一类新的Vague集间的相似度量方法及其应用

孙义阳,辛小龙
SUN Yi-yang,XIN Xiao-long

西北大学 数学系,西安 710127
Department of Mathematics,Northwest University,Xi'an 710127,China
E-mail:xlxin@nwu.edu.cn

SUN Yi-yang,XIN Xiao-long.New method of similarity measure between Vague sets and its applications.Computer Engineering and Applications,2008,44(20):70-72.

Abstract: Similarity measures are important tools for studying and applying Vague sets.In this paper,we introduce the concept of Vague sets.By using examples,we show that some existing similarity measures are not always suitable for some cases,and the reasons are analyzed as well.Based on the axiom definition of similarity measure between Vague sets,we propose a new similarity measure.Finally,the similarity measure between Vague is applied to pattern recognition.
Key words: Fuzzy set;Vague set;similarity measure;pattern recognition

摘 要:相似度量是研究和应用Vague集的重要工具。首先介绍了Vague集的概念,接着用实例说明已有的Vague集相似度量在某些情况下并不适用,并分析了造成不适用的原因,进而给出了Vague集相似度量的公理化定义,提出了一种新的Vague集之间的相似度量,并将其应用于模式识别中。
关键词: Fuzzy集;Vague集;相似度量;模式识别
DOI:10.3778/j.issn.1002-8331.2008.20.021 **文章编号:**1002-8331(2008)20-0070-03 **文献标识码:**A **中图分类号:**TP301

1 引言

Gau和Buehrer[1]于1993年提出Vague集的概念,丰富了人们对事物刻画的工具,促进了模式识别、人工智能、近似推理等领域的发展。Vague集是Zadeh[2]的Fuzzy集的一种推广形式,它等同于Atanassov[3]1986年提出的直觉模糊集。在一个Vague集A中,是用一个真隶属函数$t_A(x)$和假隶属函数$f_A(x)$来描述其隶属度的界的。这两个界构成了$[0,1]$区间上的一个子区间$[t_A(x),1-f_A(x)]$,Vague集的特点是同时考虑隶属与非隶属两方面的信息,这使得Vague集在处理不确定信息时比传统的Fuzzy集有更强的表示能力,且更为灵活。Vague的相似度量已成为研究和应用Vague集的重要工具,对于知识表述、模式识别等人工智能的研究具有重要意义[4-6]。1995年Chen[7]提出了Vague集相似性度量的方法,1997年又进一步研究了Vague集的加权相似度量[8],1999年Hong等[9]学者在文中用例子说明Chen所提出的关于Vague集相似性度量在某些情况下并不合理,并提出了新的相似性度量方法。同时Vague集与直觉模糊集只是在表现形式上有些不同,在本质上是相同的。许多学者也在探讨直觉模糊集的相似度量方法,2000年Szmidt[10]等定义了直觉模糊集的距离测度,2002年Li等[11]提出了基于距离的相似度量方法。本文回顾了已有的Vague集之间的相似度量方法,指出已有的方法在某些情况下并不适用,提出了更为合理的Vague集之间的相似度量方法,并将其应用于模式识别中。

2 基本概念

为了方便以后说明,下面首先给出Vague集理论的相关定义。

定义1[2](模糊集) 论域$X=\{x\}$上的模糊集A定义为:

$$A=\{<x,\mu_A(x)>|x\in X\}$$

其中,$\mu_A:x\to[0,1]$为模糊集A的隶属函数,$\mu_A(x)$为元素$x\in X$在A中的隶属度。

定义2[1] 设X是一论域,X对任一元素x,X中的一个Vague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,$t_A(x)$是从支持x的证据所导出的隶属度下界,$f_A(x)$则是从反对x的证据所导出的否定隶属度下界,$t_A(x)$和$f_A(x)$将区间$[0,1]$中的一个实数与X中的一个点联系起来,即

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

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

且满足$t_A(x)+f_A(x)\leqslant1,0\leqslant t_A(x)\leqslant1,0\leqslant f_A(x)\leqslant1$。

设A为一个Vague集,当X离散时,将其表示为:

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

作者简介:孙义阳(1984-),男,硕士生,主要研究方向:模糊信息论;辛小龙(1955-),男,博士,教授,博士生导师,主要研究方向:代数学,信息论,密码学等。
收稿日期:2007-10-09 **修回日期:**2008-01-16

当 X 联系时，将其表示为：

$$A=\int\frac{[t_A(x),1-f_A(x)]}{x}\mathrm{d}x,x\in X$$

定义 3[12] 一个 Vague 集 A 的补 $\overline{A}$ 定义为：

$t_{\overline{A}}(x)=f_A(x),1-f_{\overline{A}}(x)=1-t_A(x)$

定义 4 两个 Vague 集 A 和 B 相等是指 $A=B$ 当且仅当 $t_A(x)=t_B(x),1-f_A(x)=1-f_B(x)$ 时。

定义 5[12] Vague 集 A 为 B 所包含，即 $A\subseteq B$，当且仅当 $t_A(x)\leqslant t_B(x),1-f_A(x)\leqslant 1-f_B(x)$ 时。

3 Vague 集间新的相似度量

这里首先回顾 Chen、Hong 和 Kim、Li 和 Cheng 等提出的一些 Vague 集之间的相似度量方法，然后，通过例子来说明这些相似度量方法在某些情况下是不合理的，最后给出一种新的 Vague 集之间的相似度量方法，并讨论其相关性质。

设 A、B 为论域 X 上的两个 Vague 集，其中

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

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

Chen[7,8]定义的 Vague 集 A 和 B 的相似度量为：

$$T_C(A,B)=1-\frac{1}{n}\sum_{i=1}^{n}\frac{|(t_A(x_i)-t_B(x_i))-(f_A(x_i)-f_B(x_i))|}{2}\tag{1}$$

Hong 和 Kim[9]用以下例子说明 Chen 所提出的相似度量在这种情况下是不合理的。

例 1 设 A、B、C 为论域 X 上的 3 个 Vague 集，其中

$$A=\sum_{i=1}^{n}\frac{[0,1-0]}{x_i},B=\sum_{i=1}^{n}\frac{[0.5,1-0.5]}{x_i},C=\sum_{i=1}^{n}\frac{[0.49,1-0.51]}{x_i}$$

直观上，我们认为 Vague 集 B 与 C 比 A 与 B 更相似，但根据公式(1)，有 A 与 B，B 与 C 的相似度量分别为：

$T_C(A,B)=1,T_C(B,C)=0.999$

也就是说 $T_C(A,B)>T_C(B,C)$，这显然是不合理的。

Li 和 Cheng[11]定义的 Vague 集和的相似度量为：

$$T_L(A,B)=1-\sqrt[p]{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{|(t_A(x_i)-t_B(x_i))-(f_A(x_i)-f_B(x_i))|}{2}\right]^p}\tag{2}$$

其中 $1\leqslant p<+\infty$，很容易验证 Li 和 Cheng 的相似度量与 Chen 的相似度量 T_C 存在相同的缺陷，也是不合理的。

为了克服上述缺陷，Hong 和 Kim[9]对(1)做了相应的修改，而定义了 Vague 集之间新的相似度量：

$$T_H(A,B)=1-\frac{1}{n}\sum_{i=1}^{n}\frac{|t_A(x_i)-t_B(x_i)|+|f_A(x_i)-f_B(x_i)|}{2}\tag{3}$$

下面将通过另一个例子来说明式(3)在某些情况下也是不合理的。

例 2 设 A、B_1、B_2、B_3 为论域 X 上的 4 个 Vague 集，其中

$$A=\sum_{i=1}^{n}\frac{[0,1-0]}{x_i}\qquad B_1=\sum_{i=1}^{n}\frac{[0.2,1-0.8]}{x_i}$$

$$B_2=\sum_{i=1}^{n}\frac{[0.3,1-0.7]}{x_i}\qquad B_3=\sum_{i=1}^{n}\frac{[0.5,1-0.5]}{x_i}$$

由式(3)得 $T_H(A,B_1)=T_H(A,B_2)=T_H(A,B_3)=0.5$。

事实上，对 $A=\sum_{i=1}^{n}\frac{[0,1-0]}{x_i}$ 和 $B_K=\sum_{i=1}^{n}\frac{[t_{B_K}(x_i),1-f_{B_K}(x_i)]}{x_i}$，只要 $t_{B_K}(x_i)+f_{B_K}(x_i)=1(\forall x_i\in X,k=1,2,3,\cdots)$，利用式(3)总能得到 $T_H(A,B_K)=0.5$，即 $T_H(A,B_1)=T_H(A,B_2)=\cdots=0.5$。

很明显在这种情况下我们将很难做出比较和识别，这就使得在实际应用中很难做出判断。为了克服存在的这些问题，提出了一种新的 Vague 集之间相似度量的表示方法：

$$T(A,B)=1-\frac{2}{3n}\sum_{i=1}^{n}\left[\frac{|t_A(x_i)-t_B(x_i)|}{2-\max\{t_A(x_i),t_B(x_i)\}}+\frac{|f_A(x_i)-f_B(x_i)|}{2-\min\{f_A(x_i),f_B(x_i)\}}\right]\tag{4}$$

下面用本文定义的 Vague 集之间相似度量的表示方法(4)来计算例 1 和例 2 的相似度量。

对于例 1，有

$$T(A,B)=1-\frac{2}{3n}\sum_{i=1}^{n}\left[\frac{|0-0.5|}{2-\max\{0,0.5\}}+\frac{|0-0.5|}{2-\min\{0,0.5\}}\right]=0.611$$

$$T(B,C)=1-\frac{2}{3n}\sum_{i=1}^{n}\left[\frac{|0.5-0.499|}{2-\max\{0.5,0.499\}}+\frac{|0.5-0.501|}{2-\min\{0.5,0.501\}}\right]=0.999$$

对于例 2，有

$$T(A,B_1)=1-\frac{2}{3n}\sum_{i=1}^{n}\left[\frac{|0-0.2|}{2-\max\{0,0.2\}}+\frac{|0-0.8|}{2-\min\{0,0.8\}}\right]=0.659$$

$$T(A,B_2)=1-\frac{2}{3n}\sum_{i=1}^{n}\left[\frac{|0-0.3|}{2-\max\{0,0.3\}}+\frac{|0-0.7|}{2-\min\{0,0.7\}}\right]=0.649$$

$$T(A,B_3)=1-\frac{2}{3n}\sum_{i=1}^{n}\left[\frac{|0-0.5|}{2-\max\{0,0.5\}}+\frac{|0-0.5|}{2-\min\{0,0.5\}}\right]=0.611$$

因此，本文所定义的 $T(A,B)$ 是相对比较合理的。

下面证明 $T(A,B)$ 确实是 Vague 集 A 和 B 之间的相似度量。首先给出 Vague 集之间相似度量的定义。

定义 6[1] 映射：$V(X)\times V(X)\rightarrow[0,1]$ 被称为 Vague 集 A 和 B 之间的相似度量，如果 $\forall A,B,C\in V(X)$，$T(A,B)$ 满足下面性质：

(**TP1**) $0\leqslant T(A,B)\leqslant 1$；

(**TP2**) $T(A,B)=1,A\Leftrightarrow B$；

(**TP3**) $T(A,B)=T(B,A)$；

(**TP4**) 若 $A\subseteq B\subseteq C$，则有 $T(A,B)\geqslant T(A,C)$ 且 $T(B,C)\geqslant T(A,C)$。

下面证明式(4)所定义的 $T(A,B)$ 是 Vague 集 A 和 B 之间的相似度量。

定理 1 $T(A,B)$ 为论域 $X=\{x_1,x_2,\cdots,x_n\}$ 上的 Vague 集 A 和 B 之间的相似度量。

证明 显然 $T(A,B)$ 满足定义 6 中的(TP2、TP3)，这里仅需证明 $T(A,B)$ 满足(TP1)和(TP4)。

首先来证明 $T(A,B)$ 满足(TP1)：

由于 $0\leqslant t_A(x_i),t_B(x_i)\leqslant 1,0\leqslant f_A(x_i),f_B(x_i)\leqslant 1$，且 $0\leqslant t_A(x_i)+t_B(x_i)\leqslant 1,0\leqslant f_A(x_i)+f_B(x_i)\leqslant 1$，由 MATLAB 求得

$$0\leqslant\frac{|t_A(x_i)-t_B(x_i)|}{2-\max\{t_A(x_i),t_B(x_i)\}}+\frac{|f_A(x_i)-f_B(x_i)|}{2-\min\{f_A(x_i),f_B(x_i)\}}\leqslant\frac{3}{2}$$

且当 $A=B$ 时取得最小值 0，当 $A=[0,0],B=[1,1]$ 或 $A=[1,1],B=[0,0]$ 时取得最大值 $\frac{3}{2}$。因此有

$$0\leqslant\frac{2}{3n}\sum_{i=1}^{n}\left[\frac{|t_A(x_i)-t_B(x_i)|}{2-\max\{t_A(x_i),t_B(x_i)\}}+\frac{|f_A(x_i)-f_B(x_i)|}{2-\min\{f_A(x_i),f_B(x_i)\}}\right]\leqslant 1$$

从而 $0\leqslant T(A,B)\leqslant 1$。

下证 $T(A,B)$ 满足(TP4)：

对任意的 Vague 集，若 $A\subseteq B\subseteq C$，则有

$$t_A(x_i)\leqslant t_B(x_i)\leqslant t_C(x_i),\ f_A(x_i)\geqslant f_B(x_i)\geqslant f_C(x_i)$$

于是

$$\frac{|t_A(x_i)-t_B(x_i)|}{2-\max\{t_A(x_i),t_B(x_i)\}}\leqslant\frac{|t_A(x_i)-t_C(x_i)|}{2-\max\{t_A(x_i),t_C(x_i)\}}$$

$$\frac{|f_A(x_i)-f_B(x_i)|}{2-\min\{f_A(x_i),f_B(x_i)\}}\leqslant\frac{|f_A(x_i)-f_C(x_i)|}{2-\min\{f_A(x_i),f_C(x_i)\}}$$

从而

$$T(A,B)=1-\frac{2}{3n}\sum_{i=1}^{n}\left[\frac{|t_A(x_i)-t_B(x_i)|}{2-\max\{t_A(x_i),t_B(x_i)\}}+\frac{|f_A(x_i)-f_B(x_i)|}{2-\min\{f_A(x_i),f_B(x_i)\}}\right]\geqslant$$

$$1-\frac{2}{3n}\sum_{i=1}^{n}\left[\frac{|t_A(x_i)-t_C(x_i)|}{2-\max\{t_A(x_i),t_C(x_i)\}}+\frac{|f_A(x_i)-f_C(x_i)|}{2-\min\{f_A(x_i),f_C(x_i)\}}\right]=$$

$$T(A,C)$$

即 $T(A,B)\geqslant T(A,C)$。

同理 $T(B,C)\geqslant T(A,C)$。

综上 $T(A,B)$ 满足(TP4)，从而 $T(A,B)$ 为 Vague 集 A 和 B 之间的相似度量。

很容易证明由式(4)所定义的 Vague 集 A 和 B 之间的相似度量具有以下性质.

性质 1 $T(A,B)=0\Leftrightarrow A=\varnothing$ 且 $B=\overline{A}$ 或 $B=\varnothing$ 且 $A=\overline{B}$。

性质 2 若 $T(A,B)$ 为 Vague 集 A 和 B 之间的相似度量，则

$$T'(A,B)=\frac{T(A,B)}{2-T(A,B)}$$

也是 Vague 集 A 和 B 之间的相似度量。

同样地，也给出 Vague 集之间的权重相似度量。

定义 7 设 A、B 为论域 X 上的两个 Vague 集，其中

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

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

设元素 x_i 在 X 中的权重为 $\omega_i(i=1,2,\cdots,n)$，其中 $\omega_i\in[0,1]$ 且 $\sum_{i=1}^{n}\omega_i=1$，定义 A 与 B 之间的权重相似度量为：

$$T_\omega(A,B)=$$

$$1-\frac{\sum_{i=1}^{n}\frac{2}{3\omega_i}\left[\frac{|t_A(x_i)-t_B(x_i)|}{2-\max\{t_A(x_i),t_B(x_i)\}}+\frac{|f_A(x_i)-f_B(x_i)|}{2-\min\{f_A(x_i),f_B(x_i)\}}\right]}{\sum_{i=1}^{n}\omega_i}$$

定理 2 $T_\omega(A,B)$ 为论域 X 上的 Vague 集 A 与 B 之间的权重相似度量。

证明 此定理的证明类似于定理 1 中的证明。

特别的，当 $\omega_i=\frac{1}{n}(i=1,2,\cdots,n)$ 时，$T_\omega(A,B)$ 即为 $T(A,B)$。

4 Vague 集之间相似度量在模式识别中的应用

模式识别是 Vague 集间相似度量的一个应用。设 $P_1,P_2,\cdots,P_n$ 是 n 个 Vague 集，它们分别代表着 n 个模式。今有一个待识别的样本 P，其特征也用 Vague 集描述，要确定 P 对应哪一个模式。对这个问题，首先分别计算 $T(P,P_i)(i=1,2,\cdots,n)$；其次选取 $J=\{i|\max(T(P,P_i))\}$，则 $i\in J$ 对应的 P_i 即为 P 所对应的模式。当然，J 中可能有多个元素，所以 P 可能有多个模式与之对应，如果必要还可以采用进一步的原则识别。

例 3 这里考虑建筑材料的分类问题。给定四种建筑材料样本，它们分别由特征空间 $X=\{x_1,x_2,\cdots,x_{12}\}$ 上的 Vague 集 A_1、A_2、A_3、A_4 表示(参见表 1)。现给一种未知的材料 B，判断这种建筑材料应属于样本中的哪种材料？

表 1 数据表

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
$t_{A_1}(x)$	0.173	0.102	0.530	0.965	0.420	0.008	0.331	1.000	0.215	0.432	0.750	0.432
$f_{A_1}(x)$	0.524	0.818	0.326	0.008	0.351	0.956	0.512	0	0.625	0.534	0.126	0.432
$t_{A_2}(x)$	0.510	0.627	1.000	0.125	0.026	0.732	0.556	0.650	1.000	0.145	0.047	0.760
$f_{A_2}(x)$	0.365	0.125	0	0.648	0.823	0.153	0.303	0.267	0	0.762	0.923	0.231
$t_{A_3}(x)$	0.495	0.603	0.987	0.073	0.037	0.690	0.147	0.213	0.501	1.000	0.324	0.045
$f_{A_3}(x)$	0.387	0.298	0.006	0.849	0.923	0.268	0.812	0.653	0.284	0	0.483	0.912
$t_{A_4}(x)$	1.000	1.000	0.857	0.734	0.021	0.076	0.152	0.113	0.489	1.000	0.386	0.028
$f_{A_4}(x)$	0	0	0.123	0.158	0.896	0.912	0.712	0.756	0.389	0	0.485	0.912
$t_B(x)$	0.978	0.980	0.798	0.693	0.051	0.123	0.152	0.113	0.494	0.987	0.376	0.012
$f_B(x)$	0.003	0.012	0.132	0.213	0.876	0.756	0.721	0.732	0.368	0.000	0.423	0.897

从表 1 中的数据，可以得到：

$$T(A_1,B)=1-\frac{2}{3\times 12}\sum_{i=1}^{12}\left[\frac{|t_{A_1}(x_i)-t_B(x_i)|}{2-\max\{t_{A_1}(x_i),t_B(x_i)\}}+\frac{|f_{A_1}(x_i)-f_B(x_i)|}{2-\min\{f_{A_1}(x_i),f_B(x_i)\}}\right]$$

求得 $T(A_1,B)=0.586$，类似地可以得到下面的结果：

$$T(A_2,B)=0.587,T(A_3,B)=0.812,T(A_4,B)=0.964$$

从上面的结果可以看出，$J=\{i|\max(T(A_i,B))\}=\{4\}$，即 A_4 与 B 之间的差异是最小的，因而可以认为建筑材料 B 应与 A_4 一类。

5 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，本文提出的相似度量克服了 Chen 等相似度量的一些不足，该方法对于 Vague 集及其相似度量的研究具有一定意义。

参考文献：

[1] Gau Wen-Lung, Buehrer D J.Vague sets[J].IEEE Transactions on Systems, Man and cybernetics, 1993, 23(2): 610-614.

[2] Zadeh L A.Fuzzy sets[J].Information and Control, 1965, 8(3): 338-353.

[3] Bustince H, Burillo P.Vague sets are intuitionistic fuzzy sets[J]. Fuzzy Sets and Systems, 1996, 79: 403-405.

[4] Chen Shyi-Ming, Tan Jiantr Mean.Handling multicriteria fuzzy decision making problems based on vague set theory[J].Fuzzy Sets and Systems, 1994, 67(2): 163-172.

[5] Hong D H, Choi C H.Multicriteria fuzzy decision-making problems based on vague set theory[J].Fuzzy Sets and Systems, 2000, 114(1): 103-113.

[6] 冯林，罗芬，刘照鹏，等.Vague 集之间的相似度量及应用[J].计算机工程与应用，2006，42(21)：165-168.

（下转 88 页）