A Rough Set Model Based on Probabilistic Similarity Measure for Incomplete Decision Tables

Nguyen Do Van¹  Koichi Yamada²  Muneyuki Unehara³
Nagaoka University of Technology
¹s095189@stn.nagaokaut.ac.jp  ²yamada@kjs.nagaokaut.ac.jp  ³unehara@kjs.nagaokaut.ac.jp

Abstract: Rough set models in incomplete decision tables have been discussed so far. Numerous approaches to deal with missing values in incomplete information systems have been proposed. In this paper, assuming that the domain of attribute values is defined, we apply the probability of values appearing in data tables in order to measure the self-information of similarity. This is defined as the uncertainty of similarity. Based on that, set approximation is defined by giving a threshold. Finally, the merit of this similarity is clarified and compared with other approaches.

1. Introduction

Rough set theory has been developed as a mean to analyze vague description of objects. Objects in information systems with precise values have been dealt with the theory. However, problems arise when the values of object are unknown. Therefore, it is necessary to develop a theory which enables classifications of objects even if there is only partial information available. Rough set models in [1,2] deal with indiscernibility based on similarity relations. In such approaches, a missing value is considered as a special value that may take any possible value. However, this may leads to problems in handling missing values. In fact, there is an array of methods to handle the incomplete objects. In this research, we study about probability of similarity and propose a method of handling missing value in incomplete information system based on self-information of uncertainty in similarity measurement.

2. Basic concepts [1,2]

Incomplete information systems
An information system is defined as a pair \( I = (U, A) \), where \( U \) is a non-empty finite set of objects called the universe and \( A \) is a non-empty finite set of attributes such that \( f_a : U \rightarrow V_a \) for every \( a \in A \). The set \( V_a \) is called the value set of \( a \).

If \( U \) contains at least one object with an unknown (missing or null)value, then \( I \) is called an incomplete information system, otherwise complete. In the incomplete information system, objects may contain several unknown attribute values. Unknown values are denoted by special symbol "*" in the incomplete information system.

Indiscernibility relation
The relation \( IND(P) \), \( P \subseteq A \) denotes a binary relation between objects that are indiscernible in terms of values of attributes in \( P \).

\[
IND(P) = \{(x, y) \in U \times U | a \in P, f_a(x) = f_a(y)\} \quad (2.1)
\]

Let \( I_P(x) = \{y \in U | (x, y) \in IND(P)\} \) is the set of objects which are indiscernible by \( P \) with \( x \), and is called equivalence class.

Similarity relations
A similarity relation \( SIM(P) \), \( P \subseteq A \) denotes a binary relation between objects that are possibly indiscernible in terms of values of attributes. In the incomplete information system, the similarity relation is defined by:

\[
SIM(P) = \{(x, y) \in U \times U | a \in P, f_a(x) = f_a(y) \quad \text{or} \quad f_a(x) = * \quad \text{or} \quad f_a(y) = *\} \quad (2.2)
\]

Let \( S_P(x) = \{y \in U | (x, y) \in SIM(P)\} \) is the set of objects which are similar to \( x \) in terms of \( P \) in the sense of the above similarity relation. \( S_P(x) \) is called equivalence class instead of \( I_P(x) \) in the rest of this paper. In the same way, we may say that \( y \) is equivalent to \( x \) in terms of \( P \) instead of saying that \( y \) is similar to \( x \), when \( y \in S_P(x) \).

Rough sets
Rough sets in incomplete information systems are defined in the same way as in complete information systems. Let \( U/SIM(P) \) denote a classification, which is a family set \( \{S_P(x) | x \in U\} \). Note that \( U/SIM(P) \) is not a partition of \( U \), unlike \( U/IND(P) \).

Set approximations
Let \( X \subseteq U, P \subseteq A \), \( P_X \) is the lower approximation of \( X \) in terms of \( P \); if and only if

\[
P_X = \{x \in U | S_P(x) \subseteq X\} = \{x \in X | S_P(x) \subseteq X\} \quad (2.3)
\]

\( \overline{P_X} \) is the upper approximation of \( X \) in terms of \( P \); if and only if

\[
\overline{P_X} = \{x \in U | S_P(x) \cap X \neq \phi\} = \bigcup \{S_P(x) | x \in X\} \quad (2.4)
\]

Decision table
The incomplete decision table (DT) is an incomplete information system combined with a set of decision attributes. When there is only a
decision attribute, it is denoted by $\mathcal{D}T = (\mathcal{U}, A \cup \{d\})$, where $d$ is a distinguished attribute called decision, $d \notin A$, and $* \not\in V_d$, where $V_d$ is the value set of decision $d$. All elements of $A$ are called conditions.

**Generalized decision**

The function $\partial_p : \mathcal{U} \rightarrow F(V_d)$ where $P \subseteq A$, and $F(V_d)$ is the power set of $V_d$, is defined as

$$\partial_p(x) = \{i \mid f_d(y) \wedge y \in S_p(x)\}, \quad (2.5)$$

where $f_d(y)$ is the decision value of an object $y$. $\partial_p$ will be called generalized decision in DT. $\partial_p$ determines which decision classes the object $x$ may be classified to based on the available information on $x$.

**Decision rules**

Any decision table may be regarded as a set of generalized decision rules of the form:

$$\land (c, v) \rightarrow v(d, w), \quad (2.6)$$

where $c \in P, v \in V_c, w \in V_d$.

The knowledge hidden in decision data tables may be discovered and expressed in the form of decision rules. The part $\land (c, v)$ is called condition part and $\lor (d, w)$ is called decision part of rules. A rule with a single decision value in the decision part will be called definite, otherwise non-definite.

**Reduct and core**

Reduction of knowledge that preserves generalized decisions for all objects is lossless from decision making standpoint. Reduction of a decision table is defined as follows.

A reduct $P$ of decision table DT as minimal subset of $A$, such that $\partial_p(x)$ of the reduct $P$ is equal to $\partial_A(x)$ for any $x \in \mathcal{U}$.

Formally, a set $P \subseteq A$ is a reduct of decision table if and only if

$$\partial_P(x) = \partial_A(x), \forall Q \subseteq P, \partial_Q(x) \neq \partial_A(x). \quad (2.7)$$

The reduct of an information system is not unique: there may be multiple minimal subsets of attributes which preserve generalized decisions.

The core is a set of all indispensable attributes, which appear in all reducts of the decision table. Actually, core is the intersection of all reducts. The core possibly is empty. In such case, there is no indispensable attribute.

### 3. Probabilistic similarity

First, we define minimum probability that each value of an attribute appears based on the frequency in the data set. Then, the probability that two objects have the same value is calculated in order to measure the uncertainty of similarity.

The probability that a value $j \in V_a$, $j \neq *$ appears as a value of a certain object is between $\left|V_a(j)\right|/|\mathcal{U}|$ and $\left|V_a(j)\cup V_a(*)\right|/|\mathcal{U}|$, where $V_a(j)$ and $V_a(*)$ are sets of objects whose value of attribute “a” is “j” and “*”, respectively. If $f_a(x) \neq *$, that is, the attribute value of an object $x$ is not missing, the probability that $f_a(x)$ appears is between $\left|V_a(x)\right|/|\mathcal{U}|$ and $\left|\{V_a(x)\cup V_a(*)\}\right|/|\mathcal{U}|$.

Now, let us define probabilities $\rho_a(x)$ and $\rho_a(oba(j))$, which are the minimum probability that an attribute value $f_a(x) \neq *$ appears, and the minimum probability that value of attribute “a” is “j”, respectively. $oba(j)$ is an object whose value of attribute “a” is “j”.

**Definition 3.1.** Let $I = (\mathcal{U}, A)$ be an incomplete information system. $a \in A$. The minimum probability that each possible value $j \in V_a$, appears is given by

$$\rho_a(oba(j)) = \left|V_a(j)\right|/|\mathcal{U}|. \quad (3.1)$$

Then, the minimum probability that the value $f_a(x) \neq *$ appears can be defined:

$$\rho_a(x) = \left|V_a(x)\right|/|\mathcal{U}|, \text{ if } f_a(x) \neq * . \quad (3.2)$$

Now, we define probability that objects $x$ and $y$ are similar to each other on attribute “a” as the minimum estimation of probability that the two objects have the same value of “a”.

**Definition 3.2.** Let $I = (\mathcal{U}, A)$ be an incomplete information system. Given $a \in A$ and $x, y \in \mathcal{U}$, probability of similarity between $x$ and $y$ denoted by $\theta_a(x, y)$ is defined as the minimum estimation of probability that $x$ and $y$ take the same value on $a$, and is given by the following equation:

$$\theta_a(x, y) = \left\{\begin{array}{ll}
1, & \text{if } f_a(x) = f_a(y) \neq * \\
0, & \text{if } f_a(x) \neq *, f_a(y) \neq *, f_a(x) \neq f_a(y), \\
\rho_a(x), & \text{if } f_a(x) \neq *, f_a(y) = *,
\end{array}\right. \quad \theta_a(x, y) = \sum_{j \neq *, j \neq *} \left[p_a(oba(j)) \right]^2, \text{ if } f_a(x) = f_a(y) = *.$$  \quad (3.3)

In the case where two values is certain (not missing), if two values are equal, the probability of similarity is exactly 1. On the other hand, if two values are different from each other, the probability of similarity is 0. When there is a missing value, it cannot define they are the same or not. In this case, we have to employ the probability.

Following is an example of incomplete information.
Table 3.1. An incomplete information system

<table>
<thead>
<tr>
<th>(U)</th>
<th>(a_1)</th>
<th>(a_2)</th>
<th>(a_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x_1)</td>
<td>*</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>(x_2)</td>
<td>1</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>(x_3)</td>
<td>2</td>
<td>1</td>
<td>*</td>
</tr>
<tr>
<td>(x_4)</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>(x_5)</td>
<td>*</td>
<td>*</td>
<td>2</td>
</tr>
<tr>
<td>(x_6)</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 3.2. Appearing probabilities \(\rho_a(x)\)

<table>
<thead>
<tr>
<th>(U)</th>
<th>(a_1)</th>
<th>(a_2)</th>
<th>(a_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x_1)</td>
<td>NG</td>
<td>2/6</td>
<td>4/6</td>
</tr>
<tr>
<td>(x_2)</td>
<td>1/6</td>
<td>NG</td>
<td>NG</td>
</tr>
<tr>
<td>(x_3)</td>
<td>1/6</td>
<td>1/6</td>
<td>NG</td>
</tr>
<tr>
<td>(x_4)</td>
<td>2/6</td>
<td>1/6</td>
<td>4/6</td>
</tr>
<tr>
<td>(x_5)</td>
<td>NG</td>
<td>NG</td>
<td>4/6</td>
</tr>
<tr>
<td>(x_6)</td>
<td>2/6</td>
<td>2/6</td>
<td>4/6</td>
</tr>
</tbody>
</table>

NG = Not given

Consider the case where \(a = a_1\), \(x = x_1\) and \(y = x_2\) in Table 3.1. The minimum probability that the value \(f_{a_1}(x_1)\) is the same as \(f_{a_1}(x_2) = 1\) is \(\rho_{a_1}(x_2) = 1/6\). Then, the minimum probability that \(f_{a_1}(x_1) = f_{a_1}(x_5)\) is obtained as follows:

- Probability of \(f_{a_1}(x_1) = f_{a_1}(x_5) = 1\) is \(1/6 \times 1/6\).
- Probability of \(f_{a_1}(x_1) = f_{a_1}(x_5) = 2\) is \(1/6 \times 1/6\).
- Probability of \(f_{a_1}(x_1) = f_{a_1}(x_5) = 3\) is \(2/6 \times 2/6\).

So, the probability that \(f_{a_1}(x_1) = f_{a_1}(x_5)\) is \(1/36 + 1/36 + 4/36 = 6/36\).

From the probability of matching, we can measure the uncertainty of similarity information in order to define the similarity between objects.

**Definition 3.3** Let \(I = (U, A)\) be an incomplete information system, \(P \subseteq A\). Given \(x, y \in U\), the self-information of similarity measurement is denoted by

\[
\varphi_p(x, y) = -\sum_{\forall a \in P} \log_2 \theta_a(x, y).
\]  

(3.4)

Obviously \(\varphi_p(x, x) = 0\), and \(\varphi_p(x, y) = \varphi_p(y, x)\).

By the definition, the amount of self-information contained in a probabilistic similarity depends only on the probability of that similarity: the smaller its probability, the larger the self-information associated with receiving the information in that the similarity indeed occurred.

**Definition 3.4** Similarity relations based on probability: The similarity relation in incomplete information is modified as \(SIM_a(P), P \subseteq A\) and can be defined as follow:

\[
SIM_a(P) = \{(x, y) \in U \times U | \varphi_p(x, y) \leq a\}.
\]  

(3.5)

The given threshold \(a \geq 0\) is different in each system. This depends on the characteristic of attribute values of objects.

Let \(S^a_p(x) = \{y \in U | (x, y) \in SIM_a(P)\}\) is the set of objects which are similar with \(x\) on \(P\).

From Table 3.2, it can be achieved that \(\varphi_p(x_1, x_2) = 4.75489\); \(\varphi_p(x_1, x_3) \rightarrow \infty\); \(\varphi_p(x_1, x_4) \rightarrow \infty\); \(\varphi_p(x_1, x_5) = 4.16993\); \(\varphi_p(x_1, x_6) = 1.58496\). Thus, given \(a = 2\), \(S^a_p(x_1) = \{x_1, x_6\}\).

**Property 3.1** Let \(I = (U, A)\) be an incomplete information system, \(P \subseteq A\). Given \(x \in U\), if \(a = 0\) then \(S^a_p(x) = I_P(x)\).

**Property 3.2** Let \(I = (U, A)\) be an incomplete information system, \(P \subseteq A\). Given \(x, y, z \in U\), if \(\varphi_p(x, y) < \varphi_p(x, z)\), then \(y\) is more similar with \(x\) than \(z\).

4. Probabilistic Similarity-Based Rough Set Model

**PS-Rough sets**

Rough set in incomplete information systems defined in section 2 is modified. Let \(U/SIM_a(P)\) denotes classification, which is the family set \(\{S^a_p(x) | x \in U\}\).

**PS-approximations**

Let \(X \subseteq U, P \subseteq A\), \(\text{app}^a_pX\) is \(PS\)-lower approximation of \(X\), if and only if

\[
\text{app}^a_pX = \{x \in U | S^a_p(x) \subseteq X\} = \{x \in X | S^a_p(x) \subseteq X\}. \tag{4.1}
\]

\(\text{app}^a_pX\) is \(PS\)-upper approximation of \(X\), if and only if

\[
\text{app}^a_pX = \{x \in U | S^a_p(x) \cap X \neq \emptyset\} = \bigcup\{S^a_p(x) | x \in X\}. \tag{4.2}
\]

\(BN^a_pX\) is called the \(PS\)-boundary of \(X\) when

\[
BN^a_pX = \text{app}^a_pX - \text{app}^a_pX. \tag{4.3}
\]

\(AC^a_pX\) is called the accuracy of \(PS\)-approximation when

\[
AC^a_pX = \frac{\text{app}^a_pX}{|\text{app}^a_pX|}. \tag{4.4}
\]

**Properties** Let \(I = (U, A)\) be an incomplete information system, \(X \subseteq U, P \subseteq A\). Besides the properties proposed in [1,2], the additional properties are shown below:

1. If \(a \leq b \Rightarrow \text{app}^b_pX \supseteq \text{app}^a_pX\).
2. If \(a \leq b \Rightarrow \text{app}^b_pX \subseteq \text{app}^a_pX\).
3. If \(a \leq b \Rightarrow BN^b_pX \subseteq BN^a_pX\).
4. If \(a \leq b \Rightarrow AC^b_pX \geq AC^a_pX\).
5. Comparison

Numerous methods to handle missing value in incomplete information system have been discussed and tested. In such methods, several ones [3,4,5,6] concentrate on probability of matching values.

In [3], the probability that two objects \( x, y \) have the same value on an attribute \( a \) is defined as \( 1/[V_a] \) when attribute value of \( x \) or \( y \) is missing. Then an expected function of any two objects is obtained by multiplying probabilities for all those attributes. In this approach, only the cardinality of attributes, in which attribute values of objects are missing, controls the expected function. If there is no missing value, the expected function is not defined.

In [4,5], the authors proposed the “Method of possible worlds”. He replaced the missing value by all possible values in the set. Then all possible information tables are derived. Based on that, approximation set is induced as the union of all approximation of derived information tables. This possibly leads to too much consuming resource.

In [6], in a section, probability of matching values is employed to calculate the equivalence degrees. It means objects are similar to each other by some degree. However, in other section, the similarity relation is defined in a different way, in which missing values are similar to any certain values. Both equivalence degree and similarity relation contribute to define rough entropy in incomplete information systems. This may cause a potential problem because two approaches are used in one algorithm.

In our approach, frequency of values pays a vital attention to the relationship between objects. In some kind of attribute, one value may dominate other value. For example, when the age of students in the first year of undergraduate is valuated, the missing values are likely 18 years old which is the most common values. Another example is shown in the table below:

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Full</td>
<td>High-speed</td>
<td>Low</td>
</tr>
<tr>
<td>2</td>
<td>Full</td>
<td>High</td>
<td>Mid</td>
</tr>
<tr>
<td>3</td>
<td>Full</td>
<td>High</td>
<td>Mid</td>
</tr>
<tr>
<td>4</td>
<td>*</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>5</td>
<td>Compact</td>
<td>High</td>
<td>*</td>
</tr>
<tr>
<td>6</td>
<td>Full</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>7</td>
<td>Full</td>
<td>*</td>
<td>Low</td>
</tr>
<tr>
<td>8</td>
<td>Full</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>9</td>
<td>Compact</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>10</td>
<td>*</td>
<td>Low</td>
<td>High</td>
</tr>
</tbody>
</table>

In table 5.1, the set of observed objects is \( X \) (Objects with \( d = \) Good). The size of Car 4 is missing. Due to the frequency of “Full” appearing in the set “Size”, the probability of matching between missing value and “Full” is larger than that between missing value and “Compact”.

Furthermore, self-information is a measure of the information content associated with the outcome of random variable. This assists us to measure the uncertainty of similarity. The given threshold allows us to decide which level of uncertainty is acceptable with the system.

Besides that, from the self-information, we can define a family of set approximation, which may connect to the fuzzy-set. This leads to a deeper research on fuzzy rough set theory.

6. Conclusion

In this research, self-information of similarity measurement is used to define set approximation under uncertainty level. This approach can be applied to some systems, in which the frequency of values appearing is clearly recognized. Our further work is to find how self-information of similarity measurement affects to the set approximation and how this approach connects to fuzzy rough set theory.

References