

Vague 集之间的相似度量*

李凡¹, 徐章艳²

¹(华中科技大学 计算机学院, 湖北 武汉 430074);

²(广西师范大学 数学与计算机学院, 广西 桂林 541004)

E-mail: fanli@public.wh.hb.cn

摘要: 分析了Chen给出的一种度量Vague集(值)之间相似程度的度量方法,指出了这种相似度量方法所存在的缺陷.在此基础上,给出了一种新的相似度量方法.由于该方法具有较好的性质,更具合理性,因而是一种较为理想的检测Vague集(值)之间相似程度的度量方法.

关键词: fuzzy集; Vague集; 真/假隶属函数; Vague集的核; 相似度量

中图法分类号: TP301 **文献标识码:** A

Contor 创立的集合论是无法处理具有模糊性的不确定性信息和数据的,于是,Zadeh 于 1965 年提出了 Fuzzy 集理论^[1]. 在随后的几十年中,Fuzzy 集理论不断地发展和完善,并在许多领域里得到了成功的应用^[2]. Fuzzy 集最主要的特征是一个 Fuzzy 集 A 是满足某个(或几个)性质的一类对象,每一对象都有一个互不相同的隶属于 A 的程度,隶属函数 $\mu_A(x) (x \in U)$ 给每个对象分派一个 $[0, 1]$ 中的数作为它的隶属度. 这个隶属度有如下特点:该单值既包含了支持 x 的证据,也包含了反对 x 的证据,它不可能表示其中的一个,更不可能同时表示支持和反对的证据. 为了解决 Fuzzy 集无法表示和处理这类具有模糊性信息和数据的问题,Gau 和 Buehrer 于 1993 年提出了 Vague 集理论^[3,4]. 在一个 Vague 集 V 中,用一个真隶属函数 $t_v(x)$ 和假隶属函数 $f_v(x)$ 来描述其隶属度的界. 这两个界构成 $[0, 1]$ 的一个子区间 $[t_v(x), 1 - f_v(x)]$. 例如,设 A 为一个 Vague 集,假定 $[t_A(x), 1 - f_A(x)] = [0.5, 0.8]$,可知 $f_A(x) = 1 - 0.8 = 0.2$,此时,Vague 集 A 可解释为:对象 x 属于集 A 的程度为 0.5,不属于集 A 的程度为 0.2. 我们也可以用投票模型来解释 A ,即赞成票为 5 票,反对票为 2 票,弃权票为 3 票. 显然,用 Fuzzy 集是无法表示和处理这类模糊信息的.

在智能系统的推理过程中,经常需要将两个知识模式(如两个模糊断言、两个谓词公式、两个框架片断或两个语义网络片断)进行比较和耦合,即检查这两个知识模式是否完全一致或近似一致. 如果两者完全一致,或者虽不完全一致,但两者间的相似程度落在限定的阈值内,就称这两个知识模式是匹配的,否则称为不匹配.

为了度量两个 Vague 集(值)之间的相似程度,文献[5]给出了一种度量方法,经过分析我们认为,该方法具有一些不足和缺陷,于是我们提出了一种新的相似度量方法. 由于该方法具有较好的特征和度量效果,因而它是一种较为理想的检测 Vague 集(值)之间相似程度的度量方法.

1 Vague 集

定义 1. 令 U 是一个点(对象)的空间,其中的任意一个元素用 x 表示, U 中的一个 Vague 集 V

* 收稿日期: 1999-12-14; 修改日期: 2000-03-17

基金项目: 国家高性能计算基金资助项目(00303);华中科技大学科学研究基金资助项目(M99015)

作者简介: 李凡(1943-),男,湖北潜江人,教授,主要研究领域为人工智能,模糊信息处理,自动推理,遗传算法;徐章艳(1972-),男,湖北罗田人,硕士生,主要研究领域为近似推理,遗传算法.

用一个真隶属函数 t_v 和一个假隶属函数 f_v 表示, $t_v(x)$ 是从支持 x 的证据所导出的 x 的隶属度下界, $f_v(x)$ 则是从反对 x 的证据所导出的 x 的否定隶属度下界, $t_v(x)$ 和 $f_v(x)$ 将区间 $[0, 1]$ 中的一个实数与 U 中的每一个点联系起来, 即

$$\begin{aligned} t_v: U &\rightarrow [0, 1], \\ f_v: U &\rightarrow [0, 1]. \end{aligned}$$

其中 $t_v(x) + f_v(x) \leq 1$.

设 V 为一个 Vague 集, 当 U 是连续的时候, 有

$$V = \int_U [t_v(x), 1 - f_v(x)] / x, \quad x \in U.$$

当 U 为离散的时候, 有

$$V = \sum_{i=1}^n [t_v(x_i), 1 - f_v(x_i)] / x_i, \quad x_i \in U.$$

2 文献[5]中相似度量的缺陷

文献[5]中给出了如下的度量两个 Vague 集(值)之间相似程度的度量方法:

(1) 假定 $X = [t_X, 1 - f_X]$ 是论域 U 上的一个 Vague 值, 其中 $t_X \in [0, 1]$, $f_X \in [0, 1]$, 且 $t_X + f_X \leq 1$, 那么 X 的核可由如下的函数 S 进行计算:

$$S(X) = t_X - f_X. \quad (1)$$

显然, $S(X) \in [-1, 1]$.

(2) 假定 $X = [t_X, 1 - f_X]$, $Y = [t_Y, 1 - f_Y]$ 是论域 U 上的两个 Vague 值, 则 X 和 Y 之间的相似程度可由如下的函数 M 进行计算:

$$M(X, Y) = 1 - \frac{|S(X) - S(Y)|}{2}, \quad (2)$$

其中 $S(X) = t_X - f_X$, $S(Y) = t_Y - f_Y$.

(3) 假定 A 和 B 是论域 $U = \{u_1, u_2, \dots, u_n\}$ 上的两个 Vague 集, 其中

$$A = [t_A(u_1), 1 - f_A(u_1)] / u_1 + [t_A(u_2), 1 - f_A(u_2)] / u_2 + \dots + [t_A(u_n), 1 - f_A(u_n)] / u_n$$

$$= \sum_{i=1}^n [t_A(u_i), 1 - f_A(u_i)] / u_i,$$

$$B = [t_B(u_1), 1 - f_B(u_1)] / u_1 + [t_B(u_2), 1 - f_B(u_2)] / u_2 + \dots + [t_B(u_n), 1 - f_B(u_n)] / u_n$$

$$= \sum_{i=1}^n [t_B(u_i), 1 - f_B(u_i)] / u_i.$$

假定 $V_A(u_i) = [t_A(u_i), 1 - f_A(u_i)]$ 表示 Vague 集 A 中 u_i 的隶属值, $V_B(u_i) = [t_B(u_i), 1 - f_B(u_i)]$ 表示 Vague 集 B 中 u_i 的隶属值, 则 A 和 B 的核分别为 $S(V_A(u_i)) = t_A(u_i) - f_A(u_i)$ 和 $S(V_B(u_i)) = t_B(u_i) - f_B(u_i)$, 其中 $i = 1, 2, \dots, n$. Vague 集 A 和 B 的相似程度可由如下的函数 T 进行计算.

$$T(A, B) = \frac{1}{n} \sum_{i=1}^n M(V_A(u_i), V_B(u_i)) = \frac{1}{n} \sum_{i=1}^n \left(1 - \frac{|S(V_A(u_i)) - S(V_B(u_i))|}{2} \right). \quad (3)$$

文献[5]认为, 两个 Vague 集(值)之间相似度量的值 $T(A, B)$ ($M(X, Y)$) 越大, 它们之间的相似程度也越大. 按常理, 这一思想是完全正确的, 但由上述定义却得不到这样的结果. 例如, 两个 Vague 值 $[0, 1]$ 和 $[0.5, 0.5]$, 若按式(2)进行计算, 这两个 Vague 值的相似度量应为 1, 这意味着

$[0,1]$ 和 $[0.5,0.5]$ 是完全相似的,但实际上我们凭直觉就能看出这两个 Vague 值并不是完全相似的.两个 Vague 集之间的相似度量也存在类似的问题.

3 一个新的相似度量及其特征

为了更合理地度量两个 Vague 集(值)之间的相似程度,我们提出了一种新的度量方法,该方法能够较好地解决文献[5]中存在的问题.

3.1 Vague 值之间的相似度量

假定 $X=[t_x,1-f_x]$, $Y=[t_y,1-f_y]$ 是论域 U 上的两个 Vague 值, $S(X)$ 和 $S(Y)$ 的定义与文献[5]中的定义相同, X 和 Y 之间的相似程度可由下面的函数 M' 计算得出.

$$M'(X,Y)=1-\frac{|S(X)-S(Y)|}{4}-\frac{|t_x-t_y|+|f_x-f_y|}{4} \quad (4)$$

由上述定义,我们可以得到如下定理.

定理 1. $M'(X,Y) \in [0,1]$.

证明: $M'(X,Y)=1-\frac{|S(X)-S(Y)|}{4}-\frac{|t_x-t_y|+|f_x-f_y|}{4} \leq 1-\frac{0}{4}-\frac{0}{4}=1$.

由于 $|S(X)-S(Y)| \leq 2$, $|t_x-t_y| \leq 1$, $|f_x-f_y| \leq 1$, 故有

$$M'(X,Y)=1-\frac{|S(X)-S(Y)|}{4}-\frac{|t_x-t_y|+|f_x-f_y|}{4} \geq 1-\frac{2}{4}-\frac{2}{4}=0. \quad \square$$

定理 2. $M'(X,Y)=M'(Y,X)$.

证明: 由 $M'(X,Y)$ 的定义即可证明.

定理 3. $M'(X,Y)=0 \Leftrightarrow X=[0,0], Y=[1,1]$ 或 $X=[1,1], Y=[0,0]$.

证明: 若 $X=[0,0], Y=[1,1]$ 或 $X=[1,1], Y=[0,0]$, 显然有 $M'(X,Y)=0$. 若 $M'(X,Y)=0 \Rightarrow |S(X)-S(Y)|+|t_x-t_y|+|f_x-f_y|=4$, 由于 $|S(X)-S(Y)| \leq 2$, $|t_x-t_y| \leq 1$, $|f_x-f_y| \leq 1$, 故有 $|S(X)-S(Y)|=2$, $|t_x-t_y|=1$, $|f_x-f_y|=1$, 从而有

$$X=[0,0], Y=[1,1] \text{ 或 } X=[1,1], Y=[0,0].$$

由上述定理可知, 这样的相似度量是符合实际的. □

例 1: 假定论域 U 上的两个 Vague 值分别为 $X=[0.5,0.5]$, $Y=[0,1]$, 由式(4)可得 X 和 Y 之间的相似度量为

$$M'(X,Y)=1-\frac{|0-0|}{4}-\frac{|0.5-0|+|0.5-0|}{4}=1-0.25=0.75.$$

若按式(2)进行计算, 则 X 和 Y 之间的相似度量应为

$$M(X,Y)=1-\frac{|0-0|}{2}=1-0=1.$$

显然, 按式(2)计算所得到的结果是不符合我们的直觉的.

3.2 Vague 集之间的相似度量

假定 A 和 B 是论域 $U=\{u_1, u_2, \dots, u_n\}$ 上的两个 Vague 集, 其中

$$\begin{aligned}
 A &= [t_A(u_1), 1-f_A(u_1)]/u_1 + [t_A(u_2), 1-f_A(u_2)]/u_2 + \dots + [t_A(u_n), 1-f_A(u_n)]/u_n \\
 &= \sum_{i=1}^n [t_A(u_i), 1-f_A(u_i)]/u_i, \\
 B &= [t_B(u_1), 1-f_B(u_1)]/u_1 + [t_B(u_2), 1-f_B(u_2)]/u_2 + \dots + [t_B(u_n), 1-f_B(u_n)]/u_n \\
 &= \sum_{i=1}^n [t_B(u_i), 1-f_B(u_i)]/u_i.
 \end{aligned}$$

假定 $V_A(u_i) = [t_A(u_i), 1-f_A(u_i)]$ 表示 Vague 集 A 中 u_i 的隶属值, $V_B(u_i) = [t_B(u_i), 1-f_B(u_i)]$ 表示 Vague 集 B 中 u_i 的隶属值, 则 A 和 B 的核分别为 $S(V_A(u_i)) - t_A(u_i) - f_A(u_i)$ 和 $S(V_B(u_i)) - t_B(u_i) - f_B(u_i)$, 其中 $i=1, 2, \dots, n$. Vague 集 A 和 B 的相似程度可由下面的函数 T' 计算得到.

$$\begin{aligned}
 T'(A, B) &= \frac{1}{n} \sum_{i=1}^n M'(V_A(u_i), V_B(u_i)) \\
 &= \frac{1}{n} \sum_{i=1}^n \left(1 - \frac{|S(V_A(u_i)) - S(V_B(u_i))|}{4} - \frac{|t_A(u_i) - t_B(u_i)| + |f_A(u_i) - f_B(u_i)|}{4} \right). \quad (5)
 \end{aligned}$$

由上述定义, 我们可得到如下的定理.

定理 4. $T'(A, B) \in [0, 1]$.

证明: 由定理 1 即可证明.

定理 5. $T'(A, B) = T'(B, A)$.

证明: 由 $T'(A, B) = 0$ 的定义即可证明.

定理 6. $T'(A, B) = 0$

$$\Leftrightarrow A = \sum_{i=1}^n [0, 0]/u_i, B = \sum_{i=1}^n [1, 1]/u_i, \text{ 或 } B = \sum_{i=1}^n [0, 0]/u_i, A = \sum_{i=1}^n [1, 1]/u_i.$$

证明: 由定理 3 即可证明.

显然, $T'(A, B)$ 的值越大, 表明 Vague 集 A 和 B 之间的相似程度就越大. 由定理 6 可知, $T'(A, B) = 0$ 当且仅当 A 的隶属值全部为 $[0, 0]$ 且 B 的隶属值全部为 $[1, 1]$, 或者 B 的隶属值全部为 $[0, 0]$ 且 A 的隶属值全部为 $[1, 1]$.

例 2: 设 A 和 B 是论域 $U = \{u_1, u_2, u_3, u_4, u_5\}$ 上的两个 Vague 集, 其中

$$A = [0.4, 0.4]/u_1 + [0.3, 0.7]/u_2 + [0.5, 0.7]/u_3 + [0.7, 0.9]/u_4 + [0.8, 1.0]/u_5,$$

$$B = [0.3, 0.5]/u_1 + [0.4, 0.6]/u_2 + [0.4, 0.8]/u_3 + [0.7, 0.9]/u_4 + [0.9, 0.9]/u_5,$$

由式(5)可得 A 和 B 之间的相似度量为

$$\begin{aligned}
 T'(A, B) &= \frac{1}{5} \sum_{i=1}^5 M'(V_A(u_i), V_B(u_i)) \\
 &= \frac{1}{5} \sum_{i=1}^5 \left(1 - \frac{|S(V_A(u_i)) - S(V_B(u_i))|}{4} - \frac{|t_A(u_i) - t_B(u_i)| + |f_A(u_i) - f_B(u_i)|}{4} \right) \\
 &= ((1-0.2/4) + (1-0.2/4) + (1-0.2/4) + 1 + (1-0.2/4))/5 \\
 &= 4.8/5 = 0.96.
 \end{aligned}$$

3.3 Vague 集之间的加权相似度量

假定 A 和 B 是论域 $U = \{u_1, u_2, \dots, u_n\}$ 上的两个 Vague 集, 其中

$$A = \sum_{i=1}^n [t_A(u_i), 1 - f_A(u_i)] / u_i, \quad B = \sum_{i=1}^n [t_B(u_i), 1 - f_B(u_i)] / u_i,$$

假设论域 U 上的元素 u_i 的权重为 $w_i, w_i \in (0, 1], 1 \leq i \leq n$, 则 A 和 B 之间的加权相似度量按下式进行计算:

$$W(A, B) = \frac{\sum_{i=1}^n w_i M'(V_A(u_i), V_B(u_i))}{\sum_{i=1}^n w_i} = \frac{\sum_{i=1}^n w_i \left(1 - \frac{|S(V_A(u_i)) - S(V_B(u_i))|}{4} - \frac{|t_A(u_i) - t_B(u_i)| + |f_A(u_i) - f_B(u_i)|}{4} \right)}{\sum_{i=1}^n w_i}. \quad (6)$$

显然, $W(A, B) \in [0, 1]$. $W(A, B)$ 的值越大, 表明 Vague 集 A 和 B 之间的相似程度就越大.

例 3: 假定 A 和 B 为例 2 中的 A 和 B , 论域 U 上的元素 u_i 的权重分别为 0.2, 0.4, 0.8, 0.6 和 1, 此时, 由式(6)即可得到 A 和 B 之间的加权相似度量:

$$W(A, B) = [0.2(1 - 0.2/4) + 0.4(1 - 0.2/4) + 0.8(1 - 0.2/4) + 0.6(1 - 0/4) + 1(1 - 0.2/4)] \div (0.2 + 0.4 + 0.8 + 0.6 + 1.0) = 0.288/3 = 0.96.$$

这个结果与例 2 的结果完全一样, 这只是一种巧合. 因为随着权重的改变, 结果将会随之变化. 一般来说, 两个相似程度较大的 Vague 集(值)无论是采用哪一种度量方法, 所计算出的相似度量值都应该比较大. 显然, 加权相似度量方法比前一种相似度量方法更为灵活, 也更能反映两个 Vague 集(值)之间的相似程度.

4 结 论

通过对文献[5]中提出的相似度量方法进行分析, 我们指出了它存在的不足和缺陷. 在此基础上, 提出了一种新的度量 Vague 集(值)之间相似程度的度量方法. 事实表明, 该方法更符合实际, 它为检测 Vague 集(值)之间的相似程度提供了一种更好的度量方法.

References:

- [1] Zadeh, L. A. Fuzzy sets. *Information and Control*, 1965, 8(3): 338~353.
- [2] Liu, Zeng-liang. *Selected Writings of Fuzzy Technology and Application (1)*. Beijing: Beijing University of Aeronautics and Astronautics Press, 1997 (in Chinese).
- [3] Gau, Wen-lung, Buehrer, D. J. Vague sets. *IEEE Transactions on Systems, Man and Cybernetics*, 1993, 23(2): 610~614.
- [4] Li, Fan. *Fuzzy Information Processing Systems*. Beijing: Beijing University Press, 1998 (in Chinese).
- [5] Chen, S. M. Measures of similarity between vague sets. *Fuzzy Sets and Systems*, 1995, 74(2): 217~223.

附中文参考文献:

- [2] 刘增良. 模糊技术与应用选编(1). 北京: 北京航空航天大学出版社, 1997.
- [4] 李凡. 模糊信息处理系统. 北京: 北京大学出版社, 1998.

Measures of Similarity between Vague Sets *

LI Fan¹, XU Zhang-yan²

¹(College of Computer, Huazhong University of Science and Technology, Wuhan 430074, China);

²(College of Mathematics and Computer, Guangxi Normal University, Guilin 541004, China)

E-mail: fanli@public.wh.hb.cn

Abstract: In this paper, the authors analyze the measures of similarity between Vague sets given by Chen, and point out some problems in the measures. Then they revise them and propose a new kind of measures of similarity. With some good properties, the new measures are more rational and practical, thus providing a more useful way to measure the degree of similarity between Vague sets.

Key words: fuzzy set; Vague set; true/false membership function; core of Vague set; measure of similarity

* Received December 14, 1999; accepted March 17, 2000

Supported by the National High Performance Computing Foundation of China under Grant No. 00313; the Science Foundation of Huazhong University of Science and Technology of China under Grant No. M99015

2001 未来软件技术国际研讨会

International Symposium on Future Software Technology 2001 (ISFST2001)

征文通知

This is the 6th symposium dedicated to the future software technology, which will be held in Zhengzhou, China on November 5~7, 2001. It is organized by SEA (Software Engineers Association of Japan), UNU/IIST (International Institute of Software Technology of United Nations University), Zhengzhou University (China), and Henan University (China). The symposium will bring together researchers, practitioners, and educators in the leading-edge software technologies.

The theme of the ISFST2001 is: "First Step of Software Technology in 21st Century". Suggested topics include, but are not limited to: Software development method/tools; Software process (modeling/management); Web-related technologies; Database/Data warehouse; New application technologies; Software component and architecture; Formal methods/approaches; Testing and verification; System evolution and maintenance; Human aspects of computer application.

Paper Submission: Submissions should contain the type of the submission (Full (5000 words) or Extended-Abstract (1500 words)), title, author names and their affiliation and address, abstract, and a list of keywords, followed by the text. Electronic submissions (postscript or Word RTF) are welcome. Six copies are required for paper submission in case of surface mail. Submissions via FAX are not accepted. All submissions should be accompanied with cover-sheet information electronically sent to the Program Chair. The E-mail must include: title, author names, abstract, a list of keywords, the type of the submission (full paper or extended abstract), and the corresponding address of the first author (name, postal address, E-mail address, and phone and fax numbers). Paper selection will be based on originality and contribution to the topics. Accepted papers will appear in the symposium proceedings if it is presented by the author at the symposium, which will be published by SEA Japan.

Important Dates: Submission deadline: June 30, 2001;

Notification of acceptance: August 15, 2001;

Final paper due: September 20, 2001.

Communications: within China, the submissions should be received by one of the Program Chairs: Beijun Shen, Tel: 86-21-64855251, E-mail: isfst@astj.com.cn or Beijun.shen@netease.com