ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 4, Issue 9, March 2015

Algorithms of Minimal Mutual Compatible Granules and Approximations Based on Rough Set in Incomplete Information System

Chen Wu, Youquan Xu, Dandan Li, Lijuan Wang
School of Computer Science and Engineering,
Jiangsu University of Science and Technology
Zhenjiang, Jiangsu, 212003, P.R. China

***Abstract*—This paper studies a new method to obtain minimal mutual compatible granules with maximal compatible classes in complete information systems, proposes some related upper and lower approximations in rough set model and even in multi-granulation rough set model in incomplete information systems, discusses the properties and relationships of granules and relations, designs algorithms to solve minimal mutual compatible classes, the upper and lower approximations respectively. Through an example, it illustrates the correctness of algorithms. It provides some important and expanded theoretical bases for rough set theory to deal with problems in incomplete information systems.**

***Index Terms*—algorithm, incomplete information system, minimal granule, multi-granulation, rough set model.**

I. INTRODUCTION

Rough set theory (RST for short) ([1]) is put forward by Pawlak in 1982, which, as an generalization of set theory for studying intelligent systems ([2],[3]), has been applied in many scientific and technological fields such as Data Mining, Machine Learning, Knowledge Acquisition, Pattern Recognition, Intelligent Decision and so on. It uses indiscernibility relation on a universe to implement classification in complete information systems. But it may not be suitable incomplete information systems (IIS for short) because the existence of null or missing attribute values is an obstacle to hinder forming indiscernibility relation in them.

Nowadays a spirit approach or a new trend to deal with classification problems in incomplete information systems is to generalize the rough set model suggested by Pawlak into an expanded model by directly processing IIS. Tolerant RST ([4]), suggested by M. Kryszkiewicz, similarity RST ([2], [5]), put forward J.Stefanowski, limited tolerant RST, proposed by W.Guoying ([6]), and so on ([7]) are all along with this ideas. All of them are based on loosing indiscernibility relation to a general relation that can be built in IIS. The present paper puts forward a new granule called minimal granule with maximal compatible classes as primitive granules to promote the model in handling IIS based on tolerance relation. It brings even a new respect to produce a new approach to transact RST and even multi-granulation RST problems in IIS.

Accordingly, it is designs algorithms for this kind of minimal mutual compatible granules and approximations from the view points of knowledge expression systems for IIS. So the work done here is of necessity and meaning ness.

II. DEFINITIONS

An IIS is a quadruple $S=(U,AT,V,f)$ ([4]).

Definition 1. Let $x \in U$, $A \subseteq AT$. The compatible class(es) containing x is defined as

$$C(x)=\max\{X : x \in X, X^2 \subseteq COM(A)\}$$

Where max means operation is acting on by operator $\subseteq$, and $COM(A)=TOL(A)$ is the tolerance relation ([4]). $C(x)$ may be not unique.

$U/COM(A)=\{C : C \in C(x), x \in U\}$ composes of a complete covering on U. $U/COM(A)$ may be equivalently rewritten as

$$U/COM(A)=\{X \subseteq U : \max\{X^2 \subseteq COM(A)\}\}.$$

Obviously $COM(A)=\bigcap_{a \in A} COM(a)$.

Unlike $U/TOL(A)$, any two objects in a maximal compatible class $C \in U/COM(A)$ are compatible with respect to $COM(A)=TOL(A)$. Appending any other element not in C to a maximal compatible class C may cause breaking the compatibility. Any $C \in U/COM(A)$, as a maximal compatible class, is now viewed as primitive granule.

Definition 2. The minimal granules are defined by

$$SL_A(x)=\cap\{X : X \in U/COM(A), x \in X\}$$

$SL_A(x)$ is the operating results on $U/COM(A)$. Two arbitrary elements in $SL_A(x)$ are still compatible with respect to A. $\{SL_A(x) \mid x \in U\}$ constitutes a knowledge expression system. Because $C(x)$ may be not unique, a maximal compatible class for $x \in U$ from $U/COM(A)$ has alternative choices. If we select

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 4, Issue 9, March 2015

$SC_A(x) = X\{X \in U/COM(A), x \in X\}$, where X is randomly chosen, to form a knowledge expression system, then it may loss some useful information. So we remain $U/COM(A)$ as knowledge expression system, not $SC_A(x)$. This may reveal real information.

With these different knowledge expression systems, upper and lower approximations denoted by $\overline{SL}_A(x)$, $\overline{SC}_A(x)$, $\underline{SL}_A(x)$, $\underline{SC}_A(x)$, similar to $\overline{S}_A(x)$ and $\underline{S}_A(x)$ in [4], can be defined, respectively. Here we only give out one as an example.

Definition 3. The upper and lower approximations for $X \subseteq U$ in knowledge system $\{SL_A(x) \mid x \in U\}$ (Note: $SL_A(x)$ can be denoted by $SL(x)$) are defined respectively as follows:

$$\overline{SL}_A(X) = \cup\{Y : Y \in U/COM(A), X \cap Y \neq \varnothing\};$$
$$\underline{SL}_A(X) = \cap\{Y : Y \in U/COM(A), Y \subseteq X\}.$$

III. PROPERTIES AND RELATIONSHIPS

Property 1. $SL_A(x) \subseteq X_{x \in X \in U/COM(A)}$.

Theorem 1. $X \in U/COM(A)$ if and only if $X = \cap S_A(y)\{y \in X\}$.

Proof. Let $X \in U/COM(A)$. Then for any $x, z \in X$, $(x,z) \in COM(A) = TOL(A)$, $x \in S_A(z)$ So $x \in \cap S_A(y)\{y \in X\}$ and $X \subseteq \cap S_A(y)\{y \in X\}$. On the other hand, let $z \in \cap S_A(y)\{y \in X\}$ be any given, then $z \in S_A(y)$ for any $y \in X$ .That z is compatible with any element in X. Because $X \in U/COM(A)$, $X \subseteq X \cup \{z\}$ and X is maximal compatible class, it must have $z \in X$,other wise $X \cup \{z\}$ is a compatible class included in another maximal one. That contradicts to $X \in U/COM(A)$. Thus $\cap S_A(y)\{y \in X\} \subseteq X$. Therefore $X = \cap S_A(y)\{y \in X\}$.

When $X = \cap S_A(y)\{y \in X\}$, for any $w, z \in X = \cap S_A(y)\{y \in X\}$, we have $(z,w) \in COM(A)$, for $w = S_A(z), z \in S_A(w)$ .Thus $X \times X \subseteq COM(A)$. So X is a compatible class. Now we prove that X is also a maximal compatible class. If there is a $q \notin X$ such that $(X \cup \{q\})^2 \subseteq COM(A)$, then $q \in S_A(x)$ for any $x \in X$. Then $q \in \cap S_A(y)\{y \in X\} = X$. That is a contradiction. Therefore X is maximal.

Synthesizing the two directions, the theorem holds.

Theorem 2. $SL_A(x) = \cap S_A(y)\{y \in X \in U/COM(A), x \in X\} = \cap S_A(y)\{x \in S_A(y)\}$.

Proof. Because $SL_A(x) = \cap_{X \in U/COM(A) \wedge x \in X} X = X_1 \cap X_2 \cap ... \cap X_r$, where $X_i \in U/COM(A)$, $x \in X$, $i = 1, 2, ..., r$, and r is an integer. From Theorem 1, $X_i = \cap S_A(y)\{y \in X_i\}$. So

$$SL_A(x) = X_1 \cap X_2 \cap ... \cap X_r$$
$$= \cap S_A(y)\{y \in X \in U/COM(A), x \in X\}$$

Similarly, $SL_A(x) = \cap S_A(y)\{x \in S_A(y)\}$ can be proved.

IV. ALGORITHMS AND AN EXAMPLE

A. Finding maximal compatible class algorithm

After building object adjacent matrix M from IIS by relation $COM(A)$,we use a 2-dimensional and binary matrix $P_{m\times n}$ to obtain all maximal compatible classes starting from compatible classes. Let $U = \{x_i \mid i = 1, 2, ..., n\}$. $M = (m_{ij})_{n\times n}$, where m_{ij} equals 1, if $(x_i, x_j) \in COM(A)$, 0 otherwise. In the first k rows, k is the cardinality of maximal compatible classes, every row of matrix $P_{m\times n}$ will store a maximal compatible class at last, where m is an predefined positive integer, bigger than k at least and maximal number of maximal compatible classes. Note that $m <= (n*n)/2$, but m may be greater than n in some cases. $P(v, j) = 1$ means that x_j belongs to the v-th maximal compatible class, $P(v, j) = 0$ means not, v=1,2,…,k. Advantages of this algorithm are: it only use the strict lower triangle of matrix M; it eliminate duplicated classes at last automatically by the algorithm; it deal with and find singleton class(es). So it reaches computerization, not staying at theory.The algorithm is described in [8].Here it is omitted. The Output is: $P_{m\times n}$, the first k rows of it store all maximal compatible classes. The time complexity is $O(n^3)$.

B. Finding minimal granule algorithm

It is going to solve compatible class $SL_A(x)$. The description is as follows:

Input: matrix $P_{m\times n}$; $U = \{x_1, x_2, x_3, ..., x_n\}$.

Output: arrays $\{SL[x] \mid x \in U\}$. $SL[x]$, an array, is a granule containing x

Initialization: all elements of working array $SL[x]$ are set to be 0.

FOR i=1 TO n DO SL[x, i]=0;

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 4, Issue 9, March 2015

FOR x=1 TO n DO
(1) FOR i=1 TO n DO A[i]=0;
// all elements in working array A are set to 0
(2) FOR u=1 TO K DO
IF P[u, x]=1 THEN
FOR j=1 TO n DO
A[j]=A[j]*P[u,j]; // conjunction
(3) SL[x]=A;// A is copied to SL[x]

Output: SL[x]($x \in U$).

The time complexity is $O(Kn^2)$.

C. Finding upper Approximation algorithm

Let $X \subseteq U$. This algorithm finds out the upper approximation of X using $\{SL[x] | x \in U\}$. Array X[1..n] represents X , where X[i]=1 if $u_i \in X$ and X[i]=0 otherwise, i=1,…,n=|U|. Use T to stroe the upper approximation.

Input: $\{SL[x] | x \in U\}$, array *X*;
Initialization: For i=1 to n do 0→T[i];
For u=1 to n do
{ tag=1;
For i=1 to n do
If SL[u][i]*X[i]=1 then
{tag=0; break; }
If tag=0 then 1→T[u];
}
Output: T, the upper approximation of X.

The time complexity is $O(n^2)$.

Finding lower Approximation algorithm

Let $X \subseteq U$. This algorithm finds out the lower approximation of X using $\{SL[x] | x \in U\}$ .X is coded as the same as in the above.

Input : $\{SL[x] | x \in U\}$, array X;
Initialization: For i=1 to n do 0→T[i];
For u=1 to n do
{ tag=1;
For i=1 to n do
If SL[u][i]=1 then
If X[i]=0 then tag=0;
If tag=1 then 1→T[u];
}

Output: T, the lower approximation of X. The time complexity is $O(kn)$.

An Example

Let $U=\{O_1,O_2,\ldots,O_6\}$; $AT=\{a, b, c, d\}$;V={0,1, 2, 3, *}. An incomplete information system is shown in Table 1. The adjacent matrix M is constructed in Figure 1.

TABLE I. AN INCOMPLETE INFORMATION SYSTEM

U	*a*	*b*	*c*	*d*
O_1	*	2	*	0
O_2	*	2	*	1
O_3	3	*	1	*
O_4	1	*	*	*
O_5	*	2	*	*
O_6	3	2	1	*

The matrix *P* storing maximal compatible classes is in Figure 2. *U/COM*(*A*): $\{O_1,O_3,O_5,O_6\}$, $\{O_1,O_4,O_5\}$, $\{O_2,O_3,O_5,O_6\}$, $\{O_2,O_4,O_5\}$.

$$\begin{bmatrix} 1&0&1&1&1&1\\ 0&1&1&1&1&1\\ 1&1&1&0&1&1\\ 1&1&0&1&1&0\\ 1&1&1&1&1&1\\ 1&1&1&0&1&1 \end{bmatrix}$$

Fig 1. Adjancent matrix M

$\{SL[x] | x \in U\}$ is shown in Figure 3, i.e., $SL(O_1)$ $=\{O_1,O_5\}$, $SL(O_2)=\{O_2,O_5\}$, $SL(O_3)=$ $\{O_3,O_5,O_6\}$, $SL(O_4)=\{O_4,O_5\}$, $SL(O_5)=\{O_5\}$, $SL(O_6)=\{O_3,O_5,O_6\}$.

Let $X=\{O_3,O_4,O_5,O_6\}$.We compute upper and lower approximations of *X* according to their algorithms respectively. At first we represent or encode *X* into an array (0,0,1,1,1,1), then compute $\overline{SL}(X)$ and $\underline{SL}(X)$. In getting $\overline{SL}(X)$, *T*=(1,1,1,1,1,1), so, $\overline{SL}(X)$ $=\{O_1,O_2,O_3,O_4,O_5,O_6\}$. In getting $\underline{SL}(X)$, *T*=(0,0,1,1,1,1), so $\underline{SL}(X)$ $=\{O_3,O_4,O_5,O_6\}$.

$$\begin{bmatrix} 1&0&1&0&1&1\\ 1&0&0&1&1&0\\ 0&1&1&0&1&1\\ 0&1&0&1&1&0\\ 0&0&0&0&0&0\\ 0&0&0&0&0&0 \end{bmatrix}$$

Fig 2. Matrix P

$$\begin{bmatrix} 1&0&0&0&1&0\\ 0&1&0&0&1&0\\ 0&0&0&1&1&0\\ 0&0&0&0&1&0\\ 0&0&1&0&1&1\\ 0&0&0&0&0&0 \end{bmatrix}$$

Fig 3. Matrix SL

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 4, Issue 9, March 2015

V. APPLICATIONS IN MULTI-GRANULATION MODEL

In the literature ([9],[10],[11]), Qian et al. proposed the concept of multi-granulation rough set model. In our opinion, using knowledge expression systems $\{SL_A(x)\,|\,x\in U\}$ and $U/COM(A)$, we can also study multi-granulation rough set. Here we give one description.

Let $A_i \subseteq AT(i=1,2,\ldots,m)$ be an attribute subset, where m is positive integer. Then, the optimistic multi-granulation lower and upper approximations for $\forall X\subseteq U$ with respect to $\{A_i\,|\,i=1,2,\ldots,m\}$ are defined respectively as:

$$\underline{\sum_{i=1}^{m} A_i}^{o}(X)=\{x\in U:\exists i, Y\in\{SL_{A_i}(x)\,|\,x\in U\}(Y\subseteq X)\}$$

$$\overline{\sum_{i=1}^{m} A_i}^{o}(X)=\sim \underline{\sum_{i=1}^{m} A_i}^{o}(\sim X)$$

The optimistic multi-granulation boundary region is

$$Bn^{o}_{\sum_{i=1}^{m} A_i}(X)=\overline{\sum_{i=1}^{m} A_i}^{o}(X)-\underline{\sum_{i=1}^{m} A_i}^{o}(X).$$

The pessimistic multi-granulation lower and upper approximations respectively are:

$$\underline{\sum_{i=1}^{m} A_{i_\beta}}^{p}(X)=\{x\in U: Y\subseteq X, Y\in\{SL_{A_i}(x)\,|\,x\in U\}\}$$

$$\overline{\sum_{i=1}^{m} A_{i_\beta}}^{p}(X)=\sim \underline{\sum_{i=1}^{m} A_{i_\beta}}^{p}(\sim X)$$

The pessimistic multi-granulation boundary region is

$$Bn^{p}_{\sum_{i=1}^{m} A_i}(X)=\overline{\sum_{i=1}^{m} A_i}^{p}(X)-\underline{\sum_{i=1}^{m} A_i}^{p}(X).$$

It is not hard to design related algorithms in multi-granulation rough set model after having algorithms in last section.

VI. CONCLUSIONS

This paper defines minimal compatible granules $SL_A(x)$, using maximal compatible classes as primitive granules. It expands original rough set model to a generalized one. Several algorithms to solve minimal granules, upper and lower approximations are suggested through binary matrices. The correctness of the algorithms is verified by an example. This novel granular view leads to increasing study approaches in RST and in multi-granulation RST. Our next task is to study and design other algorithms to other kinds of granules.

ACKNOWLEDGMENT

This paper is sponsored by Chinese NFS (No. 61100116) and a Foundation of Graduate Department of Jiangsu University of Science and Technology.

REFERENCES

[1] Z.Pawlak, "Rough sets and intelligent data analysis", Information Sciences. 147 (2002), pp.1-12.

[2] W.Roman, Q.Swiniarski and A.Skowron, "Rough Set Method in Feature Selection and Recognition", Pattern Recognition Letters, 24 (2003), pp. 833~849.

[3] J.S.Mi, W.Z.Wu and W.X.Zhang, "Approaches to Knowledge Reduction Based on Variable Precision Rough Set Model", Information Sciences. 159 (2004), pp.255-272.

[4] M.Kryszkiewicz, "Rough Set Approach to Incomplete Information Systems", Information Sciences, 112 (1998), pp.39-49.

[5] J.Stefanowski, "Incomplete Information Tables and Rough Classification", J. Computational Intelligence. 17 (2001), pp.545-566.

[6] W.L.Chen, J.X.Cheng and C.J.Zhang, "A Generalization to Rough Set Theory Based on Tolerance Relation", J. computer engineering and applications, 16(2004), pp. 26-28.

[7] C.Wu, X.B.Yang, "Information Granules in General and Complete Covering", Proceedings of the 2005 IEEE International Conference on Granular Computing, pp. 675-678.

[8] C.Wu, "The cardinality estimation of the covering to a semi-equivalence relation and the design of its solving algorithm", J. Computer engineering and design, 4(2004), pp.89-106.

[9] Qian Y H, Liang J Y, Yao Y Y, C, et al, "MGRS: a multigranulation rough set", Information Sciences, 2010, vol.180, no.6, pp.949–970.

[10] Qian Y. H, Liang J. Y, Dang C Y. "Incomplete multigranulation rough set", IEEE Transactions on Systems, Man and Cybernetics, Part A, 2010,vol.40,no.2,pp. 420-431.

[11] Qian Y H, Liang J Y, Wei W. "Pessimistic rough decision", in: Second International Workshop on Rough Sets Theory, October 2010, Zhoushan, China, pp. 440–449.

AUTHOR BIOGRAPHY

Wu Chen, a professor with Ph.D., is with the school of computer science and engineering at the Jiangsu university of science and technology. Research area: rough set, data mining, pattern recognition and data processing and algorithm design.

Youquan Xu, a post graduate student, is studying in the school of computer science and engineering at the Jiangsu university of science and technology. Research area: rough set, data mining.

Dandan Li, a post graduate student, is studying in the school of computer science and engineering at the Jiangsu university of science and technology. Research area: rough set, decision making.

Wang Lijuan, a assistant professor with Ph.D., is with the school of computer science and engineering at the Jiangsu university of science and technology. Research area: rough set, data mining.