Attribute Reduction and Information Granularity*

Li-hong Wang
School of Computer Engineering and Science, Shanghai University,
Shanghai, 200072, P.R.C
School of Computer Science and Technology, Yantai University,
Yantai, 264005, P.R.C
and
Geng-feng Wu
School of Computer Engineering and Science, Shanghai University,
Shanghai, 200072, P.R.C

ABSTRACT

In the view of granularity, this paper analyzes the influence of three attribute reducts on an information system, finding that the possible reduct and $\mu-$decision reduct will make the granule view coarser, while discernible reduct will not change the granule view. In addition, we investigate the combination of reducts from two partial information systems in parallel or in incremental data mining and urge that the union of partial possible reducts can be regarded as a possible reduct for union of partial information systems.

Key words: Rough Set, Attribute Reduct, Information Granularity, Reducts Combination

1. INTRODUCTION

It is necessary to approximate a colossal information system in order to simplify the knowledge discovery from it. An Information System (*IS*) may be denoted as *IS*=(*O, AT*), where *O* is the non-empty set of objects, and *AT* is a non-empty set of attributes. In order to make some approximation, we collect similar objects from the set *O* to form a subset and name it as a granule. In a granule, one object is regarded as the same as the others because the inherent difference between two objects disappears when they are assigned to the same granule. The information system may be divided or covered by the set consisted of these granules, which gives an approximation to the *IS* and can be named as a granule view of it. The types of similarities range from simple equivalence relations, tolerance relations to reflective binary relations, etc. The granule view of information system depends on the similarity used to form a granule (Y.Y.Yao, 2001, Skowron, Stepaniuk, 2001).

As an important approach to simplify an *IS* with many attributes, the attribute reduct based on rough set theory can effectively discard abundant information and induct some decision rules from data, attracting many researchers to focus on it (Kryszkiewicz, 2001, Chang, 1999, Wang, 1998, Miao, 1997). Many definitions on attribute reduct have been proposed and some equivalence among them has been proved (Kryszkiewicz, 2001). However, we are not quite sure about the influence of reduct on an information system or on a granule view of it. This paper will discuss these problems in the view of information granularity.

Additionally, the restriction of memory and serial computation in single CPU make it difficult to give an effective reduct for a gigantic information system. Generally, it is done in parallel, which means to divide

This work was supported by Science and Technology Commission of Shanghai Municipality, No.705931

the whole *IS* to several parts, each of which is assigned to a node in parallel computer or to a computer in a network and then combine the results from each computer to make a reduct for the primary *IS (Scotney, McClean,1999)*. Usually it is the same case in an incremental data mining for historical and present data, that is, how to combine the existed mining result with the new information derived from present data to form the latest knowledge.

The paper is organized as follows: Section 2 gives some basic notations in rough set theory as well as three definitions for attribute reduct, and Section 3 shows the change of granularity of an information system after reduct, and Section 4 proves that the union of partial possible reducts can be regarded as a possible reduct for union of partial information systems, but it is not true for $\mu-$ decision reduct or discernible reduct.

2. ATTRIBUTE REDUCT DEFINITIONS

In this section, some basic rough set notations will be introduced, and several attribute reduct definitions will be simply described (Kryszkiewicz, 2001).

Def. 1 Indiscernible relation

For an information system *IS=(O,AT),* each subset of attributes $A \subseteq AT$ determines an indiscernible relation *IND(A)* as follows:

$$IND(A) = \{(x, y) \in O \times O \mid a \in A, a(x) = a(y)\}$$

which means object *x* and *y* have the same value at any attribute in *A*.

Def. 2 Equivalence class

According to definition 1, *IND(A)* is an equivalence relation, which can partition objects *O* into equivalence classes. Let $I_A(x)$ denote the equivalence class determined by *x* on *A*.

$$I_A(x) = \{y \in O \mid (x, y) \in IND(A)\}$$

Each equivalence class can be regarded as a granule.

Def. 3 Lower/Upper approximation of *X*

Let $X \subseteq O$, we define $\underline{A}_X = \{x \in O \mid I_A(x) \subseteq X\}$ as Lower approximation of *X*, and $\overline{A}_X = \{x \in O \mid I_A(x) \cap X \neq \phi\}$ as Upper approximation of *X*.

Def. 4 Membership function

A membership function $\mu_X^A : O \to [0,1], A \subseteq AT$ is defined as follows:

$$\mu_X^A(x) = \frac{|I_A(x) \cap X|}{|I_A(x)|} \quad \text{for any } x \in O$$

Decision Table*(DT)* is an information system $DT = (O, AT \cup \{d\})$, where *d*, $d \notin AT$ is a distinguished attribute called decision, and the elements of *AT* are called conditions. A decision class $X_i = \{x \in O \mid d(x) = i\}$ is a set, in which each element has the same decision.

Three common attribute reducts are defined as follows:

Def. 5 Possible reduct

$A \subseteq AT$ is a possible reduct of *DT* for x, $x \in O$, if and only if *A* is a set such that

$$I_A(x) \subseteq \overline{AT}_{Xd(x)} \qquad (1)$$

A is called minimal possible reduct, if it is the minimal set satisfied Eq.(1). With the decrease of the cardinality of *A*, $I_A(x)$ enlarged. However, Eq.(1) sets an upper bound for $I_A(x)$ to confine the granule which contains object *x*.

$A \subseteq AT$ is a possible reduct of *DT* if and only if *A* is a set such that Eq. (1) holds true for every *x* in *O*. There are other reducts such as approximate reduct and generalized decision reduct, which are equivalent to possible reduct (Kryszkiewicz, 2001).

Def. 6 $\mu-$ decision reduct

$A \subseteq AT$ is a $\mu-$ decision reduct of DT for x, $x \in O$, if and only if A is a set such that

$$\mu_d^A(x) = \mu_d^{AT}(x) \quad (2)$$

where $\mu_d^A(x) = (\mu_{d1}^A(x),...,\mu_{dn}^A(x))$, $\mu_{dj}^A(x)$ denotes $\mu_{Xdj}^A(x)$ and X_{dj} is the *jth* decision class. Eq.(2) requires the membership of *x* to all decision classes preserved.

A is called minimal $\mu-$ decision reduct if it is the minimal set satisfied Eq. (2).

$A \subseteq AT$ is a $\mu-$ decision reduct of *DT* if and only if *A* is a set such that Eq. (2) holds true for every *x* in *O*. There is $\mu-$ reduct which is equivalent to $\mu-$ decision reduct (Kryszkiewicz, 2001).

Def. 7 Discernible reduct

For an information system *IS=(O,AT)*, $A \subseteq AT$ is the discernible reduct for $x \in O$ if and only if $I_A(x) = I_{AT}(x)$. $A \subseteq AT$ is the discernible reduct of *DT* if and only if $I_A(x) = I_{AT}(x)$ holds true for any object *x*. Discernible reduct preserves the distinction between two different objects in *IS*.

3. INFORMATION GRANULARITY DIFFERENCE AFTER ATTRIBUTE REDUCT

Attribute reduct simplifies an information system by discarding some redundant attributes. In the view of approximation, the information system is reduced to a granule view of it. However, it is necessary to analyze the approximation by these reducts. We proved that possible reduct and $\mu-$ decision reduct lead to a coarser view of original system, while the discernible reduct does not change the granularity of an information system.

The variation of granules after possible reduct

First, we define granules as equivalence classes of objects, that is, for *x*, $x \in O$, $I_{AT\cup\{d\}}(x)$ is a granule embracing *x* before reduct while $I_{A\cup\{d\}}(x)$ is the granule after reduct. For any $y \in I_{AT\cup\{d\}}(x)$, its conditions and decision are same with those of *x*. By definition 5, possible reduct *A* should satisfy that $I_A(x) \subseteq \overline{AT}_{Xd(x)}$, while $\overline{AT}_{Xd(x)} = \{y \in O \mid I_{AT}(y) \cap X_{d(x)} \neq \phi\}$ requires the element *y* in $I_A(x)$ either has the same decision with *x* or there exist one element, which have the same conditions with those of *y* and whose decision is *d(x)*. The loose requirement makes $I_{A\cup\{d\}}(x)$ larger than $I_{AT\cup\{d\}}(x)$ in cardinality and coarser in granularity. For example, table 1 is an information system IS_1, we can get a coarser approximation for it by possible reduct.

Table 1 Information System IS_1

No.	a1	a2	a3	a4	a5	d
1	1	2	3	4	5	**1**
2	0	1	0	1	0	**2**
3	1	2	1	0	0	**1**
4	1	0	1	0	0	**2**

By definition 5, *A={a1,a2}* is a possible reduct of *AT={a1,...,a5}*. Let's regard $I_{AT\cup\{d\}}(x)$ as the granule embracing *x* before reduct and $I_{A\cup\{d\}}(x)$ as the granule after reduct. From this example we can find that the granule view of the system has been changed, the later view is coarser.

The granule view before reduct:

$$I_{AT\cup\{d\}}(1) = \{1\},\ I_{AT\cup\{d\}}(2) = \{2\},$$

$$I_{AT\cup\{d\}}(3)=\{3\},\ I_{AT\cup\{d\}}(4)=\{4\}.$$

The granule view after reduct:

$$I_{A\cup\{d\}}(1)=I_{A\cup\{d\}}(3)=\{1,3\},$$

$$I_{A\cup\{d\}}(2)=\{2\},\ I_{A\cup\{d\}}(4)=\{4\}.$$

The variation of granules after $\mu-$ decision reduct

The granule is defined as above, according to definition 6, reduct *A* should satisfy that

$$\mu_d^A(x)=\mu_d^{AT}(x)$$

that is,

$$(\mu_{d1}^A(x),...,\mu_{dn}^A(x))=(\mu_{d1}^{AT}(x),...,\mu_{dn}^{AT}(x))$$

and further

$$\frac{|I_A(x)\cap X_{di}|}{|I_A(x)|}=\frac{|I_{AT}(x)\cap X_{di}|}{|I_{AT}(x)|}$$

for *i=1..n*

Table 2 is an example for $\mu-$ decision reduct. Through definition 6,we know that *A={a1,a2}* is a $\mu-$decision reduct for *AT={a1,...,a5}*. The variation of granule view shows the approximation influence of $\mu-$decision reduct.

Table 2 Information System IS_2

No.	a1	a2	a3	a4	a5	d
1	1	2	0	1	2	**1**
2	0	1	2	3	4	**2**
3	1	2	1	0	1	**1**
4	0	1	2	5	0	**2**

The granule view before reduct:

$$I_{AT\cup\{d\}}(1)=\{1\},\ I_{AT\cup\{d\}}(2)=\{2\},$$

$$I_{AT\cup\{d\}}(3)=\{3\},\ I_{AT\cup\{d\}}(4)=\{4\}.$$

The granule view after reduct:

$$I_{A\cup\{d\}}(1)=I_{A\cup\{d\}}(3)=\{1,3\}$$

$$I_{A\cup\{d\}}(2)=I_{A\cup\{d\}}(4)=\{2,4\}$$

These examples show that possible reduct and $\mu-$decision reduct will make the granule view of an information system coarser.

Discernible reduct preserves the granule view

By definition 7, the unchanged equivalence classes of an information system after discernible reduct prevent the granule view from varying.

4. REDUCT COMBINATION

In parallel or in incremental data mining, partial mining results should be integrated to get present efficient knowledge *(Scotney, McClean,1999)*. We recommend the union of partial reducts as a reduct of whole information system, but this conclusion holds true only for possible reduct.

Theorem 1 Let $IS_1=(O_1,AT\cup\{d\})$ and $IS_2=(O_2,AT\cup\{d\})$ be two homogenous information system, and attributes set A_1/A_2 is the possible reduct for IS_1/IS_2, then $A_1\cup A_2$ is a possible reduct for $IS=(O_1\cup O_2,AT\cup\{d\})$, but this is not true for $\mu-$decision reduct or discernible reduct.

Proof:

Let A_1 be a possible reduct for IS_1, then $I_{A1}(x)\subseteq\overline{AT}_{Xd(x)}$ holds true in IS_1, which can be denoted as $I_{A1}(x)\subseteq(IS_1)\overline{AT}_{Xd(x)}$, and we know that $I_{A1\cup A2}(x)\subseteq I_{A1}(x)$ holds true for any subset A_2, and $(IS_1)\overline{AT}_{Xd(x)}\subseteq(IS_1\cup IS_2)\overline{AT}_{Xd(x)}$ holds true, then we can get $I_{A1\cup A2}(x) \subseteq (IS_1 \cup IS_2)\overline{AT}_{Xd(x)}$, which means $A_1 \cup A_2$ is a possible reduct of information system $IS = (O_1 \cup O_2, AT \cup \{d\})$.

However, this conclusion is not true for $\mu-$decision reduct or discernible reduct. Here is an example for $\mu-$decision reduct.

Table 3 Information System IS_3 and IS_4

No.	a1	a2	a3	a4	a5	d
1	1	2	1	2	3	**1**
2	1	3	1	4	6	**1**
3	1	3	2	2	8	**2**
4	0	2	1	2	3	**3**
5	1	3	1	4	6	**2**
6	1	2	7	8	9	**1**
7	1	2	0	2	3	**3**

In table 3, *{a5}* can be regarded as a $\mu-$decision reduct in *IS_3=({1,2,3},{a1,...a5,d})*, and *{a2,a3}* as a $\mu-$decision reduct in *IS_4=({4,5,6,7}, {a1,.. .a5,d}),* while $A=\{a2,a3,a5\}=\{a5\}\cup\{a2,a3\}$ is not a $\mu-$decision reduct for $IS_3 \cup IS_4$, because for object No.1

$$\mu_d^A(1) = (0.5,0,0.5) \qquad \mu_d^{AT}(1) = (1,0,0)$$

$$\mu_d^A(1) \neq \mu_d^{AT}(1)$$

which implies the distribution in decision classes has been changed by the reduct.

Similarly, table 4 shows an example for discernible reduct. By definition 7, *{a3,a4}* is a discernible reduct for *IS_5 = ({1,2,3,4},{a1,...a5,d})*, and *{a5,d}* is a discernible reduct for *IS_6=({5,6,7,8},{a1,...a5,d})*. However, $\{a3,a4,a5,d\}=\{a3,a4\}\cup\{a5,d\}$ is not a discernible reduct for $IS_5 \cup IS_6$ because object No.1 and No.5 become indiscernible after reduct.

Table 4 Information System IS_5 and IS_6

No.	a1	a2	a3	a4	a5	d
1	1	2	3	4	5	**6**
2	1	2	0	1	5	**6**
3	1	2	1	0	5	**6**
4	1	2	1	1	5	**6**
5	0	0	3	4	5	**6**
6	1	3	3	4	6	**1**
7	1	2	3	4	5	**1**
8	1	2	3	4	3	**3**

This is the end of proof.

5. CONCLUSION REMARKS

In the view of information granularity, the paper discussed the influence of attribute reducts on the granule view of an information system, proposing that the possible reduct and $\mu-$decision reduct will make the granule view coarser, while discernible reduct will not change the granule view. In addition, we gave a simple algorithm for the integration of reducts of two partial information systems in parallel or incremental data mining, and proved that the union of possible reducts is a possible reduct for union of partial information systems.

REFERENCES

[1]Andrzej Skowron, J.Stepaniuk. 2001. Information Granules: Towards Foundations of Granular Computing, International Journal of Intelligent Systems, Vol.16, p57-85

[2]B.Scotney, S.McClean. 1999. Efficient Knowledge Discovery Through the Integration of Heterogeneous Data, Information and Software Technology 41 p569-578

[3]Chang Li-yun, Wang Guo-yin, Wu Yu.1999. An Approach for Attribute Reduction and Rule Generation Based on Rough Set Theory, Journal of software, Vol.10 No.11, p1206-1211(in Chinese)

[4]Miao Duo-qian.1997. Rough Set and Its Application in Machine Learning, [Ph.D. Thesis]. Institute of Automation, The Chinese Academy of Sciences, (in Chinese)

[5]M.Kryszkiewicz. 2001. Comparative Study of

Alternative Types of Knowledge Reduction in Inconsistent Systems. International Journal of Intelligent Systems, Vol.16, p105-120

[6]Wang jue, Wang ren, Miao Duo-qian. 1998. Data Enriching Based on Rough Set Theory, Chinese Journal of Computers, 21(5), p393-400 (in Chinese)

[7]Y.Y.Yao.2001.Information Granulation and Rough Set Approximation, International Journal of Intelligent Systems, Vol.16, p87-104

Li-hong Wang, an associate professor of Yantai university, is now pursuing her Ph.D in school of computer engineering and science in Shanghai university, P.R.C. Her research interests include knowledge discovery, intelligent information processing and the architecture of computer network.

Geng-feng Wu is presently a professor of Shanghai university, an associate member of the third world academy of sciences and a councilor of Chinese computer federation. His main research interests include data mining and knowledge discovery in database, intelligent control, intelligent information processing and CSCW (computer supported cooperative work).