

DOI:10.3969/j.issn.1673-4785.201210006

网络出版地址: <http://www.cnki.net/kcms/detail/23.1538.TP.20130515.0932.007.html>

# 基于包含度的 Vague 集相似度量方法

贾伟

(宁夏大学新华学院 信息与计算机科学系, 宁夏 银川 750021)

**摘要:**针对现有相似度量方法在相似性度量中出现的度量结果不合理的问题,利用 Vague 集及其补集的包含度关系,将元素间的包含关系和补集间的包含关系相结合,提出了一个基于包含度的 Vague 集相似度量新方法.新方法考虑了未知度对相似度量的影响,设置了合理的系数,通过与现有相似度量方法的比较,说明新方法克服了现有相似度量方法的不足,能够合理有效地区分数据,度量结果符合人们的直观感受,具有一定的优越性,为研究基于包含度的相似度量方法提供了新的思路.

**关键词:** Vague 集; 相似度量; 包含度; 未知度

**中图分类号:** TP301 **文献标志码:** A **文章编号:** 1673-4785(2013)03-0271-06

中文引用格式:贾伟. 基于包含度的 Vague 集相似度量方法[J]. 智能系统学报, 2013, 8(3): 271-276.

英文引用格式:JIA Wei. A method of similarity measures between vague sets based on inclusion grade[J]. CAAI Transactions on Intelligent Systems, 2013, 8(3): 271-276.

## A method of similarity measures between vague sets based on inclusion grade

JIA Wei

(Department of Information and Computer Science, Xinhua College of Ningxia University, Yinchuan 750021, China)

**Abstract:** Through aiming at unreasonable measure results of existing similarity measure methods, the inclusion relationship between elements and inclusion relationship between complements were combined with each other by using the inclusion grade relationship between the vague set and its complement set. A new method of similarity measures between vague sets based on inclusion grade is proposed. The new method takes into account the influence of unknown degrees on similarity measure, and sets up reasonable coefficient. By comparison with the existing similarity measures, the new method overcomes the defects of the existing similarity measure methods and can distinguish data reasonably and effectively. The measure result of the new method is consistent with intuitive sense of people. A new way of thinking was provided by researching the method of similarity measures between vague sets based on inclusion grade.

**Keywords:** vague set; similarity measure; inclusion grade; unknown degree

1965 年,美国控制论专家 Zadeh 提出了模糊集理论<sup>[1]</sup>,为人们表达不确定概念提供了有力的工具. 1993 年, Gau 和 Buehrer<sup>[2]</sup>提出了 Vague 集的概念,作为模糊集的一种推广形式, Vague 集在处理不确定信息时比模糊集具有更强的表示能力,并且在本

质上与直觉模糊集一致<sup>[3]</sup>.目前, Vague 集已经成功应用于人工智能的各分支领域, Vague 集相似度量是 Vague 集在这些领域中应用的理论基础,起着重要的作用,近年来一直是研究的热点问题,很多学者从不同的角度研究了相似度量.包含度理论是一种描述不确定性关系的有效度量方法<sup>[4]</sup>,在模糊集中有着重要的应用,而且模糊集的贴适度可以由其包含度公式诱导出<sup>[5]</sup>.黄国顺等<sup>[6]</sup>将模糊集上的包含度概念扩展到 Vague 集上,指出 Vague 集的相似度

收稿日期:2012-10-08. 网络出版日期:2013-05-15.

基金项目:宁夏大学科学研究基金资助项目(ZR1147);宁夏大学新华学院科学研究基金资助项目(11XH10).

通信作者:贾伟. E-mail: jia.401@163.com.

量也可以由包含度诱导出,并且从包含度的角度对 Vague 集相似度量方法进行了研究.江伟等<sup>[7]</sup>在指出黄国顺等提出的相似度量方法存在不足之后,提出了一个新的基于包含度的相似度量方法.这些基于包含度的相似度量方法由于考虑的包含关系较少,在计算中丢失了一些重要信息,出现了相似度量结果不合理、数据无法区分的问题.本文通过研究 Vague 集的包含关系,提出一个新的相似度量方法.

## 1 Vague 集基本知识

**定义 1** 设  $U = \{x_1, x_2, \dots, x_n\}$  是一个论域,对于  $U$  的任一元素  $x$ ,  $U$  中的一个 Vague 集  $A$  是由真隶属函数  $t_A$  和假隶属函数  $f_A$  所描述:  $t_A: U \rightarrow [0, 1]$ ,  $f_A: U \rightarrow [0, 1]$ , 并且  $0 \leq t_A(x_i) + f_A(x_i) \leq 1$ , 其中  $t_A(x_i)$  是支持  $x \in A$  的证据的隶属度下界,  $f_A(x_i)$  是反对  $x \in A$  的证据的隶属度下界, 且  $t_A(x_i) + f_A(x_i) \leq 1$ . 元素  $x_i$  在 Vague 集  $A$  中的隶属度被区间  $[0, 1]$  的子区间  $[t_A(x_i), f_A(x_i)]$  所界定, 称该区域为  $x_i$  在  $A$  中的 Vague 值, 记为  $V_A(x)$ , 论域  $U$  上所有 Vague 集的全体记为  $VS(U)$ . 称  $\pi_A(x_i) = 1 - t_A(x_i) - f_A(x_i)$  为  $x$  对于 Vague 集  $A$  的不确定度(未知度), 是  $x$  相对于  $A$  的未知信息的一种度量.

**定义 2** 设论域  $U = \{x_1, x_2, \dots, x_n\}$ ,  $A$  是  $U$  上的一个 Vague 集, 则

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

$A$  的补集为

$$\bar{A} = \sum_{i=1}^n [f_A(x_i), 1 - t_A(x_i)] / x_i.$$

**定义 3** 设论域  $U = \{x_1, x_2, \dots, x_n\}$ ,  $A$  和  $B$  是  $U$  上的 2 个 Vague 集,  $A \subseteq B$ , 则对  $\forall x_i \in U$ ,  $t_A(x_i) \leq t_B(x_i)$  且  $f_A(x_i) \geq f_B(x_i)$ .

**定义 4** 称  $S_A(x) = t_A(x) - f_A(x)$  ( $-1 \leq S_A(x) \leq 1$ ) 为  $x$  的核, 它表示现有证据对元素  $x$  支持和反对 2 种力量的对比, 看作是整体支持(肯定)度.

**定义 5** 设  $A, B$  是论域  $U = \{x_1, x_2, \dots, x_n\}$  上的 Vague 集, 则  $A$  和  $B$  的并运算为

$$A \cup B = \{ \max[t_A(x_i), t_B(x_i)], \max[1 - f_A(x_i), 1 - f_B(x_i)] \}.$$

**定义 6** 设  $A, B$  是论域  $U = \{x_1, x_2, \dots, x_n\}$  上的 Vague 集, 则  $A$  和  $B$  的交运算为

$$A \cap B = \{ \min[t_A(x_i), t_B(x_i)], \min[1 - f_A(x_i), 1 - f_B(x_i)] \}.$$

**定义 7**<sup>[8]</sup> 对于任意  $A, B, C \in VS(U)$ , 由  $V(A, B)_i, V(A, B)_f$  构成 Vague 值  $V(A, B) =$

$[V(A, B)_i, V(A, B)_f]$ , 且满足:

1)  $0 \leq A(A, B)_i \leq 1, 0 \leq V(A, B)_f \leq 1$ , 即  $[0, 1] \leq V(A, B) \leq [1, 0]$ ;

2)  $A \subseteq B \Rightarrow V(A, B)_i = 1, V(A, B)_f = 0$ , 即  $V(A, B) = [1, 0]$ ;

3)  $A \subseteq B \subseteq C \Rightarrow V(C, A)_i \leq V(B, A)_i, V(C, A)_f \geq V(B, A)_f$ , 即  $V(C, A) \leq V(B, A)$ .

则称  $V(A, B)$  为  $VS(U)$  上的包含度, 其中  $V(A, B)_i$  为真隶属度,  $V(A, B)_f$  为假隶属度.

若  $V$  满足 1)、2)、3) 和如下的 4): 任意  $A, B, C \in VS(U)$ ,  $A \subseteq B \Rightarrow V(C, A) \leq V(C, B)$ , 则称  $V(A, B)$  为  $VS(U)$  上的强包含度.

对于如何判断 Vague 集相似度量方法是否合理, 目前还没有一个明确、全面的理论描述, 王伟平等<sup>[9]</sup>提出的基本准则是目前检验方法中较为合理的衡量标准, 本文以该标准衡量 Vague 集相似度量方法.

## 2 Vague 集相似度量新方法

### 2.1 现有相似度量方法的不足

现有的相似度量方法从不同角度对相似度量进行了研究, 这些方法都存在不同程度的不足.

#### 2.1.1 基于支持度和反对度距离的相似度量方法

Pei 等<sup>[10]</sup>将支持度距离和反对度距离作为度量相似度的依据, 提出的相似度量方法为

$$M_p(x, y) = 1 - \frac{|t_x - t_y| + |f_x - f_y|}{4} - \frac{\max(|t_x - t_y|, |f_x - f_y|)}{2}.$$

该方法只考虑了支持度距离和反对度距离对相似度量的影响. 但由于该方法没有充分考虑核距离、未知度因素, 在计算过程中会丢失一些信息, 因此使用该方法计算得到的度量结果缺乏合理性. 比如对于第 1 组元素  $x_1$  和  $y_1$ , 当  $t_{x_1} = 0.3, t_{y_1} = 0.2, f_{x_1} = 0.1, f_{y_1} = 0.2$  时, 得到的度量结果为 0.9, 对于第 2 组元素  $x_2$  和  $y_2$ , 当  $t_{x_2} = 0.5, t_{y_2} = 0.4, f_{x_2} = 0.3, f_{y_2} = 0.2$  时, 得到的度量结果也是 0.9. 显然, 这 2 组元素的相似度是不同的, 该方法得到的度量结果是不合理的.

#### 2.1.2 基于未知度距离的相似度量方法

目前, 一些 Vague 集相似度量方法将未知度的增加或减少作为影响相似度的主要因素, 比较典型的基于未知度距离的相似度量方法有以下几种.

Chen<sup>[11-12]</sup>、Hong 等<sup>[13]</sup>、李凡等<sup>[14]</sup>和张诚一等<sup>[15]</sup>对 Vague 集的相似度量进行了一些初步研究, 提出了一些相似度量方法, 此后, 许多文献在研究中

分析并指出了这些方法的不足,提出了一些新的相似度量方法,但是这些新的相似度量方法在实际度量中又出现了新的问题.

徐凤生<sup>[16]</sup>提出的相似度量方法为

$$M_X(x, y) = 1 - \frac{|S_x - S_y| + |t_x - t_y| + |f_x - f_y| + |\pi_x - \pi_y|}{4}.$$

该方法只考虑了未知度距离对相似度量的影响,如果仅从距离角度来解释 Vague 集的相似度,当核距离、支持度距离、反对度距离和未知度距离完全相同时,不同元素间的相似度就无法进行有效区分,尤其是当元素间存在未知度的情况下,相似度量结果不能合理解释元素间的相似度,因此,在考虑未知度距离因素的同时,还应该考虑未知度增加对相似度量的影响.

王伟平等<sup>[9]</sup>通过研究未知度增加对相似度量的影响后,提出的相似度量方法为

$$M_W(x, y) = 1 - \frac{|t_x - t_y| + |f_x - f_y| + |S_x - S_y| + (\pi_x + \pi_y)}{4}.$$

该方法只考虑了未知度增加对相似度量的影响,否定了未知度距离对相似度的影响.实际上,未知度距离表示了不同元素之间的未知度程度,在其他条件相同的情况下,未知度距离越小,相似度越大,未知度距离对元素间的相似度有着重要影响,因此,未知度距离不应该被忽视.

朱振国等<sup>[17]</sup>提出的相似度量方法为

$$M_{ZZG}(x, y) = 1 - ((2 - t_x - t_y) |t_x - t_y| + (2 - f_x - f_y) |f_x - f_y| + |S_x - S_y| + |\pi_x - \pi_y|) / ((2 - t_x - t_y) + (2 - f_x - f_y) + 2).$$

该方法考虑了衡量 Vague 集相似度量的必要参数,同时还考虑了未知度、确切的支持度和反对度对支持度距离和反对度距离的影响,具有一定的区分能力,但是对未知度的考虑还是不够充分.该方法实际上还是从距离的角度解释 Vague 集的相似度,这与徐凤生提出的方法在本质上是相同的,在相似度量过程中,当这些距离因素全部相同时,就无法对不同元素间的相似度进行有效区分.

马冯等<sup>[18]</sup>提出的 Vague 集相似度量方法为

$$M_M(x, y) = 1 - \frac{2|t_x - t_y|}{2(t_x + t_y) + 3} - \frac{2|f_x - f_y|}{2(f_x + f_y) + 3} - \frac{|\pi_x - \pi_y| + |t_x^2 - t_y^2| + |f_x^2 - f_y^2|}{\pi_x + \pi_y + 10}.$$

该方法虽然考虑了未知度距离和未知度增加对

相似度量的影响,但是没有考虑核距离对相似度量的影响.因为核距离表示的是一种元素间的相对优势,所以不应该被忽略.此外,该方法在系数的设置上存在问题,其根本原因是对于未知度因素考虑不充分,导致系数设置不合理,在实际度量计算中会出现一些不合理的度量结果.

### 2.1.3 基于未知度倾向的相似度量方法

周孟等<sup>[19]</sup>从未知度倾向角度研究相似度量方法,认为未知度中对 Vague 值  $x$  和  $y$  可能存在的支持度分别是  $\alpha_x = t_x + t_x \pi_x$  和  $\alpha_y = t_y + t_y \pi_y$ ,未知度中对 Vague 值  $x$  和  $y$  可能存在的反对度分别是  $\beta_x = f_x + f_x \pi_x$  和  $\beta_y = f_y + f_y \pi_y$ ,提出的相似度量方法是  $M_Z(x, y) = 1 - \sqrt{\frac{(\alpha_x + \alpha_y)^2 + (\beta_x + \beta_y)^2}{2}}$ .该方法只考虑了未知

度对支持度和反对度的倾向因素,没有考虑其他因素对相似度量的影响,在计算中度量效果较差.这是因为存在多个影响 Vague 集相似度的因素,比如核距离、支持度距离和反对度距离等,在研究相似度量方法时,必须考虑到这些因素,否则就会出现度量结果不准确的问题.

### 2.1.4 基于包含度角度的现有相似度量方法

黄国顺等<sup>[6]</sup>从包含度角度研究 Vague 集相似度,在研究过程中,提出了一个 Vague 集元素补集间的基于包含度的相似度量方法.他认为补集间的相似程度不适合用于描述元素间的相似程度,将元素间的包含关系和补集间的包含关系相互结合,给出了一个基于包含度的相似度量方法.但是在对数据的相似度量结果进行分析后,发现该方法的度量结果不合理,放弃了将元素间的包含关系和补集间的包含关系相互结合这一思想的进一步研究,最后给出了一个只考虑元素间包含关系的相似度量方法:

$$M_H(x, y) = \frac{\min(t_x, t_y) + \min(1 - f_x, 1 - f_y)}{\max(t_x, t_y) + \max(1 - f_x, 1 - f_y)}.$$

江伟等<sup>[7]</sup>将包含度概念和算术平均法相结合,提出了一种改进的基于包含度的 Vague 集相似度量方法:

$$M_J(x, y) = \frac{\min(t_x, t_y) + \min(1 - f_x, 1 - f_y)}{(t_x + t_y + 1 - f_x + 1 - f_y) / 2}.$$

黄国顺等提出的基于包含度的相似度量方法实际上都只考虑了元素间的包含关系,忽视了元素补集之间的包含关系,这样在计算过程中会丢失了一些重要的数据信息,导致度量结果不合理.此外,这2种相似度量方法没有充分考虑未知度因素的影响,比如对于 Vague 集的任意元素  $x$  和  $y$ ,当  $t_x = 0, t_y =$

0,  $f_x=0, f_y=0$  时, 存在未知度, 表明元素  $x$  和  $y$  之间存在不确定性, 不能认为  $x$  和  $y$  完全相似; 但是黄国顺等和江伟等提出的相似度量方法都认为  $x$  和  $y$  完全相似, 这显然不符合实际.

## 2.2 基于包含度的 Vague 值相似度量新方法

上述文献从不同角度研究相似度量方法时, 虽然提出的相似度量方法都满足相似度量基本准则, 但是在考虑影响相似度量的因素时, 都存在不足之处. 本文从包含度角度出发, 考虑了未知度因素对相似度量的影响, 根据黄国顺等提出的将元素间的包含关系和补集间的包含关系相结合的思想, 进一步研究了基于包含度的相似度量方法. 由于在 Vague 集中  $t_i + f_i + \pi_i = 1$ , 未知度因素对相似度量结果有着重要影响, 同时, 为了减少在包含关系中丢失的信息, 除了考虑元素间  $t_i$  和  $1-f_i$  的包含关系外, 还要考虑到  $t_x$  和  $1-f_x$  的补集的包含关系, 并给包含关系设置系数. 在此基础上, 本文提出了一个新的基于包含度的 Vague 值相似度量方法:

$$M(x, y) = (3\min(t_x, t_y) + \min(1 - f_x, 1 - f_y) + 3\min(f_x, f_y) + \min(1 - t_x, 1 - t_y)) / 4.$$

**定义 8** 设  $x = [t_x, 1-f_x]$ ,  $y = [t_y, 1-f_y]$  是 Vague 集  $A$  上的 2 个 Vague 值,  $t_x, t_y$  分别是  $x$  和  $y$  的支持度,  $f_x, f_y$  分别是  $x$  和  $y$  的反对度, 则 Vague 值  $x$  和  $y$  的相似度量  $M(x, y)$  定义为

$$M(x, y) = (3\min(t_x, t_y) + \min(1 - f_x, 1 - f_y) + 3\min(f_x, f_y) + \min(1 - t_x, 1 - t_y)) / 4.$$

**定理 1** 设  $A$  是论域  $U$  上的一个 Vague 集,  $x, y$  和  $z$  是 Vague 集  $A$  中的 3 个 Vague 值,  $M(x, y)$  是 Vague 值  $x$  和  $y$  的相似度量,  $M(x, z)$  是 Vague 值  $x$  和  $z$  的相似度量,  $M(y, z)$  是 Vague 值  $y$  和  $z$  的相似度量, Vague 值的相似度量应该满足下列基本准则:

- 1)  $0 \leq M(x, y) \leq 1$ ;
- 2)  $M(x, y) = M(y, x)$ ;
- 3)  $M(x, y) = M(\bar{x}, \bar{y})$ ;
- 4)  $M(x, y) = 0$ , 当且仅当  $x = [0, 0]$ ,  $y = [1, 1]$  或  $x = [1, 1]$ ,  $y = [0, 0]$ ;
- 5) 若  $x \subseteq y \subseteq z$ , 则  $M(x, y) \geq M(x, z)$ , 且  $M(y, z) \geq M(x, z)$ .

**证明**

1) 由 Vague 集的定义  $0 \leq t_x \leq 1, 0 \leq f_x \leq 1, 0 \leq t_y \leq 1, 0 \leq f_y \leq 1$ , 得到  $0 \leq 1 - t_x \leq 1, 0 \leq 1 - f_x \leq 1, 0 \leq 1 - t_y \leq 1, 0 \leq 1 - f_y \leq 1$ , 存在以下 4 种情况:

① 若  $t_x \geq t_y, 1 - f_x \geq 1 - f_y$ , 则  $M(x, y) = \frac{2+3t_y-t_x+3f_x-f_y}{4}$ , 因为  $0 \leq 2+3t_y-t_x+3f_x-f_y \leq 4$ , 所以

$0 \leq \frac{2+3t_y-t_x+3f_x-f_y}{4} \leq 1$ , 当  $t_x = 1, t_y = 0, f_x = 0, f_y = 1$  时,  $M(x, y) = 0$ , 因此,  $0 \leq M(x, y) \leq 1$ , 当且仅当  $x = [1, 1], y = [0, 0]$  时,  $M(x, y) = 0$ .

② 若  $t_x \geq t_y, 1 - f_x \leq 1 - f_y$ , 则  $M(x, y) = \frac{2+3t_y-t_x+3f_y-f_x}{4}$ , 因为  $1 \leq 2+3t_y-t_x+3f_y-f_x \leq 4$ , 所以  $\frac{1}{4} \leq \frac{2+3t_y-t_x+3f_y-f_x}{4} \leq 1$ , 即  $\frac{1}{4} \leq M(x, y) \leq 1$ .

③ 若  $t_x \leq t_y, 1 - f_x \geq 1 - f_y$ , 则  $M(x, y) = \frac{2+3t_x-t_y+3f_x-f_y}{4}$ , 因为  $1 \leq 2+3t_x-t_y+3f_x-f_y \leq 4$ , 所以  $\frac{1}{4} \leq \frac{2+3t_x-t_y+3f_x-f_y}{4} \leq 1$ , 即  $\frac{1}{4} \leq M(x, y) \leq 1$ .

④ 若  $t_x \leq t_y, 1 - f_x \leq 1 - f_y$ , 则  $M(x, y) = \frac{2+3t_x-t_y+3f_y-f_x}{4}$ , 因为  $0 \leq 2+3t_x-t_y+3f_y-f_x \leq 4$ , 所以  $0 \leq \frac{2+3t_x-t_y+3f_y-f_x}{4} \leq 1$ , 当  $t_x = 0, t_y = 1, f_x = 1, f_y = 0$  时,  $M(x, y) = 0$ , 因此,  $0 \leq M(x, y) \leq 1$ , 当且仅当  $x = [0, 0], y = [1, 1]$  时,  $M(x, y) = 0$ .

综上所述,  $M(x, y)$  满足基本准则 1) 和基本准则 4)。

2) 显然  $M(x, y)$  满足基本准则 2) 和基本准则 3)。

3) 若  $x \subseteq y \subseteq z$ , 则  $t_x \leq t_y \leq t_z, f_x \geq f_y \geq f_z$ ,

$$\begin{aligned} M(x, y) &= (3\min(t_x, t_y) + \min(1 - f_x, 1 - f_y) + 3\min(f_x, f_y) + \min(1 - t_x, 1 - t_y)) / 4, \\ M(x, z) &= (3\min(t_x, t_z) + \min(1 - f_x, 1 - f_z) + 3\min(f_x, f_z) + \min(1 - t_x, 1 - t_z)) / 4, \end{aligned}$$

$$M(x, y) = \frac{2 + 3t_x - t_y + 3f_y - f_x}{4},$$

$$M(x, z) = \frac{2 + 3t_x - t_z + 3f_z - f_x}{4}.$$

由于

$$M(x, y) - M(x, z) = \frac{t_z - t_y + 3(f_y - f_z)}{4} \geq 0,$$

所以

$$M(x, y) \geq M(x, z).$$

同理可证  $M(y, z) \geq M(x, z)$ . 因此  $M(x, y)$  满足基本准则 5)。

## 2.3 基于包含度的 Vague 集加权相似度量新方法

**定义 9** 设  $A$  和  $B$  是论域  $U = \{x_1, x_2, \dots, x_n\}$  上的

2 个 Vague 集, 其中  $A = \sum_{i=1}^n [t_A(x_i), 1-f_A(x_i)]/x_i$ ,  $B = \sum_{i=1}^n [t_B(x_i), 1-f_B(x_i)]/x_i$ ,  $\omega_i (i=1, 2, \dots, n)$  表示



元素  $x_i$  在  $U$  中的权重,其中  $\omega_i \in [0,1]$ ,  $\sum_{i=1}^n \omega_i = 1$ , 则 Vague 集  $A$  和  $B$  的加权相似度量定义为  $M(A,B) = \sum_{i=1}^n \omega_i M(A(x_i), B(x_i))$ . 如果选取  $\omega_i = \frac{1}{n}$ , 表明各元素的权重相同, 则可得到各元素的平均值.

**定理 2** 设  $A, B$  和  $C$  是论域  $U = \{x_1, x_2, \dots, x_n\}$  上的 3 个 Vague 集, Vague 集的相似度量应该满足下列基本准则:

- 1)  $0 \leq M(A, B) \leq 1$ ;
- 2)  $M(A, B) = M(B, A)$ ;
- 3)  $M(A, B) = M(\bar{A}, \bar{B})$ ;

- 4)  $M(A, B) = 0$ , 当且仅当  $A = \{[0, 0] \mid x \in U\}$ ,  $B = \{[1, 1] \mid x \in U\}$  或  $A = \{[1, 1] \mid x \in U\}$ ,  $B = \{[0, 0] \mid x \in U\}$ ;
  - 5) 若  $A \subseteq B \subseteq C$ , 则  $M(A, B) \geq M(A, C)$ , 且  $M(B, C) \geq M(A, C)$ .
- 容易证明  $M(A, B)$  满足基本准则 1) ~ 5).

3 实例分析

下面通过表 1 中的数据比较, 说明本文提出的相似度量方法  $M(x, y)$  克服了现有相似度量方法的不足.

表 1 与现有相似度量方法的比较  
Table 1 Comparison of the new method with the existing methods of similarity measurement

方 法	相似度					
	第 1 组	第 2 组	第 3 组	第 4 组	第 5 组	第 6 组
$x$	$[0, 1]$	$[0.2, 0.5]$	$[0.2, 0.5]$	$[0.3, 0.4]$	$[0.3, 0.5]$	$[0.3, 0.5]$
$y$	$[0, 1]$	$[0.3, 0.5]$	$[0.4, 0.4]$	$[0.3, 0.4]$	$[0.4, 0.6]$	$[0.5, 0.5]$
$M_x(x, y)$	1.0	0.925 0	0.825 0	1.00	0.900 0	0.850 0
$M_w(x, y)$	0.5	0.825 0	0.825 0	0.95	0.800 0	0.850 0
$M_p(x, y)$	1.0	0.925 0	0.825 0	1.00	0.900 0	0.850 0
$M_{zgc}(x, y)$	1.0	0.922 2	0.820 9	1.00	0.900 0	0.847 6
$M_m(x, y)$	1.0	0.935 7	0.814 8	1.00	0.897 5	0.877 7
$M_z(x, y)$	1.0	0.920 9	0.894 9	1.00	0.880 0	0.878 3
$M_h(x, y)$	1.0	0.875 0	0.666 7	1.00	0.800 0	0.800 0
$M_j(x, y)$	1.0	0.933 3	0.800 0	1.00	0.888 9	0.888 9
$M(x, y)$	0.5	0.825 0	0.775 0	0.95	0.800 0	0.850 0

由于第 1 组数据和第 4 组数据中都存在未知度, 表明存在不确定性, 因此不能认为完全相似, 但是徐凤生提出的相似度量方法  $M_x(x, y)$ 、Pei 等提出的相似度量方法  $M_p(x, y)$ 、朱振国等提出的相似度量方法  $M_{zgc}(x, y)$ 、马冯等提出的相似度量方法  $M_m(x, y)$ 、周孟等提出的相似度量方法  $M_z(x, y)$ 、黄国顺等提出的相似度量方法  $M_h(x, y)$  和江伟等提出的相似度量方法  $M_j(x, y)$  都认为第 1 组数据和第 4 组数据的相似度都为 1, 即完全相似, 这显然不符合实际. 只有王伟平等提出的相似度量方法  $M_w(x, y)$  和本文提出的相似度量方法  $M(x, y)$  认为第 1 组数据中支持程度与反对程度完全相同, 因为存在未知度, 所以相似度应为 0.5. 在第 4 组数据中, 由于支持程度与反对程度不同且存在的未知度较少, 因此相似度应该较高, 王伟平等提出方法  $M_w(x, y)$  和本文提出的方法  $M(x, y)$  所得到的度量结果为 0.95, 符合实际情况. 由此可见, 本文提出的方法能够合理

区分第 1 组和第 4 组 2 组数据.  
第 2 组数据和第 3 组数据的相似度明显是不同的,  $M_w(x, y)$  认为第 2 组数据和第 3 组数据的相似度相同, 这显然不符合人们的直观感受, 其他相似度量方法都可以有效区分这 2 组数据.  
第 5 组和第 6 组数据的相似度不同,  $M_h(x, y)$  和  $M_j(x, y)$  无法区分第 5 组和第 6 组数据, 其他相似度量方法都可以有效区分这 2 组数据.

通过以上实验数据的分析可以看出, 现有的相似度量方法由于对影响相似度量的因素考虑不全面, 出现了不同程度的缺陷, 只有本文提出的相似度量方法  $M(x, y)$  可以有效合理地区分所有数据.

4 结束语

本文分析了目前相似度量方法的不足之处, 在现有文献的基础上, 进一步研究了 Vague 集及其补集的包含关系, 指出了在相似度量中未知度因素的

重要性,给出了一种新的基于包含度的相似度量方法,通过实例分析证明利用新方法得到的度量结果是合理的.此外,由于 Vague 集不同于普通模糊集,存在不确定性,在提出的基于包含度的 Vague 集相似度量方法中,会遗失一些信息,这些遗失的信息可能会影响相似度量结果,对于如何有效解决这一问题还有待进一步研究.

## 参考文献:

- [1] ZADEH L A. Fuzzy sets[J]. Information and Control, 1965 (8): 338-356.
- [2] GAU W L, BUEHRER D J. Vague sets[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1993, 23(2): 610-614.
- [3] BUSTINCE H, BURILLO P. Vague sets are intuitionistic fuzzy sets[J]. Fuzzy Sets and Systems, 1996, 79(3): 403-405.
- [4] 张文修,徐宗本,梁怡,等.包含度理论[J].模糊系统与数学, 1996, 10(4): 1-9.  
ZHANG Wenxiu, XU Zongben, LIANG Yi, et al. Inclusion degree theory[J]. Fuzzy Systems and Mathematics, 1996, 10(4): 1-9.
- [5] 范九伦.包含度和贴近度的相互诱导关系[J].模糊系统与数学, 2003, 17(3): 29-37.  
FAN Jiulun. Mutual induced relations between similarity measure and subsethood measure[J]. Fuzzy Systems and Mathematics, 2003, 17(3): 29-37.
- [6] 黄国顺,刘云生.基于包含度的 Vague 集相似度量[J].小型微型计算机系统, 2006, 27(5): 873-877.  
HUANG Guoshun, LIU Yunsheng. Similarity measures of vague sets based on inclusion grades[J]. Mini-micro Systems, 2006, 27(5): 873-877.
- [7] 江伟,梁家荣.一种新的基于包含度的 Vague 相似度[J].软件导刊, 2012, 11(3): 14-16.  
JIANG Wei, LIANG Jiarong. New similarity measures vague sets based on inclusions grades[J]. Software Guide, 2012, 11(3): 14-16.
- [8] 朱英丽,杨勇,朱晓钟,等.Vague 集的包含度[J].计算机工程与应用, 2011, 47(7): 43-45.  
ZHU Yingli, YANG Yong, ZHU Xiaozhong, et al. Vague inclusion degree[J]. Computer Engineering and Applications, 2011, 47(7): 43-45.
- [9] 王伟平,吴祈宗,李玉玲.Vague 集之间相似度量的基本准则与一般方法[J].计算机工程与应用, 2008, 44(4): 73-76.  
WANG Weiping, WU Qizong, LI Yuling. Basic rules and general method on similarity measures between vague sets[J]. Computer Engineering and Applications, 2008, 44(4): 73-76.
- [10] PEI Zhenkui, LIU Jian. Research on similarity measures between vague sets[C]//Fourth International Conference on Fuzzy Systems and Knowledge Discovery. Haikou, China, 2007: 648-652.
- [11] CHEN S M. Measures of similarity between vague sets[J]. Fuzzy Sets and Systems, 1995, 74(2): 217-223.
- [12] CHEN S M. Similarity measures between vague sets and between elements[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 1997, 27(1): 153-157.
- [13] HONG D H, KIM C. A note on similarity measures between vague sets and between elements[J]. Information Sciences, 1999, 115(1/2/3/4): 83-96.
- [14] 李凡,徐章艳.Vague 集之间的相似度量[J].软件学报, 2001, 12(6): 922-927.  
LI Fan, XU Zhangyan. Measures of similarity between vague sets[J]. Journal of Software, 2001, 12(6): 922-927.
- [15] 张诚一,党平安.关于 Vague 集之间的相似度量[J].计算机工程与应用, 2003, 39(17): 92-94.  
ZHANG Chengyi, DANG Pingan. On measures of similarity between vague sets[J]. Computer Engineering and Applications, 2003, 39(17): 92-94.
- [16] 徐凤生.Vague 集的新相似度量方法[J].计算机工程与科学, 2011, 33(4): 121-123.  
XU Fengsheng. A new method for similarity measures between vague sets[J]. Computer Engineering & Science, 2011, 33(4): 121-123.
- [17] 朱振国,王匡胤.Vague 集相似度量[J].计算机科学, 2008, 35(9): 220-225.  
ZHU Zhenguo, WANG Guoyin. Similarity measure of vague set[J]. Computer Science, 2008, 35(9): 220-225.
- [18] 马冯,余建坤.一种 Vague 集相似度量的新方法[J].计算机工程与应用, 2010, 46(34): 39-42.  
MA Feng, YU Jiankun. New vague sets similarity measure method[J]. Computer Engineering and Applications, 2010, 46(34): 39-42.
- [19] 周孟,余建坤.一种新的 Vague 集相似度量方法[J].计算机应用与软件, 2012, 29(2): 138-141.  
ZHOU Meng, YU Jiankun. A new vague set similarity measurement[J]. Computer Applications and Software, 2012, 29(2): 138-141.

## 作者简介:



贾伟,男,1980年生,讲师,主要研究方向为信息系统集成、Vague 集,发表学术论文5篇。