A Model for Multi Granulation Variable Precision Rough Sets

Biqing Wang

College of Mathematics and Computer, Tongling University, Tongling, Anhui 244000, China

Abstract — **In this paper multi granulation rough sets is extended to variable precision to develop a model for multi-granulation variable precision rough sets. We define a lower approximation of the model and address the significance of granularity as heuristic information. We extend the ideas and design a granular space reduction algorithm of multi-granulation variable precision rough sets. Finally, an example is given to show the effectiveness of the model and the approach.**

Keywords - multi-granulation; variable precision; rough sets; granular space reduction

I. INTRODUCTION

Multi-granulation rough sets is a multi-view data analysis method developed in recent years. This method is proposed in [1-3] and used to overcome the drawbacks of Pawlak's rough sets model [4]. Because multi-granulation rough sets analyses problem in terms of multiple granular spaces, it can obtain an more reasonable and satisfactory solution. Then, multi-granulation rough sets has been concerned by a great many scholars. Reference [5] prolongates it to multi-granulation dominance relation rough sets. In [6], it is extended to multi-granulation fuzzy relation rough sets. Reference [7] puts forward multi-granulation neighborhood system rough sets. In [8], multi-granulation decision-theoretic rough sets is presented. The applications of multi-granulation rough sets are studied in [9-10]. All these researches promote the development of multi-granulation rough sets.

At present, multi-granulation rough sets has developed two models: optimistic multi- granulation rough sets and pessimistic multi-granulation rough sets. By analyzing these two models, we can reach the following conclusions. First, lower approximation of optimistic multi-granulation rough sets only requires that there is a granularity which makes knowledge granule contained in the category. Obviously, this kind of decision making seems to be too loose. Second, lower approximation of pessimistic multi- granulation rough sets requires that all the granularities make knowledge granule contained in the category. However, this kind of decision making may be too strict.

To solve those of problems, the concept of variable precision [11] is introduced on the basis of multi-granulation rough sets in this paper. Variable precision rough sets model proposed by Ziarko W is a generalization of Pawlak's rough sets theory. By using parameter $\beta(0\leq\beta<0.5)$, it allows error classification probability to a certain degree, thus relaxes the strict requirements for the approximate border of Pawlak's rough set model. Based on the above research, in this paper, multi-granulation variable precision rough sets model is presented, whose lower approximation requires that a certain number of granularity make the majority of objects in knowledge granule contained in the category. This model on one hand, reflects that the number of granular spaces that satisfy the requirements is moderate, and on the other, considers that knowledge granule can be partly contained in the category. Ambiguity degree of these two aspects both can be characterized by several parameters in order to achieve the optimal decision. The properties of the model are studied and the relationship among different rough sets models is discussed.

Meanwhile, how to make granular space reduction is a particular problem in research of multi-granulation rough sets. Granular space reduction means selecting a non-redundant granular space subset without affecting decision making. Because lower approximation of multi- granulation rough sets gradually decreases along with the increase of the number of granular spaces, existing attributes reduction algorithms are not applicable to granular space reduction. To overcome this problem, we propose a granular space reduction method which is suitable for multi-granulation variable precision rough sets. First, the concept of distribution reduction is introduced [12]. Then, the measurement of the significance of granularity is given. Based on this, a granular space reduction algorithm for multi-granulation variable precision rough sets is put forward. At last the algorithm is analyzed and illuminated by an example.

II. PRELIMINARIES

In this section, we review some basic concepts in rough set [1-4], [11] to be used in this paper.

Definition 1. An information system is defined as S=(U,A,V,f), where U is the set of objects; A is the set of attributes; $V=\bigcup_{a\in A} V_a$, where Va is the set of values of attribute a; $f:U\times A\rightarrow V$ is an information function, which determines values of attribute of every object u, namely, $f(u,a)\in V_a$ for every $u\in U$ and $a\in A$.

If set of attributes can be divided into condition attributes C and decision attributes D, namely, $C\cup D=A$, $C\cap D=\varnothing$, the information system is called decision system or decision table, where D has only one attribute commonly.

Definition 2. In information system S, for every attribute subset $P\subseteq A$, an indiscernibility relation IND(P) is defined as $IND(P)=\{(x,y)\in U\times U: \forall a\in P, f(x,a)=f(y,a)\}$.

Definition 3. IND(P) corresponds to a granularity. U/IND(P) corresponds to a granular space and it can be denoted by U/P. Equivalence class $[u]_P$ is called knowledge granule.

Definition 4. In information system S, for $\forall R \subseteq A$, $\forall X \subseteq U$, lower approximation is defined as $\underline{R}X=\{u \in U | [u]_R \subseteq X\}$, upper approximations is defined as $\overline{R}X=\{u \in U | [u]_R \cap X \neq \varnothing\}$.

Definition 5. In information system S, $R=\{a_1, a_2, \ldots, a_k\}$ is the set of k attribute subsets of A. For $\forall X \subseteq U$, optimistic multi-granulation rough sets lower approximation of X with respect to R is defined as

$$\underline{\sum_{i=1}^{m} a_i}^{O}(X)=\{x \in U: [x]_{a_1} \subseteq X \vee [x]_{a_2} \subseteq X \vee \ldots \vee [x]_{a_m} \subseteq X\} \quad (1)$$

optimistic multi-granulation rough sets upper approximation of X with respect to R is defined as

$$\overline{\sum_{i=1}^{m} a_i}^{O}(X) = \sim \underline{\sum_{i=1}^{m} a_i}^{O}(\sim X) \quad (2)$$

where ~X is the complement of X. Ordered pair $<\underline{\sum_{i=1}^{m} a_i}^{O}(X), \overline{\sum_{i=1}^{m} a_i}^{O}(X)>$ is called optimistic multi-granulation rough sets of X with respect to R.

Definition 6. In information system S, $R=\{a_1, a_2, \ldots, a_k\}$ is the set of k attribute subsets of A. For $\forall X \subseteq U$, pessimistic multi-granulation rough sets lower approximation of X with respect to R is defined as

$$\underline{\sum_{i=1}^{m} a_i}^{P}(X)=\{x \in U: [x]_{a_1} \subseteq X \wedge [x]_{a_2} \subseteq X \wedge \ldots \wedge [x]_{a_m} \subseteq X\} \quad (3)$$

pessimistic multi-granulation rough sets upper approximation of X with respect to R is defined as

$$\overline{\sum_{i=1}^{m} a_i}^{P}(X) = \sim \underline{\sum_{i=1}^{m} a_i}^{P}(\sim X) \quad (4)$$

where ~X is the complement of X. Ordered pair $<\underline{\sum_{i=1}^{m} a_i}^{P}(X), \overline{\sum_{i=1}^{m} a_i}^{P}(X)>$ is called pessimistic multi-granulation rough sets of X with respect to R.

Definition 7. Let (U,R) be an approximation space, where universe U is a nonempty finite set and R is an equivalent relationship on U. Let U/R be $\{E_1, E_2, \cdots, E_n\}$. For $\forall X \subseteq U$, β lower approximation of X is defined as

$$\underline{R}_{\beta}X = \cup \{E \in U/R | X \overset{\beta}{\supseteq} E\} \quad (5)$$

III. MULTI-GRANULATION VARIABLE PRECISION ROUGH SETS

This paper analyzes and compares the decision making process of optimistic multi-granulation rough sets and pessimistic multi-granulation rough sets. After that, it put forward a multi-granulation variable precision rough sets which unifies the above two models.

In multi-granulation rough sets decision theory, optimistic multi-granulation rough sets holds that if there is a granularity which makes knowledge granule $[x]_{a_i}$ contained in the category X, x belongs to lower approximation of X. Pessimistic multi-granulation rough sets holds that if all the granularities make knowledge granule $[x]_{a_i}$ contained in the category X, x belongs to lower approximation of X. But multi-granulation variable precision rough sets holds that if a certain number of granularities make knowledge granule $[x]_{a_i}$ contained in the category X, x belongs to lower approximation of X.

In addition, an obvious limitation of Pawlak's rough sets is that its classification is accurate, namely, inclusion or exclusion, not some degree of inclusion or exclusion. The multi-granulation variable precision rough sets we propose is the expansion of multi-granulation rough sets. It introduces a parameter $\beta(0 \leq \beta < 0.5)$, namely, the error rate of classification. On the one hand, it develops the concept of approximate space; on the other hand it is also beneficial for finding relevant data from the data which we considered irrelevant using rough sets theory.

Definition 8. In information system S, $R=\{a_1, a_2, \ldots, a_k\}$ is the set of k attribute subsets of A. For $0<\lambda\leq 1$, $0\leq\beta<0.5$, $\forall X \subseteq U$, multi-granulation variable precision rough sets lower approximation of X with respect to R is defined as

$$\underline{R}_{\beta}^{\lambda}X = \{u | (u \in U) \wedge (X \overset{\beta}{\supseteq} [u]_{a_i}),\ \forall a_i \in W, \\ W \subseteq R \wedge |W|/k \geq \lambda\} \quad (6)$$

multi-granulation variable precision rough sets upper approximation of X with respect to R is defined as

$$\overline{R}_{\beta}^{\lambda}X = \sim \underline{R}_{\beta}^{\lambda}(\sim X) \quad (7)$$

where ~X is the complement of X. Ordered pair $< \underline{R}_{\beta}^{\lambda}X$, $\overline{R}_{\beta}^{\lambda}X >$ is called multi-granulation variable precision rough sets of X with respect to R. The boundary region , the positive region and the negative region of X are respectively defined as

$$bnr_{\beta}^{\lambda}(X) = \overline{R}_{\beta}^{\lambda}X - \underline{R}_{\beta}^{\lambda}X \quad (8)$$

$$posr_{\beta}^{\lambda}(X) = \underline{R}_{\beta}^{\lambda}X \quad (9)$$

$$negr_{\beta}^{\lambda}(X) = U - \overline{R}_{\beta}^{\lambda}X \quad (10)$$

Theorem 1. In information system S, $R=\{a_1, a_2, \ldots, a_k\}$ is the set of k attribute subsets of A. For $\forall X \subseteq U$, multi-granulation variable precision rough sets has the following relationships with optimistic multi-granulation rough sets and pessimistic multi-granulation rough sets.

$$\underline{R}_0^{\frac{1}{k}}X = \underline{\sum_{i=1}^{m} a_i(X)}^{O} \tag{11}$$

$$\overline{R}_0^{\frac{1}{k}}X = \overline{\sum_{i=1}^{m} a_i(X)}^{O} \tag{12}$$

$$\underline{R}_0^{1}X = \underline{\sum_{i=1}^{m} a_i(X)}^{P} \tag{13}$$

$$\overline{R}_0^{1}X = \overline{\sum_{i=1}^{m} a_i(X)}^{P} \tag{14}$$

Proof. According to definition of Definition 5, Definition 6 and Definition 8, Theorem 1 holds clearly.

Theorem 1 indicates that multi-granulation variable precision rough sets degenerates to optimistic multi-granulation rough sets if $\lambda=1/k$ and $\beta=0$; multi-granulation variable precision rough sets degenerates to pessimistic multi-granulation rough sets if $\lambda=1$ and $\beta=0$. It follows from this that multi-granulation variable precision rough sets is a generalization of multi-granulation rough sets and multi-granulation rough sets is a special case of multi-granulation variable precision rough sets.

Theorem 2. In information system S, $R=\{a_1, a_2, \ldots, a_k\}$ is the set of k attribute subsets of A. For $\beta=0$, $0<\lambda\leqslant 1$, multi-granulation variable precision rough sets has the following properties.

$$\underline{\sum_{i=1}^{m} a_i(X)}^{P} \subseteq \underline{R}_0^{\lambda}X \subseteq \underline{\sum_{i=1}^{m} a_i(X)}^{O} \tag{15}$$

$$\overline{\sum_{i=1}^{m} a_i(X)}^{O} \subseteq \overline{R}_0^{\lambda}X \subseteq \overline{\sum_{i=1}^{m} a_i(X)}^{P} \tag{16}$$

Proof. According to definition of Definition 5, Definition 6 and Definition 8, Theorem 2 holds clearly.

Theorem 2 shows that lower approximation of multi-granulation variable precision rough sets is between pessimistic multi-granulation rough sets and optimistic multi-granulation rough sets; upper approximation of multi-granulation variable precision rough sets is between optimistic multi-granulation rough sets and pessimistic multi-granulation rough sets.

IV. A GRANULAR SPACE REDUCTION APPROACH TO MULTI-GRANULATION VARIABLE PRECISION ROUGH SETS

For decision making in multi-granulation rough sets, some granular spaces are of small significance. If they are removed, the decision making will not be affected. So, the problem of selecting granular spaces has arisen, namely, selecting a granular spaces set without affecting decision making. The set of selected granular spaces is called a granular space reduction.

In this paper, we introduce the concept of distribution reduction and give the definition of granular space reduction of multi-granulation variable precision rough sets.

Definition 9. In information system $S=(U,C\cup D,V,f)$, Let U/D be $\{Y_1, Y_2, \ldots, Y_j\}$. For $R\subseteq C$, $0<\lambda\leqslant 1$, $0\leq\beta<0.5$, lower approximation of multi-granulation variable precision rough sets is defined as

$$\mu(R, D, \lambda, \beta)=\{\underline{R}_{\beta}^{\lambda}Y_1, \underline{R}_{\beta}^{\lambda}Y_2, \ldots, \underline{R}_{\beta}^{\lambda}Y_j\} \tag{17}$$

In multi-granulation space, computations for lower approximation is a basic operation of granular space reduction introduced then. So, how to compute lower approximation efficiently is the key to increasing the efficiency of its relevant algorithm. Next, algorithm for computing $\mu(R, D, \lambda, \beta)$ using granular spaces is given.

Algorithm 1: Algorithm for computing $\mu(R, D, \lambda, \beta)$

Input: $S=(U,C\cup D,V,f)$. $R=\{a_1, a_2, \ldots, a_k\}$ is the set of k attribute subsets of C. $U/D=\{Y_1, Y_2, \ldots, Y_j\}$. $0<\lambda\leq 1$. $0\leq\beta<0.5$.

Output: $\mu(R, D, \lambda, \beta)$

```
for(i=1;i<=k;i++)
    {Compute U/ai={E1, E2, …, Ew};
    if (∃Ep⊄Yq)
    Delete all objects in Ep from U/D;}
μ(R, D, λ, β)=U/D;
return μ(R, D, λ, β).
```

Algorithm 1 spends most of its time dividing equivalence classes. In the worst case, time complexity for computing |R| partitions is $O(|R||U|^2)=O(|C||U|^2)$. Thus, time complexity of Algorithm 1 is $O(|C||U|^2)$.

Based on the definition 9, definition 10 is given below.

Definition 10. In information system $S=(U,C\cup D,V,f)$, Let U/D be $\{Y_1, Y_2, \ldots, Y_j\}$. For $R\subseteq C$, $0<\lambda\leqslant 1$, $0\leq\beta<0.5$, if $\mu(R, D, \lambda, \beta)=\mu(C, D, \lambda, \beta)$ and $\mu(R', D, \lambda, \beta)>\mu(C, D, \lambda, \beta)$ for any $R'\subset R$, R is called a granular space reduction of C.

Definition 10 requires that the redundant granular spaces should be removed without decreasing distinguishing ability of information system. This is in line with the definition in Pawlak's rough sets on principle.

For decision making in multi-granulation variable precision rough sets, we needed to discover which granular spaces are important and which granular spaces are redundant. So, definition of significance of granularity is given below.

Definition 11. In information system $S=(U,C\cup D,V,f)$, for $0<\lambda\leq 1$, $0\leq\beta<0.5$, $R\subseteq C$, significance of $A\subseteq C\text{-}R$ with respect to D is defined as

$$\sigma(A, R, D, \lambda, \beta)=|\mu(R, D, \lambda, \beta)|-|\mu(R\cup A, D, \lambda, \beta)| \tag{18}$$

From the above definition, we can known that $\sigma(A, R, D, \lambda, \beta)\geq 0$. Granularity A is not important with respect to D if and only if $\sigma(A, R, D, \lambda, \beta)=0$. When the value of $\sigma(A, R, D, \lambda, \beta)$ is bigger, granularity A is more important for D under given conditions. In this paper, $\sigma(A, R, D, \lambda, \beta)$

is used as the heuristic information for computing granular space reduction to reduce the search space.

So, granular space reduction algorithm of multi-granulation variable precision rough sets uses $\sigma(A, R, D, \lambda, \beta)$ as a measure of granularity significance. Each time the granularity judged to be the most important in the present situation is added to granular space reduction until lower approximation of reduction and lower approximation of the set of original attributes are identical. The obtained reduction is the granular space reduction of multi-granulation variable precision rough sets.

Then on the basis of above analysis , we design a granular space reduction algorithm of multi-granulation variable precision rough sets.

Algorithm 2: Granular space reduction algorithm of multi-granulation variable precision rough sets

Input: $S=(U, C \cup D, V, f)$, $R=\varnothing$, $C-R=\{a_1, a_2, \ldots, a_k\}$, $0<\lambda\leq 1$, $0\leq\beta<0.5$

Output: a granular space reduction of S

Compute$|\mu(R, D, \lambda, \beta)|$ and $|\mu(C, D, \lambda, \beta)|$;

while($|\mu(R, D, \lambda, \beta)| > |\mu(C, D, \lambda, \beta)|$

{for(i=1;i<=|C-R|;i++)

Compute $\sigma(a_i, R, D, \lambda, \beta)$;

Choose a_i which has the maximum $\sigma(a_i, R, D, \lambda, \beta)$, if there are several such a_i, pick one at random;

$R=R \cup \{a_i\}$;

Recalculate $|\mu(R, D, \lambda, \beta)|$;}

Return R.

In Algorithm 2, time complexity of step 1 is $O(|C||U|^2)$. Cycle number of step 2 is $|C-R|$ at most. So, time complexity of step 2 is $O(|C|^2|U|^2)$. Therefore, time complexity of Algorithm 2 is $O(|C|^2|U|^2)$.

V. EXAMPLE

In this paragraph we illustrate the notions introduced previously with a simple example. Table 1 lists a decision table, where $C=\{a_1, a_2, a_3, a_4, a_5\}$, $D=\{d\}$ and $U/D=\{Y_1, Y_2\}=\{\{u_1, u_2, u_3, u_7\}, \{u_4, u_5, u_6, u_8\}\}$. Each attribute corresponds to a granular space.

TABLE I. DECISION INFORMATION SYSTEM

U	a_1	a_2	a_3	a_4	a_5	d
u_1	0	1	0	0	2	1
u_2	1	1	2	1	2	1
u_3	2	2	0	2	3	1
u_4	3	3	2	4	1	0
u_5	4	0	2	4	3	0
u_6	3	4	4	3	4	0
u_7	0	0	1	0	0	1
u_8	3	4	3	3	1	0

The granular space reduction of Table 1 is computed by algorithm 2. When the parameters take different values, reduction results can be seen from Table 2.

The following results indicate that the granular space reduction algorithm of multi-granulation variable precision rough sets model can give reduction results under different parameter conditions. In practical application, users can flexibly choose the appropriate granular spaces and classification error in accordance with the need of different problem so as to get a better solution.

TABLE II. REDUCTION RESULTS

λ	β	$\mu(C, D, \lambda, \beta)$	R
1	0	$\{\{u_1\}, \{u_6, u_8\}\}$	$\{a_2, a_3, a_5\}$
1	1/3	$\{\{u_1\}, \{u_4, u_6, u_8\}\}$	$\{a_2, a_5\}$
4/5	0	$\{\{u_1\ u_2, u_3, u_7,\}, \{u_4, u_6, u_8\}\}$	$\{a_2, a_3, a_5\}$
4/5	1/3	$\{\{u_1\ u_2, u_3, u_7,\}, \{u_4, u_6, u_8\}\}$	$\{a_2, a_5\}$

VI. CONCLUSION

Applying the rough set model to knowledge acquiring and decision support system is one of the hot issues in research on rough sets theory and its application. Multi-granulation rough sets is improved and a multi-granulation variable precision rough sets model is constructing in this paper. We have enriched and further developed the rough sets theory and the improved method can be applied to practical problems, such as multi-domains experts decision analysis. The future work will focus on improving the calculation efficiency.

REFERENCES

[1] Qian Y H, Liang J Y, and Yao Y Y, "MGRS: A multi-granulation rough set," Information Sciences, vol. 180, pp. 949-970, 2010.

[2] Qian Y H, Liang J Y, and Dang C Y, "Incomplete multigranulation rough set," IEEE Trans on Systems, Man, and Cybernetics: Part A, vol. 40, pp. 420-431, 2010.

[3] Qian Y H, Liang J Y, and Wei W, "Pessimistic rough decision," The 2nd Int Workshop on Rough Sets Theory, pp. 440-449, 2010.

[4] Z Pawlak, "Rough sets," International Journal of Computer and Information Science, vol. 11, pp. 341-356, 1982.

[5] Xu W H, Sun W X, and Zhang X Y, "Multiple granulation rough set approach to ordered information systems," International Journal of General Systemsl, vol. 41, pp. 475-501, 2012.

[6] Yang X B, Song X N, and Dou H L, "Multi-granulation rough set: From crisp to fuzzy case," Annals of Fuzzy Mathematics and Informatics, vol. 1, pp. 55-70, 2011.

[7] Lin G P, Qian Y H, and Li J J, "NMGRS: Neighborhood-based multigranulation rough sets," International Journal of Approximate Reasoning, vol. 53, pp. 1080-1093, 2012.

[8] Qian Y H , Zhang H, and Sang Y L, "Multigranulation decision-theoretic rough sets," International Journal of Approximate Reasoning, vol. 55, pp. 225-237, 2014.

[9] Wu W Z, Leung Y, "Theory and applications of granular labelled partitions in multi-scale decision tables," Information Sciences, vol. 181, pp. 3878-3897, 2011.

[10] She Y H, He X L, "On the structure of the multigranulation rough set model," Knowledge-Based Systems, vol. 36, pp. 81-92, 2012.

[11] Ziarko W, "Variable precision roughset model," Journal of Computer and System Science, vol. 46, pp. 39-59, 1993.

[12] Zhang Wenxiu, Mi Jusheng, and Wu Weizhi, "Knowledge Reduction in Inconsistent Information Systems," Chinese Journal of Computers, vol. 26, pp. 12-18, 2003.