Notes on Rough Set Approximations and Associated Measures*

Yiyu Yao
Department of Computer Science, University of Regina
Regina, Saskatchewan, Canada S4S 0A2
E-mail: yyao@cs.uregina.ca
URL: http://www.cs.uregina.ca/∼yyao

Abstract

We review and compare two definitions of rough set approximations. One is defined by a pair of sets in the universe and the other by a pair of sets in the quotient universe. The latter definition, although less studied, is semantically superior for interpreting rule induction and is closely related to granularity switching in granular computing. Numerical measures about the accuracy and quality of approximations are examined. Several semantics difficulties are commented.

1 Introduction

We examine two fundamental issues of the rough set theory, namely, two different definitions of approximations and various measures associated with approximations. Although those issues have been well studied, there still remain a number of semantics difficulties. The main purpose of the notes is to shed some light on semantics issues of approximations and associated measures.

The definition that explicitly represents the composition of equivalence classes is semantically more appropriate for rule induction and is closely related to granularity transformation in granular computing. In measuring rough set approximations, it is essential to separate measures of the accuracy of approximations and measures of the granularity of partitions. Their combination may be useful in the search of optimal partitions for approximations.

*Journal of Zhejiang Ocean University (Natural Science), Vol. 29, No. 5, pp. 399-410, 2010.

2 Two Definitions of Rough Set Approximations

The rough set theory can be formulated based on equivalence relations defined by subsets of attributes in a data table [18]. Formally, a data table is the tuple:

$$S = (U, At, \{V_a \mid a \in At\}, \{I_a \mid a \in At\}), \tag{1}$$

where U is a finite nonempty set of objects, At is a finite nonempty set of attributes, V_a is a nonempty set of values of $a \in At$, and $I_a : U \to V_a$ is an information function that maps an object of U to exactly one value in V_a. Thus, a finite set of objects called a universe is described by a finite set of attributes.

Given a subset of attributes $A \subseteq At$, an indiscernibility relation is defined as:

$$\mathrm{IND}(A) = \{(x, y) \in U \times U \mid \forall a \in A, I_a(x) = I_a(y)\}. \tag{2}$$

That is, x and y are indiscernible with respect to A, if and only if they have exactly the same values on all attributes in A. The relation $\mathrm{IND}(A)$ is reflexive, symmetric and transitive, and hence an equivalence relation. The equivalence class containing x is denoted by $[x]_{\mathrm{IND}(A)} = \{y \in U \mid x\mathrm{IND}(A)y\}$, or simply $[x]_A$ or $[x]$ if A is understood. The equivalence relation induces a partition $U/\mathrm{IND}(A) = \{[x]_{\mathrm{IND}(A)} \mid x \in U\}$ of the universe U, namely, a family of pairwise disjoint subsets of U whose union is U. There is a one-to-one correspondence between equivalence relations on U and partitions of U. One can therefore use the terms "equivalence relation" and "partition" interchangeably.

Definition 1 *Let E denote an equivalence relation with the induced partition U/E. For a subset of objects $X \subseteq U$, Pawlak [18] introduces a pair of lower and upper approximations as follows:*

$$\begin{aligned}
\underline{apr}_E(X) &= \bigcup\{[x]_E \in U/E \mid [x]_E \subseteq X\}, \\
\overline{apr}_E(X) &= \bigcup\{[x]_E \in U/E \mid [x]_E \cap X \neq \emptyset\}.
\end{aligned} \tag{3}$$

The pair $(\underline{apr}_E(X), \overline{apr}_E(X))$ is referred to as the rough set approximation of X.

The rough set approximation $(\underline{apr}_E(X), \overline{apr}_E(X))$ induces a partition of the universe U, namely, $\{\mathrm{POS}_E(X), \mathrm{BND}_E(X), \mathrm{NEG}_E(X)\}$:

$$\begin{aligned}
\mathrm{POS}_E(X) &= \underline{apr}_E(X), \\
\mathrm{BND}_E(X) &= \overline{apr}_E(X) - \underline{apr}_E(X), \\
\mathrm{NEG}_E(X) &= (\overline{apr}_E(X))^c,
\end{aligned} \tag{4}$$

where $(\overline{apr}_E(X))^c = U - \overline{apr}_E(X)$ denotes the complement of the upper approximation. The three blocks of the partition are referred to as the positive, boundary and negative regions of X, respectively. They can be interpreted as follows. Based solely on the description of an object in a data table, an object in the positive region is certainly in X, an object in the negative region is certainly not in X, an object in the boundary region is both possibly in X and not in X. In other words, the positive region and negative region consist of objects whose descriptions allow deterministic decisions regarding their membership in X. The boundary region consists of objects whose descriptions allow non-deterministic decisions regarding their membership in X. Objects with the same description may be either in X and not in X.

By taking the union of equivalence classes, each approximation is a subset of U and the information about the composition of equivalence classes in the approximation becomes hidden and implicit. In order to explicitly express such information, an alternative definition can be given.

Definition 2 *For a subset of object $X \subseteq U$, Bryniarski [2] and Dubois and Prade [3] introduce the following pair,*

$$\begin{aligned} \underline{Apr}_E(X) &= \{[x]_E \in U/E \mid [x]_E \subseteq X\}, \\ \overline{Apr}_E(X) &= \{[x]_E \in U/E \mid [x]_E \cap X \neq \emptyset\}. \end{aligned} \tag{5}$$

as the lower and upper approximations of X in the quotient set U/E.

While $\underline{apr}_E(X)$ and $\overline{apr}_E(X)$ are subsets of U, $\underline{Apr}_E(X)$ and $\overline{Apr}_E(X)$ are subsets of U/E. The two definitions can be transformed into each other by:

$$\begin{aligned} \underline{apr}_E(X) &= \bigcup \underline{Apr}_E(X), \\ \overline{apr}_E(X) &= \bigcup \overline{Apr}_E(X); \\ \underline{Apr}_E(X) &= \{[x]_E \in U/E \mid [x]_E \subseteq \underline{apr}_E(X)\}, \\ \overline{Apr}_E(X) &= \{[x]_E \in U/E \mid [x]_E \subseteq \overline{apr}_E(X)\}. \end{aligned} \tag{6}$$

Thus, they may be considered to be equivalent [2]. On the other hand, their semantics differences have not received due attention. Although the first definition has been used in the mainstream of rough set research, the second definition is more appropriate for rule induction and is semantically superior.

In rough set based rule induction approaches [7, 10, 20, 25, 31, 33], an equivalence class in the lower/upper approximation is typically used to induce a rule whose left-hand-side is the description of the equivalence class and whose right-hand-side is the concept represented by X. Definition 2 explicitly provides the composition information about equivalence classes in the lower/upper approximation, and hence is semantically easier to interpret a rule induction process. Moreover, such composition information is useful in evaluating the induced rules. Based on Definition 2, one can similarly derive a partition of the quotient universe U/E, consisting of the positive region, the boundary region and the negative region, respectively. The physical meaning of the partition is much clearer. An equivalence class in the positive region is a subset of X, an equivalence class in the negative region is a subset of X^c, and an equivalence class in the boundary region is neither a subset of X nor a subset of X^c.

The quotient universe U/E may be viewed as a granulation or coarsening of U with each equivalence class being a coarse-grained granule. Conversely,

U may be viewed as a refinement of U/E. The equivalence relation E bridges the two universes. Definition 2 clearly shows the switching from a fine-grained ground universe U to a coarse-grained granulated universe U/E. Specifically, a subset of objects in a fine-granied universe can only be approximately defined in a coarse-grained universe [3]. On the other hand, a subset of coarse-grained universe, $Y \subseteq U/E$, can be precisely defined by a subset $\omega(Y) \subseteq U$ of the fine-grained universe as follows [3]:

$$\omega(Y) = \bigcup_{Y_i \in Y} Y_i. \tag{7}$$

An equivalence class Y_i plays dual roles. It is a subset of U, as used in the term for union, and an element of U/E, as used in the subscript of the union. By Definition 2, one can explicitly study the switching between two universes with differing granularity. This important notion makes a close connection between rough set theory and the new emerging field of granular computing [30].

Example 1 *Table 1 is an example of a data table, taken from an example from Quinlan [23]. Each object is described by three attributes. The column labeled by Class denotes an expert's classification of the objects. We use this simple example to illustrate the basic ideas discussed.*

For two subsets of attributes, {Eyes} *and* {Height, Eyes}*, they induce the following two partitions:*

$$\begin{aligned}
\pi_1 &= U/\mathrm{IND}(\{\text{Eyes}\}) = \{\{o_1, o_3, o_4, o_5, o_6\}, \{o_2, o_7, o_8\}\};\\
\pi_2 &= U/\mathrm{IND}(\{\text{Height, Eyes}\}) = \{\{o_1\}, \{o_2, o_8\}, \{o_3, o_4, o_5, o_6\}, \{o_7\}\}.
\end{aligned}$$

Consider the class $- = \{o_2, o_4, o_5, o_7, o_8\}$*, we have:*

$$\begin{aligned}
\underline{apr}_{\pi_1}(-) &= \{o_2, o_7, o_8\},\\
\overline{apr}_{\pi_1}(-) &= U;\\
\underline{Apr}_{\pi_1}(-) &= \{\{o_2, o_7, o_8\}\},\\
\overline{Apr}_{\pi_1}(-) &= \{U\};
\end{aligned}$$

and

$$\begin{aligned}
\underline{apr}_{\pi_2}(-) &= \{o_2, o_7, o_8\},\\
\overline{apr}_{\pi_2}(-) &= \{o_2, o_3, o_4, o_5, o_6, o_7, o_8\};\\
\underline{Apr}_{\pi_2}(-) &= \{\{o_2\}, \{o_7, o_8\}\},\\
\overline{Apr}_{\pi_2}(-) &= \{\{o_2, o_8\}, \{o_3, o_4, o_5, o_6\}, \{o_7\}\}.
\end{aligned}$$

For the two partitions, we have the same lower approximations according to Definition 1, but different lower approximations according to Definition 2. Approximation $\underline{Apr}_{\pi_1}$ *can be used to induce one rule for* $-$*,*

$$(\text{Eyes} = \text{brown}) \longrightarrow \text{Class} = -,$$

Object	Height	Hair	Eyes	Class
o_1	short	blond	blue	+
o_2	short	blond	brown	-
o_3	tall	red	blue	+
o_4	tall	dark	blue	-
o_5	tall	dark	blue	-
o_6	tall	blond	blue	+
o_7	tall	dark	brown	-
o_8	short	blond	brown	-

Table 1: A data table

and $\underline{Apr}_{\pi_2}$ *can be used to induce two rules,*

$$(\text{Height} = \text{short}, \text{Eyes} = \text{brown}) \longrightarrow \text{Class} = -,$$
$$(\text{Height} = \text{tall}, \text{Eyes} = \text{brown}) \longrightarrow \text{Class} = -.$$

The composition of equivalence classes provides useful structural information about rough set approximations for rule induction. In fact, the last two rules can be generalized into the first rule.

3 Measures Associated with Rough Set Approximations

Two classes of measures are investigated. One class concerns the completeness of knowledge provided by an equivalence relation for approximating a set X based on Definition 1. The other class concerns the granularity of knowledge provided by an equivalence relation.

3.1 Accuracy and roughness measures

Pawlak [18, 19] suggests two numerical measures for characterizing the imprecision of rough set approximations. The *accuracy measure* is defined by:

$$\alpha_E(X) = \frac{|\underline{apr}_E(X)|}{|\overline{apr}_E(X)|}, \tag{8}$$

where $X \neq \emptyset$ and $|\cdot|$ denotes the cardinality of a set. For the empty set $\emptyset$, we define $\alpha_E(\emptyset) = 1$. The It follows that $0 \leq \alpha_E(X) \leq 1$. Based on the accuracy measure, the *roughness measure* is defined by:

$$\rho_E(X) = 1 - \alpha_E(X). \tag{9}$$

The measure of roughness is related to the notion of boundary region as follows:

$$\rho_E(X) = \frac{|\overline{apr}_E(X) - \underline{apr}_E(X)|}{|\overline{apr}_E(X)|} = \frac{|\mathrm{BND}_E(X)|}{|\overline{apr}_E(X)|}. \tag{10}$$

Yao [28] shows that the roughness measure is in fact the well-known Marczewski-Steinhaus distance between the lower and upper approximations.

Regarding the meaningfulness and usefulness of the two measures, several crucial questions need to be answered. What are the physical interpretations of the two measures? Do the measures semantically quantify the properties to be measured? If they do not, what are new measures? Before attempting answer these questions, we first comment on the following properties satisfied by the measure $\alpha_E(\cdot)$:

$$\begin{array}{rl}
\text{(i).} & \alpha_E(X) = 1 \Longleftrightarrow \underline{apr}_E(X) = \overline{apr}_E(X), \\
\text{(ii).} & \alpha_E(X) = 0 \Longleftrightarrow \underline{apr}_E(X) = \emptyset, \\
\text{(iii).} & \text{for a fixed } \overline{apr}_E(X), \\
 & \alpha_E(X) \text{ strictly monotonically increases with } |\underline{apr}_E(X)|, \\
\text{(iv).} & \text{for a fixed } \underline{apr}_E(X) \neq \emptyset, \\
 & \alpha_E(X) \text{ strictly monotonically decreases with } |\overline{apr}_E(X)|, \\
\text{(v).} & E_1 \subseteq E_2 \Longrightarrow \alpha_{E_1}(X) \geq \alpha_{E_2}(X).
\end{array}$$

Some of these properties have been studied by many authors, for example, Gediga and Düntsch [6], Yao [29], Huynh and Nakamori [8], and Zhu [36]. Similarly properties can be given for the roughness measure $\rho_E(\cdot)$.

According to Pawlak, the accuracy measure is "intended to capture the degree of completeness of our knowledge about the set X" [19] or to "express the 'quality' of an approximation" [18]; as opposed to accuracy, the roughness measures "represents the degree of incompleteness" [19]. Unfortunately, Pawlak does not explicitly provide a definition or an interpretation for the concepts of "accuracy", "completeness", "quality" and "roughness." Consequently, they have been interpreted and used differently by many authors. If we were to interpret the accuracy and quality of approximation as a measure of the "completeness" of knowledge about X based on E and roughness as a measure of "incompleteness," some of the properties (i)-(vi) may not be entirely meaningful.

With respect to an equivalence relation E, our knowledge about X is characterized by two sets, i.e., $\underline{apr}_E(X)$ and $\overline{apr}_E(X)$, or equivalently the partition $\{\mathrm{POS}_E(X), \mathrm{BND}_E(X), \mathrm{NEG}_E(X)\}$. The use of both $\underline{apr}_E(X)$ and $\overline{apr}_E(X)$ in the measure $\alpha_E(\cdot)$ provides a good starting point. Consider first property (i), which states that the measure reaches the maximum value 1 when the lower and upper approximations are the same as X. This property is meaningful, as the state is in fact associated with the complete knowledge about X provided by the equivalence relation E. Property (ii) states that the accuracy measure reaches the minimum value 0 when the lower approximation of X is the empty set $\emptyset$, independent of the upper approximation of X. In some sense, this is not reasonable, as we know for certain that elements in $(\overline{apr}_E(X))^c$ are definitely not in X. In other words, we still have some information about X based on its upper approximation. For a state associated with no knowledge about X, it is necessary to also require that $\overline{apr}_E(X) = U$. A more reasonable property can be stated as:

$$\text{(ii}'\text{)}. \qquad \alpha_E(X) = 0 \Longleftrightarrow (\underline{apr}_E(X) = \emptyset, \overline{apr}_E(X) = U).$$

The monotonicity properties (iii) and (iv) are reasonable, as our knowledge about X becomes more accurate or complete with the increase of the size of $\underline{apr}_E(X)$ or a the decrease of the size of $\overline{apr}_E(X)$. However, property (iv) suffers from the same difficulty as that of property (ii). The monotonicity should hold when $\underline{apr}_E(X) = \emptyset$. The revised property is therefore given as:

(iv′). for a fixed $\underline{apr}_E(X)$,
$\alpha_E(X)$ strictly monotonically decreases with $|\overline{apr}_E(X)|$.

This new property seems to be more reasonable, as $\underline{apr}_E(X) = \emptyset$ is not treated differently.

Property (v) explicitly considers approximations derived from different equivalence relations. Given two equivalence relations E_1 and E_2 with $E_1 \subseteq E_2$, E_1 would produce a finer partition U/E_1 than E_2. It follows that

$$E_1 \subseteq E_2 \Longrightarrow \underline{apr}_{E_2}(X) \subseteq \underline{apr}_{E_1}(X) \subseteq X \subseteq \overline{apr}_{E_1}(X) \subseteq \overline{apr}_{E_2}(X). \qquad (11)$$

That is, E_1 produces a pair of tighter approximations than E_2. A pair of tighter approximations should provide more knowledge about the set X, and hence a large value of the measure $\alpha_E(\cdot)$. Property (v) reflects exactly this point. The accuracy measure $\alpha_E(\cdot)$ may be used to evaluate the "goodness" or "fitness" of equivalence relations in approximating X.

Properties (i), (ii′), (iii), (iv′), and (v) may be viewed as a set of axioms for a measure of the "completeness" of knowledge, or the accuracy of approximations. Another set of axioms is given by Zhu [36]. The original Pawlak accuracy measure does not satisfy properties (ii′) and (iv′). It is necessary to modify the Pawlak accuracy measure or to design new measures based on such properties.

Intuitively speaking, our knowledge about the set X depends on the set of objects that we can make deterministic decisions about their membership. Recall that we can make such decisions for objects in both $\mathrm{POS}_E(X)$ and $\mathrm{NEG}_E(X)$ and in a state of complete knowledge we can make deterministic decisions for all objects in U. The ratio of the sizes of these two may serve as a good measure of the accuracy of X. This suggests the following measure,

$$\begin{aligned}\gamma_E(X) &= \frac{|\mathrm{POS}_E(X)| + |\mathrm{NEG}_E(X)|}{|U|} \\ &= \frac{|\underline{apr}_E(X)| + |(\overline{apr}_E(X))^c|}{|U|} \\ &= \frac{|\underline{apr}_E(X)| + |\underline{apr}_E(X^c)|}{|U|}. \end{aligned} \tag{12}$$

The measure is in fact the measure of quality of approximation of the X-generated partition $\{X, X^c\}$ by E, which is proposed by Pawlak [19]. The new measure $\gamma_E(\cdot)$ satisfies the required properties (i), (ii′), (iii), (iv′), and (v), which is a desired measure of knowledge completeness about X provided by E. The corresponding measure of roughness is given by

$$\beta_E(X) = 1 - \gamma_E(X) = \frac{|\mathrm{BND}_E(X)|}{|U|} = \frac{|\overline{apr}_E(X) - \underline{apr}_E(X)|}{|U|}. \tag{13}$$

Compared with the Pawlak roughness measure $\rho_E(\cdot)$, the denominator in $\beta_E(\cdot)$ is independent of the upper approximation of X.

3.2 Measures of granularity

The measures of completeness of knowledge or the accuracy of approximations are based on rough set approximations as subsets of the universe. The information regarding the composition of equivalence classes in the approximations is not considered. In other words, a measure of accuracy only reflects one aspect of the quality or feature of the approximation. This has led to many criticisms and modifications of accuracy measures [1, 15, 26, 36].

Example 2 *Beaubouef et al. [1] give an example to illustrate the needs for considering the composition of equivalence classes in rough set approximations. Let* $U = \{1, 2, 3, 4, 5, 6, 8, 9\}$. *Suppose we have three equivalence relations* E_1, E_2 *and* E_3 *with the associated partitions:*

$$\begin{aligned} U/E_1 &= \{\{1, 2, 3, 4\}, \{5, 6, 7\}, \{8, 9\}\}, \\ U/E_2 &= \{\{1, 2\}, \{3, 4\}, \{5, 6, 7\}, \{8, 9\}\}, \\ U/E_3 &= \{\{1\}, \{2\}, \{3\}, \{4\}, \{5, 6, 7\}, \{8, 9\}\}. \end{aligned}$$

Equivalence relation E_1 *produces the coarsest partition and* E_3 *the finest partition. Consider a set* $X = \{1, 2, 3, 4, 5, 8\}$. *According to Definition 1, they give the same approximations for* X:

$$\begin{aligned} &\underline{apr}_{E_1}(X) = \underline{apr}_{E_2}(X) = \underline{apr}_{E_3}(X) = \{1, 2, 3, 4\}, \\ &\overline{apr}_{E_1}(X) = \overline{apr}_{E_2}(X) = \overline{apr}_{E_3}(X) = U. \end{aligned}$$

The information about the composition of equivalence classes is not explicitly shown. Under the Pawlak accuracy measure $\alpha_E(\cdot)$ and the new accuracy measure $\gamma_E(\cdot)$, we would have the same values for the three pairs of approximations. The granularity of partition is not reflected.

According to Definition 2, we have three different pairs of approximations:

$$\begin{aligned}
\underline{Apr}_{E_1}(X) &= \{\{1,2,3,4\}\},\\
\overline{Apr}_{E_1}(X) &= \{\{1,2,3,4\},\{5,6,7\},\{8,9\}\};\\
\underline{Apr}_{E_2}(X) &= \{\{1,2\},\{3,4\}\},\\
\overline{Apr}_{E_2}(X) &= \{\{1,2\},\{3,4\},\{5,6,7\},\{8,9\}\};\\
\underline{Apr}_{E_3}(X) &= \{\{1\},\{2\},\{3\},\{4\}\},\\
\overline{Apr}_{E_3}(X) &= \{\{1\},\{2\},\{3\},\{4\},\{5,6,7\},\{8,9\}\}.
\end{aligned}$$

The granularity information of different partitions becomes explicit in the approximations.

The example shows that Definition 2 enables us to express the granularity information about rough set approximations. A question reminds is how to design a measure to quantify granularity. Measures of granularity of a partition have been investigate by many authors, for example, Beaubouef *et al.* [1], Düntsch and Gediga [4], Wierman [24], Miao and Fan [17], Liang *et al.* [11, 14, 15], Yao [29], Liang and Shi [13], Liang and Qian [12], Mi *et al.* [16], Qian and Liang [21], Qian *et al.* [22], Xu *et al.* [27], Zhu [35], Feng *et al.* [5] and many others. A survey of measures of granularity of partitions and a general form of measures, in terms of the expected values of the granularity of blocks in a partition, are provided by Yao and Zhao [32]. As an example, we consider only an information-theoretic measure of granularity [1, 4, 24, 29].

With respect to a partition $\pi = \{X_1, X_2, \ldots, X_m\}$, we have a probability distribution:

$$P_\pi = \left(\frac{|X_1|}{|U|}, \frac{|X_2|}{|U|}, \ldots, \frac{|X_m|}{|U|}\right). \tag{14}$$

The Shannon entropy function of the probability distribution is defined by:

$$H(\pi) = H(P_\pi) = -\sum_{i=1}^{m} \frac{|X_i|}{|U|} \log \frac{|X_i|}{|U|}. \tag{15}$$

The entropy reaches the maximum value $\log |U|$ for the finest partition consisting of singleton subsets of U, and it reaches the minimum value 0 for the coarsest partition $\{U\}$. For two partitions with $\pi_1 \preceq \pi_2$, namely, π_1 is finer than or the same as π_2, we have $H(\pi_1) \geq H(\pi_2)$. That is, the value of the entropy correctly reflects the order of partitions with respect to their granularity.

We can re-express equation (15) as,

$$H(\pi) = \log |U| - \sum_{i=1}^{m} \frac{|X_i|}{|U|} \log |X_i|. \tag{16}$$

The first term is a constant independent of any partition. The quantity $\log |X_i|$ is commonly known as the Hartley measure of information of the set X_i. It has been used to measure the amount of uncertainty associated with a finite set of possible alternatives, namely, the nonspecificity inherent in the set [9]. The function $\log |X_i|$ is a monotonic increasing transformation of the size of a set. It may be used to measure the granularity of the set. Large sets result in higher degrees of granularity than small sets. The second term of the equation is basically an expectation of granularity with respect to all subsets in a partition. It follows that we can use the following function as a measure of granularity for a partition:

$$G(\pi) = \sum_{i=1}^{m} \frac{|X_i|}{|U|} \log |X_i|. \tag{17}$$

In contrast to the entropy function, for two partitions π_1 and π_2 with $\pi_1 \preceq \pi_2$, we have $G(\pi_1) \leq G(\pi_2)$. The coarsest partition $\{U\}$ has the maximum granularity value $\log |U|$, and the finest partition $\{\{x\} \mid x \in U\}$ has the minimum granularity value 0.

With respect to the two definitions of approximation, if an approximation is a non-empty set, say $\underline{apr}(X) \neq \emptyset$, then $\underline{Apr}(X)$ is a partition of $\underline{apr}(X)$. For example, in Example 2, we have $\overline{\underline{apr}_{E_2}(X)} = \{1, 2, 3, 4\}$ and $\overline{\underline{Apr}_{E_2}(X)} = \{\{1, 2\}, \{3, 4\}\}$. The latter is a partition of the former. This observation immediately suggests that one can apply the measure of granularity to an approximation defined by Definition 2, namely, $G(\underline{Apr}_E(X))$ and $G(\overline{Apr}_E(X))$. The new measure reflects the granularity of an approximation.

3.3 Composite measures of quality of approximations

In the last two subsections, we have shown that an accuracy measure can be defined based on $\underline{apr}$ and $\overline{apr}$ to reflect the completeness of knowledge and a granularity measure can be defined based on $\underline{Apr}$ and $\overline{Apr}$ to reflect the knowledge granularity. They capture entirely different aspects of the "quality" of approximations. Several authors attempt to combine them together, but still name a combined measure as a measure of roughness. It is perhaps better to keep the interpretation that a roughness measure is an inverse measure of an accuracy measure, in the sense that the summation of the two is 1. As to the composite measures, new names may be introduced. In this paper, we call a combined measure as a measure of "fitness" of an approximation, with the intended meaning that it reflects how good an approximation is.

A fitness measure is a composite measure of accuracy and granularity. There are several ways to do this. One simple method to take their product, namely,

$$F(\cdot) = \gamma(\cdot) * G(\cdot), \tag{18}$$

where F denotes a measure of fitness, γ denotes a measure of accuracy, and G denotes a measure of granularity. The use of a composite measure implies a kind of trade-off between accuracy and granularity. It should be realized that a composite measure is not always meaningful, although it has an advantage of a single number. In practice, we may be better off with a pair of numbers provided by two measures.

4 Conclusion

By examining two fundamental issues of the rough set theory, one is the rough set approximations and the other is the associated measures of approximations, we are able to report several new findings. Approximations, as subsets of the universe, are a standard definition, but fail to reflect important features of rough set theory. In contrast, approximations, as subsets of the quotient universe, are semantically superior. They connect naturally to rule induction with rough sets and granularity switching in granular computing. "Quality" of approximations can be studied and quantified in several ways, with each representing a certain aspect of approximations. An accuracy measure quantifies the completeness of knowledge about a set; a measure of granularity quantifies the grain-sizes in the knowledge. We must carefully study different types of measures, in order to ensure that we actually measure what we intended to measure.

References

[1] Beaubouef, T., Petry, F.E. and Arora, G. Information-theoretic measures of uncertainty for rough sets and rough relational databases, Information Sciences 109, 185-195, 1998.

[2] Bryniarski, E. A calculus of rough sets of the first order, Bulletin of the Polish Academy of Science, Mathematics 16, 71-78, 1989.

[3] Dubois, D. and Prade, H. Rough fuzzy sets and fuzzy rough sets, International Journal of General Systems 17, 191209, 1990.

[4] Düntsch, I. and Gediga, G. Uncertainty measures of rough set prediction, Artificial Intelligence 106 (1998) 109-137.

[5] Feng, Q.R., Miao, D.Q., Zhou, J. and Cheng, Y. A novel measure of knowledge granularity in rough sets, International Journal of Granular Computing, Rough Sets and Intelligent Systems 1, 233-251, 2010.

[6] Gediga, G. and Düntsch, I. Rough approximation quality revisited, Artificial Intelligence 132, 219-234, 2001.

[7] Grzymala-Busse, J.W. Knowledge acquisition under uncertainty – A rough set approach, Journal of Intelligent & Robotic Systems 1, 3-16, 1988.

[8] Huynh, V.N. and Nakamori, Y. A roughness measure for fuzzy sets, Information Sciences 173, 255-275, 2005.

[9] Klir, G.J. and Golger, T.A. Fuzzy sets, Uncertainty, and Information, Prentice Hall, Englewood Cliffs, 1988.

[10] Leung, Y., Wu, W.Z. and Zhang, W.X. Knowledge acquisition in incomplete information systems: a rough set approach, European Journal of Operational Research 168, 164-180, 2006.

[11] Liang, J.Y., Chin, K,S., Dang, C.Y. and Yam, R.C.M. A new method for measuring uncertainty and fuzziness in rough set theory, International Journal of General Systems 31, 331-342, 2002.

[12] Liang, J.Y. and Qian, Y.H. Axiomatic approach of knowledge granulation in information system, in: Proceedings of AI 2006, LNAI 4304, pp. 1074-1078.

[13] Liang, J.Y. and Shi, Z. The information entropy, rough entropy and knowledge granulation in rough set theory, International Journal of Uncertainty, Fuzziness and Knowledge-Based System 12, 37-46, 2004.

[14] Liang, J.Y., Shi, Z., Li, D. and Wierman, M.J. Information entropy, rough entropy and knowledge granulation in incomplete information systems, International Journal of General Systems 35, 641-654, 2006.

[15] Liang, J.Y., Wang, J.H. and Qian, Y.H. A new measure of uncertainty based on knowledge granulation for rough sets, Information Sciences 179, 458-470, 2009.

[16] Mi, J.S., Li, X.M., Zhao, H.Y. and Feng, T. Information-theoretic measure of uncertainty in generalized fuzzy rough sets, in: Proceedings of RSFDGrC 2007, LNAI 4482, pp. 63-70.

[17] Miao, D.Q. and Fan, S.D. The calculation of knowledge granulation and its application (in Chinese), Systems Engineering Theory and Practice 22, 48-56, 2002.

[18] Pawlak, Z. Rough sets, International Journal of Computer and Information Sciences 11, 341-356, 1982.

[19] Pawlak, Z. Rough Sets, Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Dordrecht, 1991.

[20] Pawlak, Z., Wong, S.K.M. and Ziarko, W. Rough sets: probabilistic versus deterministic approach, International Journal of Man-Machine Studies 29, 81-95, 1988.

[21] Qian, Y.H. and Liang, J.Y. Combination entropy and combination granulation in rough set theory, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 16, 179-193, 2008.

[22] Qian, Y.H., Liang, J.Y. and Dang, C.Y. Knowledge structure, knowledge granulation and knowledge distance in a knowledge base, International Journal of Approximate Reasoning 50, 174-188, 2009.

[23] Quinlan, J.R. Learning efficient classification procedures and their application to chess endgames, in: Michalski, J.S., Carbonell, J.G., and Michell, T.M. (Eds), Machine Learning: An Artificial Intelligence Approach, vol. 1, Morgan Kaufmann, Palo Alto, CA, 463-482, 1983.

[24] Wierman, M.J. Measuring uncertainty in rough set theory, International Journal of General Systems 28, 283-297, 1999.

[25] Wong, S.K.M. and Ziarko, W. Algorithm for inductive learning, Bulletin of the Polish Academy of Sciences, Technical Sciences 34, 271-276, 1986.

[26] Xu, B.W., Zhou, Y.M. and Lu, H.M. An improved accuracy measure for rough sets, Journal of Computer and System Sciences 71, 163-173, 2005.

[27] Xu, W.H., Zhang, X.Y. and Zhang, W.X. Knowledge granulation, knowledge entropy an knowledge uncertainty measure in ordered information systems, Applied Soft Computing 9, 1244-1251, 2009.

[28] Yao, Y.Y., Information granulation and rough set approximation, International Journal of Intelligent Systems 16, 87-104, 2001.

[29] Yao, Y.Y. Probabilistic approaches to rough sets, Expert Systems 20, 287-297, 2003.

[30] Yao, Y.Y. A partition model of granular computing, LNCS Transactions on Rough Sets I, LNCS 3100, 232-253, 2004.

[31] Yao, Y.Y. Three-way decisions with probabilistic rough sets, Information Sciences 180, 341-353, 2010.

[32] Yao, Y.Y. and Zhao, L.Q. Granularity of partitions, manuscript, 2010.

[33] Zhang, W.X., Leung, Y. and Wu, W.Z. Information System and Knowledge Discovery (in Chinese), Academic Press, Beijing, 2003.

[34] Zhao, M.Q., Yang, Q. and Gao, D.Z. Axiomatic definition of knowledge granularity and its construction method, in: Proceedings of RSKT 2008, LNAI 5009, pp. 348-354.

[35] Zhu, P. An improved axiomatic definition of information granulation, preprint, http://arxiv.org/abs/0908.3999, 2009.

[36] Zhu, P. An axiomatic approach to the roughness measure of rough sets, Fundamenta Informaticae, to appear.