Multigranulation rough set: A multiset based strategy

Xibei Yang [1,2] , Suping Xu [1] , Huili Dou [1,]* Xiaoning Song [3], Hualong Yu [1] , Jingyu Yang [4]

[1] *School of Computer Science and Engineering, Jiangsu University of Science and Technology,*
Zhenjiang, 212003, P.R. China

E-mail: yangxibei@hotmail.com, supingxu@yahoo.com, douhuili@163.com, yuhualong@just.edu.cn

[2] *School of Economics and Management, Nanjing University of Science and Technology,*
Nanjing, 210094, P.R. China

[3] *School of Internet of Things Engineering, Jiangnan University,*
Wuxi 214122, P.R. China

E-mail: xnsong@yahoo.com.cn

[4] *Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information (Nanjing University*
of Science and Technology),
Ministry of Education, Nanjing, 210094, P.R. China

E-mail: yangjy@mail.njust.edu.cn

Abstract

A simple multigranulation rough set approach is to approximate the target through a family of binary relations. Optimistic and pessimistic multigranulation rough sets are two typical examples of such approach. However, these two multigranulation rough sets do not take frequencies of occurrences of containments or intersections into account. To solve such problem, by the motivation of the multiset, the model of the multiple multigranulation rough set is proposed, in which both lower and upper approximations are multisets. Such two multisets are useful when counting frequencies of occurrences such that objects belong to lower or upper approximations with a family of binary relations. Furthermore, not only the concept of approximate distribution reduct is introduced into multiple multigranulation rough set, but also a heuristic algorithm is presented for computing reduct. Finally, multiple multigranulation rough set approach is tested on eight UCI (University of California–Irvine) data sets. Experimental results show: 1. the approximate quality based on multiple multigranulation rough set is between approximate qualities based on optimistic and pessimistic multigranulation rough sets; 2. by comparing with optimistic and pessimistic multigranulation rough sets, multiple multigranulation rough set needs more attributes to form a reduct.

Keywords: Approximate distribution reduct; Approximate quality; Multiset; Multiple multigranulation rough set.

---

* Corresponding author: No. 2, Mengxi Road, School of Computer Science and Engineering, Jiangsu University of Science and Technology, Zhenjiang, 212003, P.R. China.

1. Introduction

Back in the early 1980s, Pawlak proposed the rough set [27] for characterizing the uncertainty. Through the three decades of the development, rough set has been demonstrated to be useful in knowledge acquisition [5,10], pattern recognition [3,7,23], machine learning [6,8,9,11,26], decision support [21,44,46,57] and so on.

In Pawlak's rough set, indiscernibility relation is a basic concept, it is an intersection of some equivalence relations in knowledge base [27]. An indiscernibility relation can induce a partition on the universe of discourse. Lower, upper approximations and boundary region in rough set model are then the unions of some blocks (equivalence classes) in partition with different conditions, respectively. Obviously, Pawlak's rough set is constructed on the basis of one and only one set of the information granules (set of equivalence classes in a partition), we call such set a *granular structure* [29]. From this point of view, Pawlak's model is referred to as a single–granulation rough set approach in this paper. Nevertheless, single–granulation is not good enough for practical problem solving. For example:

1. a map can be explained from different levels of viewpoints, the coarser perspective is based on the greater information granules while the finer perspective is based on the smaller information granules, i.e. we may explore a map through different levels of granulations;

2. single feature information is not robust for authentication (e.g. fingerprint may be stolen and copied by criminals) and then multi–biometrics are needed;

3. single–granulation approach is very time–consuming in rough set theory since it needs to do the intersection on more than one binary relations.

To fill those gaps of single–granulation approach and further improve the effectiveness of Pawlak's rough set theory, Qian and Liang et al. [15,17,30,31,33,34,35,37,38] proposed the concept of the multigranulation rough set. Presently, with respect to different requirements, multigranulation rough set

progressing rapidly [2,18,41,42,45,48,49,50,51,55]. We may classify the existing multigranulation rough sets into two categories.

1. Firstly, synchronous multigranulation approach: a lot of the granulations are presented simultaneously for problem solving. For instance, in Qian et al.'s classical multigranulation rough set approach, the target is approximated through a set of partitions; Yang et al. [52] and Xu et al. [43] presented the multigranulation fuzzy rough set through a family of fuzzy relations, respectively; Lin et al. [14] presented the neighborhood multigranulation rough set by using a family of neighborhoods, i.e. neighborhood system [53]; Khan and Banerjee [12] introduced the concept of the multiple–source approximation systems, which are multigranulation fusions of Pawlak's approximation spaces; Abu–Donia [1] studied the rough approximations based on multi–knowledge; Wu and Leung [40] investigated the multi–scale information system, which reflects the explanation of same problem at different scales (levels of granulations); Dou et al. [4] integrated variable precision rough set [56] with multigranulation rough sets; She et al. [39] studied the algebraic structure of multigranulation rough set.

2. Secondly, asynchronous multigranulation approach: a granulation is constructed or obtained from the last granulation. For example, Qian et al. [33,34] proposed a positive approximation accelerator for attribute reduction, which can make universe smaller step by step; Liang et al. [19] proposed an efficient rough feature selection algorithm for large–scale data sets, which selects a valid feature subset though dividing big samples into small samples and fusing the feature selection results of small samples together, they [20] also studied the incremental feature selection mechanism by considering the monotonic increasing of samples; Wang et al. presented a dimension incremental strategy for attribute reduction, in which the current information entropy can be updated by the last computation result.

The purpose of this paper is to further explore synchronous multigranulation approach. Qian

et al.'s optimistic and pessimistic multigranulation rough sets are two typical examples of such research. Through the investigation of optimistic and pessimistic multigranulation lower approximations, we know that optimism needs at least one of the granular structures (partitions) to be satisfied with the containment between equivalence class and target, pessimism needs all of the granular structures to be satisfied with the containments between equivalence class and target. Obviously, these two multigranulation rough sets do not take frequencies of occurrences of set containments into account. We have some practical examples to illustrate such limitation.

1. Take for instance multi-subspace learning problem, each subspace is corresponding to a world which can be used to construct approximation. If Qian et al.'s optimistic approach is used, then it is confused for us to count how many subspaces have contributed to the set containments, e.g., lower approximations. Therefore, the subspaces with high contribution may be mixed with subspaces with low contribution.

2. Qian et al.'s optimistic approach is too loose (Here is a China old saying: One tree does not make a forest.) while the pessimistic approach is too strict (Here is also a China old saying: It is hard to please all.). Voting is a possible strategy to solve these problems. Therefore, it is required that the frequencies of occurrences of set containments should be counted.

To sum up, we will propose the multiple multigranulation rough set by using the concept of the multiset [13,24] in this paper. In our multiple multigranulation rough set, both lower and upper approximations are multisets, which can reflect frequencies of occurrences of objects belonging to lower and upper approximations, respectively.

To facilitate our discussions, we present the basic knowledge about optimistic and pessimistic multigranulation rough sets in Section 2. In Section 3, we propose the model of multiple multigranulation rough set, not only the basic properties of such model are studied, but also the relationships among multiple and Qian et al.'s multigranulation rough sets are explored. Since attribute reduction is one of the key problems in rough set theory, we also introduce the concept of the approximate distribution reduct into multiple multigranulation rough set. Through experimental analyses, the comparisons of approximation qualities and reducts on three different multigranulation rough sets are shown in Section 4. The paper ends with conclusions in Section 5.

2. Preliminary knowledge

2.1. Multigranulation rough sets

Formally, an information system can be denoted as a pair $I = < U, AT >$, in which U is a non–empty finite set of objects called the universe; AT is a non–empty finite set of attributes. $\forall a \in AT$, V_a is the domain of attribute a. $\forall x \in U$, let $a(x)$ denote the value of x on attribute a ($\forall a \in AT$). For an information system I, one can describe the relationship between two objects through their values on attributes. For example, suppose that $A \subseteq AT$, an indiscernibility relation $IND(A)$ may be defined as

$$IND(A) = \{(x,y) \in U^2 : a(x) = a(y), \forall a \in A\}. \quad (1)$$

Since $IND(A)$ is still an equivalence relation, $U/IND(A)$ is then denoted as the partition determined by indiscernibility relation $IND(A)$ on U. From the viewpoint of granular computing [47], each equivalence class in $U/IND(A)$ is an information granule. In other words, by a given indiscernibility relation, objects are granulated into a set of information granules, called a granular structure [29]. It should be noticed that partition is only a special granular structure, granular structure may also be a set of information granules induced by a general binary relation.

Definition 1. Let I be an information system in which $A \subseteq AT$, $\forall X \subseteq U$, the lower and upper approximations of X are denoted by $\underline{A}(X)$ and $\overline{A}(X)$, respectively, such that

$$\underline{A}(X) = \{x \in U : [x]_A \subseteq X\}; \quad (2)$$
$$\overline{A}(X) = \{x \in U : [x]_A \cap X \neq \emptyset\}; \quad (3)$$

where $[x]_A = \{y \in U : (x,y) \in IND(A)\}$ is the equivalence class of x in terms of A.

Qian et al.'s classical multigranulation rough set is different from Pawlak's rough set since the former is constructed on the basis of a family of the

binary relations instead of a single one. In this paper, to simplify our discussions, it is assumed that each attribute in an information system I is corresponding to an equivalence relation. Therefore, each attribute in I can induce a partition based granular structure on the universe of discourse and then all the attributes in I will induce a partitions based *multigranular structure*. In Qian et al.'s multigranulation rough set theory, two different models were defined. The first one is optimistic multigranulation rough set [31,37], the second one is pessimistic multigranulation rough set [35].

Definition 2. Let I be an information system in which $AT = \{a_1, a_2, \cdots, a_m\}$, $\forall X \subseteq U$, the optimistic multigranulation lower and upper approximations of X are denoted by $\underline{AT}^O(X)$ and $\overline{AT}^O(X)$, respectively, such that

$$\underline{AT}^O(X) = \{x \in U : [x]_{a_1} \subseteq X \vee \cdots \vee [x]_{a_m} \subseteq X\}; \quad (4)$$

$$\overline{AT}^O(X) = \sim (\underline{AT}^O(\sim X)); \quad (5)$$

where $[x]_{a_i}$ is the equivalence class of x in terms of a_i, $\sim X$ is the complement of set X.

The pair $[\underline{AT}^O(X), \overline{AT}^O(X)]$ is referred to as an optimistic multigranulation rough set of X.

Theorem 1. *Let I be an information system in which $AT = \{a_1, a_2, \cdots, a_m\}$, $\forall X \subseteq U$, we have*

$$\overline{AT}^O(X) = \{x \in U : [x]_{a_1} \cap X \neq \emptyset \wedge \cdots \wedge [x]_{a_m} \cap X \neq \emptyset\}. \quad (6)$$

Proof. It can be derived directly from Def. 2. □

Definition 3. Let I be an information system in which $AT = \{a_1, a_2, \cdots, a_m\}$, $\forall X \subseteq U$, the pessimistic multigranulation lower and upper approximations of X are denoted by $\underline{AT}^P(X)$ and $\overline{AT}^P(X)$, respectively, such that

$$\underline{AT}^P(X) = \{x \in U : [x]_{a_1} \subseteq X \wedge \cdots \wedge [x]_{a_m} \subseteq X\}; \quad (7)$$

$$\overline{AT}^P(X) = \sim (\underline{AT}^P(\sim X)). \quad (8)$$

The pair $[\underline{AT}^P(X), \overline{AT}^P(X)]$ is referred to as a pessimistic multigranulation rough set of X.

Theorem 2. *Let I be an information system in which*

$AT = \{a_1, a_2, \cdots, a_m\}$, $\forall X \subseteq U$, *we have*

$$\overline{AT}^P(X) = \{x \in U : [x]_{a_1} \cap X \neq \emptyset \vee \cdots \vee [x]_{a_m} \cap X \neq \emptyset\}. \quad (9)$$

Proof. It can be derived directly from Def. 3. □

Please refer to Refs. [30,31,35,37] for more details about optimistic and pessimistic multigranulation rough sets.

2.2. Multiset

Assume that U is the universe of discourse, a crisp multiset M of U is characterized by the count function such that

$$C_M : U \longrightarrow \mathbf{N} = \{0, 1, 2, \cdots\}; \quad (10)$$

in which $C_M(x)$ is the number of occurrences of the object $x \in U$ in M and $\mathbf{N}$ is the set of all natural numbers.

In this paper, for technical reasons, we consider a special multiset, which is a mapping from universe to a finite set, i.e.,

$$C_M : U \longrightarrow \{0, 1, 2, \cdots, m\}; \quad (11)$$

in which m is a fixed natural number.

Similar to classical set theory, it is not difficult to define some relations and operations on multisets. Suppose that M and N are two multisets over the same universe U, then

1. Inclusion: $M \sqsubseteq N \iff C_M(x) \leqslant C_N(x), \forall x \in U$;

2. Equality: $M = N \iff C_M(x) = C_N(x), \forall x \in U$;

3. Union: $M \sqcup N \iff C_{M \sqcup N}(x) = max\{C_M(x), C_N(x)\}, \forall x \in U$;

4. Intersection: $M \sqcap N \iff C_{M \sqcap N}(x) = min\{C_M(x), C_N(x)\}, \forall x \in U$;

5. Complement: $\neg M \iff C_{\neg M}(x) = m - C_M(x), \forall x \in U$;

6. Empty multiset: $\emptyset^0 \iff C_{\emptyset^0}(x) = 0, \forall x \in U$;

7. Full multiset: $U^m \iff C_{U^m}(x) = m, \forall x \in U$.

3. Multiple multigranulation rough set

3.1. Definition and properties

By Defs. 2 and 3, we can see that an object belongs to optimistic multigranulation lower approximation if and only if at least one of its equivalence

classes is contained in the target, an object belongs to pessimistic multigranulation lower approximation if and only if all of its equivalence classes are contained in the target. Similar conclusions of upper approximations can also be drawn by Theorems 1 and 2. In other words, for a given object, it may belong to lower approximation one or more times since it has one or more equivalence classes, which are contained in the target; it may also belong to upper approximation one or more times since it has one or more equivalence classes, which are intersected with the target. Nevertheless, optimistic and pessimistic multigranulation rough sets do not take frequencies of occurrences of such containment or intersection into consideration. Therefore, in this section, a new multigranulation rough set will be proposed to solve such problem. To achieve such goal, we need following definitions of characteristic functions.

Definition 4. Let I be an information system in which $AT = \{a_1, a_2, \cdots, a_m\}$, $\forall X \subseteq U$ and $\forall x \in U$, two characteristic functions are defined as

$$f_X^i(x) = \begin{cases} 1: [x]_{a_i} \subseteq X \\ 0: \text{otherwise} \end{cases} \quad (12)$$

$$g_X^i(x) = \begin{cases} 1: [x]_{a_i} \cap X \neq \emptyset \\ 0: \text{otherwise} \end{cases} \quad (13)$$

where $\forall a_i \in AT$.

To distinguish with optimistic and pessimistic multigranulation rough sets, our approach is referred to as multiple multigranualtion rough set in this paper.

Definition 5. Let I be an information system in which $AT = \{a_1, a_2, \cdots, a_m\}$, $\forall X \subseteq U$, the multiple multigranulation lower and upper approximations of X are denoted by $\underline{AT}^M(X)$ and $\overline{AT}^M(X)$, respectively, whose frequencies of occurrences for each $x \in U$ are:

$$C_{\underline{AT}^M(X)}(x) = \sum_{i=1}^m f_X^i(x); \quad (14)$$

$$C_{\overline{AT}^M(X)}(x) = \sum_{i=1}^m g_X^i(x). \quad (15)$$

The pair $[\underline{AT}^M(X), \overline{AT}^M(X)]$ is referred to as a multiple multigranulation rough set of X. Obviously, both $\underline{AT}^M(X)$ and $\overline{AT}^M(X)$ are multisets. $\forall x \in U$, $C_{\underline{AT}^M(X)}(x) \in \{0, 1, 2, \cdots, m\}$ is the frequency of occurrences that x in multiple multigran-

ulation lower approximation $\underline{AT}^M(X)$, $C_{\overline{AT}^M(X)}(x) \in \{0, 1, 2, \cdots, m\}$ is the frequency of occurrences that x in multiple multigranulation upper approximation $\overline{AT}^M(X)$.

Different from optimistic and pessimistic multigranulation rough sets, in our multiple multigranulation rough set model, an object may belong to multiple multigranulation lower/upper approximation more than one times.

Theorem 3. Let I be an information system in which $AT = \{a_1, a_2, \cdots, a_m\}$, $\forall X \subseteq U$ and $\forall x \in U$, we have

$$C_{\underline{AT}^M(X)}(x) \geqslant 1 \iff x \in \underline{AT}^O(X); \quad (16)$$

$$C_{\overline{AT}^M(X)}(x) = m \iff x \in \overline{AT}^O(X); \quad (17)$$

$$C_{\underline{AT}^M(X)}(x) = m \iff x \in \underline{AT}^P(X); \quad (18)$$

$$C_{\overline{AT}^M(X)}(x) \geqslant 1 \iff x \in \overline{AT}^P(X). \quad (19)$$

Proof. It can be derived directly from definitions of three multigranulation rough sets. □

In Theorem 3, formulas (16) and (17) show the relationship between multiple multigranulation rough set and optimistic multigranulation rough set, formulas (18) and (19) show the relationship between multiple multigranulation rough set and pessimistic multigranulation rough set.

Since classical set may be considered as a special multiset (given a classical set X, if $x \in X$, then $C_X(x) = 1$, otherwise, $C_X(x) = 0$), we then obtain the following theorem immediately.

Theorem 4. Let I be an information system in which $AT = \{a_1, a_2, \cdots, a_m\}$, $\forall X \subseteq U$, we have

$$\underline{AT}^O(X) \sqsubseteq \underline{AT}^M(X); \quad (20)$$

$$\underline{AT}^P(X) \sqsubseteq \underline{AT}^M(X); \quad (21)$$

$$\overline{AT}^O(X) \sqsubseteq \overline{AT}^M(X); \quad (22)$$

$$\overline{AT}^P(X) \sqsubseteq \overline{AT}^M(X). \quad (23)$$

Proof. We only prove formula (20), and the others can be proved analogously. Suppose that $\underline{AT}^O(X) \not\sqsubseteq \underline{AT}^M(X)$, then there must be $x \in U$ such that $C_{\underline{AT}^O(X)}(x) > C_{\underline{AT}^M(X)}(x)$ by the definition of in-

clusion of multiset. Since $\underline{AT}^O(X)$ is a classical set and then $C_{\underline{AT}^O(X)}(x) \in \{0,1\}$.

1. If $C_{\underline{AT}^O(X)}(x) = 0$, then $x \notin \underline{AT}^O(X)$, it follows that $C_{\underline{AT}^M(X)}(x) < 1$ by formula (16). Since $C_{\underline{AT}^M(X)}(x) \in \{0,1,2,\cdots,m\}$, then we know that $C_{\underline{AT}^M(X)}(x) = 0$.

2. If $C_{\underline{AT}^O(X)}(x) = 1$, then $x \in \underline{AT}^O(X)$, it follows that $C_{\underline{AT}^M(X)}(x) \geqslant 1$ by formula (16).

From discussions above, we obtain $C_{\underline{AT}^O(X)}(x) \leqslant C_{\underline{AT}^M(X)}(x)$, which is contradictive to assumption $C_{\underline{AT}^O(X)}(x) > C_{\underline{AT}^M(X)}(x)$ and then $\underline{AT}^O(X) \sqsubseteq \underline{AT}^M(X)$ holds. $\square$

Theorem 4 shows that both optimistic and pessimistic multigranulation lower approximations are smaller than multiple multigranulation lower approximation, both optimistic and pessimistic multigranulation upper approximations are also smaller than multiple multigranulation upper approximation.

For the readers' convenience, the relationships among three different multigranulation rough sets are shown in Fig. 1. In Fig. 1, each node denotes a multigranulation approximation or a target, and each line connects two nodes, where the lower node is a multiset inclusion of the upper node.
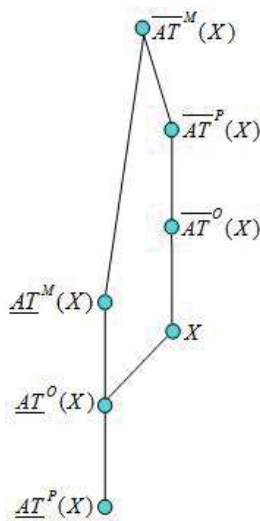

Figure 1: Relationships among three different multigranulation rough sets.

By Fig. 1, some interesting results can be obtained.

1. $\underline{AT}^M(X)$ and X are not comparable, that is to say, there is no containment between multiple multigranulation lower approximation $\underline{AT}^M(X)$ and target X. This is mainly because $\underline{AT}^M(X)$ is a multiset while X is only a crisp set.

2. Fig. 1 is a lattice, it also shows the results of Theorem 4. In such lattice, the derived partial order is the inclusion used in multisets, i.e., $\sqsubseteq$.

Theorem 5. *Let I be an information system in which* $AT = \{a_1, a_2, \cdots, a_m\}$, $\forall X, Y \subseteq U$, we have

$$\underline{AT}^M(X) \sqsubseteq \overline{AT}^M(X); \tag{24}$$

$$\underline{AT}^M(\emptyset) = \overline{AT}^M(\emptyset) = \emptyset^0; \tag{25}$$

$$\underline{AT}^M(U) = \overline{AT}^M(U) = U^m; \tag{26}$$

$$\underline{AT}^M(X) = \neg\overline{AT}^M(\sim X); \tag{27}$$

$$\overline{AT}^M(X) = \neg\underline{AT}^M(\sim X); \tag{28}$$

$$X \subseteq Y \Longrightarrow \underline{AT}^M(X) \sqsubseteq \underline{AT}^M(Y); \tag{29}$$

$$X \subseteq Y \Longrightarrow \overline{AT}^M(X) \sqsubseteq \overline{AT}^M(Y); \tag{30}$$

$$\underline{AT}^M(X \cap Y) \sqsubseteq \underline{AT}^M(X) \sqcap \underline{AT}^M(Y); \tag{31}$$

$$\underline{AT}^M(X \cup Y) \sqsupseteq \underline{AT}^M(X) \sqcup \underline{AT}^M(Y); \tag{32}$$

$$\overline{AT}^M(X \cap Y) \sqsubseteq \overline{AT}^M(X) \sqcap \overline{AT}^M(Y); \tag{33}$$

$$\overline{AT}^M(X \cup Y) \sqsupseteq \overline{AT}^M(X) \sqcup \overline{AT}^M(Y). \tag{34}$$

Proof.

1. $\forall x \in U$ and $\forall a_i \in AT$, since $[x]_{a_i} \subseteq X \Rightarrow [x]_{a_i} \cap X \neq \emptyset$, then by Eqs. (12) and (13), we have $f_X^i(x) = 1 \Rightarrow g_X^i(x) = 1$, it follows that $C_{\underline{AT}^M(X)}(x) \leqslant C_{\overline{AT}^M(X)}(x)$, i.e. $\underline{AT}^M(X) \sqsubseteq \overline{AT}^M(X)$.

2. $\forall x \in U$, since $\forall a_i \in AT$, we have $x \in [x]_{a_i}$ and then $[x]_{a_i} \not\subseteq \emptyset$, it follows that $C_{\underline{AT}^M(\emptyset)} = 0$ for each $x \in U$, i.e. $\underline{AT}^M(\emptyset) = \emptyset^0$. Similarity, it is not difficult to prove that $\overline{AT}^M(\emptyset) = \emptyset^0$.

3. $\forall x \in U$, since $\forall a_i \in AT$, we have $[x]_{a_i} \subseteq U$ and then $f_U^i(x) = 1$, it follows that $C_{\underline{AT}^M(U)} = m$ for each $x \in U$, i.e. $\underline{AT}^M(U) = U^m$. Similarity, it is not difficult to prove that $\overline{AT}^M(U) = U^m$.

4. $\forall x \in U$ and $\forall a_i \in AT$, if $f_X^i(x) = 1$, then by Eq. (12), we know that $[x]_{a_i} \subseteq X$, it follows that $[x]_{a_i} \cap (\sim X) = \emptyset$. By Eq. (13), $g_{\sim X}^i(x) = 0$. Similarity, it is not difficult to prove that if $f_X^i(x) = 0$, then $g_{\sim X}^i(x) = 1$. From discussions above, $f_X^i(x) + g_{\sim X}^i(x) = 1$ holds for each $a_i \in AT$ and then $\sum_{i=1}^{m} f_X^i(x) + \sum_{i=1}^{m} g_{\sim X}^i(x) = m$, from which we can conclude that $C_{\underline{AT}^M(X)}(x) = m - C_{\overline{AT}^M(\sim X)}(x)$, i.e. $\underline{AT}^M(X) = \neg \overline{AT}^M(\sim X)$.

5. The proof of $\overline{AT}^M(X) = \neg \underline{AT}^M(\sim X)$ is same to that of $\underline{AT}^M(X) = \neg \overline{AT}^M(\sim X)$.

6. $\forall x \in U$, since $X \subseteq Y$, then $\forall a_i \in AT$, we have $[x]_{a_i} \subseteq X \Rightarrow [x]_{a_i} \subseteq Y$, $f_X^i(x) = 1 \Rightarrow f_Y^i(x) = 1$, it follows that $C_{\underline{AT}^M(X)}(x) \leqslant C_{\underline{AT}^M(Y)}(x)$ holds for each $x \in U$, i.e. $\underline{AT}^M(X) \sqsubseteq \underline{AT}^M(Y)$.

7. The proof of $\overline{AT}^M(X) \sqsubseteq \overline{AT}^M(Y)$ is same to that of $\underline{AT}^M(X) \sqsubseteq \underline{AT}^M(Y)$.

8. Suppose that $\underline{AT}^M(X \cap Y) \not\sqsubseteq \underline{AT}^M(X) \sqcap \underline{AT}^M(Y)$, then there must be $x \in U$ such that $C_{\underline{AT}^M(X \cap Y)}(x) > C_{\underline{AT}^M(X) \sqcap \underline{AT}^M(Y)}(x)$. By Eq. (14), we know that $\sum_{i=1}^{m} f_{X \cap Y}^i(x) > min\{\sum_{i=1}^{m} f_X^i(x), \sum_{i=1}^{m} f_Y^i(x)\}$, i.e. $\sum_{i=1}^{m} f_{X \cap Y}^i(x) > \sum_{i=1}^{m} f_X^i(x)$ and $\sum_{i=1}^{m} f_{X \cap Y}^i(x) > \sum_{i=1}^{m} f_Y^i(x)$.

   $\sum_{i=1}^{m} f_{X \cap Y}^i(x) > \sum_{i=1}^{m} f_X^i(x)$ means that there must be $a_i \in AT$ such that $f_{X \cap Y}^i(x) = 1$ and $f_X^i(x) = 0$, i.e. $[x]_{a_i} \subseteq X \cap Y$ and $[x]_{a_i} \not\subseteq X$, such conclusion is contradictive to the basic property of set theory. The same is to $\sum_{i=1}^{m} f_{X \cap Y}^i(x) > \sum_{i=1}^{m} f_Y^i(x)$.

   From discussions above, we can conclude that $\underline{AT}^M(X \cap Y) \sqsubseteq \underline{AT}^M(X) \sqcap \underline{AT}^M(Y)$.

9. The proofs of formulas (32), (33) and (34) are similar to that of formula (31).

$\square$

Theorem 5 shows some basic properties of multiple multigranulation rough set.

Theorem 6. *Let I be an information system in which $AT = \{a_1, a_2, \cdots, a_m\}$, suppose that $B = \{b_1, b_2, \cdots, b_n\} \subseteq AT$, $\forall X \subseteq U$, we have*

$$\underline{B}^M(X) \sqsubseteq \underline{AT}^M(X); \qquad (35)$$

$$\overline{B}^M(X) \sqsubseteq \overline{AT}^M(X). \qquad (36)$$

Proof. It can be derived directly from Def. 5. $\square$

Theorem 6 shows the monotonic variation of multiple multigranulation lower and upper approximations with the monotonic increasing or decreasing of number of equivalence relations, the details are: if the number of used equivalence relations is increasing, then both multiple multigranulation lower and upper approximations are increasing. It should be noticed that such result is different from those of optimistic and pessimistic multigranulation rough sets. In optimistic multigranulation rough set, with the monotonic increasing of number of equivalence relations, the lower approximation is increasing while the upper approximation is decreasing; in pessimistic multigranulation rough set, with the monotonic increasing of number of equivalence relations, the lower approximation is decreasing while the upper approximation is increasing.

3.2. Approximate quality

Following Pawlak's rough set theory, Qian et al. have presented the definitions of approximate qualities based on optimistic and pessimistic multigranulation rough sets. Since in this paper, the multiple multigranulation lower approximation is a multiset rather than a classical set, we need to further present new definition of approximate quality.

Definition 6. Let $I = < U, AT \cup \{d\} >$ be a decision system in which $AT = \{a_1, a_2, \cdots, a_m\}$, partition $U/IND(\{d\}) = \{X_1, \cdots, X_k\}$ is the set of decision classes determined by decision attribute d, approximate qualities of d based on optimistic, pessimistic and multiple multigranulation rough sets are defined as $\gamma^O(AT, d)$, $\gamma^P(AT, d)$ and $\gamma^M(AT, d)$, respectively, such that

$$\gamma^O(AT, d) = \frac{|\cup_{j=1}^{k} \underline{AT}^O(X_j)|}{|U|}; \qquad (37)$$

$$\gamma^P(AT, d) = \frac{|\cup_{j=1}^{k} \underline{AT}^P(X_j)|}{|U|}; \qquad (38)$$

$$\gamma^M(AT, d) = \frac{\#(\sqcup_{j=1}^{k} \underline{AT}^M(X_j))}{\#(U^m)}; \qquad (39)$$

where $|X|$ is the cardinal number of classical set X, $\#(Y)$ is the cardinal number of multiset Y such that

$\#(Y) = C_Y(x_1) + \cdots + C_Y(x_{|U|})$.

Theorem 7. *Let $I = <U, AT \cup \{d\}>$ be a decision system in which $AT = \{a_1, a_2, \cdots, a_m\}$, we have*

$$\gamma^P(AT, d) \leqslant \gamma^M(AT, d) \leqslant \gamma^O(AT, d). \qquad (40)$$

Proof. Firstly, let us prove $C_{\cup_{j=1}^k \underline{AT}^P(X_j)}(x) \leqslant \frac{C_{\sqcup_{j=1}^k \underline{AT}^M(X_j)}(x)}{m}$ for each $x \in U$. Since pessimistic multigranulation lower approximation is a classical set, we then know that $C_{\cup_{j=1}^k \underline{AT}^P(X_j)}(x) \in \{0, 1\}$.

1. If $C_{\cup_{j=1}^k \underline{AT}^P(X_j)}(x) = 1$, then there must be $X_j \in U/IND(\{d\})$ such that $x \in \underline{AT}^P(X_j)$. By Theorem 3, $C_{\underline{AT}^M(X_j)}(x) = m$ holds. By union operation defined on multiset, we know that $C_{\sqcup_{j=1}^k \underline{AT}^M(X_j)}(x) = m$ and then $\frac{C_{\sqcup_{j=1}^k \underline{AT}^M(X_j)}(x)}{m} = 1$.

2. If $C_{\cup_{j=1}^k \underline{AT}^P(X_j)}(x) = 0$, then $\forall X_j \in U/IND(\{d\})$, $x \notin \underline{AT}^P(X_j)$. By Theorem 3, $0 \leqslant C_{\underline{AT}^M(X_j)}(x) < m$ holds and then $0 \leqslant \frac{C_{\sqcup_{j=1}^k \underline{AT}^M(X_j)}(x)}{m} < 1$.

From discussions above, we obtain $C_{\cup_{j=1}^k \underline{AT}^P(X_j)}(x) \leqslant \frac{C_{\sqcup_{j=1}^k \underline{AT}^M(X_j)}(x)}{m}$. Finally, by Eq. (38),

$$\gamma^P(AT, d) = \frac{|\cup_{j=1}^k \underline{AT}^P(X_j)|}{|U|}$$

$$= \frac{C_{\cup_{j=1}^k \underline{AT}^P(X_j)}(x_1) + \cdots + C_{\cup_{j=1}^k \underline{AT}^P(X_j)}(x_{|U|})}{|U|}$$

$$\leqslant \frac{C_{\sqcup_{j=1}^k \underline{AT}^M(X_j)}(x_1) + \cdots + C_{\sqcup_{j=1}^k \underline{AT}^M(X_j)}(x_{|U|})}{m \cdot |U|}$$

$$= \gamma^M(AT, d).$$

Similarity, it is not difficult to prove that $\gamma^M(AT, d) \leqslant \gamma^O(AT, d)$. $\qquad \square$

Theorem 7 tells us that the approximation quality based on multiple multigranulation rough set is between those based on optimistic and pessimistic multigranulation rough sets.

3.3. Approximate distribution reducts

Attribute reduction [16,25,28,36] plays a crucial role in the development of rough set theory. In Pawlak's

rough set theory, reduct is a minimal subset of attributes, which is independent and has the same discernibility power as all of the attributes. In recent years, with respect to different requirements, different types of reducts have been proposed. In this paper, we will introduce the concept of approximate distribution reduct [22,54] into our multiple multigranulation rough set. Such goal is to preserve frequencies of occurrences that objects belong to multiple multigranualtion lower or upper approximations.

Definition 7. Let $I = <U, AT \cup \{d\}>$ be a decision system in which $AT = \{a_1, a_2, \cdots, a_m\}$, partition $U/IND(\{d\}) = \{X_1, \cdots, X_k\}$ is the set of decision classes determined by decision attribute d, $ML_{AT} = \{\underline{AT}^M(X_1), \cdots, \underline{AT}^M(X_k)\}$ is the set of multiple multigranulation lower approximations of all decision classes, $MU_{AT} = \{\overline{AT}^M(X_1), \cdots, \overline{AT}^M(X_k)\}$ is the set of multiple multigranulation upper approximations of all decision classes, then

1. $B = \{b_1, b_2, \cdots, b_n\} \subseteq AT$ is referred to as a multiple multigranulation lower approximate distribution reduct in I if and only if $ML_B = ML_{AT}$ and $\forall B' \subset B, ML_{B'} \neq ML_{AT}$;

2. $B = \{b_1, b_2, \cdots, b_n\} \subseteq AT$ is referred to as a multiple multigranulation upper approximate distribution reduct in I if and only if $MU_B = MU_{AT}$ and $\forall B' \subset B, MU_{B'} \neq MU_{AT}$.

Specially, in Definition 7, if only $ML_B = ML_{AT}$ or $MU_B = MU_{AT}$ hold, then B is referred to as the multiple multigranulation lower or upper approximate distribution consistent attributes sets in I. Obviously, multiple multigranulation lower or upper approximate distribution reducts in I are the minimal subsets of attributes, which preserve multiple multigranulation lower or upper approximations of all the decision class, i.e. multiple multigranulation lower or upper approximate distribution reducts can be used to preserve the distributions of multiple multigranulation lower or upper approximations.

Theorem 8. *Let $I = <U, AT \cup \{d\}>$ be a decision system in which $AT = \{a_1, a_2, \cdots, a_m\}$, if $B = \{b_1, b_2, \cdots, b_n\} \subseteq AT$, then*

1. $ML_B = ML_{AT} \iff C_{\underline{B}^M(X_j)}(x) = C_{\underline{AT}^M(X_j)}(x)$, $\forall x \in U, \forall X_j \in U/IND(\{d\})$;

2. $MU_B = MU_{AT} \Longleftrightarrow C_{\overline{B}^M(X_j)}(x) = C_{\overline{AT}^M(X_j)}(x)$,
   $\forall x \in U, \forall X_j \in U/IND(\{d\})$.

Proof. "⇒": If $ML_B = ML_{AT}$, by Definition 7, we know that $\underline{B}^M(X_j) = \underline{AT}^M(X_j)$ for each $X_j \in U/IND(\{d\})$. By Definition 5, $C_{\underline{B}^M(X_j)}(x) = C_{\underline{AT}^M(X_j)}(x)$ holds for each $x \in U$ obviously.

"⇐": If $C_{\underline{B}^M(X_j)}(x) = C_{\underline{AT}^M(X_j)}(x)$, $\forall x \in U$, $\forall X_j \in U/IND(\{d\})$, then by Definition 5, we know that $\underline{B}^M(X_j) = \underline{AT}^M(X_j)$ holds for each $X_j \in U/IND(\{d\})$, it follows that $ML_B = ML_{AT}$ through Definition 7.

Similarity, it is not difficult to prove that $MU_B = MU_{AT} \Longleftrightarrow C_{\overline{B}^M(X_j)}(x) = C_{\overline{AT}^M(X_j)}(x), \forall x \in U, \forall X_j \in U/IND(\{d\})$. □

Theorem 8 tells us that multiple multigranulation lower or upper approximate distribution consistent attributes sets can be used to preserve frequencies of occurrences that objects belong to multiple multigranulation lower or upper approximations, repsectively.

Based on the result shown in Theorem 6, we know that multiple multigranulation rough lower and upper approximations are monotonic variations with the monotonic increasing or decreasing attributes. Therefore, let $I = <U, AT \cup \{d\}>$ be a decision system in which $AT = \{a_1, a_2, \cdots, a_m\}$, suppose that $B = \{b_1, b_2, \cdots, b_n\} \subseteq AT, \forall b \in B$, we define the following two coefficients for two approximate distribution reducts, respectively:

$$Sig_{in}^L(b,B,d) = \#(\sqcup_{j=1}^k \underline{B}^M(X_j)) - \#(\sqcup_{j=1}^k \underline{B-\{b\}}^M(X_j)); \qquad (41)$$

$$Sig_{in}^U(b,B,d) = \#(\sqcup_{j=1}^k \overline{B}^M(X_j)) - \#(\sqcup_{j=1}^k \overline{B-\{b\}}^M(X_j)); \qquad (42)$$

as the significance of b in B relative to decision d. $Sig_{in}^L(b,B,d)$ reflects the changes of multiple multigranulation lower approximations if attribute b is eliminated from B while $Sig_{in}^U(b,B,d)$ reflects the changes of multiple multigranulation upper approximations if attribute b is eliminated from B. Accordingly, we can also define

$$Sig_{out}^L(b,B,d) = \#(\sqcup_{j=1}^k \underline{B-\{b\}}^M(X_j)) - \#(\sqcup_{j=1}^k \underline{B}^M(X_j)); \qquad (43)$$

$$Sig_{out}^U(b,B,d) = \#(\sqcup_{j=1}^k \overline{B-\{b\}}^M(X_j)) - \#(\sqcup_{j=1}^k \overline{B}^M(X_j)); \qquad (44)$$

where $\forall b \in AT - B$.

$Sig_{out}^L(b,B,d)$ measures the change of multiple multigranulation lower approximations if attribute b is introduced into B, $Sig_{out}^U(b,B,d)$ measures the change of multiple multigranulation upper approximations if attribute b is introduced into B.

By above measures, a forward greedy attribute reduction algorithm for computing reduct can be designed as following.

Algorithm Attribute reduction based on multiple multigranulation rough set in I.

Input: Decision system I;

Output: A multiple multigranulation lower approximate distribution reduct B.

Step 1: $B \leftarrow \emptyset$, compute ML_{AT};

Step 2: Compute the significance of each $a_i \in AT$ with $Sig_{in}^L(a_i, AT, d)$;

Step 3: $B \leftarrow a_j$ where $Sig_{in}^L(a_j, AT, d) = max\{Sig_{in}^L(a_i, AT, d) : \forall a_i \in AT\}$;

Step 4: Do
$\quad\quad \forall a_i \in AT - B$, compute $Sig_{out}^L(a_i, B, d)$;
$\quad\quad$ If $Sig_{out}^L(a_j, AT, d) = max\{Sig_{out}^L(a_i, AT, d) : \forall a_i \in AT\}$
$\quad\quad\quad B = B \cup \{a_j\}$;
$\quad\quad$ End
$\quad\quad$ Until $ML_B = ML_{AT}$;

Step 5: $\forall a_i \in B$
$\quad\quad$ If $ML_{B-\{a_i\}} = ML_B$
$\quad\quad\quad B = B - \{a_i\}$;
$\quad\quad$ End

Step 6: Return B.

If Sig_{in}^L and Sig_{out}^L are replaced by Sig_{in}^U and Sig_{out}^U, respectively, then the above algorithm can be used to compute a multiple multigranulation upper approximate distribution reduct.

The above forward greedy attribute reduction algorithm is starting with the attribute with maximal change of significance when eliminating a single attribute, we then take the attribute with the maximal

significance into the attribute subset in each loop until the entire approximate distribution of this attribute subset satisfies the target requirement, and then we can get a attribute subset. Step 5 is to delete the redundant attribute in the obtained attribute subset.

In the above algorithm, in the worst case, all attributes should be checked for comparing the approximation equalities. Moreover, an attribute is never checked twice. Therefore the number of checking steps is bounded by $|AT|$. Moreover, the time consuming of computing approximation quality is $O(|U|^2)$, therefore, the time complexity of this algorithm is $O(|U|^2 * |AT|)$. Such complexity if same to Qian et al.'s pessimistic multigranulation rough set based attribute reduction [32] and lower than Qian et al.'s optimistic multigranulation rough set based attribute reduction $\left(O(|U|^2 * |AT| * 2^{|AT|})\right)$ [37].

4. Experimental results

In the following, through experimental analysis, we illustrate the differences among three multigranulation rough sets mentioned in this paper. All the experiments have been carried out on a personal computer with Windows 7, Intel Core 2 Duo T5800 CPU (2.00 GHz) and 2.00 GB memory. The programming language is Matlab 2010.

Table 1: Data sets description.

Data ID	Data sets	Objects	Attributes	Decision classes
1	Zoo	101	16	7
2	German Credit	1000	20	2
3	Car	1728	6	4
4	Contraceptive Method Choice	1473	9	3
5	Breast Cancer Wisconsin	569	30	2
6	Libras Movement	360	90	15
7	Glass Identification	214	9	6
8	Ionosphere	351	34	2

We have downloaded eight public data sets from UCI Repository of Machine Learning databases, which are described in Tab. 1. In our experiment,

we assume that each attribute in a data set can induce an equivalence relation and then all attributes in a data set will induce a family of equivalence relations on the universe of discourse.

4.1. Comparison among approximate qualities

Fig. 2 shows the experimental results of approximate qualities on eight data sets, each sub–figure in Fig. 2 is corresponding to the computing result on one data set. For each sub-figure, the x–coordinate pertains to the number of attributes while the y–coordinate concerns obtained approximate quality. The tagging "OMGRS" is the computing result based on optimistic multigranulation rough set, the tagging "MMGRS" is the computing result based on multiple multigranulation rough set and the tagging "PMGRS" is the computing result based on pessimistic multigranulation rough set.

It is not difficult to note from Fig. 2 that no matter how many attributes are used, approximate qualities based on multiple multigranulation rough set are between those based on optimistic and pessimistic multigranulation rough sets. Such experimental results demonstrate the theoretical result shown in Theorem 7. Moreover, it should be noticed that different from optimistic and pessimistic multigranulation rough sets, approximate quality based on multiple multigranulation rough set is not necessarily monotonic increasing or decreasing with the increasing of attributes. Though for each decision class X_j in a data set, its multiple multigranulation lower approximation is consistently increasing (see Theorem 6), the denominator of approximate quality is also increasing since $\#(U_m) = m \cdot |U|$ (m is the number of used attributes) and then such approximate quality is not necessarily monotonic.

4.2. Comparison among approximate distribution reducts

Tabs. 2–4 show the results of approximate distribution reducts and reduction ratios based on optimistic, pessimistic and multiple multigranulation rough sets, respectively.

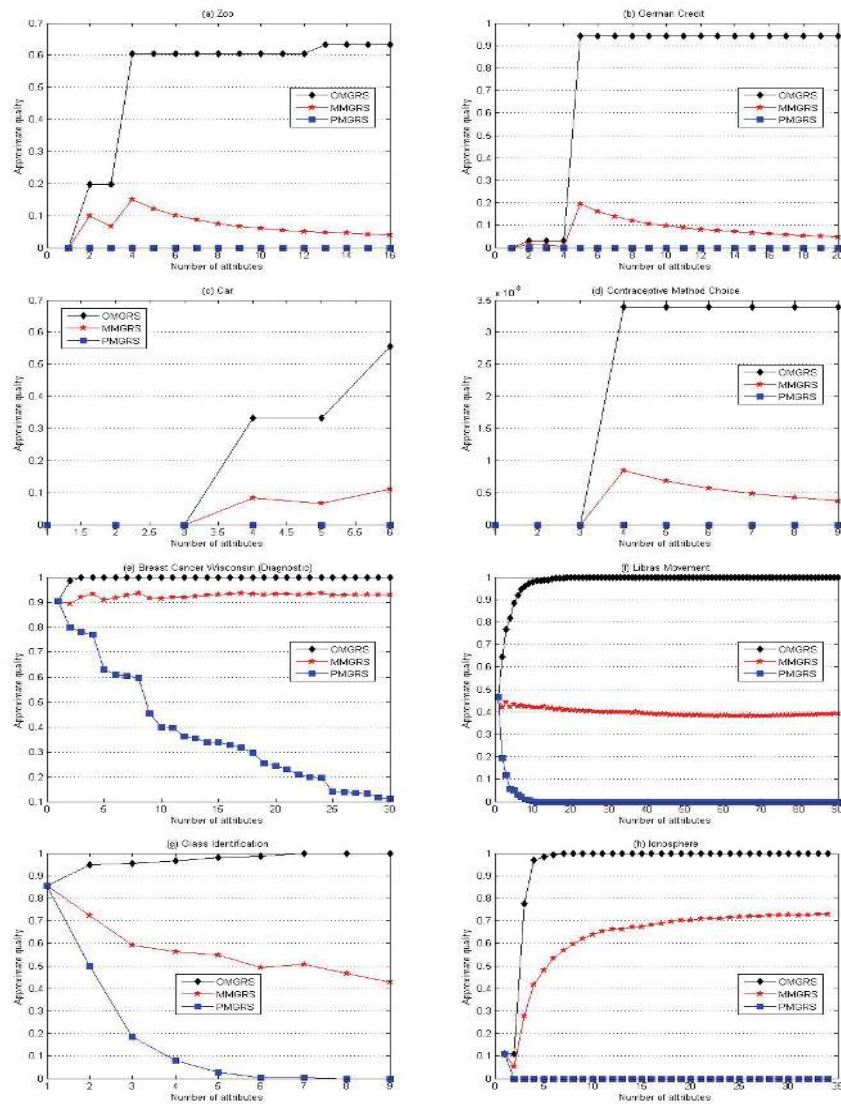By comparing with Tabs. 2 and 3, we can observe following.

Figure 2: Approximate qualities based on three multigranulation rough sets.

Table 2: Reducts of optimistic multigranulation rough set.

Data ID	Attributes in lower approximate distribution reduct	Reduction ratio	Attributes in upper approximation distribution reduct	Reduction ratio
1	2,4,14	81.25%	4,5,6,9,12,13,14	56.25%
2	5,13	90.00%	5,13	90.00%
3	4,6	66.67%	1,2,4,5,6	16.67%
4	4	88.89%	1,4	77.78%
5	1,16	93.33%	1,16	93.33%
6	1,3,12,28,65,85,89	92.22%	1,18,71	96.67%
7	1,2,4,5,7	44.44%	1,2,7	66.67%
8	1,4,6,7,18,25	82.36%	1,4,6,7,18,25	82.36%

Table 3: Reducts of pessimistic multigranulation rough set.

Data ID	Attributes in lower approximate distribution reduct	Reduction ratio	Attributes in upper approximation distribution reduct	Reduction ratio
1	1	93.75%	7	93.75%
2	1	95.00%	1	95.00%
3	1	83.33%	3	83.33%
4	1	88.89%	2	88.89%
5	1,$\cdots$,10,12,13,14,16,$\cdots$,30	6.67%	1,$\cdots$,10,12,13,14,16,$\cdots$,30	93.33%
6	1,3,10,42,52	94.44%	1,2,4,$\cdots$,7,9,10,12,13,15,$\cdots$,18,20,$\cdots$42, 44,$\cdots$,68,70,71,72,74,75,76,78,$\cdots$,89	11.11%
7	1,5,8,9	55.56%	2,$\cdots$,9	11.11%
8	2	97.06%	2	97.06%

Table 4: Reducts of multiple multigranulation rough set.

Data ID	Attributes in lower approximate distribution reduct	Reduction ratio	Attributes in upper approximation distribution reduct	Reduction ratio
1	2,4,13	81.25%	1,$\cdots$,16	0.00%
2	2,5,13	85.00%	1,$\cdots$,20	0.00%
3	4,6	66.67%	1,$\cdots$,6	0.00%
4	4	88.89%	1,$\cdots$,9	0.00%
5	1,$\cdots$,30	0.00%	1,$\cdots$,30	0.00%
6	1,$\cdots$,90	0.00%	1,$\cdots$,90	0.00%
7	1,$\cdots$,9	0.00%	1,$\cdots$,9	0.00%
8	1,3,$\cdots$,34	2.94%	1,$\cdots$,34	0.00%

1. For lower approximate distribution reduct, reduction ratios of optimistic multigranulation rough set are equal or lower than those of pessimistic multigranulation rough set except the 5th data set.

2. For upper approximate distribution reduct, reduction ratios of optimistic multigranulation rough set are equal or lower than those of pessimistic multigranulation rough set except the 5th and 6th data sets.

Through experimental analysis, though reduction ratios of pessimistic multigranulation rough set may be higher than those of optimistic multigranulation rough set, the limitation of pessimistic multigranulation rough set is stricter than that of optimistic multigranulation rough set (such case can be observed by comparing Definitions 2 and 3), it follows that we obtain empty set for pessimistic multigranulation lower approximation (see Fig. 2, the approximate quality is zero on eight data sets) and full universe for pessimistic multigranulation upper approximation frequently in our experiment. From this point of view, pessimism is meaningless since we obtain nothing of certainty or uncertainty.

Furthermore, by comparing with Tabs. 2 and 4, we can observe following.

1. For both lower and upper approximate distribution reducts, reduction ratios of multiple multigranulation rough set are lower than those of optimistic multigranulation rough set. Such difference is coming from the difference between optimistic and multiple multigranulation rough sets. Multiple multigranulation rough set is stricter than optimistic multigranulation rough set since the former needs to compute frequencies of occurrences such that objects belong to lower or upper approximations. To preserve such frequencies of occurrences for each object in the universe, more attributes are required.

2. The reducts of optimistic multigranulation rough set are included into those of multiple multigranulation rough set. This is mainly because if frequencies of occurrences for each object in lower/upper approximations are preserved, then

belonging or not belonging to optimistic multi-granulation lower/upper approximations are also preserved.

From discussions above, we may obtain the following theorem.

Theorem 9. *Let I be a decision system in which $AT = \{a_1, a_2, \cdots, a_m\}$, partition $U/IND(\{d\}) = \{X_1, \cdots, X_k\}$ is the set of decision classes determined by decision attribute d, suppose that $B = \{b_1, b_2, \cdots, b_n\} \subseteq AT$, then*
1. $ML_B = ML_{AT} \Rightarrow OL_B = OL_{AT}$;
2. $MU_B = MU_{AT} \Rightarrow OU_B = OU_{AT}$;
where $OL_{AT} = \{\underline{AT}^O(X_1), \cdots, \underline{AT}^O(X_k)\}$ and $OU_{AT} = \{\overline{AT}^O(X_1), \cdots, \overline{AT}^O(X_k)\}$.

Proof. If $ML_B = ML_{AT}$, then by Theorem 8, we have $C_{\underline{B}^M(X_j)}(x) = C_{\underline{AT}^M(X_j)}(x)$ for each $x \in U$ and each $X_j \in U/IND(\{d\})$. Therefore,
1. $\forall x \in U, \forall X_j \in U/IND(\{d\})$, if $x \in \underline{AT}^O(X_j)$, i.e. $C_{\underline{AT}^M(X_j)}(x) \geqslant 1$, then we also have $C_{\underline{B}^M(X_j)}(x) \geqslant 1$, it follows that $x \in \underline{B}^O(X_j)$ and then $\underline{AT}^O(X_j) \subseteq \underline{B}^O(X_j)$.
2. $\forall x \in U, \forall X_j \in U/IND(\{d\})$, if $x \notin \underline{AT}^O(X_j)$, i.e. $C_{\underline{AT}^M(X_j)}(x) = 0$, then we also have $C_{\underline{B}^M(X_j)}(x) = 0$, it follows that $x \notin \underline{B}^O(X_j)$ and then $\underline{B}^O(X_j) \subseteq \underline{AT}^O(X_j)$.

To sum up, we know that $\underline{AT}^O(X_j) = \underline{B}^O(X_j)$ for each $X_j \in U/IND(\{d\})$, i.e., $OL_B = OL_{AT}$.

Similarity, it is not difficult to prove that $MU_B = MU_{AT} \Rightarrow OU_B = OU_{AT}$. □

Theorem 9 tells us that multiple multigranulation lower or upper approximate distribution consistent attributes sets are also optimistic multigranulation lower or upper approximate distribution consistent attributes sets, respectively.

Finally, by comparing with Tabs. 3 and 4, we can observe following.
1. For both lower and upper approximate distribution reducts, reduction ratios of multiple multigranulation rough set are lower than those of pessimistic multigranulation rough set. Similar to optimistic case, such difference is also coming from the difference between pessimistic and multiple multigranulation rough sets. For example, suppose that $B \subseteq AT$ is a pessimistic

multigranulation lower approximate distribution reduct, if an object belongs to the lower approximation of a target, then all of the attributes in B support the containment between equivalence classes and target. However, B cannot be always satisfied with the preserving of frequencies of occurrences for such object in multiple multigranulation lower approximations. In other words, pessimistic multigranulation lower approximate distribution reduct can only guarantee all the equivalence classes (w.r.t. all attributes in reduct) of an object are contained in the target, it cannot always preserve the invariance of frequencies of occurrences for objects in multiple multigranulation lower approximations.

2. The pessimistic multigranulation upper approximate distribution reducts are included into those of multiple multigranulation upper approximate distribution reducts. This is mainly because if in multiple multigranulation rough set, frequencies of occurrences for each object in upper approximations are preserved, then belonging or not belonging to upper approximations are also preserved. For example, if the frequency of occurrences for an object in lower approximations is m (m is the number of original attributes), then to preserve such frequency of occurrences, no attribute can be deleted, which also preserve the belonging of such object to pessimistic multigranulation upper approximation.

From discussions above, similar to Theorem 9, we may also obtain the following theorem.

Theorem 10. *Let I be a decision system in which $AT = \{a_1, a_2, \cdots, a_m\}$, partition $U/IND(\{d\}) = \{X_1, \cdots, X_k\}$ is the set of decision classes determined by decision attribute d, suppose that $B = \{b_1, b_2, \cdots, b_n\} \subseteq AT$, then*

1. $ML_B = ML_{AT} \Rightarrow PL_B = PL_{AT}$;
2. $MU_B = MU_{AT} \Rightarrow PU_B = PU_{AT}$;

where $PL_{AT} = \{\underline{AT}^P(X_1), \cdots, \underline{AT}^P(X_k)\}$ and $PU_{AT} = \{\overline{AT}^P(X_1), \cdots, \overline{AT}^P(X_k)\}$.

Proof. The proof of Theorem 10 is similar to that of Theorem 9. □

Theorem 10 tells us that multiple multigranulation lower or upper approximate distribution consistent attributes sets are also pessimistic multigranulation lower or upper approximate distribution consistent attributes sets, respectively.

4.3. Related discussions

In this subsection, we summarize the differences between multiple and classical multigranulation rough set approaches.

1. Approximate quality of multiple multigranulation rough set is equal or smaller than that of optimistic multigranulation rough set; it is also equal or higher than that of pessimistic multigranulation rough set. In Section 4.1, we have also noticed that different from optimistic and pessimistic multigranulation rough set approaches, approximate quality of multiple multigranulation rough set is not necessarily monotonic with the increasing or decreasing of used attributes.

2. By comparing with Qian et al.'s two multigranulation rough sets, our multiple multigranulation rough set requires more attributes to construct a reduct. In Section 4.2, we have explained that such difference is coming from the stricter limitation of multiple multigranulation rough set. Such rough set needs to count frequencies of occurrences that objects belong to lower/upper approximations. Therefore, fewer attributes can be deleted.

3. In our experimented data sets, multiple multigranulation lower or upper approximate distribution reducts include optimistic multigranulation lower or upper approximate distribution reducts, respectively; multiple multigranulation upper approximate distribution reduct includes pessimistic multigranulation upper approximate distribution reduct. In Section 4.2, we have derived two theorems (Theorems 9 and 10) based on such experimental results.

5. Conclusions

To count frequencies of occurrences that objects belong to lower or upper approximations under multigranulation environment, we have presented a general framework for the study of multiple multigranulation rough set in this paper. Based on this framework, a general heuristic algorithm is presented to compute multiple multigranulation lower/upper approximate distribution reducts. Experimental studies pertaining to eight UCI data sets show the differences of approximate qualities and reducts between our and Qian et al.'s multigranulation rough sets.

The following research topics deserve further investigation:

1. The construction of multiple multigranulation rough set in fuzzy environment.

2. Dynamic updating of multiple multigranulation rough set and dynamic computing of reducts when multigranulation environment is dynamic variation.

3. Using multiple multigranulation rough set approach to design classifier.

Acknowledgment

References

1. H. M. Abu–Donia, Multi knowledge based rough approximations and applications, Knowledge–Based Systems 26 (2012) 20–29.

2. Y. Chen, An adjustable multigranulation fuzzy rough set, International Journal of Machine Learning and Cybernetics 7 (2016) 267–274.

3. D. G. Chen, Q. H. Hu, Y. P. Yang, Parameterized attribute reduction with Gaussian kernel based fuzzy rough sets, Information Sciences 181 (2011) 5169–5179.

4. H. L. Dou, X. B. Yang, J. Y. Fan, S. P. Xu, The models of variable precision multigranulation rough sets, in: Rough Sets and Knowledge Technology– 7th International Conference, Chengdu, China, 17–19 August, 2012, pp. 465–473.

5. L. Feng, G. Y. Wang, X. X. Li, Knowledge acquisition in vague objective information systems based on rough sets, Expert Systems 27 (2010) 129–142.

6. T. Herawan, M. M. Deris, J. H. Abawajy, A rough set approach for selecting clustering attribute, Knowledge–Based Systems 23 (2010) 220–231.

7. Q. H. Hu, S. An, D. R. Yu, Soft fuzzy rough sets for robust feature evaluation and selection, Information Sciences 180 (2010) 4384–4400.

8. Q. H. Hu, X. J. Che, L. Zhang, D. Zhang, M. Z. Guo, D. R. Yu, Rank entropy based decision trees for monotonic classification, IEEE Transactions on Knowledge and Data Engineering 24 (2012) 2052–2064.

9. Q. H. Hu, W. W. Pan, L. Zhang, D. Zhang, Y. P. Song, M. Z. Guo, D. R. Yu, Feature selection for monotonic classification, IEEE Transactions on Fuzzy Systems 20 (2012) 69–81.

10. B. Huang, D. K. Wei, H. X. Li, Y. L. Zhuang, Using a rough set model to extract rules in dominance–based interval–valued intuitionistic fuzzy information systems, Information Sciences 221 (2013) 215–229.

11. X. Y. Jia, W. H. Liao, Z. M. Tang, L. Shang, Minimum cost attribute reduction in decision–theoretic rough set models, Information Sciences 219 (2013) 151–167.

12. M. A. Khan, M. Banerjee, Formal reasoning with rough sets in multiple–source approximation systems, International Journal of Approximate Reasoning 49 (2008) 466–477.

13. D. E. Knuth, The art of computer programming, Volume 2, Addison–Wesley, 1969.

14. G. P. Lin, Y. H. Qian, J. J. Li, NMGRS: Neighborhood–based multigranulation rough sets, International Journal of Approximate Reasoning 53 (2012) 1080–1093.

15. G. P. Lin, J. Y. Liang, Y. H. Qian, Topological approach to multigranulation rough sets, International Journal of Machine Learning and Cybernetics 5 (2014) 233–243.

16. G. M. Lang, Q. G. Li, T. Yang, An incremental approach to attribute reduction of dynamic set-valued information systems, International Journal of Machine Learning and Cybernetics 5 (2014) 775–788.

17. X. Liu, Y. H. Qian, J. Y. Liang, A rule–extraction framework under multigranulation rough sets, International Journal of Machine Learning and Cybernetics 5 (2014) 319C-326.

18. J. H. Li, Y. Ren, C. L. Mei, Y. H. Qian, X. B. Yang, A comparative study of multigranulation rough sets and concept lattices via rule acquisition, Knowledge–Based Systems 91 (2016) 152–164.

19. J. Y. Liang, F. Wang, C. Y. Dang, Y. H. Qian, An efficient rough feature selection algorithm with a multi–granulation view, International Journal of Approximate Reasoning 53 (2012) 912–926.

20. J. Y. Liang, F. Wang, C. Y. Dang, Y. H. Qian, A group incremental approach to feature selection applying rough set technique, IEEE Transactions on Knowledge and Data Engineering 26 (2012) 294–308.

21. F. Y. Meng, X. H. Chen, Q. Zhang, An approach to interval-valued intuitionistic uncertain linguistic multi-attribute group decision making, International Journal of Machine Learning and Cybernetics 6 (2015) 859–871.

22. J. S. Mi, W. X. Zhang, An axiomatic characterization of a fuzzy generalization of rough sets, Information Sciences 160 (2004) 235–249.

23. D. Q. Miao, C. Gao, N. Zhang, Z. F. Zhang, Diverse reduct subspaces based co–training for partially labeled data, International Journal of Approximate Reasoning 52 (2011) 1103–1117.

24. S. Miyamoto, Generalizations of multisets and rough approximations, International Journal of Intelligent Systems 19 (2004) 639–652.

25. J. M. Ma, Y. Leung, W. X. Zhang, Attribute reductions in object-oriented concept lattice, International Journal of Machine Learning and Cybernetics 5 (2014) 789–813.

26. F. Min, W. Zhu, Attribute reduction of data with error ranges and test costs, Information Sciences 211 (2012) 48–67.

27. Z. Pawlak, Rough sets–theoretical aspects of reasoning about data, Kluwer Academic Publishers, 1992.

28. D. Pei, J. S. Mi, Attribute reduction in decision formal context based on homomorphism, International Journal of Machine Learning and Cybernetics 2 (2011) 289–293.

29. Y. H. Qian, J. Y. Liang, W. Z. Wu, C. Y. Dang, Information granularity in fuzzy binary GrC model, IEEE Transactions on Fuzzy Systems 19 (2011) 253–264.

30. Y. H. Qian, J. Y. Liang, Rough set method based on multi–granulations, in: 5th IEEE International Conference on Cognitive Informatics, Beijing, China 17–19 July, 2006, pp. 297–304.

31. Y. H. Qian, J. Y. Liang, C. Y. Dang, Incomplete multigranulation rough set, IEEE Transactions on Systems, Man and Cybernetics, Part A 20 (2010) 420–431.

32. Y. H. Qian, S. Y. Li, J. Y. Liang, Z. Z. Shi, F. Wang, Pessimistic rough set based decisions: A multigranulation fusion strategy, Information Sciences 264 (2014) 196–210.

33. Y. H. Qian, J. Y. Liang, W. Pedrycz, C. Y. Dang, Positive approximation: an accelerator for attribute reduction in rough set theory, Artificial Intelligence 174 (2010) 597–618.

34. Y. H. Qian, J. Y. Liang, W. Pedrycz, C. Y. Dang, An efficient accelerator for attribute reduction from incomplete data in rough set framework, Pattern Recognition 44 (2011) 1658–1670.

35. Y. H. Qian, J. Y. Liang, W. Wei, Pessimistic rough decision, in: Second International Workshop on Rough

Sets Theory, Zhoushan, China 19–21 Octber, 2010, pp. 440–449.

36. Y. H. Qian, J. Y. Liang, W. Wei, Consistency-preserving attribute reduction in fuzzy rough set framework, International Journal of Machine Learning and Cybernetics 4 (2013) 287–299.

37. Y. H. Qian, J. Y. Liang, Y. Y. Yao, C. Y. Dang, MGRS: a multi–granulation rough set, Information Sciences 180 (2010) 949–970.

38. Y. H. Qian, H. Zhang, Y. L. Sang, J. Y. Liang, Multigranulation decision-theoretic rough sets, International Journal of Approximate Reasoning 55 (2014) 225–237.

39. Y. H. She, X. L. He, On the structure of the multigranulation rough set model, Knowledge–Based Systems 36 (2012) 81–92.

40. W. Z. Wu, Y. Leung, Theory and applications of granular labelled partitions in multi–scale decision tables, Information Sciences 181 (2011) 3878–3897.

41. W. H. Xu, W. X. Sun, X. Y. Zhang, W. X. Zhang, Multiple granulation rough set approach to ordered information systems, International Journal of General Systems 41 (2012) 475-501.

42. W. H. Xu, X. T. Zhang, Q. R. Wang, A generalized multi–granulation rough set approach, in: 7th International Conference on Intelligent Computing, Zhengzhou, China, 11–14 August, 2011, pp. 681–689.

43. W. H. Xu, Q. R. Wang, X. T. Zhang, Multi–granulation fuzzy rough sets in a fuzzy tolerance approximation space, International Journal of Fuzzy Systems 13 (2011) 246–259.

44. Y. Y. Yao, The superiority of three–way decisions in probabilistic rough set models, Information Sciences 181 (2011) 1080–1096.

45. Y. Y. Yao, Y. H. She, Rough set models in multigranulation spaces, Information Sciences 327 (2016) 40–56.

46. Y. Y. Yao, Y. Zhao, Attribute reduction in decision–theoretic rough set models, Information Sciences 178 (2008) 3356–3373.

47. Y. Y. Yao, N. Zhang, D. Q. Miao, F. F. Xu, Set–theoretic approaches to granular computing, Fundamenta Informaticae 115 (2012) 247–264.

48. X. B. Yang, Y. H. Qian, J. Y. Yang, Hierarchical structures on multigranulation spaces, Journal of Computer Science and Technology 27 (2012) 1169–1183.

49. X. B. Yang, Y. Qi, H. L. Yu, X. N. Song, J. Y. Yang, Updating multigranulation rough approximations with increasing of granular structures, Knowledge–Based Systems 64 (2014) 59–69.

50. X. B. Yang, J. Y. Yang, Incomplete information system and rough set theory: models and attribute reductions, Science Press & Springer, 2012.

51. X.B. Yang, X.N. Song, Z.H. Chen, J.Y. Yang, On multigranulation rough sets in incomplete information system, International Journal of Machine Learning and Cybernetics 3 (2012) 223-232.

52. X. B. Yang, X. N. Song, H. L. Dou, J. Y. Yang, Multi–granulation rough set: from crisp to fuzzy case, Annals of Fuzzy Mathematics and Informatics 1 (2011) 55–70.

53. X. B. Yang, M. Zhang, H. L. Dou, J. Y. Yang, Neighborhood systems-based rough sets in incomplete information system, Knowledge–Based Systems 24 (2011) 858–867.

54. X. B. Yang, J. Y. Yang, D. J. Yu, C. Wu, Dominance–based rough set approach and knowledge reductions in incomplete ordered information system, Information Sciences 178 (2008) 1219–1234.

55. X. B. Yang, Y. Q. Zhang, J. Y. Yang, Local and global measurements of MGRS rules, International Journal of Computational Intelligence Systems 5 (2012) 1010-1024.

56. W. Ziarko, Variable precision rough set model, Journal of Computer and System Science 46 (1993) 39–59.

57. Y. H. Zhai, D. Y. Li, K. S. Qu, Decision implications: a logical point of view, International Journal of Machine Learning and Cybernetics 5 (2014) 509–516.