The Improved Similarity Measures of Vague Sets under the Influence of the Waiver

Feng Weibing¹, Guo Luna¹*, Wang Pan²

¹College of Science, Xi’an University of Science & Technology, 710054 Xi’an, China
²Xi’an Sino Huaxin Measurement & Control Co.Ltd, 710054, China

*Corresponding author:1115859330@qq.com

Keywords: similarity measures; Vague sets; the impact of waiver; pattern recognition

Abstract: The similarity measure of vague sets is an important content for studying. A reasonable structure and the way to select a similarity measure will directly affect the classification of large data and the correct judgment of the data relationships. Based on the comparison and analysis of the existing similarity measures, and aiming at the shortcomings that it does not take into the abstained part to the impact of the similarity measures and cannot effectively distinguish the data in some existing similarity measures, this paper investigates the similarity measure of vague sets (values), proposes new formulas to calculate the similarity measure of vague sets (values) and proves the completeness of the definition in theory. The distinguishability and rationality of the improved similarity measures are also illustrated by experimental data and pattern recognition.

1 Introduction

People will face more and more imprecise or fuzzy information with the development of society and science technology. It is crucial for individuals and businesses to correctly judge the relationship between things. However, Vague set theory provides an important research tool for people to express some objective and uncertain information [1]. The advantage of the Vague sets theory is that it is more effective, reliable and accurate to solve the fuzzy information. Therefore, Vague sets theory has an important significance in the fields of fuzzy control, pattern recognition and artificial intelligence.

A reasonable structure and the way to select a similarity measure will directly affect the classification of large data and the correct judgment of the data relationships. In recent years, the study of similarity measures on Vague sets are mainly based on the following two aspects: On the one hand, we are studying the thought of the true-membership functions and false-memberships functions on Vague sets. This idea can be reflected by the relative superiority, relative known information and relative unknown information. For example, in [5,11], similarity measure of Vague sets is constructed by the relative superiority and relative known information. In 2008, Wang Weiping et al. [9] classified the types of similarity measures, pointed out their drawbacks and improved them. In 2013, Peng Zuming et al. [11] put forward the new method of Vague sets to overcome the existing Vague set model which did not accord with people's intuition in practical application. On the other hand, the research of vague sets is based on the idea of distance. In 2010, Liu Li et al. [6] used the distance to reflect the size of the similarity measures of Vague sets in space. At the same time, in [12], Wang et al. presented similarity measures on Vague sets based on the idea of distance.

Based on the analysis of the experimental data by dynamic random generation, some exiting similarity measures of Vague sets cannot be reasonable and effective to distinguish the data, as well as the impact of the waiver on similarity measure is not considered. According to the above reasons, this paper gives the improved similarity measure on Vague sets(values) to overcome the drawbacks of the existing similarity measures, and proposes a new weighted similarity measure on Vague sets(values). The improved similarity measure has better validity and distinction through the analysis of the experimental data and the theoretical proof.
2 Preliminaries

2.1 The basic theory of Vague set

**Definition 1**[7]: A Vague set $A$ in the universe of discourse $U = \{x_1, x_2, ..., x_n\}$ can be expressed by a truth-membership function $t_A$ and a false-membership function $f_A$.

$$t_A : U \rightarrow [0,1]$$

$$f_A : U \rightarrow [0,1]$$

Where $t_A(x_i)$ is a lower bound on the grade of membership of $x_i$ derived from the evidence for $x_i$; $f_A(x_i)$ is a lower bound on the negation of $x_i$ derived from the evidence against $x_i$, and $0 \leq t_A(x_i) + f_A(x_i) \leq 1$. The grade of membership of $x_i$ in the Vague set $A$ is the bounded by the subinterval $[t_A(x_i), 1 - f_A(x_i)]$ can simply express by $[t_A, 1 - f_A]$.

**Definition 2**[5]: Let $A$ be a Vague set in the universe of discourse $U = \{x_1, x_2, ..., x_n\}$ and let $x, y$ be two Vague values in such that $x = [t_x, 1 - f_x], y = [t_y, 1 - f_y]$. Then $s_x = 1 - t_x - f_x$ indicates the degree of $x$ for the Vague set $A$. It expresses the degree of $x$ for the unknown information of the vague set $A$. When the value of $s_x$ is larger, the greater the degree of unknown information is. And $s_y$ also indicates the degree of $y$ for the Vague set $A$. Especially for $s_x = 1, s_y = 1$, it means that the values $x$ and $y$ are not known at all.

**Criterion 1**[6]: For two Vague sets $A, B$, if the function $M(A, B)$ satisfies the following properties:

1. $0 \leq M(A, B) \leq 1$
2. $M(A, B) = 1$ only if $A = B$
3. $M(A, B) = M(B, A)$

Then $M(A, B)$ is called a similarity measure of the Vague sets $A$ and $B$.

2.2 Conditions of similarity measure for Vague sets (values)

Similarity measure of Vague sets(values) must have the following three conditions [7]:

(1) Relative superiority.
Vague set $A$ than Vague set $B$ has a smaller relative superiority, it shows that the similarity measure on Vague sets $A$ and $B$ is greater, if the other conditions are the same.

(2) Relatively known information
Vague set $A$ than Vague set $B$ has a smaller known information, it shows that the similarity measure of Vague set $A$ and $B$ is greater if the other conditions are the same.

(3) Relatively unknown information
Vague set $A$ than Vague set $B$ has a smaller unknown information, it shows that the similarity measure of Vague set $A$ and $B$ is greater if the other conditions are the same.

3 Existing similarity measures of Vague sets

Let $x = [t_x, 1 - f_x], y = [t_y, 1 - f_y]$ be two Vague values, where $t_x \in [0,1], t_y \in [0,1]$ and $0 \leq t_x + f_x \leq 1$. The following definitions are the similarity measures between Vague values $x$ and $y$.

$$M_h(x, y) = 1 - \frac{|t_x - t_y| + |f_x - f_y|}{2}$$

$$M_i(x, y) = 1 - \frac{|t_x - t_y - (f_x - f_y)|}{4}$$

$$M_q(x, y) = 1 - \frac{|t_x - t_y| + |f_x - f_y| + |S_x - S_y| + |\pi - \pi|}{4}$$

Advances in Economics, Business and Management Research, volume 30
D H Hong et al. proposed the definition of the similarity measure (Eq. (3)), Although this method takes into account the relative known information, but there is no consideration to the two conditions that are relatively unknown information and relative superiority. However, Li Fan’s method (Eq. (4)), which takes into account the impact of relative superiority to the data. There is a good distinction when this method deals with the same data for relatively unknown information (For example, the similarity between [0.2,0.5] and [0.4,0.7] is equal to 0.9000, and the similarity between [0.3,0.6] and [0.4,0.7] is equal to 0.9500.). However, this method is not applicable when it deals with the data which is not equal to the unknown information. Wang Weiping’s method (Eq. (5)), which overcomes the shortcomings of Li Fan’s method. But this method does not take into account the impact of the waiver, so that there is no obvious distinction between some fuzzy information. Zhang Chengyi et al. proposed the definition of the similarity measure (Eq. (6)) which takes into account the second impact of the waiver on the known information. This definition has a better distinction between the second impact of the known information on the waiver of the data. It is difficult to effectively distinguish between the data if you don't have to sort out the complicated data in advance According to the distance of two points in two-dimensional space, Huang Guoshun et al. deduced the definition of similarity measure (Eq. (7)).

4 Improved similarity measure method of Vague sets (or values)

4.1 Improved similarity measure method of Vague sets(values)

**Definition 4:** Let \( x = [t_x, 1 - f_x], y = [t_y, 1 - f_y] \) be two Vague values, where \( t_x, t_y \in [0,1] \) and \( 0 \leq f_x + f_y \leq 1 \). \( t_x = 1 - f_x, t_y = 1 - f_y, s_x \) and \( s_y \) express the measure of \( x, y \) for the unknown information of the vague values. Then \( M(x, y) \) is the improved similarity measure method on Vague values as follows.

\[
M(x, y) = 1 - \frac{16}{\sqrt{\| x - y \|^2 + \| f_x - f_y \|^2}} \left( \sum_{i=1}^{n} \left( t_x + t_y, s_x + s_y, (f_x + f_y, s_x + s_y) \right) \right) \quad (8)
\]

The definition of the improved similarity measure on the Vague values (Eq. (8)) is constructed by taking into account the factors of the waiver. For the factors of the waiver, we can use the voting model to explain. In the first vote, there is a vote of support, opposition, and abstention. In the second votes, it will be based on the results of the first vote in the proportion of the votes. In the second vote, the person who abstains from the first vote will be based on the proportion of the votes in the results of the first vote. According to the definition of similarity measure of Vague value (Eq. (8)), the similarity measure between Vague sets can be introduced.

**Definition 5:** Let \( A \) and \( B \) be two vague sets of the universe \( U = \{x_1, x_2, ..., x_n\} \). \( V_a = [t_a, 1 - f_a] \) and \( V_b = [t_b, 1 - f_b] \) for \( i = 1, 2, ..., n \) where \( x_i, y_i \) are the Vague values in the Vague sets \( A, B \). The degree of similarity between the Vague sets \( A \) and \( B \) can be evaluated by the function \( M(A, B) \).

\[
M(A, B) = 1 - \frac{\sum_{i=1}^{n} \left( \| t_a - t_b \| + \| t_b - t_a \| \right) \left( f_a - f_b \right) \left( s_a - s_b \right)}{\left( f_a + f_b \right) \left( s_a + s_b \right)} \quad (9)
\]
Definition 6: The weight of values $x_i$ is $w_i$, where $w_i \geq 0$ and $\sum_{i=1}^{n} w_i > 0$. The degree of similarity between the Vague sets $A$ and $B$ can be evaluated by the function $T(A,B)$.

$$T(A,B) = 1 - \frac{\sum_{i=1}^{n} w_i M([t_{i_{v}},1-f_{i_{v}}],[t_{i_{v}},1-f_{i_{v}}])}{\sum_{i=1}^{n} w_i}$$

(10)

By the applying Eq.(9) into Eq.(10), we can get the following function:

$$T(A,B) = 1 - \frac{\sum_{i=1}^{n} w_i \left[ |t_{i_{v}}-t_{i_{v}}+t_{i_{v}}s_{i}-t_{i_{v}}s_{i}-(f_{i_{v}}-f_{i_{v}}+f_{i_{v}}s_{i}-f_{i_{v}}s_{i})| + |(t_{i_{v}}+t_{i_{v}}s_{i})-(t_{i_{v}}+t_{i_{v}}s_{i})+(f_{i_{v}}+f_{i_{v}}s_{i})-(f_{i_{v}}+f_{i_{v}}s_{i})| \right]}{\sum_{i=1}^{n} w_i}$$

(11)

Where $T(A,B) \in [0,1]$, the large the values of $T(A,B)$, the more the similarity between the Vague sets $A$ and $B$.

4.2 Proof of the definition

(1) $0 \leq M(x,y) \leq 1$

$M(x,y) = 1 - \frac{\sum_{i=1}^{n} w_i \left[ |t_{i_{v}}-t_{i_{v}}+t_{i_{v}}s_{i}-t_{i_{v}}s_{i}-(f_{i_{v}}-f_{i_{v}}+f_{i_{v}}s_{i}-f_{i_{v}}s_{i})| + |f_{i_{v}}+f_{i_{v}}s_{i}-f_{i_{v}}+f_{i_{v}}s_{i}| \right]}{\sum_{i=1}^{n} w_i}$

(12)

$1 - \frac{0}{16} - \frac{0}{16} \leq 1$; $t_{i_{v}}, t_{i_{v}} \in [0,1]$, $f_{i_{v}}, f_{i_{v}} \in [0,1]$, $s_{i}, s_{i} \in [0,1]$ then: $|t_{i_{v}}-t_{i_{v}}+t_{i_{v}}s_{i}-t_{i_{v}}s_{i}-(f_{i_{v}}-f_{i_{v}}+f_{i_{v}}s_{i}-f_{i_{v}}s_{i})| \leq 4$, in the same way: $|f_{i_{v}}+f_{i_{v}}s_{i}-f_{i_{v}}+f_{i_{v}}s_{i}| \leq 4$, then we have: $M(x,y) \geq 1 - \frac{4}{16} - \frac{4}{16} = 0$. Therefore, we have: $0 \leq M(x,y) \leq 1$.

(2) When $x = y$, we can know the $M(x,y) = 1$ by the definition. When $M(x,y) = 1$, we have:

$$M(x,y) = \frac{\sum_{i=1}^{n} w_i \left[ |t_{i_{v}}-t_{i_{v}}+t_{i_{v}}s_{i}-t_{i_{v}}s_{i}-(f_{i_{v}}-f_{i_{v}}+f_{i_{v}}s_{i}-f_{i_{v}}s_{i})| + |f_{i_{v}}+f_{i_{v}}s_{i}-f_{i_{v}}+f_{i_{v}}s_{i}| \right]}{\sum_{i=1}^{n} w_i} = 0$$

(13)

then $t_{i_{v}}+t_{i_{v}}s_{i} = t_{i_{v}}+t_{i_{v}}s_{i} = f_{i_{v}}+f_{i_{v}}s_{i}$ and when $t_{i_{v}} = t_{i_{v}}$, $f_{i_{v}} = f_{i_{v}}$, we have: $x = y$. Therefore, we have $M(x,y) = 1$ only if $x = y$.

(3) $M(x,y) = M(y,x)$. By definition 4 can be clearly demonstrated.

From the proof we can know, the definition of improved similarity measures (Eq. (8)) satisfies the three conditions of the criterion 1, therefore it is complete.

4.3 Comparison of several similarity measures

Nine groups of experimental data with dynamic random generation can be applied to the similarity measure (Eq. (3)—(8),(10)). The calculation results of several existing measures and improved similarity measures are listed in Table1.
Table 1. Comparison of similarity measures.

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>[0.2,0.5]</td>
<td>[0.3,0.5]</td>
<td>0.9500</td>
<td>0.9250</td>
<td>0.9500</td>
<td>0.9293</td>
<td>0.9250</td>
<td>0.9750</td>
</tr>
<tr>
<td>2</td>
<td>[0.4,0.8]</td>
<td>[0.4,0.9]</td>
<td>0.9500</td>
<td>0.9750</td>
<td>0.9500</td>
<td>0.9000</td>
<td>0.9150</td>
<td>0.9675</td>
</tr>
<tr>
<td>3</td>
<td>[0.3,0.8]</td>
<td>[0.4,0.7]</td>
<td>1.0000</td>
<td>0.9000</td>
<td>0.9000</td>
<td>0.8775</td>
<td>0.9200</td>
<td>0.9775</td>
</tr>
<tr>
<td>4</td>
<td>[0.3,0.6]</td>
<td>[0.4,0.7]</td>
<td>0.9500</td>
<td>0.9000</td>
<td>0.9000</td>
<td>0.8775</td>
<td>0.8700</td>
<td>0.9925</td>
</tr>
<tr>
<td>5</td>
<td>[0.2,0.5]</td>
<td>[0.4,0.4]</td>
<td>1.0000</td>
<td>0.8250</td>
<td>0.8500</td>
<td>0.8268</td>
<td>0.9050</td>
<td>0.9650</td>
</tr>
<tr>
<td>6</td>
<td>[0.2,0.4]</td>
<td>[0.3,0.4]</td>
<td>0.9500</td>
<td>0.9250</td>
<td>0.9500</td>
<td>0.9293</td>
<td>0.9250</td>
<td>0.9512</td>
</tr>
<tr>
<td>7</td>
<td>[0.4,0.8]</td>
<td>[0.5,0.8]</td>
<td>0.9500</td>
<td>0.9250</td>
<td>0.9500</td>
<td>0.9293</td>
<td>0.9450</td>
<td>0.9775</td>
</tr>
<tr>
<td>8</td>
<td>[0.4,0.81]</td>
<td>[0.5,0.81]</td>
<td>0.9500</td>
<td>0.9250</td>
<td>0.9500</td>
<td>0.9293</td>
<td>0.9450</td>
<td>0.9773</td>
</tr>
<tr>
<td>9</td>
<td>[0.21,0.51]</td>
<td>[0.31,0.51]</td>
<td>0.9500</td>
<td>0.9250</td>
<td>0.9500</td>
<td>0.9293</td>
<td>0.9260</td>
<td>0.9753</td>
</tr>
</tbody>
</table>

From table 1 we can see the following conclusion, the first group of data [0.2,0.5], we can know the certainty information (the true-membership function) $t_i = 0.2$, the negative information (false-membership function) $f_i = 0.5$, the unknown information (degree of uncertainty) $s_i = 0.3$. According to the results of Li Fan's method, the similarity measure of the first group and the sixth group data are 0.9500, but according to people's intuition, the two groups of data cannot get the same similarity measure in the case of unknown information for 0.3 and 0.2. Analysis of results can be seen that Li Fan's method cannot effectively distinguish between the second and fourth groups data. According to people's intuition, the similarity between the fifth group data cannot be 1. Wang Weiping's method cannot be effective to distinguish the third group and fourth group data. The first group, the second group and the sixth group data cannot be effectively distinguished by D H Hong's method. The third group and fourth group data cannot be effectively distinguished by Huang Guoshun's method. The first group and sixth group data cannot be effectively distinguished by Zhang Chengyi's method. The similarity measures given by Li et al, it is difficult to distinguish between the seventh group and eighth group data when the certainty information is equal and the negative information is very small. The methods of Lee et al. are limited, cannot make a reasonable distinction between the data when the certainty information and negative information simultaneously produce the tiny fluctuation, for example the first group and ninth group data.

The improved similarity measure (Eq. (8)) overcomes the drawbacks of the above existing similarity measure, is more effective to distinguish the data from table 1. There are two characteristics for the nine groups of data by dynamic random generation: arbitrary and completeness, so that the data can avoid the simple artificial structure from table 1. The completeness of the data through the three aspects of the certainty information, negative information and uncertain information, which makes the nine groups of data more representative. So we can see that the improved similarity measure method is more effective and reliable for the data.

By applying Eq. (10). Assume that the weights of $\omega_1, \omega_2$ and $\omega_3$ are $2/4, 1/4$ and $1/4$ respectively in the scheme 1. We assume that the weights of $\omega_1, \omega_2$ and $\omega_3$ are $1/4, 1/4$ and $2/4$ in the scheme 2. We assume that the weights of $\omega_1, \omega_2$ and $\omega_3$ are $1/4, 2/4$ and $1/4$ in the scheme 3. The results of calculation show that the conclusion of pattern recognition will be different when the weight coefficient $\omega_1$ is different. In the scheme 1, the weighted similarity measure has a better ability to distinguish between the nine groups data when the weighted coefficient is reasonable. We can see
that the first group and the third group data cannot be reasonably distinguished from the scheme 2. The sixth group and the seventh group data cannot be distinguished from the scheme 3. Therefore, it is crucial to select the weighted coefficients for the distinction and reliability of the data.

5 Application

The specific methods of similarity measure in pattern recognition such as the literature [15]. Assume that \( m \) patterns exist which are represented by the Vague sets \( A_i \) (\( i = 1,2,\ldots,m \)). Suppose that there is a sample to be recognized which is represented by a Vague set \( B \).

\[
M(A_i, B) = \max \{M(A_i, B)\}
\]

Where \( M(A_i, B) \) is the similarity measure between the \( A_i \) and \( B \). We conclude that the sample \( B \) belongs to the pattern \( A_i \).

**Example1.** Let index set is the \( X = \{x_1,x_2,\ldots,x_n\} \), where \( A_i \) (\( i = 1,2,3 \)) is the pattern (the standard model) and \( B \) is sample. The calculation results of the similarity measures of Vague sets \( A_i (i = 1,2,3) \) and \( B \) are listed in Table2.

<table>
<thead>
<tr>
<th>Similarity measures</th>
<th>Similarity degree</th>
<th>Classification result</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M_1 )</td>
<td>( M(A_i, B) = 0.4333, M(A_i, B) = 0.5667, M(A_i, B) = 0.5833 )</td>
<td>( \max = M(A_i, B) )</td>
</tr>
<tr>
<td>( M_2 )</td>
<td>( M(A_i, B) = 0.6667, M(A_i, B) = 0.6833, M(A_i, B) = 0.6833 )</td>
<td>( \max = M(A_i, B) )</td>
</tr>
<tr>
<td>( M_3 )</td>
<td>( M(A_i, B) = 0.8000, M(A_i, B) = 0.8500, M(A_i, B) = 0.8910 )</td>
<td>( \max = M(A_i, B) )</td>
</tr>
<tr>
<td>( M_4 )</td>
<td>( M(A_i, B) = 0.7433, M(A_i, B) = 0.9067, M(A_i, B) = 0.9250 )</td>
<td>( \max = M(A_i, B) )</td>
</tr>
<tr>
<td>( M_5 )</td>
<td>( M(A_i, B) = 0.8350, M(A_i, B) = 0.9237, M(A_i, B) = 0.9392 )</td>
<td>( \max = M(A_i, B) )</td>
</tr>
<tr>
<td>( M_{\text{improved}} )</td>
<td>( M(A_i, B) = 0.9200, M(A_i, B) = 0.9563, M(A_i, B) = 0.9717 )</td>
<td>( \max = M(A_i, B) )</td>
</tr>
<tr>
<td>( M_{\text{weigh}} )</td>
<td>( M(A_i, B) = 0.9783, M(A_i, B) = 0.9881, M(A_i, B) = 0.9703 )</td>
<td>( \max = M(A_i, B) )</td>
</tr>
<tr>
<td>( M_{\text{ted}} )</td>
<td>( M(A_i, B) = 0.9697, M(A_i, B) = 0.9825, M(A_i, B) = 0.9855 )</td>
<td>( \max = M(A_i, B) )</td>
</tr>
</tbody>
</table>

From Table2, we can see that the sample \( B \) belongs to the pattern \( A_3 \) by applying similarity measures \( M_1, M_2, M_3 \). This result does not coincide with the one obtained in [15]. For Wang Weiping’s method can be calculated the result \( \max = M(A, B) = M(A, B) \), which can be seen sample \( B \) is classified into the pattern \( A_2 \) and \( A_3 \).

In the application of the weighted similarity measure (Eq.(10)) on the pattern recognition, it can be used to adjust the weight coefficient \( \omega \) to reflect the degree of preference on each index of model. In the scheme1, assume that the weights of \( \omega_1, \omega_2, \) and \( \omega_3 \) are 1/6, 1/6, and 4/6. In the scheme2, the weights of \( \omega_1, \omega_2, \) and \( \omega_3 \) are 1/2, 2/5 and 1/10. In the scheme3, the weights of \( \omega_1, \omega_2, \) and \( \omega_3 \) are 4/6, 1/6, and 1/6. From the calculation results, we can see that the selection of the weight coefficients will directly affect the results of the pattern recognition. The sample \( B \) belongs to the pattern \( A_3 \) by scheme1, this result coincides with the one obtained in [15]. However, the calculation results of scheme 2 and scheme 3 are \( \max = M(A, B) \), namely, the sample \( B \) should belong to the pattern \( A_3 \). This result does not coincide with the one obtained in [15].

By improved similarity measure (Eq.(9)), it can be calculated the result \( \max = M(A, B) \). Namely, the sample \( B \) should belongs to the pattern \( A_2 \) and this result coincides with the one obtained in [15]. So the improved similarity measure can overcome the drawbacks of the existing similarity measures \( M_1, M_2, M_3, M_4, M_5 \) for dealing with pattern recognition problems.
6 Conclusion

In this paper, we give the improved similarity measure (Eq. (9)) and the weighted similarity measure (Eq. (10)) based on the existing similarity measures. From the experimental results shown in Tables 1-2, we can see that the improved similarity measures get correct classification results for all tables. In other words, the improved similarity measure is the best similarity measure among the existing similarity measures. It can overcome the drawbacks of the existing similarity measures for dealing with pattern recognition problems.

References