On the Similarity based on Uniqueness Measure

Masashi Emoto[1] Rolly Intan[2,3] Manami Matsumoto[2] Masao Mukaidono[2]

[1]School of Business Administration, Meiji University
[2]Department of Computer Science, Meiji University
1-1-1 Higashi-mita, Tama-ku, Kawasaki-shi, 214-8571, Japan
[3]Department of Informatics Engineering,
Petra Christian University, Surabaya, Indonesia
E-mail: {masashi, rolly, cat-tail, masao}@cs.meiji.ac.jp

Abstract: Traditionally, similarity between two objects is calculated by using only their attribute values, such as number of coincided attributes, Euclid distance, etc. A new concept of similarity dealing with a uniqueness measure is proposed in this paper by which the similarity between two objects is not only considered using their attribute values, but also a subset of objects as an important parameter. Here, the subset of objects may be regarded as knowledge of human. In this concept of similarity, if attribute values of two objects are rare in the subset, and their attribute values are the same, then their degree of similarity is high. On the other hand, if the attribute values of two objects are not rare in the subset, and their attribute values are the same, then their degree of similarity is low. Consequently, the degree of similarity between two objects will be changed depending on the subset of objects. Moreover, we discuss mathematical properties of the concept of similarity dealing with uniqueness measure. Finally, we discuss differences between the concept of similarity based on uniqueness measure and traditional similarity using some examples.
Keywords: Similarity relation, Probability, Uniqueness measure

1 Introduction

In traditional concepts of similarity, degree of similarity between two objects is calculated by using only attribute values of the objects, such as using number of coincided attributes, Euclid distance, etc. Others may also consider the similarity using weights of attributes to emphasize how important an attribute in determining the similarity of objects (see [7]). In [7], similarity is calculated by using information quantities of attribute values as weights of attributes.

In [4, 2], we discussed and proposed a new concept of similarity dealing with uniqueness measure. Simply, similarity between two objects is obtained by observing their attribute values of two objects and a subset of objects in which the objects belong to the subset. Here, we may consider the subset of objects as knowledge of human. In the concept of similarity, if attribute values of two objects are rare in the subset, and their attribute values are the same, then their degree of similarity will be high. On the other hand, if attribute values of two objects are not rare in the subset of objects, and their attribute values are the same, then their degree of similarity will be low. Consequently, the degree of similarity between two objects will be changed depending of the subset of objects. We consider this phenomenon by defining what we call it as *uniqueness measure*. Therefore, extending [4], our primary goal in this paper is to discuss mathematical properties of the concept of similarity which is dealing with the uniqueness measure. Especially, we also discuss and consider differences between the concept of similarity and the traditional concepts of similarity.

2 Human's Perception

The concept of uniqueness based similarity that we proposed is based on human's perception. In this section, first, we show an illustrative example and figures, and discuss how human's perception works in a process of recognition. Then, based on the human's perception, we propose a new concept of similarity.

2.1 Monkey vs. Human

Let us suppose that there are two photographs, a photograph of two different monkeys and a photograph of two different humans. As human, obviously we cannot recognize differences between two monkeys in the photograph well, but we can recognize differences between two humans in their photograph (if they are not identical twin). This is because we have much knowledge related to humans by having many opportunities (experiences) to see them, but we have few knowledge related to the monkeys. Hence, it can be said that it is difficult for a human to recognize a monkey as a particular monkey compared with another monkey, because of our limited knowledge about the monkey. In other words, we do not know any important features of a monkey to be recognized. It could be said that it may be difficult for monkeys to recognize a human as a particular human compared to the other human. In our point of view, there are existences of individual knowledge for both humans and monkeys representing in their features.

2.2 Human and Their Features

When a picture of two men (Fig.1) and another one of these men with a mustache (Fig.2) are compared, we may consider two cases. The first case is when a Japanese looks at those pictures and the second case is when an Arab looks at the pictures. In the first case, generally, Japanese could understand some important features needed to recognize men without a mustache. So it is easy for Japanese to distinguish men without a mustache. Furthermore, in many cases, Japanese have a rare opportunity to see men with a mustache, so they tend to have a strong impression for a mustached image. Consequently, their attention will gather for the mustache. As a result, men with a mustache gradually become more similar for Japanese than men without a mustache.

Figure 1: the case of having no mustache

Figure 2: the case of having a mustache

Next, we consider the case when Arabs look at the pictures. Since Arabs have plenty of knowledge related to men with a mustache, they could easily recognize men with a mustache without being confused by the mustache.

2.3 Summary of Discussion

From the discussion above, it can be summarized that human can recognize anything if he has a certain knowledge or experience about it, where knowledge plays important role in human's perception. The knowledge here is constituted based on subjectivity for every human, and the subjectivity is made from what has been experienced until now. When an element and their combination are unique in knowledge, human has a tendency to react to a new thing sensitively. For instance, it is applied to the features such as, the monkeys are covered with hair, the humans have a mustache, etc. Therefore, in this paper, we propose a new concept of similarity based on uniqueness (measure) in which the concept of similarity is naturally represented in human's perception. That is, human will pay much attention to a rare (unique) case in the whole data.

3 Similarity based on Uniqueness Measure

This section defines the concept of similarity based on the uniqueness measure and compares the concept to the traditional one.

First, we need to define a data table called information system as usually used in knowledge representation. Formally, an information system is defined as follows.

Definition 1 *Information system* contains data about objects of interest characterized by some attributes. An information system is defined as a quadruple $I = (U, A, V, \rho)$, 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 $\rho : (U \times a_j) \rightarrow D_j$ for every $a_j \in A$. The set $D_j \in V$ is domain or value set of attribute a_j.

Example 1 *Given an information system in Table 1 consists of ten objects, $\{u_1, u_2, \cdots, u_{10}\}$, and five attributes, $\{a_1, a_2, \cdots, a_5\}$. Simplifying the problem, we consider $\rho : (U \times a_j) \rightarrow \{0, 1\}$, $\forall a_j, j = 1, \ldots, 5$. For example in Table 1, attribute value of object u_3 in attribute a_5 is given by $\rho(u_3, a_5) = 1$. Also, attribute value of object u_7 in attribute a_1 is given by $\rho(u_7, a_1) = 0$.*

Table 1: An example of information system.

	a_1	a_2	a_3	a_4	a_5
u_1	*0*	*1*	*0*	*0*	*0*
u_2	*1*	*0*	*0*	*1*	*0*
u_3	*0*	*1*	*0*	*0*	*1*
u_4	*0*	*1*	*0*	*1*	*0*
u_5	*1*	*1*	*0*	*0*	*1*
u_6	*0*	*1*	*1*	*1*	*1*
u_7	*0*	*1*	*1*	*0*	*0*
u_8	*0*	*1*	*1*	*1*	*0*
u_9	*0*	*0*	*1*	*0*	*0*
u_{10}	*0*	*1*	*1*	*1*	*0*

Next, we define a degree of coincidence by comparing attribute values between two objects in a given attribute.

Definition 2 Degree of coincidence between two objects $u_i, u_j \in U$ in attribute $a_k \in A$ is denoted by $M(u_i, u_j, a_k)$, and defined as follows.

$$M(u_i, u_j, a_k) = \begin{cases} 1 & \text{if } \rho(u_i, a_k) = \rho(u_j, a_k) \\ 0 & \text{if } \rho(u_i, a_k) \neq \rho(u_j, a_k). \end{cases} \tag{1}$$

Simply, a traditional concept of similarity may be defined by using degree of coincidence as the following definition.

Definition 3 Traditional concept of similarity between two objects $u_i, u_j \in U$ is denoted by $S_{tra}(u_i, u_j)$, and defined by:

$$S_{tra}(u_i, u_j) = \frac{\sum_{k=1}^{|A|} M(u_i, u_j, a_k)}{|A|}, \tag{2}$$

where $a_k \in A$, and $|A|$ is cardinality (number of elements) of set of attributes A.

The above concept of traditional similarity is obtained by dividing a number of coincided attributes of two objects by a number of all attributes.

Example 2 *Using Definition 3, similarity between u_2 and u_5, and similarity between u_8 and u_{10} in Table 1 are calculated by:*

$$S_{tra}(u_2, u_5) = \frac{2}{5} = 0.4,$$
$$S_{tra}(u_8, u_{10}) = \frac{5}{5} = 1.$$

Before defining the concept of uniqueness measure, it is necessary to define probability of a certain attribute value of an object given a subset of objects.

Definition 4 Let $u_i \in U$ is an object, $X \subseteq U$, and $a_k \in A$ is an attribute. $P(u_i, a_k, X)$ is defined as probability of attribute value, $\rho(u_i, a_k)$, in a sample space X as given by:

$$P(u_i, a_k, X) = \frac{\sum_{u_j \in X} M(u_i, u_j, a_k)}{|X|}. \tag{3}$$

On the other hand, Definition 4 represents probability of a number of objects whose attribute values are the same as attribute value of object u_i at an attribute a_k in a number of objects in set X.

Example 3 *Let $X = \{u_1, \ldots, u_5\}$ be a subset of U. $P(u_2, a_1, U)$, $P(u_5, a_3, U)$, $P(u_2, a_1, X)$ and $P(u_5, a_3, X)$ are calculated as follows.*

$$P(u_2, a_1, U) = \frac{2}{10} = 0.2,$$
$$P(u_5, a_3, U) = \frac{5}{10} = 0.5,$$
$$P(u_2, a_1, X) = \frac{2}{5} = 0.4,$$
$$P(u_5, a_3, X) = \frac{5}{5} = 1.$$

Using the probability of an attribute value in a given subset of objects as defined in Definition 4, a concept of uniqueness measure is defined characterized by a function to calculate uniqueness of relationship between attribute values of two objects in a certain attribute.

Definition 5 Given $X \subseteq U$, uniqueness measure relationship between two objects, $u_i, u_j \in U$, in attribute $a_k \in A$ is characterized by a function $C(u_i, u_j, a_k, X)$.

$$\begin{aligned} & C(u_i, u_j, a_k, X) \\ = & \begin{cases} 1 - P(u_i, a_k, X)^2 \\ \quad \text{if } M(u_i, u_j, a_k) = 1 \\ 1 - 2 \times P(u_i, a_k, X) \times P(u_j, a_k, X) \\ \quad \text{if } M(u_i, u_j, a_k) = 0. \end{cases} \end{aligned} \tag{4}$$

Generally, this function calculates subtraction a probability which attribute values of two objects occur at once from one. Since all attribute values is simplified taking a value from either 0 or 1, occurrences of attribute values of two objects are divided into four cases. They are a case of 0 and 0, a case of 0 and 1, a case of 1 and 0, and a case of 1 and 1. The case which attribute values are different is two case (0 and 1, and 1 and 0). Hence, probability is twice when $M(u_i, u_j, a_k) = 0$.

Example 4 *We show two examples of a case when two attribute values are different, $C(u_2, u_5, a_2, X)$, and a case when two attribute values are the same, $C(u_2, u_5, a_3, X)$.*

Let $X = \{u_1, u_2, u_3, u_4, u_5\}$ be a subset of objects.

$$\begin{aligned} & C(u_2, u_5, a_2, X) \\ = & \ 1 - 2 \times P(u_2, a_2, X) \times P(u_5, a_2, X), \\ = & \ 1 - 2 \times 0.2 \times 0.8, \\ = & \ 1 - 0.32, \\ = & \ 0.68. \end{aligned}$$

$$\begin{aligned} & C(u_2, u_5, a_3, X) \\ = & \ 1 - P(u_2, a_3, X)^2, \\ = & \ 1 - 1^2, \\ = & \ 0. \end{aligned}$$

In the second example, attribute values of all objects in X are all 0's. Consequently, attribute values in a_3 are not rare. Therefore, $C(u_2, u_5, a_3, X) = 0$.

Now, using uniqueness measure as defined in Definition 5, we define a concept of similarity between two objects as follows.

Definition 6 Degree of similarity between two objects $u_i, u_j \in U$ for $X \subseteq U$ is calculated by the following equation.

$$S_{uni}(u_i, u_j, X) = \frac{\sum_{k=1}^{|A|}(C(u_i, u_j, a_k, X) \times M(u_i, u_j, a_k))}{\sum_{k=1}^{|A|} C(u_i, u_j, a_k, X)}. \quad (5)$$

Example 5 *Given the information system in Table 1 and let X be $\{u_1, u_2, u_3, u_4, u_5\}$. By using Eq. 5, we calculate degree of similarity between u_2 and u_5 as follow.*

$$\begin{aligned} & S_{uni}(u_2, u_5, U) \\ = & \frac{0.96 \qquad\quad + 0.75}{0.96 + 0.68 + 0.75 + 0.5 + 0.58}, \\ = & \frac{1.71}{3.47}, \\ = & 0.493. \end{aligned}$$

$$\begin{aligned} & S_{uni}(u_2, u_5, X) \\ = & \frac{0.84 \qquad\quad + 0}{0.84 + 0.68 + 0 + 0.52 + 0.52}, \\ = & \frac{0.84}{2.56}, \\ = & 0.328. \end{aligned}$$

Similarly, degree of similarity between u_8 and u_{10} is given by:

$$\begin{aligned} & S_{uni}(u_8, u_{10}, U) \\ = & \frac{0.36 + 0.36 + 0.75 + 0.75 + 0.51}{0.36 + 0.36 + 0.75 + 0.75 + 0.51}, \\ = & \frac{2.73}{2.73}, \\ = & 1. \end{aligned}$$

$$\begin{aligned} & S_{uni}(u_8, u_{10}, X) \\ = & \frac{0.64 + 0.36 + 1 + 0.84 + 0.64}{0.64 + 0.36 + 1 + 0.84 + 0.64}, \\ = & \frac{3.48}{3.48}, \\ = & 1. \end{aligned}$$

In X, all attribute values of a_3 are equal to 0. Consequently, uniqueness of 0 in attribute a_3 disappeared. Hence, it can be verified that

$$S_{uni}(u_2, u_5, U) \geq S_{uni}(u_2, u_5, X).$$

Also, in the case of $S_{uni}(u_8, u_{10}, U)$ and $S_{uni}(u_8, u_{10}, X)$, since attribute values of two objects in all attributes are the same, their similarity is equal to 1 for any subsets.

In these two cases, when a subset of objects consists of only one object and when all objects in the subset have exactly the same attribute values for all attributes, results of calculation by using Eq.5 are equal to $\frac{0}{0}$. Here, we consider $\frac{0}{0}$ as 1 in this research.

4 Mathematical Properties of Similarity based on Uniqueness Measure

As a primary goal in this paper, this section discusses mathematical properties of the concept of similarity based on the uniqueness measure. There are several well known mathematical properties of binary relations such as equivalence relation, fuzzy similarity relation [8], weak fuzzy similarity relation [5], resemblance (proximity) relation [1], etc. Equivalence relation is consider as the strongest binary relation, and it can be verified that similarity between crisp data satisfies the equivalence relation as given by the following definition.

Definition 7 It is called equivalence relation, if binary relation satisfies the following properties.

Reflexivity:
$\forall x \quad R(x, x) = 1,$
Symmetry:
$\forall x, y \quad R(x, y) = R(y, x),$
Transitivity:
$\forall x, y, z$
$R(x, y) = R(y, z) = 1 \Rightarrow R(x, z) = 1.$

Fuzzy similarity relation has a weaker transitive property than equivalence relation for dealing with fuzzy data.

Definition 8 It is called fuzzy similarity relation, if binary relation satisfies the following properties.

Reflexivity:
$\forall x \quad R(x, x) = 1,$
Symmetry:
$\forall x, y \quad R(x, y) = R(y, x),$
Max-min Transitivity:
$\forall x, y, z$
$R(x, z) \geq max_{y \in U} min[R(x, y), R(y, z)].$

The weak fuzzy similarity relation is proposed based on conditional probability relation [6]. In conditional probability relation, similarity relationship between two data is assumed similar to the relationship between two events in conditional probability.

Definition 9 It is called weak fuzzy similarity relation, if binary relation satisfies the following properties.

Reflexivity:

$\forall x \quad R(x,x) = 1,$

Conditional Symmetry:

$\forall x, y$ if $\mathrm{R(x,y)} > 0$ then $\mathrm{R(y,x)} > 0,$

Conditional Transitivity:

$\forall x, y, z$

if $\mathrm{R(x,y)} \geq \mathrm{R(y,x)} > 0$ and

$R(y,z) \geq R(z,y) > 0$ then $\mathrm{R(x,z)} \geq \mathrm{R(z,x)}.$

Definition 10 The binary relation that satisfies only reflexivity and symmetry properties is called resemblance (proximity) relation.

Theorem 1 *Similarity with uniqueness measure satisfies resemblance relation.*

It can be verified easily.

In order to explore more properties, it is necessary to define number of coincidence attributes.

Definition 11 Number of coincidence attributes is given by the following equation:

$$N(u_i, u_j) = \sum_{k=1}^{|A|} M(u_i, u_j, a_k). \tag{6}$$

Theorem 2 *In the concept of traditional similarity, all $u_i, u_j, u_k, u_l \in U$ satisfy the following property.*

$$\begin{array}{rcl} N(u_i, u_j) & \leq & N(u_k, u_l) \\ & \Updownarrow & \\ S_{tra}(u_i, u_j) & \leq & S_{tra}(u_k, u_l). \end{array} \tag{7}$$

However, in the concept of similarity with uniqueness measure, the property is not always satisfied. That is,

$$\begin{array}{rcl} N(u_i, u_j) & \leq & N(u_k, u_l) \\ & \not\Updownarrow & \\ S_{uni}(u_i, u_j, X) & \leq & S_{uni}(u_k, u_l, X). \end{array} \tag{8}$$

Proof:

In traditional similarity, it is clear to satisfy the above property from Definition 3.

In the concept of similarity with uniqueness measure, we show a counter example. Here, we consider about an information system which has four attributes a_1, a_2, a_3, a_4 and a subset X. The information system is assumed to have probability value for attribute values 0 and 1 of all four attributes in X as shown in Table 2. For instance, $a_1[0]$ means probability of 0 of attribute a_1 in subset X.

Table 2: Probabilities of 0 and 1 at 4 attributes

$a_1[0]$	$a_1[1]$	$a_2[0]$	$a_2[1]$	$a_3[0]$	$a_3[1]$	$a_4[0]$	$a_4[1]$
0.9	0.1	0.9	0.1	0.1	0.9	0.5	0.5

Table 3: u_1, u_2, u_3, u_4

	a_1	a_2	a_3	a_4
u_1	0	0	1	0
u_2	0	0	1	1
u_3	1	1	0	0
u_4	1	1	1	1

Let $X = \{u_1, u_2, u_3, u_4\}$ has attribute values as given in Table 3. We show similarity of u_1 and u_2 and similarity of u_3 and u_4 as follows.

$$\begin{array}{rl} & S_{uni}(u_1, u_2, X) \\ = & \dfrac{0.19 + 0.19 + 0.19}{0.19 + 0.19 + 0.19 + 0.5}, \\ = & \dfrac{0.57}{1.07}, \\ = & 0.532. \end{array}$$

$$\begin{array}{rl} & S_{uni}(u_3, u_4, X) \\ = & \dfrac{0.99 + 0.99}{0.99 + 0.99 + 0.82 + 0.5}, \\ = & \dfrac{1.98}{3.3}, \\ = & 0.6. \end{array}$$

$N(u_1, u_2)$ is 3 and $N(u_3, u4)$ is 2. However, $S_{uni}(u_1, u_2, X)$ is 0.532 and $S_{uni}(u_3, u_4, X)$ is 0.6. **Q.E.D.**

Definition 12 Related to the next theorem, let $u \in U$ has $\rho(u, a)$ for $a \in A$. We define that $\sim u$ has $1 - \rho(u, a)$ for $a \in A$.

Theorem 3 *In the concept of traditional similarity, all $u_i, u_j \in U$ satisfy the following property.*

$$S_{tra}(u_i, u_j) + S_{tra}(u_i, \sim u_j) = 1. \tag{9}$$

However, in the concept of similarity with uniqueness measure, all $u_i, u_j \in U$ do not always satisfy that $S_{uni}(u_i, u_j, X) + S_{uni}(u_i, \sim u_j, X)$ is equal to 1.

Proof:
First, we prove that Eq.9 is satisfied in the concept of traditional similarity

$$\begin{aligned}
& S_{tra}(u_i,u_j)+S_{tra}(u_i,\sim u_j)\\
= & \frac{\sum_{k=1}^{|A|} M(u_i,u_j,a_k)}{|A|}+\frac{\sum_{k=1}^{|A|} M(u_i,\sim u_j,a_k)}{|A|},\\
= & \frac{\sum_{k=1}^{|A|} M(u_i,u_j,a_k)}{|A|}+\frac{\sum_{k=1}^{|A|} 1-M(u_i,u_j,a_k)}{|A|},\\
= & \frac{\sum_{k=1}^{|A|} M(u_i,u_j,a_k)+1-M(u_i,u_j,a_k)}{|A|},\\
= & \frac{\sum_{k=1}^{|A|} 1}{|A|}=\frac{|A|}{|A|}=1.
\end{aligned}$$

In the concept of similarity with uniqueness measure, we show a counter example which does not satisfy

$$S_{uni}(u_i,u_j,X)+S_{uni}(u_i,\sim u_j,X)=1.$$

We use the information system in Table 1. Let $X=\{u_1,u_2,u_3,u_4,u_5\}$ be a subset of U. We consider a similarity of u_2 and u_5.

$$\begin{aligned}
& S_{uni}(u_2,u_5,X)\\
= & \frac{0.84 \qquad\quad +0}{0.84+0.68+0+0.52+0.52},\\
= & \frac{0.84}{2.56},\\
= & 0.328.
\end{aligned}$$

$$\begin{aligned}
& S_{uni}(u_2,\sim u_5,X)\\
= & \frac{\qquad 0.96 \qquad +0.84+0.64}{0.52+0.96+1+0.84+0.64},\\
= & \frac{2.44}{3.96},\\
= & 0.616.
\end{aligned}$$

From $S_{uni}(u_2,u_5,X)+S_{uni}(u_2,\sim u_5,X)=0.944\neq 1$, it can be proved that the concept of similarity with uniqueness measure does not satisfy

$$S_{uni}(u_i,u_j,X)+S_{uni}(u_i,\sim u_j,X)=1.$$

Q.E.D.

Theorem 4 *In the concept of traditional similarity, all $u_i,u_j\in U$ satisfy the following property.*

$$S_{tra}(u_i,u_j) \quad = \quad S_{tra}(\sim u_i,\sim u_j). \tag{10}$$

However, in the concept of similarity with uniqueness measure, Eq.10 is not always satisfied.

Proof:
First, we prove that Eq.10 is satisfied in the concept of traditional similarity.

$$\begin{aligned}
& S_{tra}(\sim u_i,\sim u_j)\\
= & \frac{\sum_{k=1}^{|A|} M(\sim u_i,\sim u_j,a_k)}{|A|},\\
= & \frac{\sum_{k=1}^{|A|} 1-M(\sim u_i,u_j,a_k)}{|A|},\\
= & \frac{\sum_{k=1}^{|A|} 1-(1-M(u_i,u_j,a_k))}{|A|},\\
= & \frac{\sum_{k=1}^{|A|} M(u_i,u_j,a_k)}{|A|},\\
= & S_{tra}(u_i,u_j).
\end{aligned}$$

In the concept of similarity with uniqueness measure, we show a counter example which does not satisfy Eq.10.

Again, we use the information system in Table 1. Let $X=\{u_1,u_2,u_3,u_4,u_5\}$ be a subset of U. We consider a similarity of u_2 and u_5.

$$\begin{aligned}
& S_{uni}(u_2,u_5,X)\\
= & \frac{0.84 \qquad\quad +0}{0.84+0.68+0+0.52+0.52},\\
= & \frac{0.84}{2.56},\\
= & 0.328.
\end{aligned}$$

$$\begin{aligned}
& S_{uni}(\sim u_2,\sim u_5,X)\\
= & \frac{0.64 \qquad\quad +1}{0.64+0.68+1+0.52+0.52},\\
= & \frac{1.64}{3.36},\\
= & 0.488.
\end{aligned}$$

From $S_{uni}(u_2,u_5,X)=0.328$ and $S_{uni}(\sim u_2,\sim u_5,X)=0.488$, it can be proved that the concept of similarity with uniqueness measure does not satisfy

$$S_{uni}(u_i,u_j,X)=S_{uni}(\sim u_i,\sim u_j,X).$$

Q.E.D.

Theorem 5 *For any subsets, $X1,X2\subseteq U$, similarities among objects, $u_i,u_j,u_k,u_l\in U$, do not satisfy: $S_{uni}(u_i,u_j,X_1)\le S_{uni}(u_k,u_l,X_1)$ if and only if $S_{uni}(u_i,u_j,X_2)\le S_{uni}(u_k,u_l,X_2)$. That is,*

$$\begin{aligned}
S_{uni}(u_i,u_j,X_1) & \le S_{uni}(u_k,u_l,X_1)\\
& \not\Updownarrow\\
S_{uni}(u_i,u_j,X_2) & \le S_{uni}(u_k,u_l,X_2).
\end{aligned}$$

Proof:
We show a counter example. We use an information system in Table 1. Let be $X_1 = \{u_1, u_2, u_3, u_4, u_5\}$ and $X_2 = \{u_6, u_7, u_8, u_9, u_{10}\}$. Then,

$$\begin{aligned} S_{uni}(u_4, u_5, X_1) &= 0.188, \\ S_{uni}(u_1, u_6, X_1) &= 0.329, \\ S_{uni}(u_4, u_5, X_2) &= 0.382, \\ S_{uni}(u_1, u_6, X_2) &= 0.140. \end{aligned}$$

Therefore,

$$\begin{aligned} S_{uni}(u_4, u_5, X_1) &\leq S_{uni}(u_1, u_6, X_1), \\ S_{uni}(u_4, u_5, X_2) &\geq S_{uni}(u_1, u_6, X_2). \end{aligned}$$

We proved that Similarity based on uniqueness measure does not satisfy that $S_{uni}(u_i, u_j, X_1) \leq S_{uni}(u_k, u_l, X_1)$ if and only if $S_{uni}(u_i, u_j, X_2) \leq S_{uni}(u_k, u_l, X_2)$.
Q.E.D.

Example 6 *Theorem 5 represents that order of similarity can be changed. We explain by using the following example.*

We use an information system in Table 1 and let be $X = \{u_1, u_2, u_3, u_4, u_5\}$ and $U = \{u_1, u_2, u_3, u_4, u_5, u_6, u_7, u_8, u_9, u_{10}\}$. Then,

$$\begin{aligned} S_{uni}(u_2, u_3, X) &= 0, \\ S_{uni}(u_2, u_7, X) &= 0.190, \\ S_{uni}(u_2, u_3, U) &= 0.235, \\ S_{uni}(u_2, u_7, U) &= 0.177, \end{aligned}$$

$$\begin{aligned} S_{uni}(u_2, u_3, X) &\leq S_{uni}(u_2, u_7, X), \\ S_{uni}(u_2, u_3, U) &\geq S_{uni}(u_2, u_7, U). \end{aligned}$$

In knowledge X, u_7 is more similar to u_2 than u_3. Also, in knowledge U, u_3 is more similar to u_2 than u_7. In knowledge X and U, order of similarity is changed.

5 Conclusion

In this paper, first we introduced a new concept of similarity dealing with uniqueness measure. Then, we showed mathematical properties of the similarity. Also, we compared the concept of traditional similarity to the concept of similarity with uniqueness measure. In the future, we would like to clarify more properties of the concept of similarity with uniqueness measure, and apply the concept to the real world application.

References

[1] D. Dubois, H. Prade, *Fuzzy Sets and Systems: Theory and Applications*, (Academic Press, New York, 1980).

[2] M. Matsumoto, M. Emoto, R. Intan, M. Mukaidono, "A Proposal of Similarity between Two Objects based on Uniqueness Measure", *Proceeding of Intech 2003*, pp.354-358.

[3] M. Emoto, R. Intan, M. Matsumoto, M. Mukaidono, "Mathematical Properties of Similarity Dealing with Uniqueness Measure", *Proceeding of Intech 2003*, pp.812-817.

[4] M. Matsumoto, M. Emoto, R. Intan and M. Mukaidono: Proposal of the Similarity in Consideration of Rare of an Attribute, Proceeding of 19th Fuzzy System Symposium, pp.693-696, (2003), (in Japanese).

[5] R. Intan, M. Mukaidono, "Generalized Fuzzy Rough Sets By Conditional Probability Relations", *International Journal of Pattern Recognition and Artificial Intelligence Vol. 16(7)*, World Scientific, (2002), pp. 865-881.

[6] R. Intan, M. Mukaidono, "Fuzzy Relational Database Induced by Conditional Probability Relations", *The Transaction of Institute of Electronics, Information and Communication Engineers Vol. E86-D No. 8*, (2003), pp. 1396-1405.

[7] T. Yamaguchi, K. Wakitani and M. Yachida: Learning Initial Structural Concept of Houses Based on Similarity of Relations Weighted by Information, Journal of Information Processing Society of Japan, Vol.37, No.11, pp.1906-1917, (1996), (in Japanese).

[8] L.A. Zadeh, "Similarity Relations and Fuzzy Orderings", *Inform. Sci.3(2)*, (1970), pp. 177-200.