THE RESEARCH OF SIMILARITY MEASURE ABOUT ROUGH VAGUE SET

Shengquan Ma¹§, Jinshui Ji², Jun Shen³

¹,³Department of Mathematics
Hainan Normal University
Haikou, 571158, P.R. CHINA
¹e-mail: msq6202@hainnu.edu.cn

²Faculty of Computer Science and Information Engineering
Northwest University for Nationality
P.O. Box 579, Kirkland, WA 98083-0579, USA

Abstract: In this paper the similarity measure of rough vague set is discussed in detail, we syncretize the rough set and vague set, describe the concept of rough vague set, and introduce the correlational concept of rough vague value and a new method of similarity measure. The related properties and the method of similarity measure of rough vague set are investigated.

AMS Subject Classification: 31A31
Key Words: rough set, vague set, rough vague set, similarity measure

1. Introduction

The membership function of fuzzy set assigns each object a number which is on interval [0, 1] as its membership degree. It not only includes the evidence that the element belongs to the set, but also includes the evidence that the element does not belong to the set. For overcoming the insufficiency of the information described by the single value, Zadeh [1] introduced the interval value fuzzy set in 1975, the degree of an element belongs to a set is represented by an inner closed sub-interval on [0,1], the upper end point of the sub-interval represents the necessity of the object which belongs to the set, the right end point
means the possibility of the object belonging to the set. In 1986, Atanassov [2] considered the generalization of fuzzy set from different aspects, he employed two numbers to depict the condition of an element belongs to the fuzzy set, and introduced the concept of membership degree and non-membership degree. Atanassov called the set defined by this concept as intuitionistic fuzzy set. In 1989 the Atanassov and Gargov [3] point out that the interval value fuzzy set and intuitionistic fuzzy set are two equivalent generalization in fuzzy set theorem. In 1993 Gau and Buehrer [4] explained the vague set by using the “vote model”. In fact, there is no essential distinction among the interval value fuzzy set, the intuitionistic fuzzy set and the vague set (see [5]).

In the field of computer science and its application, especially in artificial intelligence (AI), date mining and knowledge discovery in database (KDD), the theorem of rough set has important application, rough set describes the connection and the whole characteristic of things, and provides important tools for researching the inner contact of things. But vague set provides a new tool for knowledge representation. It clearly gives the representation of the degree and scope that people can know about the things, and supplies a good description about things from form to contents. Both rough set and vague set theory research the uncertainty problem in information system, the motivation for rough set theory to solve problem is the undistinguished knowledge in information system, but vague set theory focuses on the vagueness of concept and the imprecision that people get from the concept. However, in many situations, the concept is not only misty, but also is undistinguishable, it causes the impossibility for people to understand the concept accurately and totally. According to this, it needs to merge rough set theory and vague set theory to reinforce the deficiency when they handle problems alone.

In the research of the application of rough set and vague set theory, similarity measure is an important problem. It is the theory foundation in application field of fuzzy set, such as in fuzzy cluster, pattern recognition and approximate reasoning, etc. For this end, in this paper we investigate the problem of similarity measure of rough vague-set and introduce a new general measure method.

2. The Concept of Rough Sets and Vague Sets

Let $U$ be a universe of discourse, $X \subseteq U$ is an object space, a vague set $V$ defined on $X \subseteq U$ can be represented by a true membership function. $t_v(x)$ is the lower bound led by the supporting evidence of $x$, then $f_v(x)$ is the lower
bound of negative membership degree led by the opposite evidence of \(x\), \(t_v(x)\) and \(f_v(x)\) establish a contact between a real number on \([0,1]\) and every point in \(X\), i.e.

\[
t_v : X \rightarrow [0,1]; \quad f_v : X \rightarrow [0,1], \quad \forall x \in X.
\]

The membership degree of vague set \(V\) is denoted by

\[
V(x) = [t_v(x), 1 - f_v(x)],
\]

where \(0 \leq t_v(x) + f_v(x) \leq 1\).

A vague set can be denoted by \(V = \{(x, t_v(x), f_v(x))|x \in X\}\), \([t_v(x), 1 - f_v(x)]\) is called as the vague value of point \(x\) in \(V\). Note: \(x_v = [t_v(x), 1 - f_v(x)]\).

Let \(U =\{x_1, x_2, \cdots, x_n\}\) be a finite set, \(R\) be an equivalent relation on \(U\), \(U/R\) be all the equivalent family on \(U\); \([x]_R\) means equivalent family on \(R\) which contains \(x\), \(\forall x \in U, \forall X \in U\), \(R(x) = \{x \in U|[x]_R \subseteq X\}\), \(\overline{R}(x) = \{x \in U|[x]_R \neq \emptyset\}\), called \(\overline{R}(x)\) and \(\overline{R}(x)\) is called low approximate and up approximate of \(X\) respectively. When low approximate and up approximate are not equal, \(X\) be a rough set for \(R\). Simply denoted by \(\overline{R}, \overline{R}; Bn_R(X) = \overline{R}(X) - \overline{R}(X)\) be called \(R\) boundary of \(X\); \(POS_R(X) = \overline{R}(X)\) be called \(R\) positive area of \(X\), and namely the positive area of rough set; \(Neg_R(X) = U - \overline{R}(X)\) be called the negative area of \(X\).

### 3. Rough Vague Sets (RV Sets) and Rough Vague Value (RV Value)

**Definition 3.1.** Let \(U\) be a universe of discourse, \(R\) is an equivalent relation, \(V\) is a vague set. The rough vague sets constituted by \(R\) with \(V\) (the RV sets) defines as below:

\[
\overline{R}(V) = \inf\{t_v(x)|x \in [x]_R\}; \overline{R}(V) = \sup\{t_v(x)|x \in [x]_R\};
\]

\[
\overline{R}(V) = \sup\{f_v(x)|x \in [x]_R\}; \overline{R}(V) = \inf\{f_v(x)|x \in [x]_R\}.
\]

Here, \(\overline{R}, \overline{R}\) is the least and the biggest value of the true membership degree at same equivalent class, \(\overline{R}, \overline{R}\) is the least and the biggest value of the false membership degree at the same equivalent class. Up and down approximate vague set is represented by \(\overline{V} = [\overline{R}(V), 1 - \overline{R}(V)]\) and \(\overline{V} = [\overline{R}(V), 1 - \overline{R}(V)]\), then \(V = (\overline{V}, \overline{V})\) be called a rough vague set.

**Definition 3.2.** For \(X \subseteq U\), \(V = (\overline{V}, \overline{V})\) is a rough vague set of which is constituted by \(R\) and \(V\), \(\forall x \in X\), define \(\overline{V}(x) = [\overline{R}(x), 1 - \overline{R}(x)]\), and
\[ V(x) = [R_t(x), 1 - R_f(x)], \text{ then } (\mu_Rt(x), 1 - R_f(x)), (\overline{Rf}(x), 1 - R_f(x))] \text{ is called the rough vague value of } V = (V, \overline{V}), \text{ simply denoted by } x, \text{ i.e. } x = (\mu_Rt(x), 1 - R_f(x)), (\overline{Rf}(x), 1 - R_f(x)).

**Definition 3.3.** For two rough vague values \( x = (\mu_Rt(x), 1 - R_f(x)), (\overline{Rf}(x), 1 - R_f(x)) \text{ and } y = (\mu_Rt(y), 1 - R_f(y)), (\overline{Rf}(y), 1 - R_f(y)) \), \( x \) and \( y \) be called equal, if and only if \( R_t(x) = R_t(y), R_f(x) = R_f(y) \text{ and } \overline{Rf}(x) = \overline{Rf}(y) \).

**Definition 3.4.** The complement of \( x = (\mu_Rt(x), 1 - R_f(x)), (\overline{Rf}(x), 1 - R_f(x)) \) denoted by \( x^c \), if \( R_t(x^c) = R_f(x), R_f(x^c) = R_t(x), \overline{Rf}(x^c) = \overline{Rf}(x) \), \( \overline{Rf}(x^c) = \overline{Rf}(x) \). The complement of the rough vague set \( V = (V, \overline{V}) \) is denoted by \( V^c = (\overline{V}, \overline{\overline{V}}) \).

**Definition 3.5.** \( V = (V, \overline{V}) \subseteq (W, \overline{W}) \) if and only if \( V \subseteq W \text{ and } \overline{V} \subseteq \overline{W} \), i.e. \( R_{fw}(x) \subseteq R_{fw}(x) \), \( R_{fw}(x) \geq R_{fw}(x) \text{ and } \overline{R_{fw}}(x) \leq \overline{R_{fw}}(x), \overline{R_{fw}}(x) \geq \overline{R_{fw}}(x) \).

**Theorem 3.1.** Let \( V = (V, \overline{V}) \) and \( W = (W, \overline{W}) \) be two rough vague sets defined by Definition 3.1. We have:

1. \( V \subseteq V \subseteq \overline{V} \);
2. \( V \cup W = V \cup W, V \cap W = V \cap W; \)
3. \( V \cup W \subseteq V \cup W, V \cap W \subseteq V \cap W \).

For the proof see [6].

4. Similarity Measurement of Rough Vague Value

In order to study the rough vague value, we first analyze the situation of vague set. In the similarity measure of vague set, people usually adopt the similarity measure method of vague value of vague set.

For a vague value \( x = [t(x), 1 - f(x)] \), people use \( S(x) = t(x) - f(x) \) as the score of \( x \); \( \phi(x) = \frac{1}{2}(1 - f(x) + t(x)) \) as the middle point of \( x \); with \( \pi(x) = 1 - f(x) - t(x) \) as the length (the interval length) of \( x \). Use \( \pi(x) \) represent the unknowable degree of vague set \( V \). Correspondently, its knowable degree can be described by \( K(x) = 1 - \pi(x) = f(x) + t(x) \), it can also reflect the degree of supporting. Obviously, we have the following properties:

1. \( S(x) \in [-1, 1]; \)
2. \( |\phi(x) - \phi(y)| = \frac{1}{2}|S(x) - S(y)| \in [0, 1]; \)
3. \( |\pi(x) - \pi(y)| = |K(x) - K(y)| \in [0, 1]; \)
4. \( |K(x) - K(y)| = \frac{1}{2}|\phi(x) - \phi(y)| \in [0, 1]. \)

In generally speaking the characteristics of an interval has four important parameters: the left (right) point, the interval length and the middle point.
From the voting model we can see that the vague value reflected three kinds of information, namely “approven number”, “opposed number” and “abstain number”. Therefore we should consider these information and the approve tendency information when we have to measure the similarity of two vague values. In essential speaking, there are three characteristics should be considered in measuring the similarity degree of vague sets (value): comparative advantage. Under the sameness condition, the smaller comparative advantage the vague value \( x \) possesses than vague value \( y \) then the bigger the similarity degree of \( x \) and \( y \). 2) How much the comparative known information is. Under the sameness condition, the smaller known information the vague value \( x \) possesses than vague value \( y \), then the bigger the similarity degree of \( x \) and \( y \). 3) How much the comparative unknown information is. Under the sameness condition, the smaller unknown information the vague value \( x \) possesses than vague value \( y \), then the bigger the similarity degree of \( x \) and \( y \).

For two vague values \( x = [t(x), 1 - f(x)] \) and \( y = [t(y), 1 - f(y)] \), Chen [7] considered the middle point, gave the following similarity measure formula:

\[
M_c(x, y) = 1 - \frac{1}{2} |S(x) - S(y)| = 1 - |\phi(x) - \phi(y)|.
\]

Hong and Kim [8] gave the similarity measure like below:

\[
M_{HK}(x, y) = 1 - [|t(x) - t(y)| + |f(x) - f(y)|] .
\]

Li Fan and Xu Zhangyan [9] gave the formula of similarity measure based on interval end point:

\[
M_{LX}(x, y) = 1 - \frac{1}{4} [|S(x) - S(y)| + |t(x) - t(y)| + |f(x) - f(y)|] .
\]

In fact, \( M_{LX}(x, y) = \frac{1}{2}[M_c(x, y) + M_{HK}(x, y)] \).

Fan Jiulun [10] colligated the formula mentioned above and gave the following similarity measure formula:

\[
M_F(x, y) = 1 - [a|t(x) - t(y)| + b|f(x) - f(y)| + c|\pi(x) - \pi(y)| + d|\phi(x) - \phi(y)|].
\]

In the formula not only the value of a, b, c and d are nonnegative, but also the value of \( M_F(x, y) \) must be kept nonnegative by them, so this brought inconvenience in application.

According to above analysis, we first modify the definition of similarity measure about two vague values as follows.
**Definition 4.1.** Let \( x = [t(x), 1 - f(x)] \) and \( y = [t(y), 1 - f(y)] \) be two vague values, then
\[
M(x, y) = 1 - a|t(x) - t(y)| - b|f(x) - f(y)| - c|\pi(x) - \pi(y)| - d|\phi(x) - \phi(y)|
\]
is a kind of measure of the vague value \( x \) and \( y \), where, \( a, b, c, d \geq 0 \), and \( a + b + c + d = 1 \).

So the similarity measure of two vague sets \( V \) and \( W \) defined on finite universe of discourse \( X \), its similarity measure can be given as below.

**Definition 4.2.** Let \( X = (x_1, x_2, \cdots, x_n) \) be finite universe of discourse, \( V \) and \( W \) be two vague sets, then
\[
M(V, W) = \frac{1}{n} \sum_{l=1}^{n} \{1 - a|t_v(x_l) - t_w(x_l)| - b|f_v(x_l) - f_w(x_l)| - c|\pi_v(x_l) - \pi_w(x_l)|
- d|\phi_v(x_l) - \phi_w(x_l)|\} \]
is a kind of measure of the vague sets \( V \) and \( W \), where, \( a, b, c, d \geq 0 \), and \( a + b + c + d = 1 \).

According to the above discussion, for a rough vague \( X \subseteq U \) a rough vague \( V = (V, \overline{V}) \) constituted by \( R \) with \( V \) set, its rough vague value \( x = ([Rt(x), 1 - Rf(x)],[\overline{Rt}(x), 1 - \overline{Rf}(x)]) \) noted \( S(x) = Rt(x) - Rf(x) \), \( \overline{S}(x) = \overline{Rt}(x) - \overline{Rf}(x) \), \( S(x) = \alpha_1 S(x) + \alpha_2 \overline{S}(x) \), where, \( 0 \leq \alpha_1, \alpha_2 \leq 1 \), and \( \alpha_1 + \alpha_2 = 1 \).

Then, \( S(x) \) can be employed to represent the record of the rough vague value \( x \). Obviously \( S(x) \in [-1, 1] \).

Concerning the middle point of \( x \), we denote
\[
\phi(x) = \frac{1}{2}[1 - Rf(x) + Rt(x)], \overline{\phi}(x) = \frac{1}{2}[1 - \overline{Rf}(x) + \overline{Rt}(x)], \phi(x) = \beta_1 \phi(x) + \beta_2 \overline{\phi}(x),
\]
where, \( 0 \leq \beta_1, \beta_2 \leq 1 \), and \( \beta_1 + \beta_2 = 1 \).

Note
\[
\overline{\pi}(x) = 1 - Rf(x) - Rt(x), \pi(x) = 1 - \overline{Rf}(x) - \overline{Rt}(x), \pi(x) = \gamma_1 \overline{\pi}(x) + \gamma_2 \pi(x), \overline{K}(x) = \overline{Rt}(x) + Rf(x), \overline{K}(x) = \overline{Rt}(x) + \overline{Rf}(x), K(x) = \gamma_1 K(x) + \gamma_2 \overline{K}(x),
\]
where \( 0 \leq \gamma_1, \gamma_2 \leq 1 \), and \( \gamma_1 + \gamma_2 = 1 \).

Then, \( \pi(x) \) can be considered as the unknown degree of rough value \( x \), \( K(x) \) is the known degree of \( x \).
For this, we first give the concept of similarity measure between rough vague values.

**Definition 4.3.** Let X be an unempty set, R be an equivalent relation on X, Λ represents the set which includes all of rough vague values. \( \forall x, y \in \Lambda \), mapping \( M : \Lambda \times \Lambda \rightarrow [0, 1] \), if \( M(x, y) \) satisfied the following properties:

1. \( M(x, y) \in [0, 1] \);
2. \( M(x, y) = M(y, x) \);
3. if \( x = y \), then \( M(x, y) = 1 \).

Then, \( M(x, y) \) is called the similarity measure of rough vague value \( x \) and \( y \).

The similarity measure of rough vague value have the following conclusions:

**Theorem 4.1.** Let two rough vague values \( x = \langle [R_f(x), 1 - R_f(x)], [R_t(x), 1 - R_t(x)] \rangle \) and \( y = \langle [R_f(y), 1 - R_f(y)], [R_t(y), 1 - R_t(y)] \rangle \),

- \( M_1(x, y) = 1 - \frac{1}{2}([S(x) - S(y)] + [\overline{S}(x) - \overline{S}(y)]) = 1 - \frac{1}{2}[\phi(x) - \phi(y)] + |\phi(x) - \phi(y)|; \)
- \( M_2(x, y) = 1 - \frac{1}{2}([R_t(x) - R_t(y)] + [\overline{R_t}(x) - \overline{R_t}(y)] + [R_f(x) - R_f(y)] + |R_f(x) - R_f(y)|); \)
- \( M_3(x, y) = 1 - \frac{1}{2}[a|S(x) - S(y)| + 2b|K(x) - K(y)| + c|[R_t(x) - R_t(y)] + |R_f(x) - R_f(y)| + d|[\overline{R_t}(x) - \overline{R_t}(y)] + |\overline{R_f}(x) - \overline{R_f}(y)|]; \)
- \( M_4(x, y) = 1 - \frac{1}{2}[a|S(x) - S(y)| + 2b|K(x) - K(y)| + c|[R_t(x) - R_t(y)] + |R_f(x) - R_f(y)| + d|[\overline{R_t}(x) - \overline{R_t}(y)] + |\overline{R_f}(x) - \overline{R_f}(y)|]. \)

Here, \( a, b, c, d \geq 0 \), and \( a + b + c + d = 1 \). It reflects the tendency of various of information. Then \( M_1(x, y), M_2(x, y), M_3(x, y), M_4(x, y) \) are all the similarity measure of \( x \) and \( y \). Obviously, \( M_1(x, y) \) is based on the middle point, \( M_3(x, y) \) is based on end point, \( M_3(x, y) \) is based on middle point and end point, \( M_4(x, y) \) is comprehensive measure. Under the different application background they all have their own advantage. And it is very clear that

\[ M_3(x, y) = \frac{1}{3}M_1(x, y) + \frac{2}{3}M_2(x, y). \]

**Proof.** Only proof \( M_4(x, y) \), the others are similar.

First, since \( S(x) \in [-1, 1] \), then \( |S(x) - S(y)| \in [0, 2] \), \( R_t(x) \in [0, 1] \), \( R_f(x) \in [0, 1] \), \( \overline{R_t}(x) \in [0, 1] \), \( \overline{R_f}(x) \in [0, 1] \), so, \( 2|K(x) - K(y)| \in [0, 2] \), \( |R_t(x) - \overline{R_t}(y)| \in [0, 1] \), \( |R_f(x) - \overline{R_f}(y)| \in [0, 1] \), \( |\overline{R_t}(x) - \overline{R_t}(y)| \in [0, 1] \), \( |\overline{R_f}(x) - \overline{R_f}(y)| \in [0, 1] \), and \( a + b + c + d = 1 \), so \( M_4(x, y) \in [0, 1] \). In Definition 4.3 (2) and (3) is obviously.
Obviously conditions (2) and (3) in Definition 4.3 hold, then by Definition 4.3 $M_4(x, y)$ is a kind of similarity measure of rough vague value $x$ and $y$.

**Theorem 4.2.** In Theorem 4.1, $M(x, y)$ has the following properties:

1. $M(x^c, y^c) = M(x, y)$
2. When $x = ([0, 0], [0, 0]), \ y = ([1, 1], [1, 1])$ or $y = ([0, 0], [0, 0]), x = ([1, 1], [1, 1]) M(x, y) = 0$.

Proof. (1) For arbitrarily rough vague value $x$, we have $\overline{Rf}(x^c) = Rf(x)$, $\overline{Rf}(x^c) = \overline{Rf}(x)$ and $\underline{Rf}(x^c) = \underline{Rf}(x)$, so we have $S(x^c) = -S(x), |S(x^c) - S(y^c)| = |S(x) - S(y)|, |\overline{Rf}(x^c) - \overline{Rf}(y^c)| = |\overline{Rf}(x) - \overline{Rf}(y)|, |\underline{Rf}(x^c) - \underline{Rf}(y^c)| = |\underline{Rf}(x) - \underline{Rf}(y)|$.

Hence, $M(x^c, y^c) = M(x, y)$.

(2) We can get property (2) directly by the definition. \hfill $\square$

5. Similarity Measure of Rough Vague Sets

**Definition 5.1.** Let $X$ be a unempty set, $R$ is an equivalent relation defined on $X$, $\Lambda$ represents the set which includes all of rough vague sets. $\forall V, W \in \Lambda$, $M : \Lambda \times \Lambda \to [0, 1]$, if $M(V, W)$ satisfied the following properties:

1. $0 \leq M(V, W) \leq 1$;
2. if $v = w$, then $M(V, W) = 1$;

Then $M(V, W)$ is called the similarity degree of rough vague sets $V$ and $W$. Here, $V = ([\overline{Rf}(V), 1 - \overline{Rf}(V)], [\overline{Rf}(V), 1 - \overline{Rf}(V)])$ and $W = ([\overline{Rf}(W), 1 - \overline{Rf}(W)], [\overline{Rf}(W), 1 - \overline{Rf}(W)])$.

Let $X = (x_1, x_2, \ldots, x_n), \forall x_l \in X$, rough vague sets $V$ and $W$ noted by:

$V = (\overline{V}, \underline{V}) = ([\overline{Rf}(x_l), 1 - \overline{Rf}(x_l)], [\overline{Rf}(x_l), 1 - \overline{Rf}(x_l)])$;

$W = (\overline{W}, \underline{W}) = ([\overline{Rf}(x_l), 1 - \overline{Rf}(x_l)], [\overline{Rf}(x_l), 1 - \overline{Rf}(x_l)])$.

Then have the following theorem.

**Theorem 5.1.** Let $X = (x_1, x_2, \ldots, x_n)$ be finite, above mentioned given two rough vague sets, its similarity measure be given as following:

\[
M(V, W) = \frac{1}{n} \sum_{l=1}^{n} M[V(x_l), W(x_l)]
= \frac{1}{n} \sum_{l=1}^{n} \left\{ 1 - \frac{1}{2} [a S_{vw}(x_l) + 2b K_{vw}(x_l) + c R_{vw}(x_l) + d \overline{R}_{vw}(x_l)] \right\}.
\]
Here $S_{vw}(x_l) = |S_v(x_l) - S_w(x_l)|$, $K_{vw}(x_l) = |K_v(x_l) - K_w(x_l)|,$

$$R_{vw}(x_l) = |R_{tv}(x_l) - R_{tw}(x_l)| + |R_{fv}(x_l) - R_{fw}(x_l)|,$$

$$R_{vw}(x_l) = |R_{tv}(x_l) - R_{tw}(x_l)| + |R_{fv}(x_l) - R_{fw}(x_l)|, \quad i = 1, 2, \ldots, n,$$

$$a, b, c, d \geq 0, \text{ and } a + b + c + d = 1.$$

Then $M(V, W)$ is a kind of the similarity measure of rough vague sets $V$ with $W$.

Proof. The underneath proves (1) only. In fact, since $S_{vw}(x_l) = |S_v(x_l) - S_w(x_l)| \in [0, 2]$; $K_{vw}(x_l) = |K_v(x_l) - K_w(x_l)| \in [0, 1]$; $R_{vw}(x_l) = |R_{tv}(x_l) - R_{tw}(x_l)| + |R_{fv}(x_l) - R_{fw}(x_l)| \in [0, 2]$; and $a + b + c + d = 1$. So, $0 \leq aS_{vw}(x_l) + 2bK_{vw}(x_l) + cR_{vw}(x_l) + dR_{vw}(x_l) \leq 2a + 2b + 2c + 2d = 2$. Hence, $0 \leq 1 - \frac{1}{2}[aS_{vw}(x_l) + 2bK_{vw}(x_l) + cR_{vw}(x_l) + dR_{vw}(x_l)] \leq 1, i = 1, 2, \ldots, n.$

Therefore, $0 \leq M(V, W) \leq 1$.

According to Definition 5.1, $M(V, W)$ is a kind of the similarity measure of rough vague sets.

6. Conclusion

Rough vague set theory is a kind of new mathematical tool of handling uncertain information; it is very useful in processing the knowledge (concept) that undistinguishable and fuzzy. After synthesizing various circumstances that should be considered, this paper introduces a new method of the similarity measurement of rough vague set and provides theory foundation in the application field of the rough vague set.

References


