

一种新的 Vague 集相似度量方式* A Similarity Measure for Vague Sets

苏毅娟

SU Yi-juan

(广西师范学院数学与计算机科学系, 广西南宁 530001)

(Department of Mathematic and Computer Science, Guangxi University of Teachers Education, Nanning, Guangxi, 530001, China)

摘要: 在分析比较当前几种 Vague 集相似度量方法及其优缺点的基础上, 提出一种新的 Vague 集相似度量方法. 该方法充分考虑支持度以及真隶属度差距和假隶属度差距, 并且区分隶属度差距的方向. 该方法更符合实际, 它为 Vague 集之间的相似度提供了一种更好的度量方法.

关键词: Vague 集 相似度量 加权

中图分类号: TP311.12 文献标识码: A 文章编号: 1002-7378(2006)04-0260-03

Abstract: Based on the analysis and comparison of the measures of similarity of vague sets, a method to measure the similarity between Vague sets is proposed. In this method, the support, grade disparity of truth-membership, grade disparity of false-membership and direction of discrimination of the grade disparity of membership are fully considered. The result by this method is quite closer to the real situation in practice. It is revealed that it is a better method in measuring the similarity between vague sets.

Key words: Vague sets, similarity measure, weighting

在智能系统的推理过程中, 经常需要将两个模糊概念进行比较和耦合, 即检查两个知识模式是否完全一致或近似一致, 用于确定数据挖掘系统中概念间的函数依赖关系, 在实践中具有重要的意义. 为了度量两个 Vague 集(值)之间的相似程度, 文献[1~5]分别提出了不同的相似度量方法, 每一种方法侧重于不同角度, 都有各自的优缺点, 因而在实践中可以根据应用需要进行选取. 本文首先分析了上述几种方法的优缺点, 然后提出一种更为合理的 Vague 集相似度量方法并给出其性质.

1 当前 Vague 集的相似度量方法及其优缺点

目前度量 Vague 集(值)相似度的方法有很多, 假定 $X = [t_x, 1 - f_x]$ 和 $Y = [t_y, 1 - f_y]$ 是论域 U 上的两个 Vague 值, $S(x) = t_x - f_x, S(y) = t_y - f_y,$

文献[1~5]中分别定义了 M_C, M_H, M_L, M_o

$$M_C(x, y) = 1 - \frac{|S(x) - S(y)|}{2} = 1 - \frac{|(t_x - t_y) - (f_x - f_y)|}{2}, \quad (1)$$

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

$$M_L(x, y) = 1 - \frac{|S(x) - S(y)|}{4} - \frac{|t_x - t_y| + |f_x - f_y|}{4} = 1 - \frac{|(t_x - t_y) - (f_x - f_y)| + |t_x - t_y| + |f_x - f_y|}{4}, \quad (3)$$

$$M_o(x, y) = 1 - \sqrt{\frac{(t_x - t_y)^2 + (f_x - f_y)^2}{2}}. \quad (4)$$

由 M_C 的定义可知, 当 $t_x - f_x = t_y - f_y \Rightarrow M_C \equiv 1$, 即 $t_x - f_x = t_y - f_y$ 时, 用 M_C 方法进行处理过于粗糙. M_H 侧重于对两个 Vague 值间真隶属度差值和假隶属度差值的同等重视, 即真隶属度差值和假隶属度差值分别相同的 Vague 值, 其相似度相同, 但尚未区分每个 Vague 值对真隶属度间和假隶属

收稿日期: 2005-12-12

作者简介: 苏毅娟(1976-), 女, 广西灵川人, 讲师, 主要从事数据库及相关学科研究.

* 广西壮族自治区教育厅科学研究项目.

度间的正向差距和反向差距。 M_L 继承了 M_C, M_H 的优点,表示了对 Vague 值支持度和真、假隶属度值的同等重视,但由于表达式本身的原因,却增加了相似度偶然重合的可能。 M_o 同样体现了对两个 Vague 值间真隶属度差距和假隶属度差距的同等重视,但依然未考虑该差距是正向差距还是反向差距。

上述的几种相似度量方式从一定程度上解决了两个 Vague 值之间的相似程度的确定问题,但每一种都是侧重于不同的角度提出的相似度量方法,都存在反例。一般地,影响两个 Vague 值之间的相似程度的因素有三个:真隶属函数 t_x ,假隶属函数 f_x 和未知值 $1 - t_x - f_x$,由于在上述几种度量方式下赋予 $|t_x - t_y|, |f_x - f_y|$ 和 $|(t_y + f_y) - (t_x + f_x)|$ 的权重是常数,其显著特征为没有考虑该差距是正向差距还是反向差距,因而增加了相似度重合的可能性。基于这种想法,提出了一种权重可变的相似度量,该方法可以大大降低相似度重合的可能性。

2 新 Vague 集的相似度量及其性质

2.1 新 Vague 值之间的相似度量

首先,我们来考虑一个例子:在选举时有 A, B, C, D 四个候选人,10 个选民,其中支持 A 的人有 1 个,反对 A 的人有 1 个;支持 B 的人有 2 个,反对 B 的人有 1 个;支持 C 的人有 7 个,反对 C 的人有 1 个;支持 D 的人有 8 个,反对 D 的人有 1 个。可以将 A, B, C, D 四人的得票情况看做 4 个 Vague 值, $A[0.1, 0.9], B[0.2, 0.9], C[0.7, 0.9], D[0.8, 0.9]$, 下面分别比较 A, B 和 C, D 的相似情况,利用公式 (1), (2), (3), (4) 的计算,结果如表 1 所示。

表 1 新 Vague 值之间的相似度量值

	x	y	M_C	M_H	M_L	M_o	M
A, B	$[0.1, 0.9]$	$[0.2, 0.9]$	0.95	0.95	0.95	0.95	0.968
C, D	$[0.7, 0.9]$	$[0.8, 0.9]$	0.95	0.95	0.95	0.95	0.953

可以看到利用 M_C, M_H, M_L, M_o 计算的 A, B 和 C, D 的相似情况均相同,假如在选举中只有一名被选中,在不忽略弃权的情况下, D 最有可能被选中, C 被选中的可能性小于 D, 而 A 和 B 被选中的可能性极小。从直观上看, A 和 B 应该更相似,均不可能被选中,而 C 和 D 的相似性要小些,因为 D 有可能被选中,而 C 有可能选不中。在 A, B, C, D 中我们更关心的是 C, D 的相似情况,因此考虑能否在我们关注之处将相似度的差距拉大,以避免陷入难以取舍的境地。即当 $t_x - t_y$ 和 $f_x - f_y$ 相同时,相似度可能也不相同,支持度 $(t_x + t_y)$ 较大时,相似度应小些,同理

分析可得,反对度 $(f_x + f_y)$ 较小时,相似度应小些。基于这种想法,本文提出了一种权重可变的相似度量 M' :

$$M' = 1 - \frac{|(t_x - t_y)|t_x - t_y| + (2 - f_x - f_y)|f_x - f_y| + |(t_x - t_y) - (f_x - f_y)|}{(t_x + t_y) + (2 - f_x - f_y) + 2}, \quad (5)$$

其中, $|t_x - t_y|$ 的系数 $(t_x + t_y)$ 的含义为:当 $|t_x - t_y|$ 相同时,如果 $(t_x + t_y)$ 较大,相似度应小些; $|f_x - f_y|$ 的系数 $(2 - f_x - f_y)$ 的系数的含义为:当 $|f_x - f_y|$ 相同时,如果 $f_x + f_y$ 较小,相似度应小些。

令 $t_x + t_y = p, t_x - t_y = q, f_x + f_y = m, f_x - f_y = n$ 则 (5) 式可以简化为

$$M' = 1 - \frac{|pq| + |(2 - m)n| + |q - n|}{p + (2 - m) + 2}. \quad (6)$$

M' 体现了对两个 Vague 值间真隶属度差距和假隶属度差距的重视,同时也蕴含了对 Vague 值支持度的重视。因为 $|(t_x - t_y) - (f_x - f_y)|$ 的引入,使得 M' 可以区分正向差距和反向差距。由于 M' 采用了可变权值的策略,降低了相似度偶然重合的可能性,并且权值的选取也是符合实际需要的,即人们更关心支持度较高而反对度较低状况下的相似度,为了便于比较,在支持度较高而反对度较低时,拉伸相似度的差距。由上述定义,我们可以得到如下的性质:

性质 1 $M'(x, y) \in [0, 1]$.

证明 因为 $t_x \in [0, 1], t_y \in [0, 1], f_x \in [0, 1], f_y \in [0, 1]$, 所以 $|t_x - t_y| \in [0, 1], |f_x - f_y| \in [0, 1], |(t_x - t_y) - (f_x - f_y)| \in [0, 2], M' \leq 1 - \frac{(t_x + t_y) * 0 + (2 - f_x - f_y) * 0 + 0}{(t_x + t_y) + (2 - f_x - f_y) + 2}$,

$$M' \geq 1 - \frac{(t_x + t_y) + (2 - f_x - f_y) + 2}{(t_x + t_y) + (2 - f_x - f_y) + 2} = 0.$$

性质 2 $M'(x, y) = M'(y, x)$.

性质 3 $M'(x, y) = 0 \Leftrightarrow x = [0, 0]y = [1, 1]$ 或 $x = [1, 1], y = [0, 0]$.

证明 如果 $x = [0, 0]y = [1, 1]$ 或 $x = [1, 1], y = [0, 0]$, 由定义显然有 $M'(x, y) = 0$;

若 $M'(x, y) = 0 \Rightarrow t_x - t_y = 1, f_x - f_y = -1$ 或 $t_x - t_y = -1, f_x - f_y = 1$,

$$\Rightarrow x = [0, 0], y = [1, 1] \text{ 或 } x = [1, 1], y = [0, 0].$$

性质 4 $M'(x, y) = 1 \Leftrightarrow x = y$.

证明 如果 $x = y$, 由定义显然有 $M'(x, y) =$

1,

若 $M'(x, y) = 1 \Rightarrow t_x - t_y = 0, f_x - f_y = 0$, 即

$x = y$.

例 1 应用文献[5]中提出的一组数据, 将本文提出的相似度量与其它几种度量方式进行比较:

可见 M' 比 M_C, M_H, M_L, M_o 有更强的区分能力.

表 2 各种相似度量方法分析比较

	1	2	3	4	5	6	7
x	[0.3, 0.7]	[0.3, 0.6]	[0.3, 0.8]	[1, 1]	[0.5, 0.5]	[0.4, 0.8]	[0.4, 0.8]
y	[0.4, 0.6]	[0.4, 0.7]	[0.4, 0.7]	[0, 1]	[0, 1]	[0.5, 0.7]	[0.5, 0.8]
M_C	1	0.9	1	0.5	1	1	0.95
M_H	0.9	0.9	0.9	0.5	0.5	0.9	0.95
M_L	0.95	0.9	0.95	0.5	0.75	0.95	0.95
M_o	0.9	0.9	0.9	0.3	0.5	0.9	0.93
M'	0.95	0.9	0.948	0.6	0.75	0.945	0.958

2.2 新 Vague 集之间的相似度量

假设 A 和 B 是论域 $U = \{u_1, u_2 \dots u_n\}$ 上的两个 Vague 集, 其中 $V_A(u_i) = [t_A(u_i), 1 - f_A(u_i)]$ 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, $A = \sum_{i=1}^n [t_A(u_i), 1 - f_A(u_i)]/u_i, B = \sum_{i=1}^n [t_B(u_i), 1 - f_B(u_i)]/u_i$, 则 Vague 集 A 和 B 之间的相似程度可以由下面的函数 T' 计算得到:

$$T'(A, B) = \frac{1}{n} \sum_{i=1}^n M'(V_A(u_i), V_B(u_i)) = \frac{1}{n} \sum_{i=1}^n \{1 - [|p(u_i)q(u_i)| + |(2 - m(u_i))n(u_i)| + |q(u_i) - n(u_i)|] / [p(u_i) + (2 - m(u_i)) + 2] \}, \quad (7)$$

其中 $t_x(u_i) + t_y(u_i) = p(u_i), t_x(u_i) - t_y(u_i) = q(u_i), f_x(u_i) + f_y(u_i) = m(u_i), f_x(u_i) - f_y(u_i) = n(u_i)$, 由上述定义, 我们可以得到如下的性质:

性质 5 $T'(A, B) \in [0, 1]$.

性质 6 $T'(A, B) = T'(B, A)$.

性质 7 $T'(A, B) = 0 \Leftrightarrow A = \sum_{i=1}^n [0, 0]/u_i,$

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

性质 8 $T'(A, B) = 1 \Leftrightarrow A = B$.

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

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

$$B = [0.4, 0.6]/u_1 + [0.0, 1.0]/u_2 + [0.5, 0.7]/u_3 + [0.0, 1.0]/u_4,$$

由(7) 可得 A 和 B 之间的相似度量

$$T'(A, B) = \frac{1}{4} \sum_{i=1}^4 M'(V_A(u_i), V_B(u_i)) =$$

$$[(1 - 0.05) + (1 - 0.25) + (1 - 0.055) + (1 - 0.4)]/4 = 0.811.$$

2.3 新 Vague 集间的加权相似度量

假设 A 和 B 是论域 $U = \{u_1, u_2 \dots u_n\}$ 上的两个 Vague 集, ω_i 为 U 上元素的 u_i 权重, $\omega_i \in [0, 1], 1 \leq i \leq n$, 则 A 和 B 之间的加权相似度可以用 $W'(A, B)$ 计算得到:

$$W'(A, B) = \sum_{i=1}^n \omega_i M'(V_A(u_i), V_B(u_i)) /$$

$$\sum_{i=1}^n \omega_i = \{ \sum_{i=1}^n \omega_i [1 - (|p(u_i)q(u_i)| + |(2 - m(u_i))n(u_i)| + |q(u_i) - n(u_i)|) / (p(u_i) + (2 - m(u_i)) + 2)] \} / \sum_{i=1}^n \omega_i. \quad (8)$$

例 3 假设 A 和 B 为例 2 中的 A, B , 论域 U 中的元素 u_1, u_2, u_3, u_4 上的权重分别为 $0.4, 0.2, 0.8, 0.6$, 由(8) 式可得 A 和 B 之间的加权相似度量

$$W'(A, B) = [0.4(1 - 0.05) + 0.2(1 - 0.25) + 0.8(1 - 0.055) + 0.6(1 - 0.4)] / (0.4 + 0.2 + 0.8 + 0.6) = (0.38 + 0.15 + 0.756 + 0.36) / 2.0 = 0.823.$$

3 结束语

本文分析比较了当前几种基于 Vague 集的相似度量方式的不足, 提出了一种新的度量 Vague 集之间相似程度的方法. 该方法的实质是: 充分考虑支持度以及真隶属度差距和假隶属度差距, 并且区分隶属度差距的方向(正向, 负向), 同时为了更符合实际的需要, 真隶属度差距和假隶属度差距采用可变权值. 实例证明, 该方法更符合实际, 它为 Vague 集之间的相似度提供了一种更好的度量方法.

参考文献:

[1] CHEN S M. Measures of similarity between vague sets [J]. Fuzzy Sets Syst, 1995, 74(2): 217-223.
 [2] CHEN S M. Similarity measures between vague sets and between elements [J]. IEEE Trans, Syst Man, Cybern, 1997, 27(1): 153-158.
 [3] HONG D H, KIM C. A note on similarity measures between vague sets and elements [J]. Information Science, 1999, 115: 83-96.
 [4] 李凡, 徐章艳. Vague 集之间的相似度量 [J]. 软件学报, 2001, 12(6): 922-927.
 [5] 李艳红, 迟忠先, 严德勤. Vague 相似度量与 Vague 熵 [J]. 计算机科学, 2002, 29(12): 129-132.

(责任编辑: 凌汉恩 邓大玉)