POS_B(D) = \bigcup_{X \in U/D} \underline{B}(X)$ is the positive region of *D* with respect to *B* .

Definition 9: Let $S=\{U,C\bigcup D,V_a,f\}$ be an information system, if $B\subseteq C$, and for $\forall r\in B$, $POS_B(D)\neq POS_{B-\{r\}}(D)$, then B is independent; if B is independent and $POS_B(D)=POS_C(D)$, then B is a reduct of C, denoted by $RED(C)$.

Obviously, the reduct of C is not exclusive. The reducts that have the least attributes are called the minimal reduct or the optimal reduct. The attributes in all reducts of C are called the core attributes, denoted by $CORE(C)=\bigcap RED(C)$.

III. The Discussion of Certain Increment Operator in Rough Set Data Analysis

The key part of rough set data analysis is the core and reduct of rough set, which is also the main means and approaches to apply rough set to the data mining and knowledge discovering. In this part, we will discuss the performance of certain increment operator in rough set data analysis.

Let $S=\{U,C\bigcup D,V_a,f\}$ be an information system, C be a condition attribute set and D be a decision attribute set. $U/C=\{X_1,X_2\cdots X_k\}$, $U/D=\{Y_1,Y_2\cdots Y_l\}$.

If $C_0\subseteq C$, $U/C_0=\{Z_1,Z_2,\cdots Z_N\}$, information system $S_i=\{Z_i,(C-C_0)\bigcup D,V_a,f\}$ ($i=1,2\cdots N$) is called an information subsystem of information system S with respect to C_0.

We will introduce three lemmas before discussing further the theorems of certain increment operator.

Lemma 1: Supposing that attribute set $B\subseteq A$, $\forall x,y\in U$, if $[x]_B\neq[y]_B$, then $[x]_A\neq[y]_A$.

Lemma 2: $\forall x,y\in U$, if $[x]_A\neq[y]_A$, then $[x]_A\bigcap[y]_A=\phi$.

Lemma 3: If attribute sets $A\subseteq B\subseteq C$, and $POS_A(D)=POS_C(D)=M$, then we have $POS_B(D)=M$.

Theorem 1: If attribute sets $C_0\subseteq C$, $B\subseteq D$, $Y_i\in U/B$, $Y_i/C_0=\{Y_{i1},Y_{i2},\cdots Y_{im}\}$, $i=1,2\cdots l$, then $\underline{Z}_{Y_{iu}}(Y_{iv})=\phi$, where $u,v\leq m$, $u\neq v$.

Proof. $Y_i/C_0=\{Y_{i1},Y_{i2},\cdots Y_{im}\}$, hence $\forall x_s\in Y_{iu},x_t\in Y_{iv}$, $u,v\leq m$, and $u\neq v$, $[x_s]_{C_0}\neq[x_t]_{C_0}$. From lemma 1, we have $[x_s]_C\neq[x_t]_C$. From lemma 2, we can get $[x_s]_C\bigcap[x_t]_C=\phi$. For the randomicity of t, we can obtain $[x_s]_C\bigcap Y_{iv}=\phi$. According to definition 1.1, $h_{Y_{iu}}(x_s)\subseteq[x_s]_C$. Hence $h_{Y_{iu}}(x_s)\bigcap Y_{iv}=\phi$. Moreover, according to definition 1.3, $\underline{Z}_{Y_{iu}}(Y_{iv})=\phi$.

This theorem indicates that in rough set data analysis, if the definition region of certain increment operator is the equivalence class of the universe U with respect to decision attribute set D, then certain increment operator equals zero. In such case, we have the following theorem.

Theorem 2: Supposing that attribute set $C_0\subseteq C$, $Y_i\in U/D$, $Y_i/C_0=\{Y_{i1},Y_{i2},\cdots Y_{im}\}$, $i=1,2\cdots l$, then $\underline{C}(Y_{iu}\bigcup Y_{iv})=\underline{C}(Y_{iu})\bigcup\underline{C}(Y_{iv})$.

Proof. According to theorem 1, $\underline{Z}_{Y_{iu}}(Y_{iv})=\phi$ ($u,v\leq m$, $u\neq v$). From definition 1.5, we can get $\underline{C}(Y_{iu}\bigcup Y_{iv})=\underline{C}(Y_{iu})\bigcup\underline{C}(Y_{iv})$.

Inference: If attribute set $C_0\subseteq C$, $Y_i\in U/D$, $Y_i/C_0=\{Y_{i1},Y_{i2},\cdots Y_{im}\}$, then $\underline{C}(Y_i)=\bigcup_{Y_{iu}\in U/C_0}\underline{C}(Y_{iu})$.

Theorem 3: If attribute set $C_0\subseteq C$, $Y_i\in U/D$, $Y_i/C_0=\{Y_{i1},Y_{i2},\cdots Y_{im_i}\}$, $i=1,2\cdots l$, then $POS_C(D)=POS_C(D\bigcup C_0)$.

Proof. From the inference, $\underline{C}(Y_i)=\bigcup_{Y_{iu}\subseteq U/C_0}\underline{C}(Y_{iu})$, hence, $POS_C(D)=\bigcup_{Y_i\in U/D}\underline{C}(Y_i)$ = $\bigcup_{Y_i\in U/D}\bigcup_{Y_{iu}\in Y_i/C_0}\underline{C}(Y_{iu})=\bigcup_{Y_{iu}\in U/(D\bigcup C_0)}\underline{C}(Y_{iu})=POS_C(D\bigcup C_0)$.

Theorem 4: In an information subsystem S_i, for any $x\in Z_i$, we have $[x]_{C-C_0}=[x]_C$.

Proof. Since $C-C_0\subseteq C$, we have $[x]_C\subseteq[x]_{C-C_0}$.

For any $x, y \in [x]_{C-C_0}$, $\forall c \in C - C_0$, we have $c(x) = c(y)$. Furthermore, $x, y \in Z_i$, we have $c'(x) = c'(y)$, $\forall c' \in C_0$. Therefore for any $c \in C$, we can obtain $c(x) = c(y)$, then $x, y \in [x]_C$, i.e. $[x]_{C-C_0} \subseteq [x]_C$. So $[x]_{C-C_0} = [x]_C$.

Theorem 5: In information subsystem S_i, for any $Y_j \in Z_i / D$, we can get $\underline{C - C_0}(Y_j) = \underline{C}(Y_j)$.

Proof: According to the definition of lower approximation and theorem 4, $\underline{C - C_0}(Y_j) = \{x \mid x \in Z_i, [x]_{C-C_0} \subseteq Y_j\} = \{x \mid x \in Z_i, [x]_C \subseteq Y_j\} = \underline{C}(Y_j)$.

Theorem 6: If the positive region of information subsystem S_i is $POS^{(i)}_{C-C_0}(D)$, then $\bigcup_{Z_i \in U/C_0} POS^{(i)}_{C-C_0}(D) = POS_C(D)$.

Proof. $\bigcup_{Z_i \in U/C_0} POS^{(i)}_{C-C_0}(D) = \bigcup_{Z_i \in U/C_0} \bigcup_{Y_j \in Z_i/D} \underline{C - C_0}(Y_i) = \bigcup_{Y_j \in U/(D \cup C_0)} \underline{C}(Y_i) = POS_C(D \cup C_0) = POS_C(D)$.

Theorem 7: Let C_i be a reduct of the information subsystem S_i, then $\bigcup_{Z_i \in U/C_0} POS^{(i)}_{\wedge C_i}(D) = POS_C(D)$, where $\wedge C_i$ denote the conjunction of C_i.

Proof. If C_i is a reduct of information subsystem S_i, then $POS^{(i)}_{C-C_0}(D) = POS^{(i)}_{C_i}(D)$. For any C_i, we have $C_i \subseteq \wedge C_i \subseteq C - C_0$. According to lemma 3, we know $\bigcup_{Z_i \in U/C_0} POS^{(i)}_{C_i}(D) = \bigcup_{Z_i \in U/C_0} POS^{(i)}_{\wedge C_i}(D)$. From theorem 6, we have $\bigcup_{Z_i \in U/C_0} POS^{(i)}_{C-C_0}(D) = \bigcup_{Z_i \in U/C_0} POS^{(i)}_{C_i}(D) = \bigcup_{Z_i \in U/C_0} POS^{(i)}_{\wedge C_i}(D) = POS_C(D)$.

Remark: From the definition of reduct, we know that reduct is a set of some attributes. However, from the viewpoint of discernibility matrix, reduct is the conjunction of some attributes. This paper has used both of the two points of view, but it's not difficult to distinguish them according to context. Thus, in this paper, we use the same expression of reduct, without distinguishing whether it express a set or a conjunction form.

Theorem 8: Let C_i be a reduct of information subsystem S_i. If C_0 is the core of information system *S*, then $C_0 \cup (\wedge C_i)$ is a reduct of information system *S*.

Proof. To prove that $C_0 \cup (\wedge C_i)$ is a reduct of information system *S*, we only need to prove two conditions: 1) $\bigcup_{Z_i \in U/C_0} POS^{(i)}_{\wedge C_i}(D) = POS_C(D)$; 2) $\forall r \in (\wedge C_i)$, $\bigcup_{Z_i \in U/C_0} POS^{(i)}_{\wedge C_i - r}(D) \neq POS_C(D)$. As C_0 is the core of information system S, the element of C_0 can not be taken out. 1) has been proven by theorem 7, we will prove 2). $\forall r \in (\wedge C_i)$, so there would exist at least one C_i, and $r \in C_i$. If *r* is taken out from $\wedge C_i$, according to the definition of reduct, we have $POS^{(i)}_{C_i}(D) \neq POS^{(i)}_{C_i - r}(D)$, i.e. $POS^{(i)}_{\wedge C_i}(D) \neq POS^{(i)}_{C_i - r}(D)$. Hence, $\bigcup_{Z_j \in U/C_0} POS^{(i)}_{\wedge C_i}(D) \neq \bigcup_{Z_i \in U/C_0} POS^{(i)}_{C_i - r}(D)$, i.e. $\bigcup_{Z_i \in U/C_0} POS^{(i)}_{\wedge C_i - r}(D) \neq POS_C(D)$. So, $C_0 \cup (\wedge C_i)$ is a reduct of information system S.

Theorem 9: Let C_i be the minimal reduct of information subsystem S_i, then the minimal reduct of information system will need $\max\{|C_i|\}$ attributes, on the basis of attribute set C_0, denoted as $ZX(S)$, where $|C_i|$ denotes the number of attributes in C_i.

Proof. According to theorem 8, the reduct of information system can be denoted as $C_0 \cup (\wedge C_i)$. For any C_i, there is $|C_i| \leq |\wedge C_i|$. Hence, the minimal reduct of information system will need at least $\max\{|C_i|\}$ attributes, on the basis of attribute set C_0.

Theorem 10: Let $\{C_{ij}, j = 1, 2 \cdots r_i\}$ be all the minimal attribute reducts of information subsystem S_i; r_i is the number of the minimal attribute reducts, C_{ij} denotes the j-th the minimal attribute reduct of information subsystem S_i. We can construct discernibility function $\wedge_{S_i \in IS} (\vee_{j=1}^{r_i} C_{ij})$ on the basis of $\{C_{ij}, j = 1, 2 \cdots r_i\}$. According to the discernibiliy function we construct the conjunction normal forms, denoted by B_k, $k=1,2,\cdots n$. Then the minimal attribute reduct of information system *S* will still need $\min(|B_k|)$ attributes at most, on the basis of attribute set C_0, denoted as $ZD(S)$.

Proof. As $B_k = \wedge C_{ij}$, C_{ij} is a minimal attribute reduct of information subsystem S_i. From theorem 8, we know $C_0 \cup B_k$ is a reduct of information system *S*. Moreover, $|C_0 \cup B_k| \geq |C_0| + \min|B_k|$, $k=1,2,\cdots n$, thus the minimal attribute reduct of information system *S* will still need $\min(|B_k|)$ attributes, on the basis of attribute set C_0.

Theorem 11: If *B* is a reduct of information system *S*, and $|B| \leq ZD(S) + |C_0|$, then there must be a C_i, $|C_i| \leq ZD(S)$, so that $B = C_0 \cup (\wedge C_i)$.

Proof. As *B* is a reduct of information system *S*, there must be $C_i \subseteq B$, which is a reduct of information subsystem S_i. In addition, $|B| \leq ZD(S) + |C_0|$ and $C_0 \subseteq B$, we have $|C_i| \leq ZD(S)$. Hence $C_0 \cup (\wedge C_i) \subseteq B$. As C_i is a reduct of information subsystem S_i, according to theorem 8, $C_0 \cup (\wedge C_i)$ is a reduct of information system *S*. From the definition of reduct, we can obtain $C_0 \cup (\wedge C_i) = B$.

Ⅳ. The Minimal Reduct Algorithm Based on HAR

The information system based on rough set can be separated into information system with decision (DIS) and information system without decision (NDIS). Computing the reduct of NDIS is to obtain the minimal attribute set who can distinguish the records according to information system, the basis of which is maintenance indiscernibility. The purpose of computing the reduct of DIS is to obtain the most simplified decision. The basis of the computing the reduct of NDIS is the rule of compatibility. If we consider the attribute set of NDIS as the decision set of DIS, then NDIS can be considered as DIS, thus we deal with DIS only in this paper. According to whether or not having core, the information system based on rough set can be separated into core-information system (CIS) and nor-core-information system (NCIS). We will discuss them respectively as following.

A. The Minimal Reduct Algorithm of CIS

Let $S=(U, C \cup D, V_a, f)$ is a information system, where $U=\{x_1, x_2, \cdots x_n\}$ is universe; *C* is the condition attribute set, *D* is the decision attribute set, V_a is the set constructed by all attribute values, f : $U \to V_a$ is the information function.

The process of the minimal reduct of CIS is as follows:

1. To get the core of *S*, denoted by C_0.
2. Classify *U* according to C_0, $U/C_0 = \{Z_1, Z_2, \cdots Z_k\}$. Separate *S* into *k* sub-information systems (SIS) $S_i = (Z_i, (C-C_0) \cup D, V_{a'}, f)$, $i=1,2,\cdots k$, where $a' \in C - C_0$.
3. Deduce all the reduct of $S_i = (Z_i, (C-C_0) \cup D, V_{a'}, f)$, $i=1,2,\cdots k$.
4. Use theorem 9 and theorem 10 to get $ZX(S)$ and $ZD(S)$.
5. Judge whether or not $ZX(S)$ equals to $ZD(S)$. If they are not equal, then we compute next step. If they are equal, stop. Then the collection of the minimal reduct of *S* is the union of the core and the reduct having $ZX(S)$ attributes.
6. Select the reducts of S_i, $i=1,2,\cdots k$, whose number of attributes is less than $ZD(S)$ in step 3, and then use them to construct discernibility function and then solve it. In the process of solving, all conjunction normal forms whose number of elements is greater than $ZD(S)$ are deleted. Then the collection of the minimal reduct of *S* is the union of the core and the reduct which having the least number of elements in the conjunction normal forms are selected in results.

In step 3, if there is a SIS S_j which leads $\forall x_m, x_n \in Z_j, \forall d \in D$, then we have $d(x_m) = d(x_n)$, that is to say, there is a $Y_k \in U/D$, which leads $Z_j \subseteq Y_k$, that is to say that Z_i relates with C_0 only, and does not relate with $C - C_0$. Thus we can suppose that the reduct of S_i are empty set ϕ.

Using this method we can compute the minimal reduct of information system, but they are not all the minimal reducts. If we want obtain all the minimal reducts, we need only get rid of step 5, then change “less than” into “less than or equal to” in step 6.

SIS S_i is NCIS. Now we use discernibility matrix to prove it. Let S_i have core. Then there exit two records $x_u, x_v \in Z_i$. There is one and only one attribute a in $C - C_0$ that leads to $a(x_u) \neq a(x_v)$. And let $a' \in C - C_0 - \{a\}$, then $a'(x_u) = a'(x_v)$. Because $Z_i \in U / C_0$, then $\forall a_0 \in C_0$, $a_0(x_u) = a_0(x_v)$. That is, there is only one attribute a leads $a(x_u) \neq a(x_v)$, thus a is the core of information system. This is conflict with $a \in C - C_0$. So S_i is NCIS.

Because SIS is NCIS and some information systems are the NCISs, so it is necessary to discuss how to compute the reduct of NCIS.

B. The Minimal Reduct Algorithm of NCIS

The character of NCIS is that there are many attributes but few records in information system. If NCIS is small, then we can use discernibility matrix to compute the minimal reduct directly. In fact, we can not assure the scale of NCIS does not exceed the applied range of discernibility matrix, and thus under the enlightenment of computing the reduct of CIS, we get the minimal reduct algorithm of CIS.

According to discernibility matrix, the cores of information system are the entries having single attribute in discernibility matrix. Because the relativity positive region will change if we get rid of these entries, thus they must exit in every reduct.

However the entries having multi-attribute in discernibility matrix do not have this character. But there is one attribute at least in every entry having multi-attribute, which is the element of some minimal reduct.

The algorithm for NCIS is as follows:

1. To get the discernibility matrix of NCIS.
2. Find out a entry that has the least attributes in discernibility matrix.
3. Extract every attribute from this entry in return as preparative core, then according to the minimal reduct algorithm of CIS to compute the minimal reduct.
4. The reduct having the least number in all the minimal reduct based on every preparative core is the minimal reduct of NCIS.

In the minimal reduct algorithm of CIS, we propose the concept of preparative core. The minimal reduct of information system based on preparative core is equivalent to the minimal reduct of information system based on hypothetic condition, when it is assumed that preparative core exits in the minimal reduct. From the above analysis, we know that this condition is not always satisfied. However, if we extract every attribute in the entry having the least number of attributes in discernibility matrix in return as preparative core, then this condition will be satisfied once at least. That is to say, the results that are computed by step 3 and step 4 in above algorithm must be the minimal reduct.

In CIS, if the SISs is still large, then it will produce data overflow in the process of computing. From above section we know SISs is the NCISs. So it can be separated further by the minimal reduct algorithm of NCIS until it can be computed.

Ⅴ. Example

Take CTR (Car Test Results) database [5] for example. Using the minimal reduct algorithm of CIS proposed in the paper to compute its minimal reduct.

In CRT database, 10 attributes are given in Table 1: ***a***—the whole length, ***b***—the number of bicycle pump, ***c***—existence of turbocharger, ***d***—types of fuel system, ***e***—tonnage of engine, ***f***—compressibility, ***g***—power, ***h***—types of driver, ***i***—weight, ***y***—miles.

Attribute values: c—roadlouse, s—mini car, sm—small, y—yes, n—no, E—2-BBL, m—middle, ma—manual, h—high, he—weight, l—light, lo—low, a—auto

Information system is $S=\{U,C\bigcup D,V_a,f\}$, where $U=\{1,2,\cdots 21\}$, $C=\{a,b,\cdots i\}$, $D=\{y\}$.

1. To get core of *S*. $core(C)=\{d,i\}$.
2. According to $\{d,i\}$ we separate *S*, $U/\{d,i\}$ ={{1,2,3,5,12,14,16,17,18,19, 20} $\{4,11\},\{6\},\{7\},\{8\},\{9,10,13,15,21\}\}$. According to theorem 8, we only need to consider SISs constructed by the equivalence class whose records are larger than or equal to 2. In this example, there are 3 equivalence classes whose records larger than 2, mark them Z_1,Z_2,Z_6 respectively. And because all records in equivalence class Z_2 are equal, we only need to consider two SISs $S_1=(Z_1,C-\{d,i\},V_a,f)$ and $S_6=(Z_6,C-\{d,i\},V_a,f)$.
3. The reducts of S_1 are: $\{a,e\}$, $\{a,f,g\}$, $\{a,b,f\}$, $\{a,b,g\}$, $\{b,e,g\}$, $\{b,f,g\}$, $\{b,c,e,h\}$, $\{b,c,f,h\}$, $\{a,b,c,h\}$.

   The reducts of SIS S_6 are: $\{a\},\{e,f\},\{f,g\}$.
4. According to theorem 9, we have *ZX*=2. According to theorem 10, we have $a\wedge ae=ae$, *ZD*=2. Thus *ZD*=*ZX*, hence $\{d,i\}\bigcup\{a,e\}=\{a,d,e,i\}$ is a minimal reduct of *S*.

If we want to get all the minimal reduct, it need still the reduct whose attributes are lower than or equal to *ZD* in SISs S_1 and S_6 to construct discernibility function $(a\vee ef\vee fg)\wedge ae=ae\vee aef\vee aefg$, the reduct whose the number of attributes equal to 2 have only one. Thus, there is only one minimal reduct in *S*, that is $\{d,i\}\bigcup\{a,e\}=\{a,d,e,i\}$.

Table 1. CRT Information System

U/C	*a*	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*y*
1	c	6	y	E	m	h	h	a	m	M
2	c	6	n	E	m	m	h	ma	m	M
3	c	6	n	E	m	h	h	ma	m	M
4	c	4	y	E	m	h	h	ma	l	H
5	c	6	n	E	m	h	m	ma	m	M
6	c	6	n	B	m	m	m	a	he	Lo
7	c	6	n	E	m	m	h	ma	he	Lo
8	s	4	n	B	sm	m	lo	ma	l	H
9	c	4	n	B	sm	h	lo	ma	m	M
10	c	4	n	B	sm	h	m	a	m	M
11	s	4	n	E	sm	h	lo	ma	l	H
12	s	4	n	E	m	m	m	ma	m	H
13	c	4	n	B	m	m	m	ma	m	M
14	s	4	y	E	sm	h	h	ma	m	H
15	s	4	n	B	sm	m	lo	ma	m	H
16	c	4	y	E	m	m	h	ma	m	M
17	c	6	n	E	m	m	h	a	m	M
18	c	4	n	E	m	m	h	a	m	M
19	s	4	n	E	sm	h	m	ma	m	H
20	c	4	n	E	sm	h	m	ma	m	H
21	c	4	n	B	sm	h	m	ma	m	M

From this example we can see that using the concepts of *ZD*, *ZX* to compute the minimal reduct can minish the searching range conspicuously. In this example there are only two SISs. If there are more SISs, then the degree of decrease will be more conspicuously.

Ⅵ. Conclusion

In this paper, we find the condition that the certain increment operator equals zero in the aspect of rough set data analysis and furthermore develop the theory and introduce a algorithm of rough set reduct based on this condition. Firstly, transform the huge information system into smaller manageable information subsystem; secondly, reduce the information subsystem; and lastly, synthesize the results of information subsystem reduct, and obtain the reduct of the huge information system. All of which have been proven sufficiently by the 11 theorems in this paper. We believe that the discussion of certain increment operator is worthwhile, which would propel the development of rough set data analysis theory. However, for some special information system, the users or decision makers may consider some special conditions which lead the minimal reduct is not always necessary for users. For this problem, we will use the idea of hierachical reduct and the concepts of the minimum and the maximum to discuss it according to user's requirement in latter works.

References

[1] Z. Pawlak, *Rough Sets*. Theoretical Aspects of Reasoning about Data. Kluwer, Dordrecht (1991).

[2] Z. Pawlak, "Rough Set Theory and Its Applications to Data Analysis". Cybernetics and Systems: An International Journal, Vol. 29. (1998) 661-688.

[3] H. Zhang, H. Liang, and D.Liu, "Two New Operators in Rough Set Theory with Applications to Fuzzy Sets". Information Science, Vol. 166. (2004) 147-165.

[4] H. Liang, H. Zhang, and D. Liu, "Roughness of Fuzzy Set Based On Two New Operators". Proceedings of the IEEE International Conference on Fuzzy Systems, Budapest, Hungary (2004) 583-586.

[5] W. Zhang, W. Wu, J. Liang, and D. Li, *Rough Set Theory and Methods*. Science Press (2001) (in Chinese).

[6] A. Skowron, and C. Rauszer, "The Discernibility Matrices and Functions in Information Systems". In: R. Slowinski (ed), Intelligent Decision Support. Handbook of Applications and Advanced of the Rough Set Theory. Kluwer, Dordrecht (1992) 331–362.

[7] A. Skowron, and L. Polkowski, "Synthesis of Decision Systems from Data Tables". In: T.Y. Lin and N. Cercone (eds.), Rough Sets and Data Mining: Analysis for Imprecise Data, Kluwer Academic Publishers, Boston (1997) 259–300.

[8] A. Skowron, and R. Swiniarski, "Information Granulation and Pattern Recognition". In: Polkowski and Skowron (eds.), Rough-Neuro Computing, Physica-Verlag, Berlin (2002) 636–670.

[9] J. Stepaniuk, "Approximation Spaces, Reducts and Representatives". In: Polkowski and Skowron (eds), Rough Sets in Knowledge Discovery, Physica-Verlag, Heidelberg (1998) 109–126.

[10] H. Tanaka, and Y.Maeda, "Reduction Methods for Medical Data". In: Polkowski and Skowron (eds), Rough Sets in Knowledge Discovery, Physica-Verlag, Heidelberg (1998) 295–306.

Pinjia Zhang was Born in Jilin City, Jilin Province, in 1984. He currently is a undergraduate student in Electrical Engineering Department, Tsinghua University. His current research interest incudes: rough set,system control, signal processing.

Hongli.liang is a PhD student in the Chinese Academy of Sciences, working in the Laboratory of Complex Systems and Intellgence Science, Institute of Automation. His research interests include rough and fuzzy set theory, data mining and knowledge discovery, pattern recognition, learning theory, and soft computing.