

Vague 集的模糊熵

金 兰, 杨 勇, 牛彩云

(西北师范大学 数学与信息科学学院, 甘肃 兰州 730070)

摘 要: Vague 集模糊熵的度量必须兼顾 Vague 集的未知性和它的不确定性. 文中反例证实现有方法在某些情况下得出的结果与实际结论不一致, 并提出了一个新的 Vague 集模糊熵计算方法. 通过定理和实例分析证明: 该方法同时考虑到了影响 Vague 集模糊熵的两个因素.

关键词: Vague 集; 隶属函数; 模糊熵; 不确定性

中图分类号: TP18

文献标识码: A

文章编号: 1000-7180(2010)04-0133-04

Fuzzy Entropy of Vague Sets

JIN Lan, YANG Yong, NIU Cai-yun

(College of Mathematics and Information Science, Northwest Normal University, Lanzhou 730070, China)

Abstract: The fuzzy entropy of vague sets should include the unknown and uncertainty. Some counterexamples are given to show that the result of the present methods in some cases and real conclusions are contrary, a new method of fuzzy entropy of vague sets is presented and a theorem proves that the method actually takes the two factors into account simultaneously.

Key words: vague set; membership function; fuzzy entropy; uncertainty

1 引言

1965 年, Zadeh 提出了 Fuzzy 集理论^[1]. 但模糊集的隶属函数值是一个单一的值, 它不能同时表示支持和反对的证据, 为此, Gau 和 Buehrer 于 1993 年提出了 Vague 集理论^[2]. 在一个 Vague 集 A 中, 用一个真隶属函数 $t_A(x)$ 和一个假隶属函数 $f_A(x)$ 来描述其隶属度的界, 这两个界构成区间 $[0, 1]$ 的一个子区间 $[t_A(x), 1 - f_A(x)]$. 其中一个对象的支持度、反对度和未知度分别为 $t_A(x)$ 、 $f_A(x)$ 和 $1 - t_A(x) - f_A(x)$. 这使 Vague 集在处理不确定性信息时比传统的模糊集有更强的表示能力且更具灵活性, 但这也导致了它的模糊性不但来自数据的未知信息 $1 - t_A(x) - f_A(x)$, 而且还来自数据本身的不确定性^[3]. 比如 Vague 集元素取值 $[0.5, 0.5]$ 时, 尽管此时并不存在任何证据的未知度, 但由于肯定和

否定的证据各占 50%, 使得人们很难做出合理的判断, 此时应具有最大的模糊性.

熵是信息论中的一个概念, 它主要是用来刻画一个对象所蕴含的平均信息量. 目前, 熵的概念已被成功应用到了很多的领域, 为一些问题的解决提供了更有效的工具. 模糊熵的概念是 Zadeh 于 1969 年首次提出来的, 它主要用来刻画一个模糊集的模糊性, 对于模糊熵的研究, 很多学者给出了不同的表示和计算方法. 文中在研究现有 Vague 集模糊熵度量方法的基础上, 提出了一种新的 Vague 集模糊熵度量方法, 并通过数据分析与现有方法作比较, 验证了该方法的有效性和合理性.

2 基础知识

定义 1 (Vague 集) 设 X 是一个点(对象)空间, 其中任意一个元素用 x 表示, X 上的一个 Vague

收稿日期: 2009-03-31

基金项目: 国家自然科学基金项目(10771171); 兰州市科技攻关项目(2008-1-34)

集 A 用一个真隶属函数 $t_A(x)$ 和一个假隶属函数 $f_A(x)$ 表示, $t_A(x)$ 表示从支持 x 的证据所导出的 x 的肯定隶属度下界, $f_A(x)$ 表示从反对 x 的证据导出的 x 的否定隶属度的下界, $t_A(x)$ 和 $f_A(x)$ 将区间 $[0,1]$ 中的一个实数和 X 中每一个点联系起来, 即:

$$t_A: X \rightarrow [0,1];$$

$$f_A: X \rightarrow [0,1].$$

元素 x 与 A 的关系表示为 $[t_A(x), 1 - f_A(x)]$, 其中 $t_A(x) + f_A(x) \leq 1$.

设 A 为一个 Vague 集, 当 X 为连续时, Vague 集 A 可以表示为 $A = \int_{x \in X} [t_A(x), 1 - f_A(x)]/x$; 当 X 为离散时, Vague 集 A 则表示为 $A = \sum_{x \in X} [t_A(x), 1 - f_A(x)]/x$.

Vague 集 A 中, 差值 $\pi_A(x) = 1 - t_A(x) - f_A(x)$ 描述了 x 对于 Vague 集 A 的不确定程度, 它刻画了 x 相对于 Vague 集 A 的踌躇程度, 是 x 相对于 A 的未知信息的一种度量. 若 $1 - f_A(x) = t_A(x) \in [0,1]$, 则 Vague 集退化为模糊集; 若 $1 - f_A(x) = t_A(x) \in \{0,1\}$, 则 Vague 集将退化为经典集合.

定义 2 一个 Vague 集 A 的补 $\bar{A}$ 定义为 $\bar{t}_A = f_A, 1 - \bar{f}_A = 1 - t_A$.

定义 3 Vague 集 A 和 $B, A \subseteq B$ 当且仅当对 $\forall x \in X$ 有 $t_A(x) \leq t_B(x), 1 - f_A(x) \leq 1 - f_B(x)$.

定义 4 对 $\forall x \in X$, 若有 $V_A(x) = [0,0]$ 或 $[1,1]$, 则称 A 为非模糊集.

Szmidt 和 Kacprzyk 给出了一个直觉模糊集嫡的公理化定义, 因文献[4] 的原因, 这里以 Vague 集的形式引述如下:

定义 5 称函数 $E: VS_S(X) \rightarrow [0,1]$ 为 Vague 集 A 的嫡^[5], 如果它满足如下条件:

(SK₁) $E(A) = 0$ 当且仅当 $V_A(x_i) = [0,0]$ 或 $[1,1]$;

(SK₂) $E(A) = 1$ 当且仅当 $\forall x_i \in X$, 有 $t_A(x_i) = f_A(x_i)$;

(SK₃) 当 $t_B(x_i) \leq f_B(x_i)$ 时, 若有 $t_A(x_i) \leq t_B(x_i)$ 且 $f_A(x_i) \geq f_B(x_i)$; 或当 $t_B(x_i) \geq f_B(x_i)$ 时, 若有 $t_A(x_i) \geq t_B(x_i)$ 且 $f_A(x_i) \leq f_B(x_i)$ 则都有 $E(A) \leq E(B)$;

(SK₄) $E(A) = E(\bar{A})$.

3 现有 Vague 集模糊嫡方法的缺陷

很多学者提出了不同的方法来计算 Vague 集的模糊嫡, 但通过下面的一个例子说明它们计算 Vague 集的模糊嫡的结果是不符合实际的.

例 1 令两个 Vague 集 $A = [0.4, 0.4]/x_1$ 和 $B = [0.4, 0.9]/x_1$, 使用文献[5] 中的方法计算可得出相同的结果: $E(A) = 0.667, E(B) = 0.667$, 很明显这与事实是不符的.

例 2 令两个 Vague 集 $A = [0.3, 0.3]/x_1$ 和 $B = [0.2, 0.3]/x_1$, 使用已有的一些方法得出的模糊嫡结果如表 1 所示.

表 1 现有几种 Vague 模糊嫡计算方法比较

	$E_1^{[5]}$	$E_2^{[6]} E_3^{[7]} E_4^{[8]}$	$E_5^{[9]}$	$E_6^{[10]}$
A	0.429	0.429	0.300	0.429
B	0.333	0.375	0.300	0.419

用投票模型来解释上面所得结果, 即 A 是 3 人支持、7 人反对、无人中立, 而 B 是 2 人支持、7 人反对、1 人中立. 很显然, A 中的所有入都有自己的见解, 要么同意, 要么反对, 而 B 中还有中立的情况. 那么, 按照直觉判断, A 的模糊性肯定比 B 的要小. 然而, 由上表可以看出, 已有的这些方法计算出的 A 与 B 两者间的模糊性与上面所分析的刚好相反. 这说明已有的方法存在着一定的局限性和不足. 因此, 文中提出了一种新的 Vague 集模糊嫡计算方法, 使计算结果更直观有效.

4 Vague 集的新模糊嫡

定理 1 设 A 是论域 X 上的一个 Vague 集, $V_A(x) = [t_A(x), 1 - f_A(x)]$ 是 A 中的一个 Vague 值, 则它的模糊嫡可由以下函数计算:

$$E(V_A(x)) = \frac{\pi_A(x) + 2 - |t_A(x) - f_A(x)| - (t_A(x) + f_A(x))^2}{\pi_A(x) + 2 + |t_A(x) - f_A(x)| - (t_A(x) + f_A(x))^2}$$

为了进一步说明所提出的新的嫡的公式的有效性, 给出以下证明, 证实它符合定义 5 中的四个约束条件.

命题 1 设 A 是论域 X 上的一个 Vague 集, E 如上定义, 则 $E(A) = 0$ 当且仅当 $V_A(x) = [0,0]$ 或 $[1,1]$.

证明:

$E(A) = 0$, 有 $\pi_A(x) + 2 - |t_A(x) - f_A(x)| - (t_A(x) + f_A(x))^2 = 0$, 则 $1 + (1 - (t_A(x) +$

$f_A(x))) (2 + (t_A(x) + f_A(x))) - |t_A(x) - f_A(x)| = 0$, 有 $\pi_A(x) (2 + (t_A(x) + f_A(x))) = |t_A(x) - f_A(x)| - 1$. 因为 $0 \leq |t_A(x) - f_A(x)| \leq 1$, $0 \leq \pi_A(x) \leq 1$, $2 \leq (2 + (t_A(x) + f_A(x))) \leq 3$, 所以要使等式成立必然有 $|t_A(x) - f_A(x)| = 0$ 且 $\pi_A(x) = 0$, 即 $t_A(x) = 0, f_A(x) = 1$ 或 $t_A(x) = 1, f_A(x) = 0$.

由 $V_A(x) = [0, 0]$ 或 $[1, 1]$, 显然有 $E(A) = 0$.

命题2 设 A 是论域 X 上的一个 Vague 集, E 如上定义, 则 $E(A) = 1$ 当且仅当 $\forall x \in X$, 有 $t_A(x) = f_A(x)$.

证明:

$E(A) = 1$, 有 $\pi_A(x) + 2 - |t_A(x) - f_A(x)| - (t_A(x) + f_A(x))^2 = \pi_A(x) + 2 + |t_A(x) - f_A(x)| - (t_A(x) + f_A(x))^2$, 则 $-|t_A(x) - f_A(x)| = |$

$$F'_t = \frac{-2(t+f)(3-2t-(t+f)^2) - (-2-2(t+f))(3-2f-(t+f)^2)}{(3-2t-(t+f)^2)^2}$$

$$= \frac{2(t-f)^2 + 6 - 4f(1+2f)}{(3-2t-(t+f)^2)^2} \geq 0,$$

$$F'_f = \frac{(-2-2(t+f))(3-2t-(t+f)^2) - (-2(t+f))(3-2f-(t+f)^2)}{(3-2t-(t+f)^2)^2}$$

$$= \frac{-6 + 4t(1+2t) - 2(t-f)^2}{(3-2t-(t+f)^2)^2} \leq 0,$$

即 F 是关于 t 的增函数, 关于 f 的减函数.

于是由条件 $t_A(x) \leq t_B(x) \leq f_B(x) \leq f_A(x)$, 有结论 $E(A) \leq E(B)$.

同理, 当 $t_A(x) \geq t_B(x) \geq f_B(x) \geq f_A(x)$ 时, $E(A) \geq E(B)$.

命题4 设 A 是论域 X 上的一个 Vague 集, E 如上定义, 则 $E(A) = E(\bar{A})$. 显然成立.

定理2 已知论域 $X = \{x_1, x_2, \dots, x_n\}$, 设 $A = \sum_{i=1}^n [t_A(x_i), 1 - f_A(x_i)]/x_i$ 是论域 X 上的一个 Vague 集, 则它的模糊熵可由以下函数计算:

$$E(A) = \frac{1}{n} \sum_{i=1}^n \cdot$$

$$\frac{\pi_A(x_i) + 2 - |t_A(x_i) - f_A(x_i)| - (t_A(x_i) + f_A(x_i))^2}{\pi_A(x_i) + 2 + |t_A(x_i) - f_A(x_i)| - (t_A(x_i) + f_A(x_i))^2}$$

5 实例分析

对于例1中的 Vague 集 A 和 B 用定理2中的方法进行计算:

$$E(A) = 0.667 \quad E(B) = 0.765$$

$t_A(x) - f_A(x)|$ 即 $t_A(x) = f_A(x)$.

由 $t_A(x) = f_A(x)$, 显然有 $E(A) = 1$.

命题3 设 A, B 是论域 X 上的两个 Vague 集, E 如上定义, $\forall x \in X$, 则当 $t_B(x) \leq f_B(x)$ 时, 若有 $t_A(x) \leq t_B(x)$ 且 $f_A(x) \geq f_B(x)$; 或当 $t_B(x) \geq f_B(x)$ 时, 若有 $t_A(x) \geq t_B(x)$ 且 $f_A(x) \leq f_B(x)$ 则都有 $E(A) \leq E(B)$.

证明:

$t_A(x) \leq t_B(x) \leq f_B(x) \leq f_A(x)$, 有

$$\frac{\pi_A(x) + 2 - |t_A(x) - f_A(x)| - (t_A(x) + f_A(x))^2}{\pi_A(x) + 2 + |t_A(x) - f_A(x)| - (t_A(x) + f_A(x))^2} = \frac{3 - 2f_A(x) - (t_A(x) + f_A(x))^2}{3 - 2t_A(x) - (t_A(x) + f_A(x))^2}$$

设有函数 $F = \frac{3 - 2f - (t+f)^2}{3 - 2t - (t+f)^2}$, 由 $t \leq f$ 及

$t + f \leq 1$ 可得 $t \leq \frac{1}{2}$. 对 F 中的 t 与 f 分别求导得

显然这个结果纠正了例1中所使用方法的计算结果, 使 Vague 集的模糊熵的计算更符合实际^[6-8].

对于例2中的 Vague 集 A 和 B 用定理2中的方法进行计算:

$$E(A) = 0.429 \quad E(B) = 0.441$$

显然这个结果纠正了例2中所提方法的计算结果, 使 Vague 集的模糊熵的计算更符合人们的直觉.

例3 针对文献[8]和文献[10]中的例1数据, 令两个 Vague 集 $A = [0.1, 0.2]/x_1$ 和 $B = [0.2, 0.2]/x_1$, 根据定理2中的方法进行计算:

$$E(A) = 0.296 \quad E(B) = 0.250$$

显然这个结果不存在文献[8]和文献[10]中所说的缺陷, 是符合实际的.

例4 令 $X = \{x_1, x_2\}$, $A = [0.0, 0.6]/x_1 + [0.4, 1]/x_2$, 用定理2中的方法进行计算有:

$$E(A) = 0.718$$

本例表明 Vague 集的模糊熵的度量不仅要考虑到 Vague 集的未知性, 还要考虑到它的不确定性, 即由其真隶属度与假隶属度之间的距离所产生的不确定性. 明显可以看出, 本例的未知性为 0.6,

而真/假隶属度之间的差为 0.4. 因此可以确定, 肯定存在不确定性, 即它的模糊熵一定是大于 0.6 的, 从而证实本文所提出的模糊熵计算方法的正确性和有效性.

6 结束语

Vague 集^[11-12]的理论研究还在不断的深入, 对于 Vague 集模糊熵的研究更是得到了大量的关注, 许多学者也不断的提出新的 Vague 集模糊熵的度量方法. 文中在研究现有方法的基础之上, 给出了 Vague 集的一般定义以及 Vague 集模糊熵的公理化定义, 提出了一个新的 Vague 值以及 Vague 集的模糊熵的度量方法, 并结合公理化定义对该方法给予了证明, 证实其可用性. 最后通过实验数据进一步说明了该方法的有效性和合理性.

参考文献:

- [1] Zadeh L A. Fuzzy sets[J]. Information and Control, 1965, 8(3):338-357.
- [2] Gau W L, Buehrer D J. Vague sets[J]. IEEE Transactions on Systems, Man and Cybernetics, 1993, 23(2):610-614.
- [3] 马志锋, 邢汉承, 郑小妹. 不完整 Vague 决策表中的近似学习方法[J]. 计算机研究与发展, 2000, 37(9):1050-1057.

- [4] Bustince H, Burillo P. Vague sets are intuitionistic fuzzy sets[J]. Fuzzy Sets and Systems, 1996, 79(3):403-405.
- [5] 刘云生, 黄国顺. Vague 集的熵[J]. 小型微型计算机系统, 2006, 27(11):2115-2119.
- [6] 黄国顺, 刘云生. 关于 Vague 集的模糊熵[J]. 计算机工程与应用, 2005, 41(33):48-50.
- [7] Szmidt Eulalia, Kacprzyk J. Entropy for thtutionistic fuzzy sets[J]. Fuzzy sets and systems, 2001, 118(3):467-477.
- [8] 李天志, 梁家荣, 范平, 等. Vague 集的模糊熵[J]. 计算机应用研究, 2007, 24(10):93-98.
- [9] 李艳红, 迟忠先, 阎德勤. Vague 相似度量与 Vague 熵[J]. 计算机科学, 2002, 29(12):129-132.
- [10] 范平, 梁家荣, 李天志. Vague 集的新模糊熵[J]. 计算机工程与应用, 2007, 43(13):179-181.
- [11] 吴成茂. 模糊熵的修改及其在图像分割中的应用[J]. 微电子学与计算机, 2008, 25(1):25-28.
- [12] 王新程, 杨蓉蓉. 基于最大熵的 P2P 文件污染防治机制[J]. 微电子与计算机, 2009, 26(1):190-192.

作者简介:

金 兰 女, (1981-), 硕士研究生. 研究方向为数据挖掘、粗糙集理论.

杨 勇 男, (1967-), 博士, 副教授. 研究方向为粗糙集理论.

牛彩云 女, (1981-), 硕士研究生. 研究方向为数据挖掘、粗糙集理论.

(上接第 132 页)

义. 文中利用这种虚拟化的思想, 在以冯诺依曼体系结构为基础的操作系统上实现了哈佛内存体系结构, 这种方法的优点是对操作系统的修改十分轻微. 实验结果表明, 这是一种有效的防内存扫描的 Rootkit 技术.

参考文献:

- [1] 孙淑华, 马恒太, 张楠, 等. 内核级木马隐藏技术研究与实践[J]. 微电子学与计算机, 2004, 21(3):76-80.
- [2] 孙淑华, 马恒太, 卿斯汉. 内存映射型内核级木马的研究与改进[J]. 微电子学与计算机, 2004, 21(11):15-19.
- [3] Fuzen, FU rootkit[EB/OL]. (2008-01-25)[2009-03-10]. <http://www.rootkit.com/project.php?id=12>.

- [4] John G, Levine, Julian B, et al. Detecting and categorizing kernel-level rootkits to aid future detection[J]. IEEE Security & Privacy, 2006, 4(1):24-32.
- [5] Sparks S, Butler J. "Shadow walker": raising the bar for windows rootkit detection [C]// Black Hat Japan. Japan, Tokyo, 2005.
- [6] Chen S, Xu J, Sezer E C, et al. Noncontrol-data attacks are realistic threats[C]// Proc. USENIX Security Symposium. USA: Baltimore, 2005.

作者简介:

易再尧 男, (1972-), 博士研究生. 研究方向为网络与信息安全.