VPT-OMGRS: Variable Precision Optimistic Multigranulation Rough Set Based on Tolerance Relations

Sheng Yao*, Longshu Li and Yi Xu

Dept. of Computer Science and Technology, University of AnHui University, TK 230601-China

Received 15 September 2013; Accepted 23 November 2013

Abstract

In recent years, multigranulation rough set has been an emerging direction in rough set theory, which uses multiple structures. To study the incomplete information system under multigranulation environment, this paper first proposes the variable precision optimistic multigranulation rough set based on tolerance relations combining the advantages of tolerance relation and variable precision rough set. Then, three elementary measures are introduced for the new model, and the properties of the new model and the measures are discussed. Finally, some experimental results show the variation of the three elementary measures with the value threshold.

Keywords: Tolerance Relation, Variable Precision Rough Set, Multigranulation

1. Introduction

Rough set theory [1,3], introduced by Pawlak, has become a useful tool to deal with various types of data. It has often applied in artificial intelligence [11], machine learning, pattern recognition, granular computing image processing, decision support and knowledge discovery.

In the past, most rough set model and their extensions were constructed on one granulation structure, however, in fact, viewing a problem only in one granulation structure is one-sided, for example, when making decisions, different decision makers should be independent, so doing the interaction of their decisions is not in accordance with actual needs. To solve the above problem, Qian et al. introduced multigranulation rough set theory (MGRS) [2], in which the set approximations are defined by multiple equivalence relations on a universe.

Recently, extensions to multigranulation rough set have been widely proposed. Zhang [4] et al. defined a variable precision multigranulation rough set, in which the optimistic multigranulation rough sets can be considered as two extreme cases. Lin [5] et al. proposed a covering-based pessimistic multigranulation rough set. Qian [6] et al. continue to establish a rough-set model based on multiple tolerance relations in incomplete information.

In this paper, our objective is to develop a new multigranulation rough set through combining the idea of the variable precision and the multigranulation, called the variable precision optimistic multigranulation rough set based on tolerance relations. In this new model, due to the control of the value threshold, it broaden the strict definition of the boundary region, and it is more suitability to deal with incomplete and noisy data.

This study is organized as follows. Some basic concepts such as multigranulation rough set, tolerance relation and variable precision rough set are briefly reviewed in Section 2. In Section 3, the variable precision optimistic multigranulation rough set based on tolerance relations is presented, some of its important properties are investigated. In Section 4, three elementary measures are introduced for the new model, some of the important properties of the three measures are discussed. The variation of the three elementary measures with the value threshold are shown through experiment in Section 5. Finally, Section 6 concludes this paper by giving some discussions.

2. Preliminary

In this section, we review some basic concepts. In this full paper, we suppose the universe of discourse U is a finite non-empty set.

Definition 1. [2] Let $I = (U, AT, V, f)$ be an information system in which $A_i, A_j \subseteq AT$ and $X \subseteq U$. The optimistic multigranulation lower and upper approximations are denoted by $\sum_{i=1}^{n} A_i(X)$ and $\sum_{i=1}^{n} A_i^O(X)$, respectively,

$$\sum_{i=1}^{n} A_i^O(X) = \{x \in U : \forall x \in X \text{ s.t. } x \in L \cup \bigvee_{i=1}^{n} [X]_{A_i} \subseteq X\}$$
3. Variable Precision Optimistic Multigranulation Rough Set Based on Tolerance Relations

Definition 6. Let \( S = (U, \mathcal{R}, f) \) be an incomplete information system in which \( R_1, R_2, L, R_m \subseteq \mathcal{R} \) be m-tolerance relations on \( U \), then \( \forall X \subseteq U \) and \( \beta \in [0, 0.5] \), the variable precision optimistic multigranulation lower approximation of \( X \) based on tolerance relations and
Theorem 1 says that variable precision optimistic multigranulation upper approximation of $X$ based on tolerance relations is a collection of elements, in which the relative degree of misclassification of its tolerance class with respect to the approximate target is less than $1 - \beta$.

Theorem 2. Let $S = (U, \mathcal{R}, f)$ be an incomplete information system in which $R_1, R_2, L, R_m \subseteq \mathcal{R}$ be $m$-tolerance relations on $U$, then $\forall X \subseteq U$ and $\beta \in [0, 0.5)$, the following properties hold.

\[
\sum_{i=1}^{m} R_i^{\beta} (X) = \bigcup_{i=1}^{m} R_i^{T\beta} (X) \tag{1}
\]

\[
\sum_{i=1}^{m} R_i^{\beta} (X) = \bigcap_{i=1}^{m} R_i^{T\beta} (X) \tag{2}
\]

Proof.
(1) $\forall x \in \sum_{i=1}^{m} R_i^{\beta} (X)$, by Definition 6, we have $c(T_{R_i}(x), X) \leq \beta \lor c(T_{R_i}(x), X) \leq \beta \lor c(T_{R_i}(x), X) \leq \beta$. It follows that $x \in R_i^{T\beta} (X) \lor R_i^{T\beta} (X) \lor R_i^{T\beta} (X)$, i.e. $x \in \bigcup_{i=1}^{m} R_i^{T\beta} (X)$.

Conversely, $\forall x \in \bigcup_{i=1}^{m} R_i^{T\beta} (X)$, we have $x \in R_i^{T\beta} (X) \lor R_i^{T\beta} (X) \lor R_i^{T\beta} (X)$, then according to Definition 6, $x \in \sum_{i=1}^{m} R_i^{\beta} (X)$.

From the discussions above, it follows that

\[
\sum_{i=1}^{m} R_i^{\beta} (X) = \bigcup_{i=1}^{m} R_i^{T\beta} (X).
\]

(2) $\forall x \in \sum_{i=1}^{m} R_i^{\beta} (X)$, by (1) from Theorem 1, we have $c(T_{R_i}(x), X) \leq 1 - \beta \lor c(T_{R_i}(x), X) < 1 - \beta$, thus $x \in R_i^{T\beta} (X) \lor R_i^{T\beta} (X) \lor R_i^{T\beta} (X)$, i.e. $x \in \bigcap_{i=1}^{m} R_i^{T\beta} (X)$.

Conversely, $\forall x \in \bigcap_{i=1}^{m} R_i^{T\beta} (X)$, thus $x \in R_i^{T\beta} (X) \lor R_i^{T\beta} (X) \lor R_i^{T\beta} (X)$, then according to (1) of Theorem 1, $x \in \sum_{i=1}^{m} R_i^{\beta} (X)$.

From discussions above, it follows that

\[
\sum_{i=1}^{m} R_i^{\beta} (X) = \bigcap_{i=1}^{m} R_i^{T\beta} (X).
\]

In Theorem 2, (1) says that variable precision optimistic multigranulation lower approximation of $X$ based on tolerance relations can be viewed as the union of all the single variable precision lower approximations based on tolerance relations, (2) says that variable precision optimistic multigranulation upper approximation of $X$ based on tolerance relations can be viewed as the intersection of all the single variable precision upper approximations based on tolerance relations.

Directly from Definition 6, we also obtain the following properties of the variable precision optimistic multigranulation lower and upper approximations of $X$ based on tolerance relations.

Proposition 1. Let $S = (U, \mathcal{R}, f)$ be an incomplete information system in which $R_1, R_2, L, R_m \subseteq \mathcal{R}$ be $m$-tolerance relations on $U$, then $\forall X \subseteq U$ and $\beta \in [0, 0.5)$, the variable precision optimistic multigranulation lower and upper approximations of $X$ based on tolerance relations have the following properties:

\[
(1) \sum_{i=1}^{m} R_i^{\beta} (U) = \bigcup_{i=1}^{m} R_i^{T\beta} (U) \tag{1}
\]

\[
(2) \sum_{i=1}^{m} R_i^{\beta} (\phi) = \bigcap_{i=1}^{m} R_i^{T\beta} (\phi) \tag{2}
\]

\[
(3) \beta_1 \leq \beta_2 \Rightarrow \sum_{i=1}^{m} R_i^{\beta_1} (X) \subseteq \sum_{i=1}^{m} R_i^{\beta_2} (X) \tag{3}
\]
(4) $\beta_1 \leq \beta_2 \Rightarrow \sum_{i=1}^{m} R_i (X) \subseteq \sum_{i=1}^{m} R_i (X)$

(5) $\sum_{i=1}^{m} R_i (X) = \sum_{i=1}^{m} R_i (X)$

(6) $\sum_{i=1}^{m} R_i (X) = \sum_{i=1}^{m} R_i (X)$

Proof:

(1a) $\forall i \in \{1,2,L,m\}$, we have $c(T(x),U)=1-\frac{|T(x)|}{|T(x)|}$.

(1b) Since $0 \leq \beta < 0.5$, then $c(T(x),U) \leq 1$.

By Definition 6, it follows that $\sum_{i=1}^{m} R_i (U) = U$.

(2a) By Definition 6, $\forall i \in \{1,2,L,m\}$, we have $c(T(x),\phi) > \beta$.

Since $0 \leq \beta < 0.5$, then $c(T(x),\phi) > \beta$. Hence, $\sum_{i=1}^{m} R_i (\phi) = \phi$.

(2b) Since $0 \leq \beta < 0.5$, then $0.5 < 1 - \beta \leq 1$, then by the proof of (2a), $\forall i \in \{1,2,L,m\}$, we have $c(T(x),\phi) > \beta$.

Hence, $\sum_{i=1}^{m} R_i (\phi) = \phi$.

(3) $\forall X \subseteq \sum_{i=1}^{m} R_i (X)$, there must be $i \in \{1,2,L,m\}$ such that $c(T(x),X) \leq \beta_1$, since $\beta_1 \leq \beta_2$, then we have $c(T(x),X) \leq \beta_2$. i.e., $x \in \sum_{i=1}^{m} R_i (X)$. Hence, $\sum_{i=1}^{m} R_i (X) \subseteq \sum_{i=1}^{m} R_i (X)$.

(4) $\forall x \subseteq \sum_{i=1}^{m} R_i (X)$, then according to (1) of Theorem 1, $\forall i \in \{1,2,L,m\}$, we have $c(T(x),X) \leq 1 - \beta_2$, then since $\beta_1 \leq \beta_2$, we have $1 - \beta_1 \leq 1 - \beta_2$, hence, $c(T(x),X) \leq 1 - \beta_1$.

(5) From Definition 6, let $X \sim X$, we have $\sum_{i=1}^{m} R_i (X) = \sum_{i=1}^{m} R_i (X)$.

(6) In the above formula, let $X \sim X$, we have $\sum_{i=1}^{m} R_i (X) = \sum_{i=1}^{m} R_i (X)$.

Remark. It should be noticed that in the variable precision optimistic multigranulation based on tolerance relations rough set model, $\sum_{i=1}^{m} R_i (X) \subseteq \sum_{i=1}^{m} R_i (X)$ does not always hold.

For discovering the relationship between the variable precision optimistic multigranulation lower and upper approximations based on tolerance relations of a single set and the corresponding variable precision optimistic multigranulation lower and upper approximations based on tolerance relations of two sets described by using m-tolerance relations on the universe, the following properties are given.

Proposition 2. Let $S = (U, T, f)$ be an incomplete information system in which $R_1, R_2, L, R_m \subseteq R$ be m-tolerance relations on $U$, then $\forall Y \subseteq U$ and $\beta \in \{0,0.5\}$, the following properties hold:

(1) $X \subseteq Y \Rightarrow \sum_{i=1}^{m} R_i (X) \subseteq \sum_{i=1}^{m} R_i (Y)$

(2) $X \subseteq Y \Rightarrow \sum_{i=1}^{m} R_i (X) \subseteq \sum_{i=1}^{m} R_i (Y)$

(3) $\sum_{i=1}^{m} R_i (X \cap Y) \subseteq \sum_{i=1}^{m} R_i (X) \cap \sum_{i=1}^{m} R_i (Y)$

(4) $\sum_{i=1}^{m} R_i (X \cup Y) \supseteq \sum_{i=1}^{m} R_i (X) \cup \sum_{i=1}^{m} R_i (Y)$

(5) $\sum_{i=1}^{m} R_i (X \cap Y) \supseteq \sum_{i=1}^{m} R_i (X) \cup \sum_{i=1}^{m} R_i (Y)$
By the existing knowledge, some objects cannot be decided which class they belong to when they are classified. These objects all form the called boundary region. Due to the existence of the boundary region, any set is uncertainty. The bigger the boundary region of a set, the bigger the uncertainty of a set. To better exactly express the degree of the uncertainty of a set, we introduce three kinds of measures, which are accuracy measure, approximation quality and precision of approximation.

Definition 7. Let \( S = (U, \mathcal{R}, f) \) be an incomplete information system in which \( R_1, R_2 \subseteq \mathcal{R} \) be m-tolerance relations on \( U \), then \( \forall X \subseteq U \) and \( \beta \in [0, 0.5) \), the accuracy measure of \( X \) in VPT-MGRS can be denoted by \( \alpha^{O,T}_{\mathcal{R}}(X) \), where

\[
\alpha^{O,T}_{\mathcal{R}}(X) = \frac{\sum_{i=1}^{m} R_{i}^{O,T}(X)}{\sum_{i=1}^{m} R_{i}^{O,T}(X)}
\]

From Definition 7, we can derive the following theorem.

Theorem 3. Let \( S = (U, \mathcal{R}, f) \) be an incomplete information system in which \( R_1, R_2 \subseteq \mathcal{R} \) be m-tolerance relations on \( U \), then \( \forall X \subseteq U \) and \( \beta \in [0, 0.5) \), the following properties hold:

1. \( \alpha^{O,T}_{\mathcal{R}}(X) \leq \alpha^{O,T}_{\mathcal{R}}(X) \)
2. \( \alpha^{O,T}_{\mathcal{R}}(X) \leq \alpha^{O,T}_{\mathcal{R}}(X) \)

Definition 8. Let \( S = (U, \mathcal{R}, f) \) be an incomplete information system in which \( R_1, R_2 \subseteq \mathcal{R} \) be m-tolerance relations on \( U \), then \( \forall X \subseteq U \) and \( \beta \in [0, 0.5) \), the approximation quality of \( X \) in VPT-MGRS also called a degree of dependence, can be denoted by \( \gamma^{O,T}_{\mathcal{R}}(X) \), where

\[
\gamma^{O,T}_{\mathcal{R}}(X) = \frac{\sum_{i=1}^{m} \left| \sum_{j=1}^{n} R_{i}^{O,T}(D_j) \right|}{|U|}
\]

From Definition 8, we can derive the following theorem.

Theorem 4. Let \( S = (U, \mathcal{R}, f) \) be an incomplete information system in which \( R_1, R_2 \subseteq \mathcal{R} \) be m-tolerance relations on \( U \), then \( \forall X \subseteq U \) and \( \beta \in [0, 0.5) \), the following properties hold:

1. \( \gamma^{O,T}_{\mathcal{R}}(X) \leq \gamma^{O,T}_{\mathcal{R}}(X) \)
2. \( \gamma^{O,T}_{\mathcal{R}}(X) \leq \gamma^{O,T}_{\mathcal{R}}(X) \)

Definition 9. Let \( S = (U, \mathcal{R}, f) \) be an incomplete information system in which \( R_1, R_2 \subseteq \mathcal{R} \) be m-tolerance relations on \( U \), then \( \forall X \subseteq U \) and \( \beta \in [0, 0.5) \), the
precision of approximation of \( X \) in VPT-MGRS can be denoted by \( \pi_{\sum R_\delta}^{OT}(X) \), where

\[ \pi_{\sum R_\delta}^{OT}(X) = \frac{\sum_{i=1}^{m} \pi_{R_\delta}^{T}(X)}{|X|} \]

From Definition 9, we can derive the following theorem.

**Theorem 5.** Let \( S = (U, R, \mathcal{F}) \) be an incomplete information system in which \( R_1, R_2, \ldots, R_m \subseteq R \) be m-tolerance relations on \( U \), then \( \forall X \subseteq U, \forall R \subseteq R \) and \( \beta \in [0, 0.5] \), the following properties hold.

1. \[ \pi_{\sum R_\delta}^{OT}(X) \leq \pi_{\sum R_\delta}^{OT}(X) \]
2. \[ \pi_{R_\delta}^{T}(X) \leq \pi_{\sum R_\delta}^{OT}(X) \]

### 5. Experimental Analysis

In what follows, through experiment we validate the properties and some theorems of the variable precision multigranulation rough set in incomplete rough set. We have selected four public data sets from UCI Repository of machine learning databases, which are described in Table 1.

<table>
<thead>
<tr>
<th>Data Sets</th>
<th>Samples</th>
<th>Numerical features</th>
<th>Decision Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automobile</td>
<td>205</td>
<td>25</td>
<td>5</td>
</tr>
<tr>
<td>Hepatitics</td>
<td>155</td>
<td>19</td>
<td>2</td>
</tr>
<tr>
<td>Mammographic</td>
<td>961</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Pima-indians</td>
<td>768</td>
<td>7</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 1: DATA SETS DESCRIPTION

According to decision attributes, data sets Automobile is divided into five decision classes, which are \( D_1, D_2, D_3, D_4, D_5 \), respectively. We select attribute set \( C = \{ C_1, C_2, C_3, C_4, C_5, C_6 \} \), and assume that the first attribute to the fourth attribute are \( C_1 \), the fifth attribute to the eighth attribute are \( C_2 \), the ninth attribute to the 12th attribute are \( C_3 \), the 13th attribute to the 16th attribute are \( C_4 \), the 17th attribute to the 20th attribute are \( C_5 \), and the 21st attribute to the 25th attribute are \( C_6 \).

Here, in the variable precision multigranulation rough set based on tolerance relations, when \( \beta \) changes, we compare the variation of the value of the approximation measure of decision classes of the four incomplete decision systems, the variation of the value of the precision of approximation measure of decision classes of the four incomplete decision systems and the variation of the value of the degree of dependence of the four incomplete decision systems. Specifically, the approximation measure in the variable precision multigranulation rough set based on tolerance relations is denoted by \( OT_\alpha \), which is shown in Figs. 1-4, the precision of approximation measure in the variable precision multigranulation rough set based on tolerance relations is denoted by \( OT_\pi \), which is shown in Figs. 5-8, and the degree of dependence measure in the variable precision multigranulation rough set based on tolerance relations is denoted by \( OT_\gamma \), which is shown in Fig. 9.
Fig. 4. Variation of $OT_{\alpha}$ of each decision class with the value of $\beta$ (data set Pima-indians)

Fig. 5. Variation of $OT_{\pi}$ of each decision class with the value of $\beta$ (data set Automobile)

Fig. 6. Variation of $OT_{\pi}$ of each decision class with the value of $\beta$ (data set Hepatitics)

Fig. 7. Variation of $OT_{\pi}$ of each decision class with the value of $\beta$ (data set Mammographic)

Fig. 8. Variation of $OT_{\gamma}$ of each decision class with the value of $\beta$ (data set Pima-indians)

Fig. 9. Comparison of $OT_{\gamma}$ on the four data sets

It is shown in Figs.1-4 that the value of $OT_{\alpha}$ increases as the value of $\beta$ does in the same data. From Fig.5 to Fig.8, we also see that the value of $OT_{\pi}$ increases as the value of $\beta$ does in the same data. In Fig.9, on the horizontal axis, the term A represents data set Automobile, the term B represents data set Hepatitics, the term C represents data set Mammographic, and the term D represents data set Pima-indians. The same results can be concluded by Fig.9, which is that the value of $OT_{\gamma}$ increases as the value of $\beta$ does in the same data.

Specially, in Fig.1, we only see that there are only two lines, which represents the variation of the value of $OT_{\alpha}$ of the decision class $D_1$ with $\beta$ and the variation of the value of $OT_{\alpha}$ of the decision class $D_2$ with $\beta$. But according to the assumption above, there are five decision classes in data set Automobile. It is because that the variable precision optimistic multigranulation lower and upper approximation of the decision class $D_3$ and the decision class $D_5$ based on tolerance relations are both equal to zero when the value of $\beta$ is one of the four values, which are 0.1, 0.2, 0.3 and 0.4, so the values of $OT_{\alpha}$ of the decision class $D_3$ are not exist. In addition, since the variable precision optimistic multigranulation upper approximation of the decision class $D_4$ based on tolerance relations is equal to zero when the value of $\beta$ is selected 0.1, 0.2, 0.3 and 0.4, so the values of $OT_{\alpha}$ of the decision class $D_4$ are infinite. By the above analyses, the variation of the value of $OT_{\alpha}$ of the decision
class \(D_3\) with \(\beta\), the variation of the value of \(OT_\alpha\) of the decision class \(D_3\) with \(\beta\) and the variation of the value of \(OT_\alpha\) of the decision class \(D_3\) with \(\beta\) cannot be reflected in Fig.1.

6.Conclusions

Multigranulation rough set is an emerging direction in rough set theory, which uses multiple structures. To study the incomplete information system under multigranulation environment, we first proposed a new multigranulation rough set model called the variable precision optimistic multigranulation rough set based on tolerance relations through combining the ideas of tolerance relation and variable precision rough set. Then, we introduced three elementary measures for measuring the uncertainty of the new model, and discussed some properties. Finally, some experimental results indicate the variation of the three elementary measures with the value threshold.

Acknowledgements

This work was supported by National Natural Science Foundation of China (60273043), Natural Science Foundation of Anhui province (090412054), Natural Science Foundation of Anhui Higher Education Institutions (KJ2011Z020) and Anhui Provincial Natural Science Foundation (1208085QF120).

References