

一个基于未知度的 Vague 集(值)相似度量新方法

刘 庆,辛小龙,张 磊

(西北大学 数学系,陕西 西安 710069)

摘 要:相似度是研究和应用 Vague 集(值)的重要工具。分析了现有 Vague 集(值)之间的相似度量方法,揭示了这些方法不能准确刻画 Vague 集(值)之间相似度量的本质,并总结了相似度量的一般准则,指出了影响 Vague 集相似度量的3个因素。在此基础上充分考虑了未知度,它为 Vague 集之间的相似度提供了一种更好的度量方法。最后用实例说明其应用以及该方法的有效性和直观性。

关键词:Vague 集(值);相似度量;真/假隶属度函数;未知度

中图分类号:TP301.6

文献标识码:A

文章编号:1673-629X(2009)03-0148-03

A New Method of Similarity Measures Based on Hesitancy Degree Between Vague Sets

LIU Qing, XIN Xiao-long, ZHANG Lei

(Department of Mathematics, Northwest University, Xi'an 710069, China)

Abstract: Similarity measures are important tools for studying and applying Vague sets. Analyzed the existing methods of similarity measures between Vague sets, and pointed out that these measures methods couldn't describe the nature of similarity measures between Vague sets. Furthermore, the same standard of these measures are pointed out. It is shown that three factors influencing Vague sets should be taken into account while calculating the similarity degree. Based on them, hesitancy degree is fully considered. It is revealed that it is a better method in measuring the similarity between Vague sets. Finally, the application of the similarity measures and its validity and intuition are illustrated by an example.

Key words: Vague sets; similarity measures; true/false membership function; hesitancy degree

0 引 言

Zadeh 于 1965 年首先提出了 Fuzzy 集理论^[1],在随后的几十年中,它不断地得到完善和发展,并在多个领域得到成功应用,但是 Fuzzy 集的隶属度函数是一个单一值,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],在这一理论中,论域中的元素与论域上的集合之间的关系不再是 Fuzzy 集中的“一定程度上属于”的关系,而是“在一定程度范围之内属于”的关系,这种对论域中元素与论域上集合之间关系的新界定,可以描述人们对未知事物的未知性而表现出来的思维的不确定性。

目前,Vague 集已经被成功应用于智能系统、信息融合等,在这些应用中,相似度量起到非常重要的作用,因此研究 Vague 集的相似度量具有重要的意义。

当前,已经有多种度量方法^[3~7],它们的侧重点不同,都有各自的优点与局限,在实践中可以根据需要进行选取。文中充分考虑了未知信息的重要性,认为不能像大多学者考虑的那样,把未知部分看作肯定和否定各占一半,提出了一种充分考虑未知度的 Vague 集之间相似度量的新方法。

1 基础知识

定义 1^[2] 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 集 A 用一个真隶属度函数 $t_A(x)$ 和一个假隶属度函数 $f_A(x)$ 表示, $t_A(x)$ 是从支持 x 的证据导出的 x 的肯定隶属度的下界, $f_A(x)$ 是从反对 x 的证据所导出的 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 的每一点联系起来。即

$$t_A(x): X \rightarrow [0,1], f_A(x): X \rightarrow [0,1]$$

x 关于 A 的隶属度 $V(x)$ 表示为 $[t_A(x), 1 - f_A(x)]$,其中 $t_A(x) + f_A(x) \leq 1$,特别当 $t_A(x) +$

收稿日期:2008-06-29

基金项目:陕西省自然科学基金(2007A19)

作者简介:刘 庆(1978-),男,硕士,研究方向为模糊代数、信息安全;辛小龙,博士,教授,博士生导师,研究领域为代数学、信息安全。

$f_A(x) = 1$ 时,它就退化为普通模糊集。

设 A 为一个 Vague 集,当 X 离散时将其表示为:

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

当 X 连续时将其表示为:

$$A = \int [t_A(x_i), 1 - f_A(x_i)] / x_i, \quad x_i \in X$$

定义 2 称 $S_A(x) = t_A(x) - f_A(x)$ ($-1 \leq S_A(x) \leq 1$) 为 x 的核,它表征现有证据对元素 x 支持和反对两种力量的对比,看作整体支持(肯定)度。

定义 3 称 $\pi_A(x) = 1 - t_A(x) - f_A(x)$ ($0 \leq \pi_A(x) \leq 1$) 为元素 x 对 Vague 集 A 的未知度,它表征对事物认识的不确定度,未知度越大,认识的精度越低。

2 当前 Vague 集(值)之间的相似度量方法和一般准则

2.1 已有 Vague 集之间的相似度量方法及不足

目前 Vague 集之间的相似度量方法很多,有代表性的如下:

方法 1:在文献[3]中,chen 等定义了 Vague 值 x, y 之间相似度量方法如下:

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

方法 2:在文献[4]中,Hong 等定义了 Vague 值 x, y 之间相似度量方法如下:

$$M_H = 1 - \frac{|t_x - t_y| + |f_x - f_y|}{2}$$

方法 3:在文献[5]中,李凡等定义了 Vague 值 x, y 之间相似度量方法如下:

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

方法 4:在文献[6]中,李艳红等定义了 Vague 值 x, y 之间相似度量方法如下:

$$M_0 = \sqrt{\frac{(t_x - t_y)^2 + (f_x - f_y)^2}{2}}$$

方法 5:在文献[7]中,夏少云等定义了 Vague 值 x, y 之间相似度量方法如下:

$$M_X = 1 - |\mu_A(x) - \mu_B(y)|, \text{ 其中 } \mu_A(x), \mu_B(y) \text{ 为元素 } x, y \text{ 对 Vague 集 } A \text{ 的精确隶属度,且 } \mu_A(x) \in [t_A(x), 1 - f_A(x)], \mu_B(x) \in [t_B(x), 1 - f_B(x)]$$

以上度量方法分别是不同角度提出的 Vague 集之间的相似度量方法,都能够解决一定的实际问题,但都不能适合一般情况,比较明显的缺陷是认为未知度不为零的两个相同元素相似度为 1。造成上述缺陷的

原因是没有充分考虑未知度大小对相似度的影响。目前还尚未找到一种完全符合客观事实和直观感觉,不存在反例的相似度量方法,在实践中需要根据需要进行选取。文中充分考虑未知度大小对相似度的影响,从另一角度提出了一种新的度量方法。

2.2 Vague 集之间的相似度量方法的一般准则

文献[2~7]列举了众多学者在研究相似度量方面逐渐形成的一些共识,但是目前尚无一种明确、全面的理论描述,文中将这些共识总结为以下一般准则。

设 $x = [t_A(x), 1 - f_A(x)], y = [t_A(y), 1 - f_A(y)]$ 是 Vague 集 A 中的两个 Vague 值, $M(x, y)$ 表示两个 Vague 值之间的相似度,则 $M(x, y)$ 应该满足下列准则:

准则 1 $0 \leq M(x, y) \leq 1$

准则 2 $M(x, y) = M(y, x)$

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

准则 4 $M(x, y) = 1 \Leftrightarrow x = y$

3 新的 Vague 集(值)之间的相似度量方法

3.1 新的 Vague 值之间的相似度量方法

文中从另一角度,充分考虑了未知度 $\pi_A(x) = 1 - t_A(x) - f_A(x)$,避免了传统的把未知信息一分为二,肯定和否定各占一半的思想,使比较的结果更合理,更接近实际情况。衡量了 Vague 集(值)之间的相似度量方法要考虑以下三个方面的影响:

(1) 相对优势。即 $t_x - t_y$ 和 $f_x - f_y$,在其他条件相同的情况下,如果 Vague 值 x 比 y 具有越小的相对优势,则 Vague 值 x 和 y 的相似度越大;

(2) 相对已知信息的多少。即 $(t_x + t_y) - (f_x + f_y)$,在其他条件相同的情况下,如果 Vague 值 x 比 y 具有越少的已知信息,则 Vague 值 x 和 y 的相似度越大;

(3) 相对未知信息的多少。即 $\pi_A(x) - \pi_B(x) = (1 - t_A(x) - f_A(x)) - (1 - t_B(x) - f_B(x))$,在其他条件相同的情况下,如果 Vague 值 x 比 y 具有越少的已知信息,则 Vague 值 x 和 y 的相似度越大。

据此,文中提出新的 Vague 值之间的相似度量方法,设 $x = [t_x, 1 - f_x]$ 和 $y = [t_y, 1 - f_y]$ 是 Vague 集 A, B 上的两个 Vague 值,则 x, y 间的相似度量公式为:

$$\overline{M(x, y)} = [1 - |\pi_A(x) - \pi_B(y)|] \times (1 - \frac{|t_x - t_y| + |f_x - f_y|}{2})$$

定理 1 $\overline{M(x, y)}$ 满足 Vague 集之间的相似度量方法的一般准则。

证明 准则 2 和 3 是显然的,只证准则 1 和 4。

准则 1 证明 由 Vague 集定义

$$0 \leq |t_x - t_y| \leq 1, 0 \leq |f_x - f_y| \leq 1, 0 \leq |\pi_A(x) - \pi_B(y)| \leq 1$$

$$\text{所以 } 0 \leq |t_x - t_y| + |f_x - f_y| \leq 2$$

由此可得

$$0 \leq 1 - \frac{(t_x - t_y) - (f_x - f_y)}{2} \leq 1 \quad (*)$$

$$0 \leq 1 - |\pi_A(x) - \pi_B(y)| \leq 1 \quad (**)$$

$$\text{故得 } 0 \leq [1 - |\pi_A(x) - \pi_B(y)|] \times (1 - \frac{|t_x - t_y| + |f_x - f_y|}{2}) \leq 1$$

$$\text{即 } 0 \leq \overline{M(x, y)} \leq 1$$

准则 4 证明 由 (*), (**) 与 $\overline{M(x, y)} = 1$

$$\Leftrightarrow 1 - |\pi_A(x) - \pi_B(y)| = 1 \text{ 和 } 1 -$$

$$\frac{|t_x - t_y| + |f_x - f_y|}{2} = 1$$

$$\Leftrightarrow \pi_A(x) - \pi_B(y) = 0 \text{ 且 } |t_x - t_y| + |f_x - f_y| =$$

0

$\Leftrightarrow x = y$ 证毕。

综上所述,公式满足 Vague 集之间的相似度量的一般准则,可见是 Vague 值相似度量的又一公式,且充分体现了未知信息,使得结果更加接近于实际情况。

例 1^[8] 设 $x = [0.3, 0.7]$ 和 $y = [0.4, 0.6]$ 分别是 Vague 集 A 和 B 中的两个 Vague 值,利用上面的方法,可得 $\overline{M(x, y)} = 0.72$ 。

3.2 新的 Vague 集之间的相似度量方法

假设 A, B 是论域 X 上的两个 Vague 集,其中 $V_A(x_i) = [t_A(x_i), 1 - f_A(x_i)]$ 表示 Vague 集 A 中 x_i 的隶属值, $V_B(x_i) = [t_B(x_i), 1 - f_B(x_i)]$ 表示 Vague 集 B 中 x_i 的隶属值,

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

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

则 Vague 集 A, B 之间的相似度可由如下函数 \bar{T} 进行计算:

$$\begin{aligned} \bar{T}(A, B) &= \frac{1}{n} \sum_{i=1}^n \overline{M(V_A(x_i), V_B(x_i))} \\ &= \frac{1}{n} \sum_{i=1}^n [1 - |\pi_A(x_i) - \pi_B(x_i)|] \times (1 - \frac{|t_A(x_i) - t_B(x_i)| + |f_A(x_i) - f_B(x_i)|}{2}) \end{aligned}$$

定理 2 Vague 集 A, B 之间的相似度量满足下面的性质:

$$(1) 0 \leq \bar{T}(A, B) \leq 1$$

$$(2) \bar{T}(A, B) = \bar{T}(B, A)$$

$$(3) \bar{T}(A, B) = 0 \Leftrightarrow A(x_i) = [1, 1], B(x_i) = [0, 0] \text{ 或 } A(x_i) = [0, 0], B(x_i) = [1, 1]$$

$$(4) \bar{T}(A, B) = 1 \Leftrightarrow t_A(x_i) = t_B(x_i), f_A(x_i) = f_B(x_i) \quad i = 1, 2, \dots, n$$

证明类似于定理 1 的证明,这里不再赘述。

例 2 设 A, B 是论域 $X = \{x_1, x_2, x_3, x_4, x_5\}$ 上的两个 Vague 集,其中

$$A = [0.2, 0.4] / x_1 + [0.3, 0.7] / x_2 + [0.5, 0.7] / x_3 + [0.7, 0.9] / x_4 + [0.8, 1.0] / x_5$$

$$B = [0.3, 0.5] / x_1 + [0.4, 0.6] / x_2 + [0.6, 0.8] / x_3 + [0.7, 0.9] / x_4 + [0.9, 1.0] / x_5$$

则可得 $\bar{T}(A, B) = 0.875$

3.3 新的 Vague 集之间的加权相似度量

假设 A, B 是论域 $X = \{x_1, x_2, \dots, x_n\}$ 上的两个 Vague 集, w_i 为 X 上元素 x_i 的权重, $x_i \in [0, 1], 0 \leq i \leq n$, 则 A, B 之间的加权相似度量可以用 $\bar{W}(A, B)$ 计算:

$$\begin{aligned} \bar{W}(A, B) &= \frac{1}{n} \sum_{i=1}^n w_i \overline{M(V_A(x_i), V_B(x_i))} / \sum_{i=1}^n w_i = \\ &= \frac{1}{n} \sum_{i=1}^n w_i [1 - |\pi_A(x_i) - \pi_B(x_i)|] \times (1 - \frac{|t_A(x_i) - t_B(x_i)| + |f_A(x_i) - f_B(x_i)|}{2}) / \sum_{i=1}^n w_i \end{aligned}$$

例 3 借用文献[9]中的一个例子说明两个 Vague 集之间新度量方法的合理性、有效性。设有三种定义在 $X = \{x_1, x_2, x_3\}$ 上的已知模式,它们由 Vague 集表示的特征如下:

$$A_1 = [0.1, 0.9] / x_1 + [0.5, 0.9] / x_2 + [0.1, 0.1] / x_3$$

$$A_2 = [0.5, 0.5] / x_1 + [0.7, 0.7] / x_2 + [0.1, 0.2] / x_3$$

$$A_3 = [0.7, 0.8] / x_1 + [0.1, 0.2] / x_2 + [0.4, 0.6] / x_3$$

B 是一个待识别的样本,其 Vague 集特征描述为:

$$B = [0.4, 0.6] / x_1 + [0.6, 0.8] / x_2 \times [0.0, 0.2] / x_3, \text{ 试问样本 B 到底该属于哪种模式?}$$

解:易得 $\bar{T}(B, A_1) = 0.573, \bar{T}(B, A_2) = 0.788, \bar{T}(B, A_3) = 0.627$, 根据度量最大原则, Vague 集 B 与 A_2 最相似,即样本 B 应该归属于 A_2 模式,与文献[9]中结果完全一样,相似度数据的大小合理且比较直观。

4 结束语

文中在现有相似度量方法的基础上,总结出了 Vague 集(值)之间相似度量的一般准则,并指出影响

(下转第 154 页)

(2) 调整界面的大小。

平台与插件应用程序中,一些界面中的交互文字经过翻译后会扩大很多。即在中文中的交互界面比较合适,而编译后交互文字已经不适应以前的中文界面。比如,中文语言“编辑”翻译成德文“Bearbeiten”,其德文文字长度增加了很多。因此需要调整界面大小,有时甚至要重新设计。

(3) 调整界面的格式。

日期格式、日历格式、字符串格式、货币和数字格式、字体、图片标识等,这些格式针对不同的国家与地区,肯定有所不同,所以在进行本地化时,要根据当地的风俗习惯进行相应的修改与调整。

4 程序实例

平台+插件模式的应用程序的实现方式有很多,例如基于 COM 组件开发的方式,基于 DLL 开发的方式等等。无论用什么样的语言与开发工具实现,基本思想都是一样的。文中在实验时,给出了一个在 VC++2005 编译器下用 DLL 实现的平台+插件的简单的应用程序。

在该程序中,采用了三个层次。

(1)最上层:总框架平台程序——EXE,单文档—视图程序。实现图像与图形的显示。

(2)中间层:界面接口模块——DLL 实现,包含图像图形显示视图的基类,与插件 Frame 的基类。

(3)最底层:各种功能的具体实现模块——DLL 实现,一些平级的基本的插件,实例中提供了两个基本的插件,一个是提供操作栅格数据功能,另一个提供矢量数据编辑的功能。

该程序实例中,总框架平台程序,界面接口模块,

栅格操作插件,矢量编辑插件,四个工程,全部实现了 Unicode 化,具备中文软件国际化的能力。而后,又利用纯资源 DLL,分别把这四个工程文件中的资源文件都分离到中文纯资源 DLL 与英文纯资源文件中。最终实现了中文软件的多国语言版本。

5 结束语

平台/插件模式已广泛应用在软件开发,而中文软件国际化是当前中国软件业应对经济全球化竞争的必然趋势。因此,在平台/插件的模式中文软件开发时,应该根据软件国际化的理论及规则进行软件开发,以适应国际市场,提高中文软件的国际竞争能力。

参考文献:

- [1] 吴亮. 基于插件技术的 GIS 应用程序框架的研究与实现[J]. 中国地质大学学报, 2006, 9(5): 609-614.
- [2] 马林, 黄文培. Struts 应用程序中文显示方案及其国际化技术[J]. 计算机技术与发展, 2007, 17(2): 31-34.
- [3] Yeo A. Software Internationalization and Localization[C]// 6th Australian Conference on Computer - Human Interaction (OZCHI'96). Washington, DC, USA: IEEE Computer Society, 1996.
- [4] Adams G. Internationalization and character set standards[J]. Standard View, 1993, 1(1): 20-22.
- [5] 张阳, 蒋胜平. 中文软件的国际化方法研究[J]. 电脑与信息技术, 2006, 12(6): 44-46.
- [6] Jiang Shengping. Globalization of Non-English Software[J]. International Journal of INFORMATION, 2006, 9(4): 36-41.
- [7] Belge M. The next step in software internationalization[J]. Interactions, 1995, 2(1): 54-62.
- [8] Set and Systems, 1995, 74(2): 213-217.
- [4] Hong D H, Kim C. A note on similarity measures between vague sets and between elements[J]. Information Sciences, 1999, 115(1): 83-96.
- [5] 李凡, 徐章艳. Vague 集之间的相似度量[J]. 软件学报, 2001, 12(6): 922-926.
- [6] 李艳红, 迟忠先, 严德勤. Vague 集相似度量与 Vague 熵[J]. 计算机科学, 2002, 29(12): 129-132.
- [7] 夏少云. Vague 集之间相似度量的分析与研究[J]. 北方交通大学学报, 2004, 28(1): 95-99.
- [8] 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(7): 92-95.
- [9] Li Zhi-zhen, S Peng - Fei. Similarity measure on intuitionistic fuzzy sets[J]. Pattern Recognition letters, 2003, 24(15): 2687-2693.
- [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 Trans SMC, 1993, 23(2): 610-614.
- [3] Chen S M. Similarity measures between Vaguess set[J]. Fuzzy

(上接第 150 页)

相似度量的三个方面的因素,在此基础上,充分考虑了未知信息在相似度量中的作用,解决了传统相似度量方法的中把未知信息看作肯定和否定各占一半的问题,提出了一种新的相似度量方法,并证明其满足公理化定义的所有条件,从而说明了它的正确性,并举出实例论证了它的有效性。

参考文献: