Information Sciences 180 (2010) 949–970

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

MGRS: A multi-granulation rough set ☆

Yuhua Qian [a,c], Jiye Liang [a,*], Yiyu Yao [b], Chuangyin Dang [c]

[a] Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Taiyuan, 030006 Shanxi, China
[b] Department of Computer Science, University of Regina, Regina, Saskatchewan, Canada S4S 0A2
[c] Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:
Received 31 October 2008
Received in revised form 3 November 2009
Accepted 13 November 2009

Keywords:
Rough sets
Multi-granulation
Measures
Attribute reduction
Rule extraction

ABSTRACT

The original rough set model was developed by Pawlak, which is mainly concerned with the approximation of sets described by a single binary relation on the universe. In the view of granular computing, the classical rough set theory is established through a single granulation. This paper extends Pawlak's rough set model to a *multi-granulation rough set model* (MGRS), where the set approximations are defined by using multi equivalence relations on the universe. A number of important properties of MGRS are obtained. It is shown that some of the properties of Pawlak's rough set theory are special instances of those of MGRS.

Moreover, several important measures, such as *accuracy measure* α, *quality of approximation* γ and *precision of approximation* π, are presented, which are re-interpreted in terms of a classic measure based on sets, the Marczewski–Steinhaus metric and the inclusion degree measure. A concept of *approximation reduct* is introduced to describe the smallest attribute subset that preserves the lower approximation and upper approximation of all decision classes in MGRS as well. Finally, we discuss how to extract decision rules using MGRS. Unlike the decision rules ("AND" rules) from Pawlak's rough set model, the form of decision rules in MGRS is "OR". Several pivotal algorithms are also designed, which are helpful for applying this theory to practical issues. The multi-granulation rough set model provides an effective approach for problem solving in the context of multi granulations.

1. Introduction

Rough set theory, originated by Pawlak [16,17], has become a well-established mechanism for uncertainty management in a wide variety of applications related to artificial intelligence [2,5,6,11,20,23,24,31,32,41]. One of the strengths of rough set theory is that all its parameters are obtained from the given data. This can be seen in the following paragraph from [16]: "The numerical value of imprecision is not pre-assumed, as it is in probability theory or fuzzy sets – but is calculated on the basis of approximations which are the fundamental concepts used to express imprecision of knowledge". In other words, instead of using , *the rough set data analysis* (RSDA) utilizes solely the granularity structure of the given data, expressed as classes of suitable equivalence relations.

Knowledge representation in the rough set model is realized via *information systems* (IS) which are a tabular form of an OBJECT → ATTRIBUTE VALUE relationship, similar to relational databases. An information system is an ordered triplet $S = (U, AT, f)$, where U is a finite non-empty set of objects, AT is a finite non-empty set of attributes (predictor features), and $f_a : U \to V_a$ for any $a \in AT$ with V_a being the domain of an attribute a. For any $x \in U$, an information vector of x is given

☆ This is an extended version of the paper presented at the 5th IEEE International Conference on Cognitive Informatics, Beijing, China.

* Corresponding author. Tel./fax: +86 0351 7018176.
E-mail addresses: jinchengqyh@126.com (Y. Qian), ljy@sxu.edu.cn (J. Liang), yyao@cs.uregina.ca (Y. Yao), mecdang@cityu.edu.hk (C. Dang).

0020-0255/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2009.11.023

by $Inf(x) = \{(a, f_a(x)) | a \in AT\}$. In particular, a target information system is given by $S = (U, AT, f, D, g)$, where D is a finite non-empty set of decision attributes and $g_d : U \to V_d$ for any $d \in D$ with V_d being the domain of a decision attribute d.

If Q is a set of predictor features and d a decision attribute, then RSDA generates rules of the form

$$\bigwedge_{q \in Q} x^q = m_q \Rightarrow x^d = m_d^0 \vee x^d = m_d^1 \vee \cdots \vee x^d = m_d^k, \tag{1}$$

where $x^q = m_q$ denotes that the attribute value of object x under attribute q is equal to m_q, and $x^d = m_d^r$ $(r = 1, 2, \ldots, k)$ represents that the attribute value of object x under decision attribute d equals to m_d^r $(r = 1, 2, \ldots, k)$. Clearly, the form of decision rules is "AND" rules, i.e., conjunction operations in between the descriptions of condition attributes should be performed.

In the past 10 years, several extensions of the rough set model have been proposed in terms of various requirements, such as the variable precision rough set (VPRS) model (see [43]), the rough set model based on tolerance relation (see [7–9]), the Bayesian rough set model (see [33]), the fuzzy rough set model and the rough fuzzy set model (see [1,34,35]). In the view of granular computing (proposed by Zadeh [40]), a general concept described by a set is always characterized via the so-called upper and lower approximations under a single granulation, i.e., the concept is depicted by known knowledge induced from a single relation (such as equivalence relation, tolerance relation and reflexive relation) on the universe. However, this approach to describing a target concept is mainly based on the following assumption:

If P and Q are two sets from predictor features and $X \subseteq U$ is a target concept, then the rough set of X is derived from the quotient set $U/(P \cup Q)$. In fact, the quotient set is equivalent to the formula

$$\widehat{P \cup Q} = \{P_i \cap Q_j : P_i \in U/P,\ Q_j \in U/Q,\ P_i \cap P_j \neq \emptyset\}.$$

It implies the following two ideas:

(1) We can perform an intersection operation between any P_i and Q_j.
(2) The target concept is approximately described by using the quotient set $U/(P \cup Q)$.

In fact, the target concept is described by using a finer granulation (partitions) formed through combining two known granulations (partitions) induced from two-attribute subsets. Although it generates a much finer granulation and more knowledge, the combination/fining destroys the original granulation structure/partitions.

In general, the above assumption cannot always be satisfied or required in practice. In the following, several practical cases are given to illustrate its restrictions.

Case 1. In some data analysis issues, for the same object, there is a contradiction or inconsistent relationship between its values under one attribute set P and those under another attribute set Q. In other words, we cannot perform the intersection operations between their quotient sets and the target concept cannot be approximated by using $U/(P \cup Q)$.

Case 2. In the process of some decision making, the decision or the view of each of decision makers may be independent for the same project (or a sample, object and element) in the universe. In this situation, the intersection operations between any two quotient sets will be redundant for decision making.

Case 3. To extract decision rules from distributive information systems[1] and groups of intelligent agents[2] through using rough set approaches, knowledge representation and rough set approximations should be investigated. For the reduction of the time complexity of rule extractions, it is unnecessary for us to perform the intersection operations in between all the sites in the context of distributive information systems.

In these circumstances, we often need to describe concurrently a target concept through multi binary relations (e.g. equivalence relation, tolerance relation, reflexive relation and neighborhood relation) on the universe according to a user's requirements or targets of problem solving. In the literature [18], to more widely apply rough set theory to practical issues, a simple multi-granulation rough set model is proposed, based on multi equivalence relations. Furthermore, Qian et al. illuminated several basic views for establishing a multi-granulation rough set model in the context of incomplete information systems [19].

In the view of granular computing, an equivalence relation on the universe can be regarded as a granulation, and a partition on the universe can be regarded as a granulation space [13,14,20,22,37–39]. Hence, the classical rough set theory is based on a single granulation (only one equivalence relation). Note that any attribute set can induce a certain equivalence relation in an information system.

The main objective of this paper is to extend Pawlak's single-granulation rough set model to a *multi-granulation rough set model* (MGRS), where the set approximations are defined by using multi equivalence/tolerance relations on the universe. The

[1] In data mining and knowledge discovery, we are often faced with such a problem in which there are many information systems from various site and terminal unit. These information systems can be called distributive information systems [27–30].

[2] In artificial intelligence, an intelligent agent is an autonomous entity which observes and acts upon an environment and directs its activity towards achieving goals (i.e. it is rational). Intelligent agents are often described schematically as an abstract functional system similar to a computer program [25,26].

rest of the paper is organized as follows. Some preliminary concepts in Pawlak's rough set theory such as the lower approximation, upper approximation, accuracy measure and degree of dependency are briefly reviewed in Section 2.

In Section 3, for a complete information system, based on multi equivalence relations, an extension of Pawlak's rough set model is obtained, where a target concept is approximated by using the equivalence classes induced from multi-granulations. A detailed algorithm is designed to compute the lower approximation of a target concept in complete information systems. And a number of important properties of MGRS are investigated. It is shown that some of the properties of Pawlak's rough set theory are special instances of those of MGRS. Several important measures in MGRS, such as *accuracy measure* α, *quality of approximation* γ and *precision of approximation* π, are presented, which are re-interpreted in terms of a classic measure based on sets, the Marczewski–Steinhaus metric, and the inclusion degree measure. An importance measure of lower approximation and an importance measure of upper approximation are introduced for evaluating the significance of a condition attribute with respect to the decision attribute in complete target information systems.

In Section 4, a concept of approximation reduct is introduced to describe the smallest attribute subset that preserves the lower approximation and upper approximation of all decision classes in MGRS. The notion is based on the so-called upper approximation reduct and lower approximation reduct. An approximation core is employed to describe the intersection set of all approximation reducts. Based on this concept, we discuss how to extract decision rules from a complete target information system. Unlike decision rules ("AND" rules) from Pawlak's rough set model, the form of decision rules in MGRS is "OR". Furthermore, their computational methods are presented, which are helpful for applying this theory in practical issues.

Finally, Section 5 concludes the paper.

2. Pawlak's rough set theory

Throughout this paper, we assume that the universe U is a finite non-empty set.

Let us recall a few facts about partitions and equivalence relations. Suppose that $\widehat{P}$ is a partition of U induced from the attribute set P in an information system. If $x \in U$, we let $\widehat{P}(x)$ be the class of $\widehat{P}$ containing x, and $\theta_{\widehat{P}}$ the equivalence relation associated with $\widehat{P}$, i.e.,

$$x\theta_{\widehat{P}}y \Longleftrightarrow \widehat{P}(x) = \widehat{P}(y). \tag{2}$$

Rough set data analysis is based on the conviction that knowledge about the world is available only up to a certain granularity, and that granularity can be expressed mathematically by partitions and their associated equivalence relations [13].

If $X \subseteq U$ and $\widehat{P}$ is a partition of U, then the *lower approximation* (of X by $\widehat{P}$) is defined as

$$\underline{X}_{\widehat{P}} = \bigcup\{Y \in \widehat{P} : Y \subseteq X\} \tag{3}$$

and the *upper approximation* as

$$\overline{X}^{\widehat{P}} = \bigcup\{Y \in \widehat{P} : Y \cap X \neq \emptyset\}. \tag{4}$$

A pair of the form $\langle \underline{X}_{\widehat{P}}, \overline{X}^{\widehat{P}} \rangle$ is called a *rough set*. Obviously, $\overline{X}^{\widehat{P}} = U \setminus (\underline{-X}_{\widehat{P}})$, i.e., the upper approximation can be expressed by using the set complement and the lower approximation.

The *area of uncertainty or boundary region* is defined as

$$\partial_{\widehat{P}}(X) = \overline{X}_{\widehat{P}} \setminus \underline{X}_{\widehat{P}}. \tag{5}$$

To measure the imprecision of a rough set, Pawlak [16] recommended for $X \neq \emptyset$ the ratio

$$\alpha(\widehat{P}, X) = \frac{|\underline{X}_{\widehat{P}}|}{|\overline{X}^{\widehat{P}}|} = \frac{|\underline{X}_{\widehat{P}}|}{|U| - |\underline{(\sim X)}_{\widehat{P}}|}, \tag{6}$$

which is called the *accuracy measure* of X by $\widehat{P}$. It characterizes the degree of completeness of our knowledge about X, given the granularity of $\widehat{P}$. This measure depends not only on the approximation of X, but on the approximation of $\sim X$ as well.

Suppose that two views of the world are given by the partitions $\widehat{P}$ and $\widehat{Q}$ of the universe U, with associated equivalence relations $\theta_{\widehat{P}}$ and $\theta_{\widehat{Q}}$. If a class P_i of $\widehat{P}$ is a subset of a class Q_j of $\widehat{Q}$, then P_i is called deterministic with respect to $\widehat{Q}$, or just deterministic, if $\widehat{Q}$ is understood.

A frequently applied measure for this situation is the *quality of approximation* of $\widehat{Q}$ by $\widehat{P}$, also called the *degree of dependency*. It is defined as

$$\gamma(\widehat{P}, \widehat{Q}) = \frac{\sum\left\{\left|\underline{X}_{\widehat{P}}\right| : X \in \widehat{Q}\right\}}{|U|}, \tag{7}$$

which evaluates the deterministic part of the rough set description of $\widehat{Q}$ by counting those elements that can be re-classified to blocks of $\widehat{Q}$ with the knowledge given by $\widehat{P}$.

In the rough set data analysis, the measure of importance of condition attributes $B \subseteq AT$ with respect to decision attributes D is defined as $\gamma(AT,D) - \gamma(AT \setminus B, D)$. In particular, when $B = \{a\}, \gamma(AT,D) - \gamma(AT \setminus a, D)$ is the *measure of importance of attribute* $a \in AT$ with respect to D.

For an information system $S = (U, AT, f)$ and $B \subseteq AT$, if $\widehat{B} = \widehat{AT}$ and $\widehat{B \setminus \{b\}} \neq \widehat{AT}$ for any $b \in B$, then B is called a *reduct* of S. Furthermore, let $\{B_i : i \leqslant l\}$ be the set of all the reducts of S, then we call $\mathbf{B} = \bigcap_{i=1}^{l} B_i$ the *core* of this information system.

For a target information system $S = (U, AT, f, D, g), Pos_{AT}(D) = \bigcup_{X \in \widehat{D}} \underline{X}_{\widehat{AT}}$ is called a *positive region* of D with respect to AT. For $B \subseteq AT$, if $Pos_B(D) = Pos_{AT}(D)$ and $Pos_{B \setminus \{a\}}(D) \neq Pos_{AT}(D)$ for any $a \in B$, then B is called a *relative reduct* of S. Furthermore, let $\{B_i : i \leqslant l\}$ be the set of all the relative reducts of S, then we call $\mathbf{B} = \bigcap_{i=1}^{l} B_i$ the *relative core* of this target information system.

If every class of $\widehat{Q}$ is a union of classes of $\widehat{P}$, i.e. $\theta_{\widehat{P}} \subseteq \theta_{\widehat{Q}}$, then we say that $\widehat{P}$ is *finer than a partition* $\widehat{Q}$, and write $\widehat{P} \preceq \widehat{Q}$. In particular, the *identity partition* is the partition containing only singleton sets, the *universal partition* only has the universe set. The former is the finest partition on any non-empty set, and the latter is the roughest partition on the universe U.

3. MGRS in complete information systems

For an information system, any attribute domain V_a may contain special symbol "$*$" to indicate that the value of an attribute is unknown. Here, we assume that an object $x \in U$ possesses only one value for an attribute $a, a \in AT$. Thus, if the value of an attribute a is missing, then the real value must be from the set $V_a \setminus \{*\}$. Any domain value different from "$*$" will be called regular. A system in which values of all attributes for all objects from U are regular (known) is called complete, and it is called incomplete otherwise [7–10,12]. In particular, $S = (U, AT, f, D, g)$ is called a complete target information system if values of all attributes AT and D for all objects from U are regular (known), where AT is called the conditional attributes and D is called the decision attributes.

Let $S = (U, AT, f)$ be a complete information system. Each subset of attributes $P \subseteq AT$ determines a binary indiscernibility relation $IND(P)$ on U:

$$IND(P) = \{(x,y) \in U \times U : \forall a \in P, f_a(x) = f_a(y)\}. \tag{8}$$

The relation $IND(P), P \subseteq AT$, is an equivalence relation $\theta_{\widehat{P}}$ and constructs a partition $\widehat{P}$ of U.

Example 3.1. Here, we employ an example to illustrate some concepts of a complete target information system and computations involved in our proposed MGRS. Table 1 depicts a complete target information system containing some information about an emporium investment project. *Locus, Investment* and *Population density* are the conditional attributes of the system, whereas *Decision* is the decision attribute. (In the sequel, L, I, P and D will stand for *Locus, Investment*, *Population density* and *Decision*, respectively.) The attribute domains are as follows: $V_L = \{good, common, bad\}, V_I = \{high, low\}$, $V_P = \{big, small, medium\}$ and $V_D = \{Yes, No\}$.

3.1. Rough set approximation

In this subsection, we first discuss the approximation of a set by using two equivalence relations on the universe, i.e., the target concept is described by two granulation spaces.

Definition 3.1 [18]. Let $S = (U, AT, f)$ be a complete information system, $\widehat{P}, \widehat{Q}$ be two partitions on the universe U, and $X \subseteq U$. The *lower approximation* and the *upper approximation* of X in U are defined by the following

$$\underline{X}_{\widehat{P}+\widehat{Q}} = \{x : \widehat{P}(x) \subseteq X \text{ or } \widehat{Q}(x) \subseteq X\} \tag{9}$$

and

$$\overline{X}^{\widehat{P}+\widehat{Q}} = \sim \underline{(\sim X)}_{\widehat{P}+\widehat{Q}}, \tag{10}$$

Table 1
A complete target information system about emporium investment project.

Project	Locus	Investment	Population density	Decision
e_1	Common	High	Big	Yes
e_2	Bad	High	Big	Yes
e_3	Bad	Low	Small	No
e_4	Bad	Low	Small	No
e_5	Bad	Low	Small	No
e_6	Bad	High	Medium	Yes
e_7	Common	High	Medium	No
e_8	Good	High	Medium	Yes

The *area of uncertainty* or*boundary region* is defined as

$$\partial_{\widehat{P}+\widehat{Q}}(X) = \overline{X}_{\widehat{P}+\widehat{Q}} \setminus \underline{X}_{\widehat{P}+\widehat{Q}}.$$

Remark. From Eqs. (9) and (10), it can be seen that the lower approximation in MGRS is defined through using the equivalence classes induced by multi independent equivalence relations, whereas standard rough lower approximation is represented via those derived by only one equivalence relation. Each of the upper approximations in MGRS and SGRS can be characterized by the complementary set of the lower approximation of the complementary set of the target concept. In fact, if we perform an intersection operation between two equivalence partitions and use these new obtained classes to approximation a given target concept, then MGRS will generate standard rough set model. That is to say, the difference between standard rough lower and upper approximations and multi-granulation lower and upper approximations is precisely caused by two different approximation methods.

We will illuminate the rough set approximation based on multi-granulations and the difference between MGRS and Pawlak's rough set theory through the following example and Proposition 3.1.

Example 3.2 (*Continued from Example 3.1*). Let $X = \{e_1, e_2, e_6, e_8\}$. Three partitions are induced from Table 1 as follows:

$$\widehat{L} = \{\{e_1, e_7\}, \{e_2, e_3, e_4, e_5, e_6\}, \{e_8\}\},$$

$$\widehat{P}\{\{e_1, e_2\}, \{e_3, e_4, e_5\}, \{, e_6, e_7, e_8\}\}$$

and

$$\widehat{L \cup P} = \{\{e_1\}, \{e_2\}, \{e_3, e_4, e_5\}, \{e_6\}, \{e_7\}, \{e_8\}\}.$$

By computing, we have that

$$\underline{X}_{\widehat{L}+\widehat{P}} = \{x : \ \widehat{L}(x) \subseteq X \ or \ \widehat{P}(x) \subseteq X\} = \{e_8\} \cup \{e_1, e_2\} = \{e_1, e_2, e_8\},$$

$$\overline{X}^{\widehat{L}+\widehat{P}} = \sim \underline{(\sim X)}_{\widehat{L}+\widehat{P}} = \sim \{\emptyset \cup \{e_3, e_4, e_5\}\} = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8\} \cap \{e_1, e_2, e_6, e_7, e_8\} = \{e_1, e_2, e_6, e_7, e_8\}.$$

But, the lower approximation and the upper approximation of X based on Pawlak's rough set theory are as follows:

$$\underline{X}_{\widehat{L \cup P}} = \{Y \in \widehat{L \cup P} : Y \subseteq X\} = \{e_1, e_2, e_6, e_8\},$$

$$\overline{X}^{\widehat{L \cup P}} = \{Y \in \widehat{L \cup P} : Y \cap X \neq \emptyset\} = \{e_1, e_2, e_6, e_8\}.$$

Obviously,

$$\underline{X}_{\widehat{L}+\widehat{P}} = \{e_1, e_2, e_8\} \subseteq \{e_1, e_2, e_6, e_8\} = \underline{X}_{\widehat{L \cup P}},$$

$$\overline{X}^{\widehat{L}+\widehat{P}} = \{e_1, e_2, e_6, e_7, e_8\} \supseteq \{e_1, e_2, e_6, e_8\} = \overline{X}^{\widehat{L \cup P}}.$$

As a result of this example, we have the following proposition.

Proposition 3.1. *Let $S = (U, AT, f)$ be a complete information system, $\widehat{P}, \widehat{Q}$ be two partitions induced from the attributes P and Q, respectively, and $X \subseteq U$. Then, $\underline{X}_{\widehat{P}+\widehat{Q}} \subseteq \underline{X}_{\widehat{P \cup Q}}$ and $\overline{X}^{\widehat{P}+\widehat{Q}} \supseteq \overline{X}^{\widehat{P \cup Q}}$.*

Proof

(1) For any $x \in \underline{X}_{\widehat{P}+\widehat{Q}}$, from Definition 3.1, it follows that $x \in \widehat{P}(x)$ and $x \in \widehat{Q}(x)$. Hence, $x \in \widehat{P}(x) \cap \widehat{Q}(x)$. But $\widehat{P}(x) \cap \widehat{Q}(x) \in \widehat{P \cup Q}$ for any $x \in U$, and $\underline{X}_{\widehat{P \cup Q}} = \bigcup\{Y \in \widehat{P \cup Q} : Y \subseteq X\}$ from the definition. Therefore, $x \in \underline{X}_{\widehat{P \cup Q}}$, i.e., $\underline{X}_{\widehat{P}+\widehat{Q}} \subseteq \underline{X}_{\widehat{P \cup Q}}$.

(2) From Pawlak's rough set theory, we know $\overline{X}^{\widehat{P \cup Q}} = \sim \underline{(\sim X)}_{\widehat{P \cap Q}}$. Applying the result of (1), we have that $\overline{X}^{\widehat{P \cup Q}} = \sim \underline{(\sim X)}_{\widehat{P \cup Q}} \subseteq \sim \underline{(\sim X)}_{\widehat{P}+\widehat{Q}} = \overline{X}^{\widehat{P}+\widehat{Q}}$, i.e., $\overline{X}^{\widehat{P}+\widehat{Q}} \supseteq \overline{X}^{\widehat{P \cup Q}}$.

This completes the proof. □

Corollary 3.1. $\partial_{\widehat{P}}(X) \subseteq \partial_{\widehat{P}+\widehat{Q}}(X)$ *and* $\partial_{\widehat{Q}}(X) \subseteq \partial_{\widehat{P}+\widehat{Q}}(X)$.

The following Fig. 1 shows that the difference between Pawlak's rough set model and the multi-granulation rough set model.

In Fig. 1, the bias region is the lower approximation of a set X obtained by a single granulation $P \cup Q$, which is expressed by the equivalence classes in the quotient set $U/(P \cup Q)$, and the shaded region is the lower approximation of X induced by two granulations $P + Q$, which is characterized by the equivalence classes in the quotient set U/P and the quotient set U/Q together.

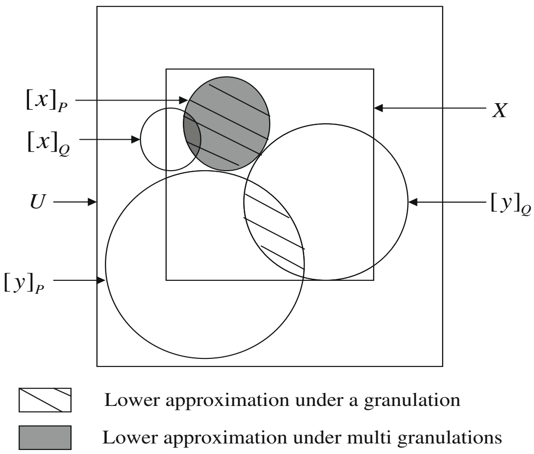

Fig. 1. Difference between Pawlak's rough set model and MGRS.

Just from the definition of approximations, one can get the following properties of the lower and upper approximations.

Proposition 3.2. *Let $S = (U, AT, f)$ be a complete information system, $\widehat{P}, \widehat{Q}$ be two partitions induced by the attributes P and Q, respectively, and $X \subseteq U$. The following properties hold*

(1) $\underline{X}_{\widehat{P}+\widehat{Q}} \subseteq X \subseteq \overline{X}^{\widehat{P}+\widehat{Q}}$;
(2) $\underline{\emptyset}_{\widehat{P}+\widehat{Q}} = \overline{\emptyset}^{\widehat{P}+\widehat{Q}} = \emptyset$ and $\underline{U}_{\widehat{P}+\widehat{Q}} = \overline{U}^{\widehat{P}+\widehat{Q}} = U$;
(3) $\underline{(\sim X)}_{\widehat{P}+\widehat{Q}} = \sim \overline{X}^{\widehat{P}+\widehat{Q}}$ and $\overline{(\sim X)}^{\widehat{P}+\widehat{Q}} = \sim \underline{X}_{\widehat{P}+\widehat{Q}}$;
(4) $\underline{\underline{X}_{\widehat{P}+\widehat{Q}}}_{\widehat{P}+\widehat{Q}} = \overline{\underline{X}_{\widehat{P}+\widehat{Q}}}^{\widehat{P}+\widehat{Q}} = \underline{X}_{\widehat{P}+\widehat{Q}}$;
(5) $\overline{\overline{X}^{\widehat{P}+\widehat{Q}}}^{\widehat{P}+\widehat{Q}} = \underline{\overline{X}^{\widehat{P}+\widehat{Q}}}_{\widehat{P}+\widehat{Q}} = \overline{X}^{\widehat{P}+\widehat{Q}}$;
(6) $\underline{X}_{\widehat{P}+\widehat{Q}} = \underline{X}_{\widehat{P}} \cup \underline{X}_{\widehat{Q}}$;
(7) $\overline{X}^{\widehat{P}+\widehat{Q}} = \overline{X}^{\widehat{P}} \cap \overline{X}^{\widehat{Q}}$;
(8) $\underline{X}_{\widehat{P}+\widehat{Q}} = \underline{X}_{\widehat{Q}+\widehat{P}}$ and $\overline{X}^{\widehat{P}+\widehat{Q}} = \overline{X}^{\widehat{Q}+\widehat{P}}$.

Proof. If $P = Q (P, Q \subseteq AT)$, then (9) degenerates into $\underline{X}_{\widehat{P}} = \{Y \in \widehat{P} : Y \subseteq X\}$, and (10) degenerates into $\overline{X}^{\widehat{P}} = \{Y \in \widehat{P} : Y \cap X \neq \emptyset\}$. Obviously, they are the same as the lower approximation and the upper approximation of Pawlak's rough set theory [16], respectively. Hence, the terms (1)–(8) hold.

If $P \neq Q (P, Q \subseteq AT)$, we prove them as follows:

(1a) Let $x, y \in \underline{X}_{\widehat{P}+\widehat{Q}} (x, y \in U)$. Then, $\widehat{P}(x) \subseteq X$ and $\widehat{Q}(x) \subseteq X$. But $x \in \widehat{P}(x)$ and $y \in \widehat{Q}(y)$. Hence, $x, y \in X$ and $\underline{X}_{\widehat{P}+\widehat{Q}} \subseteq X$.
(1b) Let $x, y \in X$. Then, $x \in \widehat{P}(x) \cap X$ and $y \in \widehat{Q}(y) \cap X$, i.e., $\widehat{P}(x) \cap X \neq \emptyset$ and $\widehat{Q}(y) \cap X \neq \emptyset$. Hence, $x, y \in \overline{X}^{\widehat{P}+\widehat{Q}}$ and $X \subseteq \overline{X}^{\widehat{P}+\widehat{Q}}$.
(2a) From (1), we know that $\underline{\emptyset}_{\widehat{P}+\widehat{Q}} \subseteq \emptyset$ and $\emptyset \subseteq \underline{\emptyset}_{\widehat{P}+\widehat{Q}}$ (because the empty set is included in every set). Therefore, $\underline{\emptyset}_{\widehat{P}+\widehat{Q}} = \emptyset$.
(2b) Suppose $\overline{\emptyset}^{\widehat{P}+\widehat{Q}} \neq \emptyset$. Then, there exists x such that $x \in \overline{\emptyset}^{\widehat{P}+\widehat{Q}} \neq \emptyset$. Hence, $\widehat{P}(x) \cap \emptyset \neq \emptyset$. But $\widehat{P}(x) \cap \emptyset = \emptyset$. It contradicts the assumption. So, $\overline{\emptyset}^{\widehat{P}+\widehat{Q}} = \emptyset$.
(2c) From (1), we know that $\underline{U}_{\widehat{P}+\widehat{Q}} \subseteq U$. And if $x \in U$, then $\widehat{P}(x) \subseteq U$ and $\widehat{Q}(x) \subseteq U$. Hence, $x \in \underline{U}_{\widehat{P}+\widehat{Q}}$ and $U \subseteq \underline{U}_{\widehat{P}+\widehat{Q}}$. Thus, $\underline{U}_{\widehat{P}+\widehat{Q}} = U$.
(2d) From (1), one can get that $U \subseteq \overline{U}^{\widehat{P}+\widehat{Q}}$. And $\overline{U}^{\widehat{P}+\widehat{Q}} \subseteq U$ hold clearly. Thus, $\overline{U}^{\widehat{P}+\widehat{Q}} = U$.
(3) From (10), $\underline{(\sim X)}_{\widehat{P}+\widehat{Q}} = \sim \overline{X}^{\widehat{P}+\widehat{Q}}$ is obvious. Let $X = \sim X$. Then, $\overline{(\sim X)}^{\widehat{P}+\widehat{Q}} = \sim \underline{(\sim (\sim X))}_{\widehat{P}+\widehat{Q}} = \sim \underline{X}_{\widehat{P}+\widehat{Q}}$.
(4a) From (1), we know that $\underline{\underline{X}_{\widehat{P}+\widehat{Q}}}_{\widehat{P}+\widehat{Q}} \subseteq \underline{X}_{\widehat{P}+\widehat{Q}}$. If $x \in \underline{X}_{\widehat{P}+\widehat{Q}}$, then $\widehat{P}(x), \widehat{Q}(x) \subseteq X$. Hence, $\underline{\widehat{P}(x)}_{\widehat{P}+\widehat{Q}} \subseteq \underline{X}_{\widehat{P}+\widehat{Q}}$ and $\underline{\widehat{Q}(x)}_{\widehat{P}+\widehat{Q}} \subseteq \underline{X}_{\widehat{P}+\widehat{Q}}$. But $\underline{\widehat{P}(x)}_{\widehat{P}+\widehat{Q}} = \widehat{P}(x)$ and $\underline{\widehat{Q}(x)}_{\widehat{P}+\widehat{Q}} = \widehat{Q}(x)$. Thus, $\widehat{P}(x), \widehat{Q}(x) \subseteq \underline{X}_{\widehat{P}+\widehat{Q}}$ and $x \in \underline{\underline{X}_{\widehat{P}+\widehat{Q}}}_{\widehat{P}+\widehat{Q}}$. Hence, we have that $\underline{X}_{\widehat{P}+\widehat{Q}} = \underline{\underline{X}_{\widehat{P}+\widehat{Q}}}_{\widehat{P}+\widehat{Q}}$.

(4b) From (1), $\underline{X}_{\widehat{P}+\widehat{Q}} \subseteq \overline{\underline{X}_{\widehat{P}+\widehat{Q}}}^{\widehat{P}+\widehat{Q}}$ hold. If $x \in \overline{\underline{X}_{\widehat{P}+\widehat{Q}}}^{\widehat{P}+\widehat{Q}}$, then $\widehat{P}(x) \cap \underline{X}_{\widehat{P}+\widehat{Q}} \neq \emptyset$ and $\widehat{Q}(x) \cap \underline{X}_{\widehat{P}+\widehat{Q}} \neq \emptyset$, i.e., there exist $y \in \widehat{P}(x)$ and $z \in \widehat{Q}(x)$ such that $y \in \underline{X}_{\widehat{P}+\widehat{Q}}$ and $z \in \underline{X}_{\widehat{P}+\widehat{Q}}$. Hence, $\widehat{P}(y) \subseteq X, \widehat{Q}(z) \subseteq X$. But $\widehat{P}(y) = \widehat{P}(x)$ and $\widehat{Q}(z) = \widehat{Q}(x)$. Thus, $\widehat{P}(x) \subseteq X, \widehat{Q}(x) \subseteq X$ and $x \in \underline{X}_{\widehat{P}+\widehat{Q}}$. Hence, we have that $\underline{X}_{\widehat{P}+\widehat{Q}} \supseteq \overline{\underline{X}_{\widehat{P}+\widehat{Q}}}^{\widehat{P}+\widehat{Q}}$. Therefore, $\underline{X}_{\widehat{P}+\widehat{Q}} = \overline{\underline{X}_{\widehat{P}+\widehat{Q}}}^{\widehat{P}+\widehat{Q}}$.

(5a) From (1), $\overline{X}^{\widehat{P}+\widehat{Q}} \subseteq \overline{\overline{X}^{\widehat{P}+\widehat{Q}}}^{\widehat{P}+\widehat{Q}}$ hold. If $x \in \overline{\overline{X}^{\widehat{P}+\widehat{Q}}}^{\widehat{P}+\widehat{Q}}$, then $\widehat{P}(x) \cap \overline{X}^{\widehat{P}+\widehat{Q}} \neq \emptyset$ and $\widehat{Q}(x) \cap \overline{X}^{\widehat{P}+\widehat{Q}} \neq \emptyset$. For some $y \in \widehat{P}(x), y \in \overline{X}^{\widehat{P}+\widehat{Q}}$, and some $z \in \widehat{Q}(x), z \in \overline{X}^{\widehat{P}+\widehat{Q}}$, we have that $\widehat{P}(y) \cap X \neq \emptyset$ and $\widehat{Q}(z) \cap X \neq \emptyset$. But $\widehat{P}(x) = \widehat{P}(y)$ and $\widehat{Q}(x) = \widehat{Q}(z)$. Thus, $\widehat{P}(x) \cap X \neq \emptyset, \widehat{Q}(x) \cap X \neq \emptyset$. That is to say, $x \in \overline{X}^{\widehat{P}+\widehat{Q}}$ hold, which yields $\overline{X}^{\widehat{P}+\widehat{Q}} \supseteq \overline{\overline{X}^{\widehat{P}+\widehat{Q}}}^{\widehat{P}+\widehat{Q}}$. Therefore, we have that $\overline{X}^{\widehat{P}+\widehat{Q}} = \overline{\overline{X}^{\widehat{P}+\widehat{Q}}}^{\widehat{P}+\widehat{Q}}$.

(5b) From (1), we know $\overline{X}^{\widehat{P}+\widehat{Q}} \supseteq \underline{\overline{X}^{\widehat{P}+\widehat{Q}}}_{\widehat{P}+\widehat{Q}}$. If $x, y \in \overline{X}^{\widehat{P}+\widehat{Q}}$, then $\widehat{P}(x) \cap X \neq \emptyset, \widehat{Q}(y) \cap X \neq \emptyset$. Thus, $\widehat{P}(x) \subseteq \overline{X}^{\widehat{P}+\widehat{Q}}$ and $\widehat{Q}(y) \subseteq \overline{X}^{\widehat{P}+\widehat{Q}}$ (because if $x' \in \widehat{P}(x)$, then $\widehat{P}(x') \cap X = \widehat{P}(x) \cap X \neq \emptyset$, i.e., $x' \in \overline{X}^{\widehat{P}+\widehat{Q}}$). And $x \in \underline{\overline{X}^{\widehat{P}+\widehat{Q}}}_{\widehat{P}+\widehat{Q}}$, we have that $\underline{\overline{X}^{\widehat{P}+\widehat{Q}}}_{\widehat{P}+\widehat{Q}} \supseteq \overline{X}^{\widehat{P}+\widehat{Q}}$. Therefore, we get that $\underline{\overline{X}^{\widehat{P}+\widehat{Q}}}_{\widehat{P}+\widehat{Q}} = \overline{X}^{\widehat{P}+\widehat{Q}}$.

(6) From (9), we easily know that for $\forall x \in U$, if $\widehat{P}(x) \subseteq X$ then $x \in \underline{X}_{\widehat{P}+\widehat{Q}}$, and if $\widehat{Q}(x) \subseteq X$ then $x \in \underline{X}_{\widehat{P}+\widehat{Q}}$. That is, $\underline{X}_{\widehat{P}} \subseteq \underline{X}_{\widehat{P}+\widehat{Q}}$ and $\underline{X}_{\widehat{Q}} \subseteq \underline{X}_{\widehat{P}+\widehat{Q}}$. And, if there exists $y \in X$ with $y \in \underline{X}_{\widehat{P}+\widehat{Q}} - \bigcup_{x \in U} \widehat{P}(x) - \bigcup_{x \in U} \widehat{Q}(x) = \emptyset$, then $\widehat{P}(y) = \emptyset$ and $\widehat{Q}(y) = \emptyset$. Therefore, we have that $\underline{X}_{\widehat{P}+\widehat{Q}} = \underline{X}_{\widehat{P}} \cup \underline{X}_{\widehat{Q}}$.

(7) From (10) and (6), one can obtain that

$$\overline{X}^{\widehat{P}+\widehat{Q}} = \sim \underline{(\sim X)}_{\widehat{P}+\widehat{Q}} = \sim (\underline{(\sim X)}_{\widehat{P}} \cup \underline{(\sim X)}_{\widehat{Q}}) = \sim (\sim \overline{X}^{\widehat{P}} \cup \sim \overline{X}^{\widehat{Q}}) = \overline{X}^{\widehat{P}} \cap \overline{X}^{\widehat{Q}}.$$

(8) They are straightforward from Definition 3.1.

This completes this proof. □

In order to discover the relationship between the approximations of a single set and the approximations of two sets described by using two equivalence relations (granulations) on the universe, the following properties are given.

Proposition 3.3. *Let $S = (U, AT, f)$ be a complete information system, $\widehat{P}, \widehat{Q}$ be two partitions induced by the attributes P and Q, respectively, and $X, Y \subseteq U$. The following properties hold*

(1) $\underline{(X \cap Y)}_{\widehat{P}+\widehat{Q}} = (\underline{X}_{\widehat{P}} \cap \underline{Y}_{\widehat{P}}) \cup (\underline{X}_{\widehat{Q}} \cap \underline{Y}_{\widehat{Q}})$;
(2) $\overline{(X \cup Y)}^{\widehat{P}+\widehat{Q}} = (\overline{X}^{\widehat{P}} \cup \overline{Y}^{\widehat{P}}) \cap (\overline{X}^{\widehat{Q}} \cup \overline{Y}^{\widehat{Q}})$;
(3) $\underline{(X \cap Y)}_{\widehat{P}+\widehat{Q}} \subseteq \underline{X}_{\widehat{P}+\widehat{Q}} \cap \underline{Y}_{\widehat{P}+\widehat{Q}}$;
(4) $\overline{(X \cup Y)}^{\widehat{P}+\widehat{Q}} \supseteq \overline{X}^{\widehat{P}+\widehat{Q}} \cup \overline{Y}^{\widehat{P}+\widehat{Q}}$;
(5) $X \subseteq Y \Rightarrow \underline{X}_{\widehat{P}+\widehat{Q}} \subseteq \underline{Y}_{\widehat{P}+\widehat{Q}}$;
(6) $X \subseteq Y \Rightarrow \overline{X}^{\widehat{P}+\widehat{Q}} \subseteq \overline{Y}^{\widehat{P}+\widehat{Q}}$;
(7) $\underline{(X \cup Y)}_{\widehat{P}+\widehat{Q}} \supseteq \underline{X}_{\widehat{P}+\widehat{Q}} \cup \underline{Y}_{\widehat{P}+\widehat{Q}}$;
(8) $\overline{(X \cap Y)}^{\widehat{P}+\widehat{Q}} \subseteq \overline{X}^{\widehat{P}+\widehat{Q}} \cap \overline{Y}^{\widehat{P}+\widehat{Q}}$.

Proof. If $P = Q$ $(P, Q \subseteq AT)$, then (9) degenerates into $\underline{X}_{\widehat{P}} = \{Y \in \widehat{P} : Y \subseteq X\}$ and (10) degenerates into $\overline{X}^{\widehat{P}} = \{Y \in \widehat{P} : Y \cap X \neq \emptyset\}$. Obviously, they are the same as the lower approximation and the upper approximation of Pawlak's rough set theory [16], respectively. Hence, (1)–(8) hold.

If $P \neq Q (P, Q \subseteq AT)$, we prove them as follows:

(1) $\underline{(X \cap Y)}_{\widehat{P}+\widehat{Q}} = \underline{(X \cap Y)}_{\widehat{P}} \cup \underline{(X \cap Y)}_{\widehat{Q}} = (\underline{X}_{\widehat{P}} \cap \underline{Y}_{\widehat{P}}) \cup (\underline{X}_{\widehat{Q}} \cap \underline{Y}_{\widehat{Q}})$.
(2) $\overline{(X \cup Y)}^{\widehat{P}+\widehat{Q}} = \overline{(X \cup Y)}^{\widehat{P}} \cap \overline{(X \cup Y)}^{\widehat{Q}} = (\overline{X}^{\widehat{P}} \cup \overline{Y}^{\widehat{P}}) \cap (\overline{X}^{\widehat{Q}} \cup \overline{Y}^{\widehat{Q}})$.
(3) It follows from (1) that

$$\begin{aligned}
\underline{(X \cap Y)}_{\widehat{P}+\widehat{Q}} &= (\underline{X}_{\widehat{P}} \cap \underline{Y}_{\widehat{P}}) \cup (\underline{X}_{\widehat{Q}} \cap \underline{Y}_{\widehat{Q}}) = ((\underline{X}_{\widehat{P}} \cap \underline{Y}_{\widehat{P}}) \cup \underline{X}_{\widehat{Q}}) \cap ((\underline{X}_{\widehat{P}} \cap \underline{Y}_{\widehat{P}}) \cup \underline{Y}_{\widehat{Q}}) \\
&= ((\underline{X}_{\widehat{P}} \cup \underline{X}_{\widehat{Q}}) \cap (\underline{Y}_{\widehat{P}} \cup \underline{X}_{\widehat{Q}}) \cap ((\underline{X}_{\widehat{P}} \cup \underline{Y}_{\widehat{Q}}) \cap (\underline{Y}_{\widehat{P}} \cup \underline{Y}_{\widehat{Q}}))) \\
&= \underline{X}_{\widehat{P}+\widehat{Q}} \cap \underline{Y}_{\widehat{P}+\widehat{Q}} \cap ((\underline{Y}_{\widehat{P}} \cup \underline{X}_{\widehat{Q}}) \cap (\underline{X}_{\widehat{P}} \cup \underline{Y}_{\widehat{Q}})) \subseteq \underline{X}_{\widehat{P}+\widehat{Q}} \cap \underline{Y}_{\widehat{P}+\widehat{Q}}.
\end{aligned}$$

(4) It follows from (2) that

$$\begin{aligned}\overline{(X\cup Y)}^{\widehat{P}+\widehat{Q}} &= (\overline{X}^{\widehat{P}}\cup\overline{Y}^{\widehat{P}})\cap(\overline{X}^{\widehat{Q}}\cup\overline{Y}^{\widehat{Q}}) = ((\overline{X}^{\widehat{P}}\cup\overline{Y}^{\widehat{P}})\cap\overline{X}^{\widehat{Q}})\cup((\overline{X}^{\widehat{P}}\cup\overline{Y}^{\widehat{P}})\cap\overline{Y}^{\widehat{Q}})\\ &= ((\overline{X}^{\widehat{P}}\cap\overline{X}^{\widehat{Q}})\cup(\overline{Y}^{\widehat{P}}\cap\overline{X}^{\widehat{Q}})\cup((\overline{X}^{\widehat{P}}\cap\overline{Y}^{\widehat{Q}})\cup(\overline{Y}^{\widehat{P}}\cap\overline{Y}^{\widehat{Q}})) = \overline{X}^{\widehat{P}+\widehat{Q}}\cup\overline{Y}^{\widehat{P}+\widehat{Q}}\cup((\overline{Y}^{\widehat{P}}\cap\overline{X}^{\widehat{Q}})\cup(\overline{X}^{\widehat{P}}\cap\overline{Y}^{\widehat{Q}}))\\ &\supseteq \overline{X}^{\widehat{P}+\widehat{Q}}\cup\overline{Y}^{\widehat{P}+\widehat{Q}}.\end{aligned}$$

(5) If $X\subseteq Y$, then $X\cap Y=X$. It follows from (3) that

$$\underline{(X\cap Y)}_{\widehat{P}+\widehat{Q}} = \underline{X}_{\widehat{P}+\widehat{Q}} \subseteq \underline{X}_{\widehat{P}+\widehat{Q}}\cap\underline{Y}_{\widehat{P}+\widehat{Q}} \Rightarrow \underline{X}_{\widehat{P}+\widehat{Q}} = \underline{X}_{\widehat{P}+\widehat{Q}}\cap\underline{Y}_{\widehat{P}+\widehat{Q}} \Rightarrow \underline{X}_{\widehat{P}+\widehat{Q}}\subseteq\underline{Y}_{\widehat{P}+\widehat{Q}}.$$

(6) If $X\subseteq Y$, then $X\cup Y=Y$. It follows from (4) that

$$\overline{(X\cup Y)}^{\widehat{P}+\widehat{Q}} = \overline{Y}^{\widehat{P}+\widehat{Q}} \supseteq \overline{X}^{\widehat{P}+\widehat{Q}}\cup\overline{Y}^{\widehat{P}+\widehat{Q}} \Rightarrow \overline{Y}^{\widehat{P}+\widehat{Q}} = \overline{X}^{\widehat{P}+\widehat{Q}}\cup\overline{Y}^{\widehat{P}+\widehat{Q}} \Rightarrow \overline{X}^{\widehat{P}+\widehat{Q}}\subseteq\overline{Y}^{\widehat{P}+\widehat{Q}}.$$

(7) It is clear that $X\subseteq X\cup Y$ and $Y\subseteq X\cup Y$. It follows that $\underline{X}_{\widehat{P}+\widehat{Q}}\subseteq\underline{X\cup Y}_{\widehat{P}+\widehat{Q}}$ and $\underline{Y}_{\widehat{P}+\widehat{Q}}\subseteq\underline{X\cup Y}_{\widehat{P}+\widehat{Q}}$. Hence, $\underline{X}_{\widehat{P}+\widehat{Q}}\cup\underline{Y}_{\widehat{P}+\widehat{Q}} \subseteq\underline{X\cup Y}_{\widehat{P}+\widehat{Q}}$.

(8) It is clear that $X\cap Y\subseteq X$ and $X\cap Y\subseteq Y$. It follows that $\overline{X}^{\widehat{P}+\widehat{Q}}\supseteq\overline{X\cap Y}^{\widehat{P}+\widehat{Q}}$ and $\overline{Y}^{\widehat{P}+\widehat{Q}}\supseteq\overline{X\cap Y}^{\widehat{P}+\widehat{Q}}$. Hence, $\overline{(X\cap Y)}^{\widehat{P}+\widehat{Q}}\subseteq\overline{X}^{\widehat{P}+\widehat{Q}} \cap\overline{Y}^{\widehat{P}+\widehat{Q}}$.

This completes this proof. □

Based on the above conclusions, we here extend Pawlak's rough set model to a multi-granulation rough set model (MGRS), where the set approximations are defined through using multi equivalence relations on the universe.

Definition 3.2. Let $S=(U,AT,f)$ be a complete information system, $X\subseteq U$ and $\widehat{P_1},\widehat{P_2},\ldots,\widehat{P}_m$ be m partitions induced by the attributes $P_1,P_2,\ldots,P_m$, respectively. The lower approximation and the upper approximation of X related to $\widehat{P_1},\widehat{P_2},\ldots,\widehat{P}_m$ are defined by the following:

$$\underline{X}_{\sum_{i=1}^m\widehat{P}_i} = \{x: \bigvee\widehat{P}_i(x)\subseteq X, i\leqslant m\} \tag{11}$$

and

$$\overline{X}^{\sum_{i=1}^m\widehat{P}_i} = \sim\underline{(\sim X)}_{\sum_{i=1}^m\widehat{P}_i}. \tag{12}$$

Similarly, the *area of uncertainty* or*boundary region* in MGRS can be extended as

$$\partial_{\sum_{i=1}^m\widehat{P}_i}(X) = \overline{X}_{\sum_{i=1}^m\widehat{P}_i}\setminus\underline{X}_{\sum_{i=1}^m\widehat{P}_i}.$$

From the definition we obtain the following interpretations:

- The lower approximation of a set X with respect to $\sum_{i=1}^m\widehat{P}_i$ is the set of all elements, which can certainly be classified as X using $\sum_{i=1}^m\widehat{P}_i$ (are certainly X in view of $\sum_{i=1}^m\widehat{P}_i$).
- The upper approximation of a set X with respect to $\sum_{i=1}^m\widehat{P}_i$ is the set of all elements, which can possibly be classified as X using $\sum_{i=1}^m\widehat{P}_i$ (are possibly X in view of $\sum_{i=1}^m\widehat{P}_i$).
- The boundary region of a set X with respect to $\sum_{i=1}^m\widehat{P}_i$ is the set of all elements, which can be classified neither as X nor as not-X using $\sum_{i=1}^m\widehat{P}_i$.

To apply this approach to practical issues, we here present an algorithm for computing the lower approximation of a set X in the rough set model based on multi equivalence relations.

Algorithm 1. Let $S=(U,AT,f)$ be a complete information system, $X\subseteq U$ and $P\subseteq 2^{AT}$, where $P=\{P_1,P_2,\ldots,P_m\}$.
This algorithm gives the lower approximation of X by P: $\underline{X}_{\sum_{i=1}^m\widehat{P}_i}=\{x|\widehat{P}_i(x)\subseteq X, i\leqslant m\}$.
We use the following pointers:
$i=1,2,\ldots,m$ points to $\widehat{P}_i$,
$j=1,2,\ldots,|\widehat{P}_i|$ points to $Y_i^j\in\widehat{P}_i$, and
L records the computation of the lower approximation.

For every i and every j, we check whether or not $Y_i^j\cap X=Y_i^j$. If $Y_i^j\cap X=Y_i^j$, then we put Y_i^j into the lower approximation of X: $L\leftarrow L\cup Y_i^j$.

(1) Compute m partitions: $\widehat{P_1},\widehat{P_2},\ldots,\widehat{P}_m$ (see Algorithm E in [4]);
(2) Set $i\leftarrow 1, j\leftarrow 1, L=\emptyset$;

(3) For $i = 1$ to m Do
  For $j = 1$ to $|U|$ Do
    If $Y_i^j \cap X = Y_i^j$, then
      let $L \leftarrow L \cup \{u_j\}$,
    Endif
  Endfor
  Set $j \leftarrow 1$,
Endfor
(4) The computation of the lower approximation X by P is completed. Output the set L.

We know that the time complexity of computing m partitions is $\mathbf{O}(m|U|^2)$ (see Algorithm E in [4]). The time complexity of (I3) is also $\mathbf{O}(m|U|^2)$ since there are $\sum_{i=1}^m|\widehat{P}_i|$ ($\leqslant |U| \times |U|$) intersections $Y_i^j \cap X$ to be calculated. Hence, the time complexity of Algorithm 1 is $\mathbf{O}$ $(m|U|^2)$.

This algorithm can be run in parallel mode to compute concurrently all corresponding classifications and intersections from many attributes. Its time complexity will be $\mathbf{O}$ $(|U|^2)$. Similar to this idea, the algorithm for computing the upper approximation of a set also can be designed correspondingly.

Just from the definitions of above approximations, one can get the following properties of the lower and upper approximations.

Proposition 3.4. *Let $S = (U, AT, f)$ be a complete information system, $X \subseteq U$ and $\widehat{P_1}, \widehat{P_2}, \ldots, \widehat{P}_m$ be m partitions induced by the attributes $P_1, P_2, \ldots, P_m$, respectively. Then, the following properties hold*

(1) $\underline{X}_{\sum_{i=1}^m \widehat{P_i}} = \bigcup_{i=1}^m \underline{X}_{\widehat{P_i}}$;
(2) $\overline{X}^{\sum_{i=1}^m \widehat{P_i}} = \bigcap_{i=1}^m \overline{X}^{\widehat{P_i}}$;
(3) $\underline{(\sim X)}_{\sum_{i=1}^m \widehat{P_i}} = \sim \overline{X}^{\sum_{i=1}^m \widehat{P_i}}$;
(4) $\overline{(\sim X)}^{\sum_{i=1}^m \widehat{P_i}} = \sim \underline{X}_{\sum_{i=1}^m \widehat{P_i}}$.

Proof. If $i = 1$, they are straightforward. If $i > 1$, we prove them as follows:

(1) It can be easily proved from the formula (11).
(2) From (1) and the formula (12), we have that

$$\overline{X}^{\sum_{i=1}^m \widehat{P_i}} = \sim \underline{(\sim X)}_{\sum_{i=1}^m \widehat{P_i}} = \sim \bigcup_{i=1}^m \underline{(\sim X)}_{\widehat{P_i}} = \sim \bigcup_{i=1}^m \left(\sim \overline{X}^{\widehat{P_i}}\right) = \bigcap_{i=1}^m \overline{X}^{\widehat{P_i}}.$$

(3) It can be easily proved from the formula (12).
(4) Let $X = \sim X$ in the formula (12). Then, we have that $\overline{(\sim X)}^{\sum_{i=1}^m \widehat{P_i}} = \sim \underline{X}_{\sum_{i=1}^m \widehat{P_i}}$.

This completes this proof. $\square$

Proposition 3.5. *Let $S = (U, AT, f)$ be a complete information system, $X_1, X_2, \ldots, X_n \subseteq U$, and $\widehat{P_1}, \widehat{P_2}, \ldots, \widehat{P}_m$ be m partitions induced by the attributes $P_1, P_2, \ldots, P_m$, respectively. Then, the following properties hold*

(1) $\underline{\left(\bigcap_{j=1}^n X_j\right)}_{\sum_{i=1}^m \widehat{P_i}} = \bigcup_{i=1}^m \left(\bigcap_{j=1}^n \underline{X_j}_{\widehat{P_i}}\right)$;
(2) $\overline{\left(\bigcup_{j=1}^n X_j\right)}^{\sum_{i=1}^m \widehat{P_i}} = \bigcap_{i=1}^m \left(\bigcup_{j=1}^n \overline{X_j}^{\widehat{P_i}}\right)$;
(3) $\underline{\left(\bigcap_{j=1}^n X_j\right)}_{\sum_{i=1}^m \widehat{P_i}} \subseteq \bigcap_{j=1}^n (\underline{X_j}_{\sum_{i=1}^m \widehat{P_i}})$;
(4) $\overline{\left(\bigcup_{j=1}^n X_j\right)}^{\sum_{i=1}^m \widehat{P_i}} \supseteq \bigcup_{j=1}^n (\overline{X_j}^{\sum_{i=1}^m \widehat{P_i}})$;
(5) $\underline{\left(\bigcup_{j=1}^n X_j\right)}_{\sum_{i=1}^m \widehat{P_i}} \supseteq \bigcup_{j=1}^n (\underline{X_j}_{\sum_{i=1}^m \widehat{P_i}})$;
(6) $\overline{\left(\bigcap_{j=1}^n X_j\right)}^{\sum_{i=1}^m \widehat{P_i}} \subseteq \bigcap_{j=1}^n (\overline{X_j}^{\sum_{i=1}^m \widehat{P_i}})$.

Proof. Similar to Proposition 3.3, one can prove these properties.

(1) $\underline{\left(\bigcap_{j=1}^n X_j\right)}_{\sum_{i=1}^m \widehat{P_i}} = \bigcup_{i=1}^m \underline{\left(\bigcap_{j=1}^n X_j\right)}_{\widehat{P_i}} = \bigcup_{i=1}^m \left(\bigcap_{j=1}^n \underline{X_j}_{\widehat{P_i}}\right)$.

(2) $\overline{\left(\bigcup_{j=1}^{n} X_j\right)}^{\sum_{i=1}^{m}\widehat{P_i}} = \bigcap_{i=1}^{m} \overline{\left(\bigcup_{j=1}^{n} X_j\right)}^{\widehat{P_i}} = \bigcap_{i=1}^{m} \left(\bigcup_{j=1}^{n} \overline{X_j}^{\widehat{P_i}}\right)$.

(3) $\underline{\left(\bigcap_{j=1}^{n} X_j\right)}_{\sum_{i=1}^{m}\widehat{P_i}} = \bigcup_{i=1}^{m} \left(\bigcap_{j=1}^{n} \underline{X_j}_{\widehat{P_i}}\right) = \bigcap_{j=1}^{n} \left(\bigcup_{i=1}^{m} \underline{X_j}_{\widehat{P_i}}\right) \cap \cdots = \bigcap_{j=1}^{n} (\underline{X_j}_{\sum_{i=1}^{m}\widehat{P_i}}) \cap \cdots \subseteq \bigcap_{j=1}^{n} (\underline{X_j}_{\sum_{i=1}^{m}\widehat{P_i}})$.

(4) $\overline{\left(\bigcup_{j=1}^{n} X_j\right)}^{\sum_{i=1}^{m}\widehat{P_i}} = \bigcap_{i=1}^{m} \left(\bigcup_{j=1}^{n} \overline{X_j}^{\widehat{P_i}}\right) = \bigcup_{j=1}^{n} \left(\bigcap_{i=1}^{m} \overline{X_j}^{\widehat{P_i}}\right) \cup \cdots = \bigcup_{j=1}^{n} (\overline{X_j}^{\sum_{i=1}^{m}\widehat{P_i}}) \cup \cdots \supseteq \bigcup_{j=1}^{n} (\overline{X_j}^{\sum_{i=1}^{m}\widehat{P_i}})$.

(5) It follows from $X_j \subseteq \bigcup_{j=1}^{n} X_j$ that $\underline{X_j}_{\sum_{i=1}^{m}\widehat{P_i}} \subseteq \underline{\bigcup_{i=1}^{m} X_j}_{\sum_{i=1}^{m}\widehat{P_i}}$. Hence, we have that $\underline{\left(\bigcup_{j=1}^{n} X_j\right)}_{\sum_{i=1}^{m}\widehat{P_i}} \supseteq \bigcup_{j=1}^{n} (\underline{X_j}_{\sum_{i=1}^{m}\widehat{P_i}})$.

(6) It follows from $\left(\bigcap_{j=1}^{n} X_j \subseteq X_j (j \in \{1,2,\ldots,n\})\right)$ that $\overline{X_j}^{\sum_{i=1}^{m}\widehat{P_i}} \supseteq \overline{\bigcap_{j=1}^{n} X_j}^{\sum_{i=1}^{m}\widehat{P_i}}$. Hence, it has that $\overline{\left(\bigcap_{j=1}^{n} X_j\right)}^{\sum_{i=1}^{m}\widehat{P_i}} \subseteq \bigcap_{j=1}^{n} (\overline{X_j}^{\sum_{i=1}^{m}\widehat{P_i}})$.

This completes the proof. □

Proposition 3.6. *Let $S = (U, AT, f)$ be a complete information system, $X_1, X_2, \ldots, X_n \subseteq U$ with $X_1 \subseteq X_2 \subseteq \cdots X_n$, and $\widehat{P_1}, \widehat{P_2}, \ldots, \widehat{P}_m$ be m partitions induced by the attributes $P_1, P_2, \ldots, P_m$, respectively. Then, the following properties hold*

(1) $\underline{X_1}_{\sum_{i=1}^{m}\widehat{P_i}} \subseteq \underline{X_2}_{\sum_{i=1}^{m}\widehat{P_i}} \subseteq \cdots \subseteq \underline{X_n}_{\sum_{i=1}^{m}\widehat{P_i}}$;
(2) $\overline{X_1}^{\sum_{i=1}^{m}\widehat{P_i}} \subseteq \overline{X_2}^{\sum_{i=1}^{m}\widehat{P_i}} \subseteq \cdots \subseteq \overline{X_n}^{\sum_{i=1}^{m}\widehat{P_i}}$.

Proof. Suppose $1 \leqslant i \leqslant j \leqslant n$. Then, $X_i \subseteq X_j$ holds.

(1) Clearly, $X_i \cap X_j = X_i$. Hence, it follows from (3) in Proposition 3.5 that $\underline{X_i}_{\sum_{i=1}^{m}\widehat{P_i}} = \underline{(X_i \cap X_j)}_{\sum_{i=1}^{m}\widehat{P_i}} \subseteq \underline{X_i}_{\sum_{i=1}^{m}\widehat{P_i}} \cap \underline{X_j}_{\sum_{i=1}^{m}\widehat{P_i}} \Rightarrow \underline{X_i}_{\sum_{i=1}^{m}\widehat{P_i}} = \underline{X_i}_{\sum_{i=1}^{m}\widehat{P_i}} \cap \underline{X_j}_{\sum_{i=1}^{m}\widehat{P_i}} \Rightarrow \underline{X_i}_{\sum_{i=1}^{m}\widehat{P_i}} \subseteq \underline{X_j}_{\sum_{i=1}^{m}\widehat{P_i}}$.

Therefore, we have that

$\underline{X_1}_{\sum_{i=1}^{m}\widehat{P_i}} \subseteq \underline{X_2}_{\sum_{i=1}^{m}\widehat{P_i}} \subseteq \cdots \subseteq \underline{X_n}_{\sum_{i=1}^{m}\widehat{P_i}}$.

(2) Obviously, $X_i \cup X_j = X_j$. Hence, it follows from (4) in Proposition 3.5 that

$$\overline{X_j}^{\sum_{i=1}^{m}\widehat{P_i}} = \overline{(X_i \cup X_j)}^{\sum_{i=1}^{m}\widehat{P_i}} \supseteq \overline{X_i}^{\sum_{i=1}^{m}\widehat{P_i}} \cup \overline{X_j}^{\sum_{i=1}^{m}\widehat{P_i}} \Rightarrow \overline{X_j}^{\sum_{i=1}^{m}\widehat{P_i}} = \overline{X_i}^{\sum_{i=1}^{m}\widehat{P_i}} \cup \overline{X_j}^{\sum_{i=1}^{m}\widehat{P_i}} \Rightarrow \overline{X_i}^{\sum_{i=1}^{m}\widehat{P_i}} \subseteq \overline{X_j}^{\sum_{i=1}^{m}\widehat{P_i}}.$$

Therefore, we have that
$\overline{X_1}^{\sum_{i=1}^{m}\widehat{P_i}} \subseteq \overline{X_2}^{\sum_{i=1}^{m}\widehat{P_i}} \subseteq \cdots \subseteq \overline{X_n}^{\sum_{i=1}^{m}\widehat{P_i}}$.

This completes the proof. □

3.2. Several measures in MGRS

Uncertainty of a set (category) is due to the existence of a borderline region. The greater the borderline region of a set, the lower is the accuracy of the set (and vice versa). Similar to $\alpha(\widehat{P}, X)$ in (2), in order to more precisely express this idea, we introduce another accuracy measure as follows.

Definition 3.3. Let $S = (U, AT, f)$ be a complete information system, $X \subseteq U$ and $\widehat{P_1}, \widehat{P_2}, \ldots, \widehat{P}_m$ be m partitions induced by the attributes $P_1, P_2, \ldots, P_m$, respectively. The *accuracy measure* of X by $\sum_{i=1}^{m}\widehat{P}_i$ is defined as

$$\alpha\left(\sum_{i=1}^{m}\widehat{P}_i, X\right) = \frac{\left|\underline{X}_{\sum_{i=1}^{m}\widehat{P_i}}\right|}{\left|\overline{X}^{\sum_{i=1}^{m}\widehat{P_i}}\right|}, \tag{13}$$

where $X \neq \emptyset$ and $|X|$ denotes the cardinality of a set X.

Proposition 3.7. *Let $S = (U, AT, f)$ be a complete information system, $X \subseteq U$ and $\mathbf{P} = \{\widehat{P_1}, \widehat{P_2}, \ldots, \widehat{P}_m\}$ be m partitions induced by the attributes $P_1, P_2, \ldots, P_m$, respectively. If $P' \subseteq \mathbf{P}$, then*

$$\alpha\left(\sum_{i=1}^{m}\widehat{P}_i, X\right) \geqslant \alpha\left(\sum_{\widehat{P}_i \subseteq P'}\widehat{P}_i, X\right) \geqslant \alpha(\widehat{P}_i, X), \quad i \leqslant m.$$

Proof. Since $P' \subseteq \mathbf{P}$ is a subset of $\mathbf{P}$, it follows from Definition 3.2 that

$$\bigcup_{i=1}^{m} \underline{X}_{\widehat{P_i}} \supseteq \bigcup_{\widehat{P_i} \in P'} \underline{X}_{\widehat{P_i}} \text{ and } \bigcap_{i=1}^{m} \overline{X}^{\widehat{P_i}} \subseteq \bigcap_{\widehat{P_i} \in P'} \overline{X}^{\widehat{P_i}}.$$

Then, it is clear that

$$\left|\bigcup_{i=1}^{m} \underline{X}_{\widehat{P_i}}\right| \geqslant \left|\bigcup_{\widehat{P_i} \in P'} \underline{X}_{\widehat{P_i}}\right| \text{ and } \left|\bigcap_{i=1}^{m} \overline{X}^{\widehat{P_i}}\right| \leqslant \left|\bigcap_{\widehat{P_i} \in P'} \overline{X}^{\widehat{P_i}}\right|.$$

Hence, we have that

$$\alpha\left(\sum_{i=1}^{m} \widehat{P}_i, X\right) = \frac{\left|\underline{X}_{\sum_{i=1}^{m} \widehat{P_i}}\right|}{\left|\overline{X}^{\sum_{i=1}^{m} \widehat{P_i}}\right|} = \frac{\left|\bigcup_{i=1}^{m} \underline{X}_{\widehat{P_i}}\right|}{\left|\bigcap_{i=1}^{m} \overline{X}^{\widehat{P_i}}\right|} \geqslant \frac{\left|\bigcup_{\widehat{P_i} \in P'} \underline{X}_{\widehat{P_i}}\right|}{\left|\bigcap_{\widehat{P_i} \in P'} \overline{X}^{\widehat{P_i}}\right|} = \frac{\left|\sum_{\widehat{P_i} \in P'} \underline{X}_{\widehat{P_i}}\right|}{\left|\sum_{\widehat{P_i} \in P'} \overline{X}^{\widehat{P_i}}\right|} = \alpha\left(\sum_{\widehat{P_i} \subseteq P'} \widehat{P}_i, X\right).$$

Similarly, we have $\alpha\left(\sum_{\widehat{P_i} \subseteq P'} \widehat{P}_i, X\right) \geqslant \alpha(\widehat{P}_i, X)$ $(i \leqslant m)$.

Thus, for any $P' \subseteq \mathbf{P}$ and $P_i \in \mathbf{P}$, the inequality $\alpha\left(\sum_{i=1}^{m} \widehat{P}_i, X\right) \geqslant \alpha\left(\sum_{\widehat{P_i} \subseteq P'} \widehat{P}_i, X\right) \geqslant \alpha(\widehat{P}_i, X)(i \leqslant m)$ hold. This completes the proof. □

Proposition 3.7 shows that the accuracy measure of a set enlarges as the number of granulations for describing the concept increases.

Note that the accuracy measure of a set described by using multi granulations is always better than that of the set described by using a single granulation. The former is suitable for more accurately describing a target concept and solving problems according to a user's requirements.

In particular, if $\widehat{P}_i \preceq \widehat{P}_j$, then $\alpha(\widehat{P}_i + \widehat{P}_j, X) = \alpha(\widehat{P}_i, X)$. It can be understood by the following proposition.

Proposition 3.8. *Let $S = (U, AT, f)$ be a complete information system, $X \subseteq U$ and $\mathbf{P} = \{\widehat{P_1}, \widehat{P_2}, \ldots, \widehat{P}_m\}$ with $\widehat{P_1} \preceq \widehat{P_2} \preceq \cdots \preceq \widehat{P}_m$ be m partitions induced by the attributes $P_1, P_2, \ldots, P_m$, respectively. Then,*

$$\underline{X}_{\sum_{i=1}^{m} \widehat{P_i}} = \underline{X}_{\widehat{P_1}} \quad \text{and} \quad \overline{X}^{\sum_{i=1}^{m} \widehat{P_i}} = \overline{X}^{\widehat{P_1}}.$$

Proof. Suppose $1 \leqslant j \leqslant k \leqslant m$ and $\widehat{P}_j \preceq \widehat{P_k}$. From the definition of $\preceq$, we know that for any $\widehat{P_j}(x) \in \widehat{P}_j$, there exists $\widehat{P_k}(x) \in \widehat{P_k}$ such that $\widehat{P_j}(x) \subseteq \widehat{P_k}(x)$. Therefore, we have that $\underline{X}_{\widehat{P_k}} \subseteq \underline{X}_{\widehat{P_j}}$, i.e., $\underline{X}_{\widehat{P_j}+\widehat{P_k}} = \underline{X}_{\widehat{P_j}} \cup \underline{X}_{\widehat{P_k}} = \underline{X}_{\widehat{P_j}}$. Since $\widehat{P_1} \preceq \widehat{P_2} \preceq \cdots \preceq \widehat{P}_m$, we have that $\underline{X}_{\sum_{i=1}^{m} \widehat{P_i}} = \underline{X}_{\widehat{P_1}}$.

Similarly, we also have that $\overline{X}^{\widehat{P_j}} \subseteq \overline{X}^{\widehat{P_k}}$, i.e., $\overline{X}^{\widehat{P_j}+\widehat{P_k}} = \overline{X}^{\widehat{P_j}} \cap \overline{X}^{\widehat{P_k}} = \overline{X}^{\widehat{P_j}}$. Thus, we have that $\overline{X}^{\sum_{i=1}^{m} \widehat{P_i}} = \overline{X}^{\widehat{P_1}}$.

This completes the proof. □

Let $S = (U, AT, f)$ be a complete information system, $\widehat{Q}$ be the partition induced by the attribute set Q, and $\mathbf{P} = \{\widehat{P_1}, \widehat{P_2}, \ldots, \widehat{P}_m\}$ m partitions induced by the attributes $P_1, P_2, \ldots, P_m$, respectively. The *quality of approximation* of $\widehat{Q}$ by $\mathbf{P}$, also called the *degree of dependency*, is defined by

$$\gamma\left(\sum_{i=1}^{m} \widehat{P}_i, \widehat{Q}\right) = \frac{\sum\left\{\left|\underline{Y}_{\sum_{i=1}^{m} \widehat{P_i}}\right| : Y \in \widehat{Q}\right\}}{|U|}, \tag{14}$$

and is used to evaluate the deterministic part of the rough set description of $\widehat{Q}$ by counting those elements which can be reclassified to blocks of $\widehat{Q}$ with the knowledge given by $\sum_{i=1}^{m} \widehat{P}_i$.

Corollary 3.2. *If $\widehat{Q} \preceq \widehat{R}$, then $\gamma\left(\sum_{i=1}^{m} \widehat{P}_i, \widehat{Q}\right) \leqslant \gamma\left(\sum_{i=1}^{m} \widehat{P}_i, \widehat{R}\right)$.*

Corollary 3.3. *Let $\mathbf{P} = \{\widehat{P_1}, \widehat{P_2}, \ldots, \widehat{P}_m\}$ be m partitions. If $P' \subseteq \mathbf{P}$, then $\gamma\left(\sum_{i=1}^{m} \widehat{P}_i, \widehat{Q}\right) \geqslant \gamma\left(\sum_{\widehat{P_i} \subseteq P'} \widehat{P}_i, \widehat{Q}\right) \geqslant \gamma(\widehat{P}_i, \widehat{Q})$.*

Gediga and Düntsch [3] *introduced a simple statistic $\pi(\widehat{P}, X) = \frac{|\underline{X}_{\widehat{P}}|}{|X|}$ for the precision of (deterministic) approximation of $X \subseteq U$ given $\widehat{P}$, which is not affected by the approximation of $\sim X$. This is just the relative number of elements in X which can be approximated by $\widehat{P}$. Clearly, $\pi(\widehat{P}, X) \geqslant \alpha(\widehat{P}, X)$. It is important to point out that $\pi(\widehat{P}, X)$ requires complete knowledge of X, whereas α does not, since the latter uses only the rough set $(\underline{X}, \overline{X})$. In MGRS, it can be extended and becomes the formula*

$$\pi\left(\sum_{i=1}^{m} \widehat{P}_i, X\right) = \frac{\left|\underline{X}_{\sum_{i=1}^{m} \widehat{P_i}}\right|}{|X|}. \tag{15}$$

Obviously, $\pi\left(\sum_{i=1}^{m} \widehat{P}_i, X\right) \geqslant \alpha\left(\sum_{i=1}^{m} \widehat{P}_i, X\right)$.

In fact, this measure denotes the relative number of elements in X which can be approximated by $\left(\sum_{i=1}^{m}\widehat{P}_i\right)$.

Corollary 3.4. *Let* $\mathbf{P}=\{\widehat{P_1},\widehat{P_2},\ldots,\widehat{P}_m\}$ *be m partitions. If* $P'\subseteq\mathbf{P}$, *then* $\pi\left(\sum_{i=1}^{m}\widehat{P}_i,X\right)\geqslant\pi\left(\sum_{\widehat{P}_i\subseteq P'}\widehat{P}_i,X\right)\geqslant\pi(\widehat{P}_i,X)$.

However, if $X\subseteq Y\subseteq U$, the inequality $\pi\left(\sum_{i=1}^{m}\widehat{P}_i,X\right)\leqslant\pi\left(\sum_{i=1}^{m}\widehat{P}_i,Y\right)$ cannot be established in general.

For each class X of $\widehat{Q}$, the accuracy of approximation of $\sum_{i=1}^{m}\widehat{P}_i$ with respect to X is weighted by the cardinality of X relative to the number of elements in U, and we get that

$$\gamma\left(\sum_{i=1}^{m}\widehat{P}_i,\widehat{Q}\right)=\sum_{X\in\widehat{Q}}\frac{|X|}{|U|}\pi\left(\sum_{i=1}^{m}\widehat{P}_i,X\right)=\sum_{X\in\widehat{Q}}p(X)\pi\left(\sum_{i=1}^{m}\widehat{P}_i,X\right).$$

Therefore, $\gamma\left(\sum_{i=1}^{m}\widehat{P}_i,\widehat{Q}\right)$ is the mean precision of the approximation of $\widehat{Q}$ by $\sum_{i=1}^{m}\widehat{P}_i$.

Using α as a basis, we have

$$\gamma\left(\sum_{i=1}^{m}\widehat{P}_i,\widehat{Q}\right)=\sum_{X\in\widehat{Q}}\frac{|\overline{X}|}{|U|}\alpha\left(\sum_{i=1}^{m}\widehat{P}_i,X\right)=\sum_{X\in\widehat{Q}}p(\overline{X})\alpha\left(\sum_{i=1}^{m}\widehat{P}_i,X\right).$$

Thus, $\gamma\left(\sum_{i=1}^{m}\widehat{P}_i,\widehat{Q}\right)$ also can be regarded as the weighted mean of the accuracies of approximation of the sets $X\in\widehat{Q}$ by $\sum_{i=1}^{m}\widehat{P}_i$.

Yao [38] connected rough set approximation with a classic distance measure based on sets, called *Marczewski–Steinhaus metric (MZ)*, which is defined by

$$MZ(X,Y)=\frac{|X\cup Y|-|X\cap Y|}{|X\cup Y|}.$$

Gediga and Düntsch [3] redefined the measures α,γ and π using MZ, which discovers the relationships between these measures and MZ in Pawlak's rough set theory.

In the multi-granulations rough set model, the above measures α,γ and π presented can be redefined through using MZ as

$$\alpha\left(\sum_{i=1}^{m}\widehat{P}_i,X\right)=1-MZ\left(\underline{X}_{\sum_{i=1}^{m}\widehat{P}_i},\overline{X}^{\sum_{i=1}^{m}\widehat{P}_i}\right),$$
$$\pi\left(\sum_{i=1}^{m}\widehat{P}_i,X\right)=1-MZ\left(\underline{X}_{\sum_{i=1}^{m}\widehat{P}_i},X\right),$$

and

$$\gamma\left(\sum_{i=1}^{m}\widehat{P}_i,\widehat{Q}\right)=1-MZ\left(\bigcup_{X\in\widehat{Q}}\underline{X}_{\sum_{i=1}^{m}\widehat{P}_i},\bigcup_{X\in\widehat{Q}}\overline{X}^{\sum_{i=1}^{m}\widehat{P}_i}\right)=1-MZ\left(\bigcup_{X\in\widehat{Q}}\underline{X}_{\sum_{i=1}^{m}\widehat{P}_i},U\right).$$

In addition, Xu and Liang [36] introduced two forms of inclusion degree concept D_0 and D_1 in rough set theory as follows:

$$D_0(Y/X)=\frac{|X\cap Y|}{|X|}(X\neq\emptyset)\text{ and}$$
$$D_1(\widehat{P},\widehat{Q})=\frac{\left|\left(\bigcup_{X\in\widehat{P}}X\right)\cap\left(\bigcup_{Y\in\widehat{Q}}Y\right)\right|}{\left|\bigcup_{Y\in\widehat{Q}}Y\right|}.$$

It is easy to see that

$$\alpha\left(\sum_{i=1}^{m}\widehat{P}_i,X\right)=\frac{\left|\underline{X}_{\sum_{i=1}^{m}\widehat{P}_i}\cap\overline{X}^{\sum_{i=1}^{m}\widehat{P}_i}\right|}{\left|\overline{X}^{\sum_{i=1}^{m}\widehat{P}_i}\right|}=D_0\left(\underline{X}_{\sum_{i=1}^{m}\widehat{P}_i}\Big/\overline{X}^{\sum_{i=1}^{m}\widehat{P}_i}\right),$$
$$\pi\left(\sum_{i=1}^{m}\widehat{P}_i,X\right)=\frac{\left|\underline{X}_{\sum_{i=1}^{m}\widehat{P}_i}\cap X\right|}{|X|}=D_0\left(\underline{X}_{\sum_{i=1}^{m}\widehat{P}_i}\Big/X\right)$$

and

$$\gamma\left(\sum_{i=1}^{m}\widehat{P}_i,\widehat{Q}\right)=\frac{\sum\left\{\left|\underline{Y}_{\sum_{i=1}^{m}\widehat{P}_i}\right|:Y\in\widehat{Q}\right\}}{|U|}=\frac{\left|\left(\bigcup_{Y\in\widehat{Q}}\underline{Y}_{\sum_{i=1}^{m}\widehat{P}_i}\right)\cap\left(\bigcup_{Y\in\widehat{Q}}Y\right)\right|}{\left|\bigcup_{Y\in\widehat{Q}}Y\right|}=D_1\left(\bigcup_{Y\in\widehat{Q}}\underline{Y}_{\sum_{i=1}^{m}\widehat{P}_i}\Big/\bigcup_{Y\in\widehat{Q}}Y\right).$$

Thus, these three measures α, π and γ can be reduced to inclusion degree.

Since the multi-granulations rough set model mainly considers the lower approximation and the upper approximation of a target concept by multi equivalence relations, in the following, we only introduce a measure of importance of condition attributes with respect to decision attributes in a complete target information system.

Let $S = (U, AT, f, D, g)$ be a complete target information system, a measure of importance of condition attributes $P \subseteq AT$ with respect to decision attributes D in MGRS in terms of the under approximation and the upper approximation can be divided into two forms: an *importance measure of the lower approximation* and an *importance measure of the upper approximation*.

Let $S = (U, AT, f, D, g)$ be a complete target information system and P be a non-empty subset of AT: $\emptyset \subset P \subseteq AT$. Given a condition attribute $a \in P$ and $X \in \widehat{D}$. Firstly, we give two preliminary definitions in the following:

Definition 3.4. We say that a is *lower approximation significant* in P with respect to X if $\underline{X}_{\sum_{i=1}^{|P|}\widehat{P_i}} \supset \underline{X}_{\sum_{i=1,P_i \neq a}^{|P|}\widehat{P_i}}$ $(P_i \in P)$, and that a is not lower approximation significant in P with respect to X if $\underline{X}_{\sum_{i=1}^{|P|}\widehat{P_i}} = \underline{X}_{\sum_{i=1,P_i \neq a}^{|P|}\widehat{P_i}}$ $(P_i \in P)$, where $|P|$ is the cardinality of attribute set P.

Definition 3.5. We say that a is *upper approximation significant* in P with respect to X if $\overline{X}^{\sum_{i=1}^{|P|}\widehat{P_i}} \subset \overline{X}^{\sum_{i=1,P_i \neq a}^{|P|}\widehat{P_i}}$ $(P_i \in P)$, and that a is not upper approximation significant in P with respect to X if $\overline{X}^{\sum_{i=1}^{|P|}\widehat{P_i}} = \overline{X}^{\sum_{i=1,P_i \neq a}^{|P|}\widehat{P_i}}$ $(P_i \in P)$, where $|P|$ is the cardinality of attribute set P.

We introduce a quantitative measure for the significance as follows:

The *importance measure of the lower approximation* of condition attributes $P \subseteq AT$ with respect to decision attributes D in MGRS is defined as

$$S_P(D) = \frac{\sum\left\{\left|\underline{X}_{\sum_{i=1}^{m}\widehat{P_i}} \setminus \underline{X}_{\sum_{i=1,P_i \notin P}^{m}\widehat{P_i}}\right| : X \in \widehat{D}\right\}}{|U|}, \tag{16}$$

where the attributes $AT = \{P_1, P_2, \ldots, P_m\}$, $\widehat{P_i} \in AT$ is the partition induced by the condition attribute P_i, and $\widehat{D}$ is the partition induced by the decision attributes D.

The *importance measure of the upper approximation* of condition attributes $P \subseteq AT$ with respect to decision attributes D in MGRS is defined as

$$S^P(D) = \frac{\sum\left\{\left|\overline{X}^{\sum_{i=1,P_i \notin P}^{m}\widehat{P_i}} \setminus \overline{X}^{\sum_{i=1}^{m}\widehat{P_i}}\right| : X \in \widehat{D}\right\}}{|U|}, \tag{17}$$

where the attributes $AT = \{P_1, P_2, \ldots, P_m\}$, $\widehat{P_i} \in AT$ is the partition induced by the condition attribute P_i, and $\widehat{D}$ is the partition induced by the decision attributes D.

In particular, when $P = \{a\}$, $S_a(D)$ is the importance measure of the lower approximation of the attribute $a \in AT$ with respect to D and $S^a(D)$ is the importance measure of the upper approximation of the attribute $a \in AT$ with respect to D.

To compute the significance of an attribute a in P with respect to D, we need to compute $|P|$ partitions $\widehat{P}_i$ $(i \leqslant |P|)$. The time complexity for computing each partition is $\mathbf{O}(|U|^2)$. So, the time complexity for computing $|P|$ partitions is $\mathbf{O}(|P||U|^2)$. Therefore, the time complexity of computing a lower approximation of $X \in \widehat{D}$ $(|\widehat{D}| \leqslant |U|)$ by P is $\mathbf{O}(|P||U|^3)$.

From the above two definitions, we know the following:

- $S_P(D) \geqslant 0$ and $S^P(D) \geqslant 0$;
- attributes P with respect to D is the lower approximation significant if and only if $S_P(D) = 0$; and
- attributes P with respect to D is the upper approximation significant if and only if $S^P(D) = 0$.

Example 3.3 (*Continued from Example 3.1*). Compute the importance measure of each condition attribute with respect to the decision attribute d.

By computing, we have that

$$S_L(d) = \frac{|\{1,2,8\} \setminus \{1,2\}| + |\{3,4,5\} \setminus \{3,4,5\}|}{8} = \frac{1}{8},$$
$$S_I(d) = \frac{|\{1,2,8\} \setminus \{1,2,8\}| + |\{3,4,5\} \setminus \{3,4,5\}|}{8} = 0 \text{ and}$$
$$S_P(d) = \frac{|\{1,2,8\} \setminus \{8\}| + |\{3,4,5\} \setminus \{3,4,5\}|}{8} = \frac{2}{8}$$

and

$$S^L(d) = \frac{|\{1,2,6,7,8\} \setminus \{1,2,6,7,8\}| + |\{3,4,5,6,7,8\} \setminus \{3,4,5,6,7\}|}{8} = \frac{1}{8},$$
$$S^I(d) = \frac{|\{1,2,6,7,8\} \setminus \{1,2,6,7,8\}| + |\{3,4,5,6,7\} \setminus \{3,4,5,6,7\}|}{8} = 0 \text{ and}$$
$$S^P(d) = \frac{|\{1,2,6,7,8\} \setminus \{1,2,6,7,8\}| + |\{1,2,3,4,5,6,7\} \setminus \{3,4,5,6,7\}|}{8} = \frac{2}{8}.$$

It follows from Example 3.3 that $S_P(d) > S_L(d) > S_I(d)$ and $S^P(d) > S^L(d) > S^I(d)$. That is to say, the important measures of the lower approximation and upper approximation of condition attribute P are all maximum, and the important measures of the lower approximation and upper approximation of condition attribute I are all minimum. In fact, from $S_I(d) = S^I(d) = 0$, one can remove the condition attribute I in terms of the approximation representation of all decision classes in Table 1.

In the following, through experimental analyses, we illustrate the deference between the MGRS and Pawlak's rough set model. We have downloaded three public data sets (complete target information systems) with practical applications from UCI Repository of machine learning databases, which are described in Table 2. All condition attributes and decision attributes in these three data sets are discrete.

Here, we compare the degree of dependency in MGRS with that in Pawlak's rough set model on these three practical data sets. The comparisons of values of two measures with the numbers of features in these three data sets are shown in Tables 3–5 and Figs. 2–4.

In Figs. 2–4, the term MGRS is the complete multi-granulations rough set framework proposed in this paper, and the term SGRS is Pawlak's rough set model. It can be seen from Figs. 2–4 that the value of the degree of dependency in MGRS is not bigger than that in Pawlak's rough set model for the same number of selected features, and this value increases as the number of selected features becomes bigger in the same data set. In particular, from Figs. 2 and 3, it is easy to see that the values of the degree of dependency in MGRS are equal to zero. In this situation, the lower approximation of the target decision equals an empty set in the decision table. In essence, it is because that the equivalence classes induced by a singleton

Table 2
Data sets description.

Data sets	Samples	Condition features	Decision classes
Tie-tac-toe	958	9	2
Dermatology	366	33	6
Car	1728	6	4

Table 3
The two degrees of dependency with different numbers of features in the data set Tie-tac-toe.

Measure	Features								
	1	2	3	4	5	6	7	8	9
γ in SGRS	0.0000	0.0000	0.1253	0.1628	0.4186	0.7766	0.9436	1.0000	1.0000
γ in MGRS	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 4
The two degrees of dependency with different numbers of features in the data set Dermatology.

Measure	Features										
	3	6	9	12	15	18	21	24	27	30	33
γ in SGRS	0.0437	0.6066	0.8552	0.8962	0.9809	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
γ in MGRS	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 5
The two degrees of dependency with different numbers of features in the data set Car.

Measure	Features					
	1	2	3	4	5	6
γ in SGRS	0.0000	0.1875	0.1875	0.4583	0.4809	1.0000
γ in MGRS	0.0000	0.0000	0.0000	0.3333	0.3333	0.3333

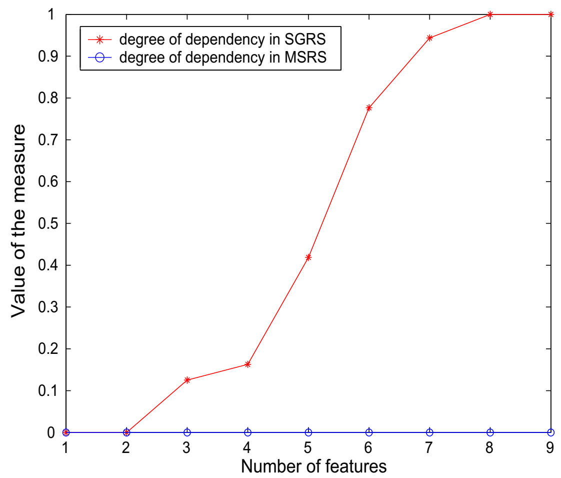

Fig. 2. Variation of the two degrees of dependency with the numbers of features (data set Tie-tac-toe).

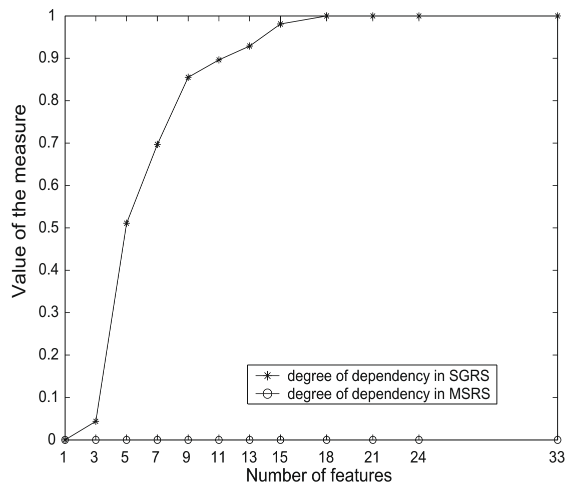

Fig. 3. Variation of the two degrees of dependency with the numbers of features (data set Dermatology).

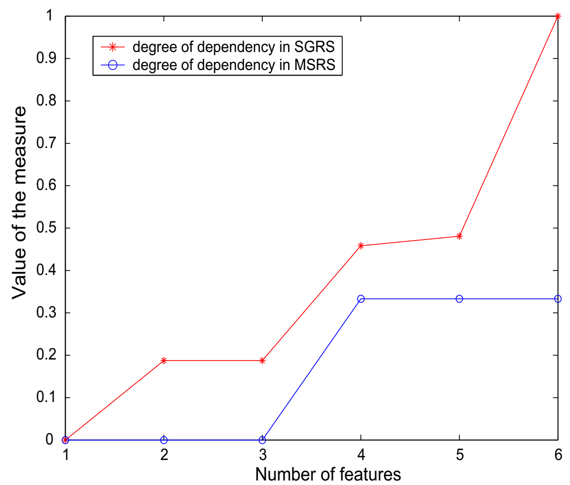

Fig. 4. Variation of the two degrees of dependency with the numbers of features (data set Car).

attribute are all coarser than those induced by multi attributes. One can draw the same conclusion from Fig. 4. Therefore, although the MGRS cannot obtain much bigger approximation measure and degree of dependency than Pawlak's rough set model, this approach can be used in concept representation, rule extraction and data analysis from data sets under multi granulations on the basis of keeping original granulation structure.

4. Attribute reduction

Intuitively, some attributes are not significant in a representation and their removal has no real impact on the value of the representation of elements [15,41,42]. If it is not significant, one can simply remove an attribute from further consideration.

Let $S=(U,AT,f,D,g)$ be a complete target information system, $P\subseteq AT$ and the decision partition by D be $\widehat{D}=\{X_1, X_2,\ldots,X_r\}$. The lower approximation distribution function and upper approximation distribution function are defined as follows:

$$\underline{D}_P=\left(\underline{X_1}_{\sum_{P_i\in P}\widehat{P_i}},\underline{X_2}_{\sum_{P_i\in P}\widehat{P_i}},\ldots,\underline{X_r}_{\sum_{P_i\in P}\widehat{P_i}}\right)$$

and

$$\overline{D}^P=\left(\overline{X_1}^{\sum_{P_i\in P}\widehat{P_i}},\overline{X_2}^{\sum_{P_i\in P}\widehat{P_i}},\ldots,\overline{X_r}^{\sum_{P_i\in P}\widehat{P_i}}\right).$$

Definition 4.1. Let $S=(U,AT,f,D,g)$ be a complete target information system and P be a non-empty subset of AT.

(1) If $\underline{D}_P=\underline{D}_{AT}$, we say that P is a *lower approximation consistent set* of S. If P is a lower approximation consistent set, and no proper subset of P is lower approximation consistent, then P is called a *lower approximation reduct* of S.
(2) If $\overline{D}^P=\overline{D}^{AT}$, we say that P is an *upper approximation consistent set* of S. If P is an upper approximation consistent set, and no proper subset of P is upper approximation consistent, then P is called an *upper approximation reduct* of S.
(3) If P is not only a lower approximation reduct but also a upper approximation reduct, then P is called an *approximation reduct* of S.

Let $S=(U,AT,f,D,g)$ be a complete target information system. If $\widehat{AT}\preceq\widehat{D}$, then we say S is consistent, otherwise it is inconsistent [10].

It is easy to prove that an upper approximation consistent set must be a lower approximation consistent set, but the converse is not true in an inconsistent target information system. Clearly, P is a lower approximation consistent set iff P is a upper approximation consistent set in a consistent target information system.

In particular, If $\widehat{D}=\{X\}$, we regard the above P as a lower approximation reduct, an upper approximation reduct and an approximation reduct of a set X, respectively.

Let $S=(U,AT,f,d,g)$ be a complete target information system, where $U=\{e_1,e_2,\ldots,e_{|U|}\},AT=\{P_1,P_2,\ldots,P_{|AT|}\}$ and $\widehat{d}=\{X_1,X_2,\ldots,X_r\}$. In the rough set model based on multi equivalence relations, we here develop an algorithm for computing all lower approximation reducts-that is, all subsets $AT_0:AT_{01},AT_{02},\ldots,AT_{0s}$ of AT such that:

(1) $\underline{d}_{AT_0}=\underline{d}_{AT}$; and
(2) if $AT'\subset AT_0$, then $\underline{d}_{AT'}\neq\underline{d}_{AT}$.

Algorithm 2. This algorithm gives all lower approximation reducts of the target information system S.

Let us denote the binomial coefficients by $C^k_{|AT|}=|AT|!/k!(|AT|-k)!$.

(1) Let us denote $C^1_{|AT|}=|AT|$ singletons, one-attribute subsets, by
$$AT_{11}=\{P_1\},AT_{12}=\{P_2\},\ldots,AT_{1j}=\{P_j\},\ldots,AT_{1C^1_{|AT|}}=\{P_{|AT|}\}.$$
(2) Let us denote $C^2_{|AT|}=|AT|(|AT|-1)/2!$ two-attribute subsets by
$$AT_{21}=\{P_1,P_2\},\ldots,AT_{2j}=\{P_1,P_j\},\ldots,AT_{2C^2_{|AT|}}=\{P_{|AT|-1},P_{|AT|}\}.$$
(3) Generally, let us denote $C^k_{|AT|}=|AT|!/k!(|AT|-k)!$ k-attribute subsets by
$$AT_{k1}=\{P_1,P_2,\ldots,P_k\},\ldots,AT_{kj},\ldots,AT_{kC^k_{|AT|}}=\{P_{|AT|-k+1},\ldots,P_{|AT|-1},P_{|AT|}\}.$$
(4) Notice that $C^{|AT|}_{|AT|}=1$ $|AT|$-attribute subset is $AT_{|AT|1}=AT$.

The algorithm is to search subsets of AT as follows: singletons, two-attribute subsets, ..., t-attribute subsets, and so on. Continue up to the unique $|AT|$-attribute subset AT itself.

We use the following variables:

- s – the number of the lower approximation reducts we have already found,
- t – counting from 1 to s,
- k – we are currently searching k-attribute subset AT_{kj}, and
- j – we are currently searching the jth subset AT_{kj} in all k-attribute subsets $AT_{k1}, \ldots, AT_{kj}, \ldots, AT_{kC^k_{|AT|}}$.

(II1) Set $j \leftarrow 1, s \leftarrow 0, k \leftarrow 1$;
(II2) While $k \leqslant |AT|$ Do
    $j \leftarrow 1$;
    While $j \leqslant C^k_{|AT|}$ Do
      for $t = 1$ to s Do
        If $AT_{0t} \subset AT_{kj}$, then break;
        Endif
      Endfor
      if $\underline{d}_{AT_{kj}} = \underline{d}_{AT}$, then
        $s \leftarrow s + 1, AT_{0s} \leftarrow AT_{kj}$;
      Endif
      $j \leftarrow j + 1$;
    Endwhile
    $k \leftarrow k + 1$;
  Endwhile
(II3) Output $AT_{01}, AT_{02}, \ldots, AT_{0s}$ (s lower approximation reducts).

The time complexity of this algorithm for finding all lower approximation reducts is exponential since it checks all subsets in 2^{AT}, and $|2^{AT}| = 2^{|AT|}$. We know that the time complexity of computing $|AT|$ partitions is $\mathbf{O}(|AT||U|^2)$ and the time complexity of computing a lower approximation of every $X \in \widehat{d}$ ($|\widehat{d}| \leqslant |U|$) by AT_{kj} ($k \leqslant |AT|$) is $\mathbf{O}(|AT||U|^3)$. Thus, the time complexity of Algorithm 2 is

$$2^{|AT|} \times \mathbf{O}(|AT||U|^2 + |AT||U|^3) = \mathbf{O}(2^{|AT|}|AT||U|^3).$$

Example 4.1 (*Continued from Example 3.1*). Compute all lower approximation reducts for the complete target information system about emporium investment project.

One can find all lower approximation reducts for this target information system in Table 1 by using the above Algorithm 2 (see Table 6).

Similar to the idea of Algorithm 2, one can design an algorithm to compute all upper approximation reducts in a complete target information system.

However, the time complexity of Algorithm 2 is exponential so that it cannot be applied efficiently to practical applications. We here provide a heuristic algorithm based on the importance measure of lower approximation of a condition attribute with respect to the decision attribute d to find a lower approximation reduct in complete target information systems.

Algorithm 3. Let $S = (U, AT, f, d, g)$ be a complete target information system, where $U = \{e_1, e_2, \ldots, e_{|U|}\}, AT = \{P_1, P_2, \ldots, P_{|AT|}\}$ and $\widehat{d} = \{X_1, X_2, \ldots, X_r\}$.

This algorithm finds a lower approximation reduct through using a heuristic information.

Table 6
A lower approximation from Table 1.

Project	Locus	Population density	Decision
e_1	Common	Big	Yes
e_2	Bad	Big	Yes
e_3	Bad	Small	No
e_4	Bad	Small	No
e_5	Bad	Small	No
e_6	Bad	Medium	Yes
e_7	Common	Medium	No
e_8	Good	Medium	Yes

We use the following variables:

- AT_0 – it is used to record a lower approximation reduct, and
- i – we are currently searching the ith condition attribute AT'_i in the given sequence.

(III1) Compute $|AT|$ condition partitions and a decision partition $\widehat{d}$;
(III2) Sort $AT = \{P'_1, P'_2, \ldots, P'_{|AT|}\}$, where $S_{P'_i}(d) \geqslant S_{P'_{i+1}}(d)$;
(III3) Set $i \leftarrow 1, AT_0 = \emptyset$.
(III4) If $\underline{d}_{AT_0} \neq \underline{d}_{AT}$, then
$AT_0 \leftarrow AT_0 \cup P'_i$,
$i \leftarrow i + 1$;
Endif
(III5) Found a lower approximation reduct: AT_0. Output the set AT_0.

The time complexity of this algorithm for computing $|AT|$ condition partitions and a decision partition $\widehat{d}$ is $\mathbf{O}((|AT|+1)|U|^2)$. The time complexity of computing $|AT|$ importance measures is $\mathbf{O}(|AT||U|^3)$ and the time complexity of sorting is $\mathbf{O}(|AT|log_2|AT|)$. And the time complexity for running $|AT|$ comparisons $\underline{d}_{AT_0} = \underline{d}_{AT}$ is $\mathbf{O}(|AT||U|^3)$. Thus, the time complexity of Algorithm 3 is

$$\mathbf{O}((|AT|+1)|U|^2 + |AT||U|^3 + |AT|log_2|AT| + |AT||U|^3) = \mathbf{O}(|AT||U|^3).$$

Let $\mathbf{A}$ be the set of all lower approximation reducts and $\mathbf{B}$ be the set of all upper approximation reducts. It is obvious that the approximation reducts $\mathbf{C} = \mathbf{A} \cap \mathbf{B}$.

Let $S = (U, AT, f, D, g)$ be a complete target information system, where $U = \{e_1, e_2, \ldots, e_{|U|}\}, AT = \{P_1, P_2, \ldots, P_{|AT|}\}$ and $\widehat{D} = \{X_1, X_2, \ldots, X_r\}$. We denote all lower approximation reducts of $X \in \widehat{D}$ by $\mathbf{A}(X)$ and all upper approximation reducts of $X \in \widehat{D}$ by $\mathbf{B}(X)$ and all approximation reducts of $X \in \widehat{D}$ by $\mathbf{C}(X)$, respectively. And, we call $Core(\mathbf{A}(X))$ the lower approximation core of $X, Core(\mathbf{B}(X))$ the upper approximation core of X and $Core(\mathbf{C}(X))$ the approximation core of X, respectively.

Proposition 4.1. *Let $S = (U, AT, f, D, g)$ be a complete target information system and $\widehat{D} = \{X_1, X_2, \ldots, X_r\}$. Then*

$$\mathbf{A} = \bigcap_{k=1}^{r} \mathbf{A}(X_k) \text{ and } \mathbf{B} = \bigcap_{k=1}^{r} \mathbf{B}(X_k).$$

Proof. They are straightforward from Definition 4.1.

We call $Core(\mathbf{A}) = \bigcap \mathbf{A}_i(\mathbf{A}_i \in \mathbf{A})$, $Core(\mathbf{B}) = \bigcap \mathbf{B}_i(\mathbf{B}_i \in \mathbf{B})$ and $Core(S) = \bigcap \mathbf{C}_i(\mathbf{C}_i \in \mathbf{C})$ the lower approximation core, the upper approximation core and the approximation core of a complete target information system S, respectively. □

Proposition 4.2. *Let $S = (U, AT, f, D, g)$ be a complete target information system and $\widehat{D} = \{X_1, X_2, \ldots, X_r\}$. Then*

$$Core(\mathbf{A}) = \bigcap_{k=1}^{r} Core(\mathbf{A}(X_k)) \text{ and } Core(\mathbf{B}) = \bigcap_{k=1}^{r} Core(\mathbf{B}(X_k)).$$

Proof. They are straightforward. □

Clearly, we have that $Core(S) = Core(\mathbf{A}) \cap Core(\mathbf{B})$. In fact, the core is indispensable attribute to construct an approximation reduct. One can find $Core(S) = \{Locus, Populationdensity\}$ from the target information system S in Table 1. Fig. 5 shows the relationship between the approximation reducts and the approximation core of a target information system.

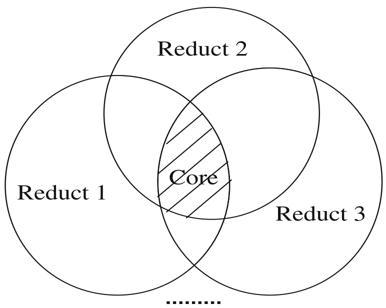

Fig. 5. Relationship between the approximation reducts and the approximation core.

In the following, we discuss the definition of decision rule and several rule extracting methods based on MGRS in a complete target information system.

If Q is a set of predictor features and d a decision attribute, then MGRS generates rules of the form

$$\bigvee_{q\in Q} x^q = m_q \Rightarrow x^d = m_d^0 \vee x^d = m_d^1 \vee \cdots \vee x^d = m_d^k, \tag{18}$$

where x^r is the attribute value of object x with respect to attribute r.

Unlike the decision rules ("AND" rules) from Pawlak's rough set theory [16,17,20,21], the form of these decision rules is "OR". That is to say, they can be decomposed to many decision rules. In essence, the restriction of this kind of decision rules is weaker than that of decision rules from Pawlak's rough set theory, since intersection operations among equivalence classes need not be performed in MGRS.

In the following, we present an algorithm for rule extracting in the rough set model based on multi equivalence relations.

Algorithm 4. Let $S = (U, AT, f, d, g)$ be a complete target information system, where $U = \{e_1, e_2, \ldots, e_{|U|}\}, AT = \{P_1, P_2, \ldots, P_{|AT|}\}$ and $\widehat{d} = \{X_1, X_2, \ldots, X_r\}$.

This algorithm extracts some certain "OR" decision rules from a complete target information system on the basis of lower approximation reduct of a system.

We use the following variables:

- i – counting from 1 to $|AT_0|$,
- j – we are currently searching the jth equivalence class P_i^j in the partition $\widehat{P}_i$,
- k – counting from 1 to r,
- *Rule* – it is used to record decision rules extracted, and
- *Ruleset* – it is used to record the set containing all decision rules extracted.

(IV1) Compute a lower approximation reduct $AT_0 = \{P_1, P_2, \ldots, P_{|AT_0|}\}$.
(IV2) Set $i \leftarrow 1, j \leftarrow 1, k \leftarrow 1$, $Rule = \emptyset$ and $Ruleset = \emptyset$.
(IV3) While $k \leqslant r$ Do //all decision classes have not been checked;
While $i \leqslant |AT_0|$ Do //all condition attributes in the lower approximation reduct have not been checked;
While $j \leqslant |\widehat{P}_i|$ Do //all equivalence classes in the the partition $\widehat{P}_i$ have not been checked;
If $P_i^j \subseteq X_k$
then $Rule \leftarrow Rule \cup dex(P_i^j)$,
otherwise we ignore it; //it cannot form a certain rule;
Endif
$j \leftarrow j+1$; // to check next equivalence classes $P_i^j \in \widehat{P}_i$;
Endwhile
$i \leftarrow i+1, j \leftarrow 1$;// to check next attribute P_i;
Endwhile
$Rule \leftarrow Rule \Rightarrow dex(X_k)$, put *Rule* into the set *Ruleset* and $Rule \leftarrow \emptyset$;
$k \leftarrow k+1, i \leftarrow 1$; // to check next X_k;
Endwhile
(IV4) Output the decision rule set *Ruleset*.

We know that the time complexity of computing a lower approximation reduct is $\mathbf{O}(|AT||U|^3)$. The time complexity of (IV3) is $\mathbf{O}(|AT_0||U|^2)$ since it performs intersection operations between each P_i^j and X_k (see Algorithm 1). Thus, the time complexity of Algorithm 4 is

$$\mathbf{O}(|AT_0||U|^2 + |AT||U|^3) = \mathbf{O}(|AT||U|^3).$$

Example 4.2 (*Continued from Example 3.1*). Extract certain "OR" decision rules from Table 1 by using Algorithm 4.

One can find a lower approximation reduct $\widehat{AT_{02}} = \{Locus, Populationdensity\}$ for this target information system in Table 1 by using the above Algorithm 2.

In Table 1, the decision partition is $\widehat{decision} = \{X_1, X_2\} = \{\{e_1, e_2, e_6, e_8\}, \{e_3, e_4, e_5, e_7\}\}$. By computing, their lower approximations by two granulations $\widehat{L} + \widehat{P}$ are as

$$\underline{X_1}_{\widehat{L}+\widehat{P}} = \{e_1, e_2, e_8\} \text{ and } \underline{X_2}_{\widehat{L}+\widehat{P}} = \{e_3, e_4, e_5\}.$$

There are two certain "OR" decision rules extracted from Table 1 as follows:

$$(Locus = good) \vee (Populationdensity = big) \Rightarrow (Decision = Yes)$$

and

$(Populationdensity = small) \Rightarrow (Decision = No)$.

If we check that whether $P_i^j \cap X_k = \emptyset$ or not based on a upper approximation reduct obtained, the algorithm for extracting uncertain decision rules from a complete target information system also can be designed analogously.

However, Since the time complexity of Algorithm 4 based on an approximation reduct is $\mathbf{O}(|AT||U|^3)$, it is inconvenient to be used in practical issues. In the following, we present an improved algorithm for rule extracting in the rough set model based on multi equivalence relations. It is worth noting that this algorithm need not compute an approximation reduct.

Algorithm 5. Let $S = (U, AT, f, D, g)$ be a complete target information system, where $U = \{e_1, e_2, \ldots, e_{|U|}\}, AT = \{P_1, P_2, \ldots, P_{|AT|}\}$ and $\widehat{D} = \{X_1, X_2, \ldots, X_r\}$.

This algorithm directly extracts some certain "OR" decision rules from a complete target information system.

We use the following variables:

- i – counting from 1 to $|AT|$,
- j – we are currently searching the jth equivalence class P_i^j in the partition $\widehat{P}_i$,
- k – counting from 1 to r,
- *Rule* – it is used to record decision rules extracted, and
- *Ruleset* – it is used to record the set containing all decision rules extracted.

(IV1) Compute $|AT|$ partitions $\{\widehat{P_1}, \widehat{P_2}, \ldots, \widehat{P_{|AT|}}\}$.
(IV2) Set $i \leftarrow 1, j \leftarrow 1, k \leftarrow 1$, $Rule = \emptyset$ and $Ruleset = \emptyset$.
(IV3) While $k \leqslant r$ Do //all decision classes have not been checked;
    While $i \leqslant |AT_0|$ Do //all condition attributes have not been checked;
      While $j \leqslant |\widehat{P}_i|$ Do // all equivalence classes in the partition $\widehat{P}_i$ have not been checked;
        If $P_i^j \subseteq X_k$ then $Rule \leftarrow Rule \cup dex(P_i^j)$,
        otherwise we ignore it; //it cannot form a certain rule;
      Endif
      $j \leftarrow j + 1$; //to check next equivalence classes $P_i^j \in \widehat{P}_i$;
    Endwhile
    $i \leftarrow i + 1, j \leftarrow 1$; //to check next attribute P_i;
    Endwhile
    $Rule \leftarrow Rule \Rightarrow dex(X_k)$, put *Rule* into the set *Ruleset* and $Rule \leftarrow \emptyset$; $k \leftarrow k + 1, i \leftarrow 1$; //to check next X_k;
  Endwhile
(IV4) Output the decision rule set *Ruleset*.

We know that the time complexity of computing $|AT|$ partitions is $\mathbf{O}(|AT||U|^2)$. The time complexity of computing intersection operations between a partition and $X_i \in \widehat{D}$ is $\mathbf{O}(|U||X_i|)$. Thus, the time complexity of Algorithm 5 is

$$\begin{aligned}\mathbf{O}(|AT||U|^2 + |AT||U||X_1| + |AT||U||X_2| + \cdots + |AT||U||X_r|) &= \mathbf{O}(|AT||U|^2 + |AT||U|(|X_1| + |X_2| + \cdots + |X_r|)) \\ &= \mathbf{O}(|AT||U|^2 + |AT||U|^2) = \mathbf{O}(|AT||U|^2).\end{aligned}$$

This algorithm also can be run in parallel mode to compute concurrently all corresponding classifications and intersections between each partition and decision classes from many attributes. This time complexity will be $\mathbf{O}(|U|^2)$.

Example 4.3 (*Continued from Example 3.1*). Extract the "OR" decision rules from Table 1 through using Algorithm 5.

There are two certain "OR" decision rules extracted from Table 1 by using Algorithm 5 as follows:

$(Locus = good) \vee (Populationdensity = big) \Rightarrow (Decision = Yes)$

and

$(Populationdensity = small) \vee (Investment = low) \Rightarrow (Decision = No)$.

From Examples 4.2 and 4.3, we know that these decision rules extracted from the same target information systems are dissimilar. For Example 4.3, unlike Example 4.2, the rule $(Populationdensity = small) \vee (Investment = low) \Rightarrow (Decision = No)$ has two parts:

$(Populationdensity = small) \Rightarrow (Decision = No)$

and

$(Investment = low) \Rightarrow (Decision = No)$.

In fact, for the target set $X_2 = \{e_3, e_4, e_5, e_7\}$, it has two lower approximation reducts *Populationdensity* and *Investment*. However, the lower approximation reduct of this target system does not contain the attribute *Investment*. This difference can be easily understood by Proposition 4.1, i.e., the lower approximation reduct of a target system S is the intersection set of the lower approximation reducts of every decision classes induced by the decision attributes.

What we want to point out is that: like the idea of Algorithm 5, one can extract uncertain decision rules from a complete target information system through using a upper approximation reduct of every decision class.

The above results and analyses give a tentative study for knowledge discovery from multi information systems and data analysis through multi granulations in the framework of rough set theory. For some practical applications, the two methods SGRS and MGRS can be combined to solve problems. As mentioned in footnotes 1 and 2, distributive information systems and groups of intelligent agents are all data analysis and problem solving from multi granulations (each information system or intelligent agent can be regarded as a granulation or viewpoint). Rasiowa and Marek [25,26] gave mechanical proof systems for logic of reaching consensus by groups of intelligent agents. Rauszer [27–30] established rough logic for multi agent systems and proposed approximation methods for information systems, in which there exists an ordering relation between two information systems. MGRS proposed in this paper does not concern on logic reasoning from multi granulations, but try to establish a study framework based on rough set theory through using multi granulations. Hence, one can say that the multi-granulation rough set model will provide a novel approach to knowledge discovery from multi information systems and data analysis through multi intelligent agents.

5. Conclusions and discussion

In this paper, the classical single-granulation rough set theory has been significantly extended. As a result of this extension, a multi-granulation rough set model (MGRS) has been developed. In this extension, the approximations of sets are defined by using multi equivalence relations on the universe. These equivalence relations can be chosen according to a user's requirements or targets of problem solving. This extension has a number of useful properties. In particular, some of the properties of Pawlak's rough set model have become special instances of those of MGRS.

Under MGRS, we have developed several important measures, such as the accuracy measure α, the quality of approximation γ and the precision of approximation π, and re-interpreted them in terms of a classic measure based on sets, the Marczewski–Steinhaus metric and the inclusion degree measure. An importance measure of upper approximation and an importance measure of lower approximation have been introduced to measure the importance of a condition attribute with respect to decision attributes as well.

In order to acquire brief rules from a target complete information system, the attribute reduction and rule extraction have been discussed. A concept of approximation reduct has been used to describe the smallest attribute subset that preserves the lower approximation and upper approximation of all decision classes in MGRS. Unlike decision rules ("AND" rules) from Pawlak's rough set model, the form of decision rules in MGRS is "OR". Several key attribute reduction algorithms and rule extracting algorithms have been designed as well, which will be helpful for applying this theory to practical issues. The multi-granulation rough set model provides an effective approach for problem solving in the context of multi granulations.

Standard rough set theory and multi-granulation rough set framework are complementary in many practical applications. When two attribute sets in information systems possesses a contradiction or inconsistent relationship, or efficient computation is required, MGRS will display its advantage for rule extraction and knowledge discovery; when there is a consistent relationship between its values under one-attribute set and those under another attribute set, standard rough set theory (SGRS) will hold dominant position. In particular, for some practical applications in which the above two cases occur concurrently, these two concepts can be combined to solve problems.

Further research includes how to evaluate the MGRS method in comparison with Pawlak's original approaches and how to extend other rough set methods in the context of multi granulations.

Acknowledgements

The authors wish to thank the anonymous reviewers for their constructive comments on this study. This work was supported by National Natural Science Fund of China (Nos. 60903110, 60773133, 70971080), High Technology Research and Development Program of China (No. 2007AA01Z165), National Key Basic Research and Development Program of China (973) (No. 2007CB311002) and Natural Science Fund of Shanxi Province, China (Nos. 2009021017-1, 2008011038).

References

[1] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, International Journal of General Systems 17 (1990) 191–209.
[2] I. Düntsch, G. Gediga, Uncertainty measures of rough set prediction, Artificial Intelligence 106 (1998) 109–137.
[3] G. Gediga, I. Düntsch, Rough approximation quality revisited, Artificial Intelligence 132 (2001) 219–234.
[4] J.W. Guan, D.A. Bell, Rough computational methods for information systems, Artificial Intelligence 105 (1998) 77–103.
[5] R. Jensen, Q. Shen, Fuzzy-rough sets assisted attribute selection, IEEE Transactions on Fuzzy Systems 15 (1) (2007) 73–89.
[6] G. Jeon, D. Kim, J. Jeong, Rough sets attributes reduction based expert system in interlaced video sequences, IEEE Transactions on Consumer Electronics 52 (4) (2006) 1348–1355.
[7] A. Skowron, J. Stepaniuk, Tolerance approximation spaces, Fundamenta Informaticae 27 (2-3) (1996) 245–253.
[8] M. Kryszkiewicz, Rough set approach to incomplete information systems, Information Sciences 112 (1998) 39–49.

[9] M. Kryszkiewicz, Rules in incomplete information systems, Information Sciences 113 (1999) 271–292.
[10] Y. Leung, D.Y. Li, Maximal consistent block technique for rule acquisition in incomplete information systems, Information Sciences 153 (2003) 85–106.
[11] J.Y. Liang, C.Y. Dang, K.S. Chin, C.M. Yam Richard, A new method for measuring uncertainty and fuzziness in rough set theory, International Journal of General Systems 31 (4) (2002) 331–342.
[12] J.Y. Liang, D.Y. Li, Uncertainty and Knowledge Acquisition in Information Systems, Science Press, Beijing, China, 2005.
[13] J.Y. Liang, Y.H. Qian, Axiomatic approach of knowledge granulation in information systems, Lecture Notes in Artificial Intelligence 4304 (2006) 1074–1078.
[14] J.M. Ma, W.X. Zhang, Y. Leung, X.X. Song, Granular computing and dual Galois connection, Information Sciences 177 (2007) 5365–5377.
[15] J.S. Mi, W.Z. Wu, W.X. Zhang, Approaches to knowledge reductions based on variable precision rough sets model, Information Sciences 159 (2004) 255–272.
[16] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, System Theory, Knowledge Engineering and Problem Solving, vol. 9, Kluwer, Dordrecht, 1991.
[17] Z. Pawlak, A. Skowron, Rudiments of rough sets, Information Sciences 177 (2007) 3–27.
[18] Y.H. Qian, J.Y. Liang, Rough set method based on multi-granulations, Proceedings of 5th IEEE Conference on Cognitive Informatics 1 (2006) 297–304.
[19] Y.H. Qian, C.Y. Dang, J.Y. Liang, Incomplete multi-granulations rough set, IEEE Transactions on Systems, Man and Cybernetics-Part A, 2009, doi:10.1109/TSMCA.2009.2035436.
[20] Y.H. Qian, J.Y. Liang, D.Y. Li, H.Y. Zhang, C.Y. Dang, Measures for evaluating the decision performance of a decision table in rough set theory, Information Sciences 178 (2008) 181–202.
[21] Y.H. Qian, J.Y. Liang, C.Y. Dang, Converse approximation and rule extraction from decision tables in rough set theory, Computers and Mathematics with Applications 55 (2008) 1754–1765.
[22] Y.H. Qian, J.Y. Liang, C.Y. Dang, Knowledge structure, knowledge granulation and knowledge distance in a knowledge base, International Journal of Approximate Reasoning 50 (1) (2009) 174–188.
[23] Y.H. Qian, C.Y. Dang, J.Y. Liang, H.Y. Zhang, J.M. Ma, On the evaluation of the decision performance of an incomplete decision table, Data& Knowledge Engineering 65 (3) (2008) 373–400.
[24] Y.H. Qian, C.Y. Dang, J.Y. Liang, Consistency measure, inclusion degree and fuzzy measure in decision tables, Fuzzy Sets and Systems 159 (2008) 2353–2377.
[25] H. Rasiowa, Mechanical proof systems for logic of reaching consensus by groups of intelligent agents, International Journal of Approximate Reasoning 5 (4) (1991) 415–432.
[26] H. Rasiowa, V. Marek, Mechanical proof systems for logic II, consensus programs and their processing, Journal of Intelligent Information Systems 2 (2) (1992) 149–164.
[27] C. Rauszer, Distributive knowledge representation systems, Foundations of Computing and Decision Sciences 18 (1993) 307–332.
[28] C. Rauszer, Communication systems in distributed information systems, Proceedings of the Intelligent Information Systems (1993) 15–29.
[29] C. Rauszer, Approximation methods for knowledge representation systems, Lecture Notes in Computer Science 689 (1993) 326–337.
[30] C. Rauszer, Rough logic for multiagent systems, Lecture Notes in Artificial Intelligence 808 (1994) 161–181.
[31] M. Rebolledo, Rough intervals-enhancing intervals for qualitative modeling of technical systems, Artificial Intelligence 170 (2006) 667–685.
[32] Q. Shen, A. Chouchoulas, A rough-fuzzy approach for generating classification rules, Pattern Recognition (35) (2002) 2425–2438.
[33] D. Ślezak, W. Ziarko, The investigation of the Bayesian rough set model, International Journal of Approximate Reasoning 40 (2005) 81–91.
[34] W.Z. Wu, W.X. Zhang, Constructive and axiomatic approaches of fuzzy approximation operators, Information Sciences 159 (2004) 233–254.
[35] W.Z. Wu, J.S. Mi, W.X. Zhang, Generalized fuzzy rough sets, Information Sciences 152 (2003) 263–282.
[36] Z.B. Xu, J.Y. Liang, C.Y. Dang, K.S. Chin, Inclusion degree: a perspective on measures for rough set data analysis, Information Sciences 141 (2002) 227–236.
[37] Y.Y. Yao, Granular computing basis issues and possible solutions, in: Proceedings of the Fifth International Conference on Computing and Information, vol. I, 2000, pp. 186–189.
[38] Y.Y. Yao, Information granulation and rough set approximation, International Journal of Intelligent Systems 16 (2001) 87–104.
[39] Y.Y. Yao, Perspectives of granular computing, in: Proceedings of 2005 IEEE International Conference on Granular Computing, vol. I, 2005, pp. 85–90.
[40] L.A. Zadeh, Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, Fuzzy Sets and Systems 90 (1997) 111–127.
[41] A. Zeng, D. Pan, Q.L. Zheng, H. Peng, Knowledge acquisition based on rough set theory and principal component analysis, IEEE Intelligent Systems (2006) 78–85.
[42] W.X. Zhang, J.S. Mi, W.Z. Wu, Approaches to knowledge reductions in inconsistent systems, International Journal of Intelligent Systems 21 (9) (2003) 989–1000.
[43] W. Ziarko, Variable precision rough sets model, Journal of Computer System Science 46 (1) (1993) 39–59.