

Application of Similarity Measure between Vague Sets in Network Information Filtering

Xiaofang Fu¹, Ming Liu¹ & Jing Chen¹

¹School of Electronics and Information Engineering, Qiongzhou University, Sanya, 572022, China

Keywords: Vague sets, three-dimensional representation, fuzzy set operation, similarity measure

Abstract. This paper presents three series of formula formulas of similarity measure between Vague sets based on the three-dimensional representation of Vague values and fuzzy set operation. The method of network information filtering and searching based on the similarity measure between Vague sets and relevant examples are proposed. Examples suggest these formulas are practical.

Introduction

Network information filtering is the method and process to keep desired information or delete undesired information out of dynamic network information with related tools subject to certain standards based on user demand. As malicious network information becomes overwhelming, information filtering can be as well used as a technical means to cope with malicious information. Since Vague sets can better describe the uncertainty of objective things than classical sets and most network information are fuzzy, it's more confident and practical to describe network information with Vague sets. In subjection degree, Vague sets are more specific than the one-dimensional representation $(\mu_A(x))$ of the fuzzy set theory or the two-dimensional representation $([t_A(x), 1 - f_A(x)])$. This paper adopts the three-dimensional representation $(t_A(x), f_A(x), \pi_A(x))$, so that fuzzy information can be represented to people straightway. Vague set theory has been applied in approximate reasoning, pattern recognition, decision-making analysis and other fields of intelligent systems. Similarity measure is an indispensable tool in its application. In the similarity measure of Vague sets, the application of fuzzy operation is a practical train of thought. This paper presents three series of formulas of similarity measure between Vague sets, tries to represent network information with Vague sets and provides a method of network information filtering and searching with the similarity measure between Vague sets.

Concept of Vague Set

Let A be a Vague set on domain X , a true-membership function t_A and a false-membership function f_A can be used to describe $t_A: X \rightarrow [0, 1]$, $f_A: X \rightarrow [0, 1]$, where $t_A(x)$ is the lower bound of positive membership derived from evidences supportive of x , while $f_A(x)$ is the lower bond of negative membership derived from evidences against x ; the Vague value of x for A can be denoted as $A(x) = [t_A(x), 1 - f_A(x)]$ ($x = [t_x, 1 - f_x]$ for short), and the constraint condition $t_A(x) + f_A(x) \leq 1$ is satisfied. Let $\pi_A(x) = 1 - f_A(x) - t_A(x)$ ($\pi_x = 1 - f_x - t_x$ for short), and the three-dimensional representation of Vague value $A(x)$ will be $A(x) = (t_A(x), f_A(x), \pi_A(x))$ or $x = (t_x, f_x, \pi_x)$, where t_x , f_x and π_x are the affirmation degree, dissenting degree and uncertainty of x for A , respectively.

The two-dimensional representation and three-dimensional representation of Vague set A on the limited domain $X = \{x_1, x_2, \dots, x_m\}$ are (respectively):

$$A = \sum_{i=1}^m [t_A(x_i), 1 - f_A(x_i)] / x_i, \quad x_i \in X.$$

$$\text{Or: } A = \{[t_A(x_1), 1 - f_A(x_1)], [t_A(x_2), 1 - f_A(x_2)], \dots, [t_A(x_m), 1 - f_A(x_m)]\}.$$

$$A = \sum_{i=1}^m (t_A(x_i), f_A(x_i), \pi_A(x_i)) / x_i, \quad x_i \in X,$$

$$\text{Or: } A = \{(t_A(x_1), f_A(x_1), \pi_A(x_1)), (t_A(x_2), f_A(x_2), \pi_A(x_2)), \dots, (t_A(x_m), f_A(x_m), \pi_A(x_m))\}$$

Extension of Vague Value

Definition 1: The three-dimensional representation of the Vague value on domain X is $x = (t_x, f_x, \pi_x)$, define:

$$\begin{aligned} t_x^{(0)} &= t_x, \quad f_x^{(0)} = f_x, \quad \pi_x^{(0)} = \pi_x; \\ t_x^{(n+1)} &= t_x^{(n)} + \pi_x^{(n)} t_x, \quad f_x^{(n+1)} = f_x^{(n)} + \pi_x^{(n)} f_x, \quad \pi_x^{(n+1)} = \pi_x^{(n)} \pi_x, \quad n = 0, 1, 2, 3, \dots \end{aligned}$$

(1)

Theorem 1: if $n = 0, 1, 2, 3, \dots$:

1) $[t_x^{(n)}, 1 - f_x^{(n)}]$ is the Vague value on domain X ;

2) The three-dimensional representation of $x^{(n)}$ is

$$x^{(n)} = (t_x^{(n)}, f_x^{(n)}, \pi_x^{(n)}).$$

Note: let $x^{(n)} = [t_x^{(n)}, 1 - f_x^{(n)}]$ and the three-dimensional representation is $x^{(n)} = (t_x^{(n)}, f_x^{(n)}, \pi_x^{(n)})$, which is denoted as the n th Vague value of Vague value x after n extensions. It can be explained with the voter model that: if there is no winner for the first voting, the second round of voting should be carried out. In the second round of voting, those who abstain from voting may still vote by the ratio between affirmative votes and dissenting votes in the first round of voting, and thus $x^{(n)}$ can be got. Such idea of extension comes from Reference [3], in which the definitions are:

$$\begin{aligned} t_x^{(n)} &= t_x + t_x(1 + \pi_x + \pi_x^2 + \dots + \pi_x^n), \\ f_x^{(n)} &= f_x + f_x(1 + \pi_x + \pi_x^2 + \dots + \pi_x^n), \\ \pi_x^{(n)} &= \pi_x^{n+1}. \end{aligned}$$

It can be seen from the comparison with extension formula (1) that, the recursive definition of formula (1) can reflect the idea of three-dimensional representation in a more straight-forward way. $t_x^{(n+1)}$, $f_x^{(n+1)}$ and $\pi_x^{(n+1)}$ are deduced from $t_x^{(n)}$, $f_x^{(n)}$ and $\pi_x^{(n)}$, respectively. It's also noteworthy that formula (1) is an "exploration" into data $x = (t_x, f_x, \pi_x)$, while formula (1) represents the process of exploration. The closed interval of the two-dimensional representation of $x^{(n)}$ has the inclusion relation below:

$$x^{(+\infty)} \subseteq \dots \subseteq x^{(n+1)} \subseteq x^{(n)} \subseteq \dots \subseteq x^{(1)} \subseteq x^{(0)} \subseteq [0, 1]$$

Where $x^{(+\infty)} = [t_x^{(+\infty)}, 1 - f_x^{(+\infty)}] = [t_x / (t_x + f_x), 1 - f_x / (t_x + f_x)]$. The result of exploration is $x^{(+\infty)}$, a fuzzy value converted from Vague value x .

4. Similarity Measure between Vague Sets

Definition 2: Let a specific Vague set (three-dimensional representation) be:

$$A = \sum_{i=1}^m (t_A(x_i), f_A(x_i), \pi_A(x_i)) / x_i = \sum_{i=1}^m (t_{x_i}, f_{x_i}, \pi_{x_i}) / x_i, \quad x_i \in X,$$

$$B = \sum_{i=1}^m (t_B(x_i), f_B(x_i), \pi_B(x_i)) / x_i = \sum_{i=1}^m (t_{y_i}, f_{y_i}, \pi_{y_i}) / x_i, x_i \in X.$$

They are Vague sets on domain $X = \{x_1, x_2, \dots, x_m\}$. If $n = 0, 1, 2, \dots$, the function between A and B can be denoted as:

$$R_{k,n}(A, B) = \frac{1}{m} \sum_{i=1}^m R_{k,n}(A(x_i), B(x_i)), \quad k = 1, 2, 3.$$

Where $R_{1,n}(x_i, y_i) = (t_{x_i}^{(n)} \wedge t_{y_i}^{(n)}) + (f_{x_i}^{(n)} \wedge f_{y_i}^{(n)}) + (\pi_{x_i}^{(n)} \wedge \pi_{y_i}^{(n)}), \quad (2)$

$$R_{2,n}(x_i, y_i) = \frac{(t_{x_i}^{(n)} \wedge t_{y_i}^{(n)}) + (f_{x_i}^{(n)} \wedge f_{y_i}^{(n)})}{(t_{x_i}^{(n)} \vee t_{y_i}^{(n)}) + (f_{x_i}^{(n)} \vee f_{y_i}^{(n)})}, \quad (3)$$

$$R_{3,n}(x_i, y_i) = \frac{(t_{x_i}^{(n)} \wedge t_{y_i}^{(n)}) + (f_{x_i}^{(n)} \wedge f_{y_i}^{(n)}) + (\pi_{x_i}^{(n)} \wedge \pi_{y_i}^{(n)})}{(t_{x_i}^{(n)} \vee t_{y_i}^{(n)}) + (f_{x_i}^{(n)} \vee f_{y_i}^{(n)}) + (\pi_{x_i}^{(n)} \vee \pi_{y_i}^{(n)})} \quad (4)$$

" \vee " and " \wedge " are the maximizing and minimizing operation of the fuzzy set, i.e., if $a, b \in [0, 1]$, $a \vee b = \max\{a, b\}$, $a \wedge b = \min\{a, b\}$.

Theorem 2: If $i = 1, 2, \dots, m; n = 0, 1, 2, \dots; k = 1, 2, 3$:

- 1) $R_{k,n}(x_i, y_i) \in [0, 1]$;
- 2) $R_{k,n}(x_i, y_i) = R_{k,n}(y_i, x_i)$;
- 3) If $x_i = y_i$, $R_{k,n}(x_i, y_i) = 1$;
- 4) If $x_i = (0, 1, 0)$, $y_i = (1, 0, 0)$
or if $x_i = (1, 0, 0)$, $y_i = (0, 1, 0)$ ($1 \leq i \leq m$), $R_{k,n}(x_i, y_i) = 0$.

Definition 3: Denote the following function as the similarity measure between Vague set A and B , if $M(A, B)$ satisfies each axiom below:

$$M(A, B) = \begin{cases} \text{Arbitrary value, if } A = B = \{(0, 0, 1), (0, 0, 1), \dots, (0, 0, 1)\} \\ R(A, B), & \text{others} \end{cases}.$$

Axiom 1 Boundedness $M(A, B) \in [0, 1]$;

Axiom 2 Symmetry $M(A, B) = M(B, A)$;

Axiom 3 Boundary Condition 1:

if $A = B$ and $\neq \{(0, 0, 1), (0, 0, 1), \dots, (0, 0, 1)\}$, $M(A, B) = 1$;

Axiom 4 Boundary Condition 2: If $A(x_i) = (0, 1, 0)$, $B(x_i) = (1, 0, 0)$ or: if $A(x_i) = (1, 0, 0)$, $B(x_i) = (0, 1, 0)$, ($i = 1, 2, \dots, m$), $M(A, B) = 0$.

And function $R(A, B)$ is denoted as the adjoint function of the similarity measure between Vague set A and B .

Theorem 3 If $n = 1, 2, \dots$, the function below is the similarity measure between Vague set A and B and the weighted similarity measure.

$$M_{k,n}(A, B) = \begin{cases} \text{Arbitrary value, if } A = B = \{(0, 0, 1), (0, 0, 1), \dots, (0, 0, 1)\} \\ R_{k,n}(A, B), & \text{others} \end{cases},$$

$k = 1, 2, 3$

And function $R_{1,n}(A, B)$, $R_{2,n}(A, B)$ and $R_{3,n}(A, B)$ are adjoint functions of the similarity measure between Vague set A and B .

Application of Similarity Measure between Vague Sets in Network Information Filtering

Reference [9] proposes a method of text search based on fuzzy semantic correlation. Since Vague sets, especially the three-dimensional representation, can describe fuzzy information more comprehensively, it will be more convenient and practical to apply Vague sets in network information description. Therefore, this search method can be popularized as the method and network filtering and searching of the similarity measure. To be specific:

Step1: Establish malicious information base and denote them as B_1, B_2, \dots, B_k . For texts kept in the malicious information bases, extract the set of keywords of Vague sets in accordance with the method proposed by [9]; establish an indicator connection between the set of keywords of Vague sets and the original text, so that the corresponding original text can be found easily according to the indicator of the set of keywords of Vague sets.

Step 2: Now a user wants to check text T to check it is malicious information. Extract the set of keywords A of Vague sets of text T in accordance with the method proposed by [9]; search with A , e.g., to calculate the similarity measure $M(A, B_i)$ between A and information base B_i . Under the given threshold λ ($\in (0, 1]$), if $M(A, B_i) > \lambda$, text T , represented by A , is T_i -type malicious information represented by B_i . If $1 \leq i \leq k$, $M(A, B_i) \leq \lambda$, the text represented by A , is not malicious information.

E.g., if there are texts of malicious information T_1 , T_2 and T_3 represented by three-dimensional representation of Vague sets, and the keywords information bases of them are B_1 , B_2 and B_3 , respectively; data of the set of keywords A of text A is as shown below:

$$\begin{aligned} B_1 &= \{(1.0, 0.0, 0.0), (0.7, 0.0, 0.3), (0.6, 0.2, 0.2)\}, \\ B_2 &= \{(0.7, 0.2, 0.1), (1.0, 0.0, 0.0), (0.9, 0.1, 0.0)\}, \\ B_3 &= \{(0.7, 0.0, 0.3), (0.8, 0.1, 0.1), (1.0, 0.0, 0.0)\}, \\ A &= \{(0.6, 0.2, 0.2), (0.7, 0.2, 0.1), (0.6, 0.1, 0.3)\}. \end{aligned}$$

Table 1: Search and Calculation

Formula of A Similarity Measure	A B_1	A B_2 B_3
$M_{1,3}(A, B_i)$		0.7955
0.8806	0.8427	
$M_{2,3}(A, B_i)$		0.6813
0.7973	0.7343	
$M_{3,3}(A, B_i)$		0.6631
0.7974	0.7535	

It can be seen from Table 1 that, let $\lambda = 0.9$, text T , represented by A , is not malicious information; let $\lambda = 0.85$, as indicated by data of $M_{1,3}(A, B_2)$, text T , represented by A , is malicious information of text T_2 , which is represented by B_2 . Yet other data still indicates it is not malicious information.

It is suggested that, the value of threshold λ determines the search result directly. The value of λ is connected with the selected series of similarity measure formula, the value of parameter n ,

malicious information base B_1, B_2, \dots, B_k and the searching object A , and it can be acquired only based on a lot of experiment inductions. The selection principle of value of threshold λ is aimed at lowering the rate of false judgment to an allowable error range.

Conclusions

The three-dimensional representation of Vague values can reflect the idea of Vague sets in a more straight-forward way. This paper extends Vague value $A(x)$ with the three-dimensional representation of Vague values and presents the formula of three classes of similarity measures between Vague sets with the basic maximizing and minimizing operation of fuzzy sets based on the definition of piecewise function. In accordance with the definition of piecewise function, three series of formulas of similarity measure between Vague sets are presented and it is suggested that Vague sets should be applied to refer to network information. Based on this, the method of network information filtering and searching based on the similarity measure between Vague sets is provided. Experiments suggest that these formulas are practical. Thus, the range of choice of similarity measure formula is broadened for the wider application of Vague sets in engineering and theoretical research.

Acknowledgements

The research work was supported by Natural Science Foundation of Hainan Provincial under Grant No. 114011.

References

- [1] L A Zadeh. Fuzzy Sets [J]. Information and Control, 1965; 8(3): 338~353.
- [2] W L Gau, D J Buehrer. Vague Sets [J]. IEEE Transactions on Systems, Man, and Cybernetics. 1993; 23(2): 610-614.
- [3] Liu Huawen & Wang Fengying. The transformation of Vague set and its similar measurement [J]. Computer engineering and application, 2004; 40(32): 79-81, 84.
- [4] Zhang Fujin & Wang Hongxu. Similarity Measure between Vague Sets [J]. Computer Science, 2006, 33 (5): 197-199
- [5] Qiu Weigen. Vague Sets and its Similarity Measure [J]. Computer Science, 2007, 34 (1): 156-158.
- [6] Li Yanhong, Chi Zhongxian & Yan Dele. Similarity Measure between Vague Sets and Vague Entropy [J]. Computer Science, 2002, 29 (12): 129-132
- [7] Liu Huawen. Basis of Fuzzy Pattern Recognition- Similarity Measure [J]. Pattern Recognition and Artificial Intelligence, 2004, 17 (2): 141-145
- [8] Fu Xiaofang, Wang Hongxu & Wu Yan. The Similar Measurement of the Vague Set in the Same Kind of Ternary Array [J]. Computer engineering and application, 2009, Vol. 45, No. 31: 46-48
- [9] He Xingui. Semantic Correlation between Fuzzy Sets and Application [J]. Journal of Software, 1994, 5 (6): 19-24
- [10] Wang Haifeng, Wang Hongxu & Zhang Kun. The application of the optimal arithmetic of the Vague set scheme in the optimal transmission scheme [J]. Computer science, 2013, 40(6A): 103-104, 119.