

Vague 集的相似度量^{*})

闫德勤

(辽宁师范大学计算机系 大连 116029)

摘要 Vague 集的相似性度量对于知识表达、模式识别等人工智能研究具有重要意义。Chen 给出的关于 Vague 集的相似性度量体现出注重确定 Vague 集中总体肯定度相似的思想,但相似度量的公式存在度量上的不足,也蕴含了在 Vague 集的不可知部分按支持度和反对度各占一半的概率计算,这样既不客观也不合理。为此,提出了一种改进的相似度量。

关键词 Vague 集,相似性度量,度量意义

Similarity Measures between Vague Sets

YAN De-Qin

(Department of Computer Science, Liaoning Normal University, Dalian 116029)

Abstract Similarity measures between vague sets are important for knowledge expression and pattern recognition. In this paper, similarity measure proposed by Chen is studied, and a modified measure is proposed.

Keywords Vague sets, Similarity measure, Mean of similarity measure

1 引言

Gau 等人于 1993 年提出 Vague 集^[2]的概念丰富了人们对事物刻画的工具,促进了模式识别、机器学习等领域的发展。Vague 集的相似性度量是 Vague 集研究与应用的一个重要方面,对于知识表达、模式识别等人工智能研究具有重要意义。1995 年 Chen^[3]提出了 Vague 集的相似性度量的方法。1999 年 Hong 等学者在文[5]中用例子说明 Chen 所提出的关于 Vague 集的相似性度量的方法不合理,并提出了新的关于 Vague 集的相似性度量的方法。继 Hong 之后一些学者从不同的角度对 Vague 集的相似性度量进行了研究,作者在文[6]中对已有的主要研究结果进行了说明。在文[6]中作者结合模糊相似的思想强调了 Vague 集的相似性度量的“距离”性。进一步地作者在文[7]文中明确指出 Hong 提出的关于 Vague 集的相似性度量与 Chen 提出的相似度量思想出发点是不同的,度量意义是不同的。

在本文中作者认为 Chen 给出的关于 Vague 集的相似性度量体现出注重确定 Vague 集中总体肯定度相似的思想,但相似度量的公式存在度量上的不足,也蕴含了在 Vague 集的不可知部分按肯定度和否定度各占一半的概率计算,这样既不客观也不合理。为此,提出了一种改进的相似度量。

2 关于 Chen 提出的相似度量

Vague 集是对模糊集的一个补充发展,为便于说明,下面首先给出模糊集与 Vague 集的定义。在以下定义 1 与定义 2 中,以 x 表示论域 U 中的任一元素。

定义 1^[1] 设在论域 U 上给定了一个映射 $A: U \rightarrow [0, 1], x \rightarrow A(x)$

则称 A 为 U 上的模糊(Fuzzy)集, $A(x)$ 称为 A 的隶属度函数

(或称为 x 对 A 的隶属度)。

定义 2^[2] 令 U 是一个(对象)空间, U 中的一个 Vague 集 V 用一个真隶属函数 $t_v(x)$ 和一个假隶属函数 $f_v(x)$ 表示, $t_v(x)$ 是从支持 x 的证据所导出的对 x 肯定隶属度下界, $f_v(x)$ 则是从反对 x 的证据所导出的对 x 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,简写即

$$t_v: U \rightarrow [0, 1];$$

$$f_v: U \rightarrow [0, 1].$$

其中 $t_v + f_v \leq 1$ 。

令 $x = [t_x, 1 - f_x], y = [t_y, 1 - f_y]$ 是论域 U 上的两个 Vague 集,Chen^[3]定义了 x, y 之间的相似度量 M_C 如下:

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

其中, $S(x) = t_x - f_x, S(y) = t_y - f_y$ 分别称为 x, y 的记分。

这里可以看出 Chen 要表达的是两个 vague 集以“记分”的差异决定相似度。记分的实质就是确定 vague 集中总体肯定(肯定隶属度减去否定隶属度)的程度。这是 Chen 关于 vague 集相似度量的意义所在。

Vague 集中 t_v 和 f_v 把区间 $[0, 1]$ 分成了三部分,为了便于说明,记中间部分为:

$$m_v = 1 - t_v - f_v.$$

沿用 Vague 集的定义, m_v 属于不可知隶属度,即是从既不支持又不反对的证据所导出的既不肯定又不否定的程度。从投票模型来解释 Vague 集的角度看, m_v 表示弃权的比例。从对事物可知程度的表示角度看, m_v 表示对事物的不可知程度。

^{*} 基金项目:国家自然科学基金(No. 60372071);辽宁省教育厅高等学校科学研究基金(2004C031);辽宁师范大学校基金。闫德勤 博士,教授,主要从事模式识别、数据挖掘和粗糙集等方面的研究。

相似度量 $M_C(x, y)$ 的出发点基于 $S(x) = t_x - f_x$ 和 $S(y) = t_y - f_y$ 。这表明 $M_C(x, y)$ 强调在纯支持度(把反对或否定看作是对支持和肯定的一种抵消)意义下两个集合的相似程度。当 $t_x - t_y = f_x - f_y$ 时,由(1)式知 $M_C(x, y) = 1$ 。但此时 $m_x - m_y$ 可能较大,例如对 $x = [0.1, 0.9], y = [0.4, 0.6]$, 有 $m_x = 0.8, m_y = 0.2$, 但 $M_C(x, y) = 1$ 。这是引起讨论的一个方面^[5~7]。作者认为,对于 Chen 提出的相似度量 $M_C(x, y)$ 的合理性不能简单使用几何相似的直观来判定。

然而,下面的一个重要的例子说明了相似度量 $M_C(x, y)$ 存在着缺陷:

对于 $x = [0, 1], y = [0.5, 0.5]$, 有 $M_C(x, y) = 1$ 。

在这个例子中,说明 vague 集 $x = [0, 1]$ 没有表示任何对事物确定的信息,即对于肯定或否定的情况完全不可知。或者说是百分之百的不确定。而 vague 集 $y = [0.5, 0.5]$ 则是百分之五十的肯定和百分之五十的否定,更重要的是表示对事物是完全可知的。 $M_C(x, y) = 1$ 说明对事物的完全不确定等价于百分之五十的肯定和百分之五十的否定。这在有些情况下是可以这样处理的。但是在极大部分情况则不能这样处理。更大的问题是这个例子说明公式 $M_C(x, y)$ 把完全不可知和完全可知等价,这对于知识与信息的表示和处理会带来错误。同时,公式也蕴含了在 Vague 集的不可知部分按肯定度和否定度各占一半的概率来计算,这是一种人为的规定性,不能客观地体现出事物间的关系。

为此,下面我们提出一种改进的相似度量。

3 新的相似度量

我们把关于 vague 集的 $S(x) = t_x - f_x$ 和 $S(y) = t_y - f_y$ 看作总体支持(肯定)度,同时令 $K(x) = t_x + f_x, K(y) = t_y + f_y$ 。称 $K(x)$ 和 $K(y)$ 分别为关于 vague 集 x, y 的总体可知度。我们提出的新的相似度量同时考虑到支持度和可知度两个方面,其公式如下:

$$M_Y(x, y) = 1 - \frac{|S(x) - S(y)|}{2} - \frac{|K(x) - K(y)|}{2}$$

对于 $x = [0, 1], y = [0.5, 0.5]$, 利用新的相似度量计算得到: $M_Y(x, y) = 0.5$ 。这个结果可以看出对于区分用 vague 集表示的事物 $M_Y(x, y)$ 比 $M_C(x, y)$ 更精确。

定理 1 相似度量 $M_Y(x, y)$ 有以下性质:

- (1) $0 \leq M_Y(x, y) \leq 1$
- (2) $M_Y(x, y) = M_Y(y, x)$
- (3) $M_Y(x, y) = 1$ 当且仅当 $x = y$

证明:定理的(1)和(2)容易证明,这里略去。下面证明性质(3):由 $M_Y(x, y)$ 的表达式可以得到, $M_Y(x, y) = 1$ 当且仅当 $S(x) = S(y)$ 以及 $K(x) = K(y)$ 。由此可以推出 $t_x = t_y$ 和 $f_x = f_y$, 即 $x = y$ 。证毕。

设 A 和 B 为在论域 $U = \{u_1, u_2, \dots, u_n\}$ 上的两个 vague 集,其表示式为:

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

此时,令

$$\begin{aligned} S(V_A(u_i)) &= t_A(u_i) - f_A(u_i) \\ S(V_B(u_i)) &= t_B(u_i) - f_B(u_i) \\ S(V_A(u_i)) &= t_A(u_i) + f_A(u_i) \\ S(V_B(u_i)) &= t_B(u_i) + f_B(u_i) \end{aligned}$$

则 vague 集 A 和 B 的相似度量为:

$$\begin{aligned} T_Y(A, B) &= \frac{1}{n} \sum_{i=1}^n M_Y(V_A(u_i), V_B(u_i)) \\ &= \frac{1}{n} \sum_{i=1}^n \left(1 - \frac{|S(V_A(u_i)) - S(V_B(u_i))|}{2} \right) \\ &\quad - \frac{1}{n} \sum_{i=1}^n \left(\frac{|K(V_A(u_i)) - K(V_B(u_i))|}{2} \right) \end{aligned}$$

利用 vague 集的相似度量可以进行模式识别、信息处理、数据挖掘等方面的研究与应用。相关应用例子见文[3~5]及其它相关文献,这里略去。

结论 相似度量的合理与科学性对其理论研究和应用具有深刻的意义。本文从 vague 集相似度量的内涵上论证了 Chen 提出的相似度量的意义和不足,给出了一种改进的相似度量方法。该方法对于 vague 集及其相似度量的研究具有新的意义。

参考文献

- 1 Zadeh L A. Fuzzy sets, Inform. And Control, 1965, 8: 338~356
- 2 Gau W L, Buehrer D J. Vague sets. IEEE Transactions on Systems, Man, and Cybernetics, 1993, 23(2): 610~614
- 3 Chen S M. Measures between vague sets. Fuzzy sets Syst., 1995, 74(2): 217~223
- 4 Chen S M. Similarity measure between Vague sets and elements. IEEE Transactions on Systems, Man, and Cybernetics, 1997, 27(1): 153~158
- 5 Hong D H, Kim C. A note on similarity measures between vague sets and between elements. Information Science, 1999, 115: 83~96
- 6 闫德勤,迟忠先,李艳红. 关于 Vague 集的相似度量. 模式识别与人工智能, 2004, 17(1): 22~26
- 7 闫德勤,迟忠先. 关于 Vague 集相似度量的一个注记. 计算机科学, 2005, 32(10)