文章编号:1001 - 9081(2008)06 - 1595 - 03

基于 (t_x, f_x) 扩展的 Vague 集之间的相似度量及其应用

符晓芳[1,4],张福金[2],王鸿绪[1,3]

(1. 琼州学院 数学系,海南 五指山 572200; 2. 琼州学院 物理系,海南 五指山 572200;
3. 琼州学院 计算机科学与技术系,海南 五指山 572200; 4. 海南大学 信息科学技术学院,海口 570228)
(fuxiaofang2008@126.com)

摘　要:基于 Vague 值的三维表示,(t_x, f_x) 扩展和模糊集运算,给出 Vague 集间的相似度量的三个系列公式。提出 Vague 集间的相似度量在网络信息过滤问题中的应用思路及例子。此例表明这些新公式是实用的。

关键词:Vague 集;三维表示;(t_x, f_x) 扩展;模糊集运算;相似度量

中图分类号: TP18　　**文献标志码**:A

Similarity measures between Vague sets based on (t_x, f_x) expansion and its application

FU Xiao-fang[1,4], ZHANG Fu-jin[2], WANG Hong-xu[1,3]

(1. *Department of Mathematics, Qiongzhou University, Wuzhishan Hainan 572200, China*;
2. *Department of Physics, Qiongzhou University, Wuzhishan Hainan 572200, China*;
3. *Department of Computer Science and Technology, Qiongzhou University, Wuzhishan Hainan 572200, China*;
4. *College of Information Science and Technology, Hainan University, Haikou Hainan 570228, China*)

Abstract: Three series of similarity measures formulas between Vague sets based on three-dimensional representation and (t_x, f_x) expansion and fuzzy set operations were presented. The thoughts and an example of applying the similarity measures between Vague sets to the network information filtered problem were given. This example shows that the new formulas are practicable.

Key words: Vague sets; three-dimensional representation; (t_x, f_x) expand; fuzzy set operations; similarity measures

0　引言

Fuzzy Sets[1]用精确的数学语言来刻画模糊概念,用多值逻辑的思想巧妙地处理模糊信息,已经在包括模糊控制、数据库、模式识别、人工智能、决策分析在内的众多领域得到广泛应用,方法日渐成熟。

Vague 集理论[2]在隶属度上比模糊集理论的一维表示 $(\mu_A(x))$ 更为详尽,采用二维 $([t_A(x), 1-f_A(x)])$ 表示。但是为何不直接用三维表示 $((t_A(x), f_A(x), \pi_A(x)))$ 呢?三维表示可将模糊信息全部直接地呈现在人们的面前。鉴于 Vague 集理论已经在包括近似推理、模式识别、决策分析及其他智能系统在内的领域中得到应用,并且在应用时相似度量成为最主要的工具之一,本文将提出三个系列的 Vague 集间的相似度量公式。

1　Vague 集的概念

设 A 为论域 X 上的一个 Vague 集,可用一个真隶属函数 t_A 和一个假隶属函数 f_A 刻画:$t_A: X\to[0,1]$,$f_A: X\to[0,1]$,其中 $t_A(x)$ 是从支持 x 的证据所导出的肯定隶属度的下界,$f_A(x)$ 是从反对 x 的证据所导出的否定隶属度的下界。x 对 A 的 Vague 值可记为 $A(x) = [t_A(x), 1-f_A(x)]$(简记为 $x = [t_x, 1-f_x]$),且满足约束条件:$t_A(x)+f_A(x)\leqslant 1$。规定 $\pi_A(x) = 1-f_A(x)-t_A(x)$(简记为 $\pi_x = 1-f_x-t_x$)。则 Vague 值 $A(x)$ 也可用三维表示为 $A(x) = (t_A(x), f_A(x), \pi_A(x))$ 或者 $x = (t_x, f_x, \pi_x)$,其中 t_x、f_x、π_x 分别称为 x 对 A 的赞成度、反对度、不确定度。

对于有限论域 $X = \{x_1, x_2, \cdots, x_m\}$ 上的 Vague 集 A,其二维表示为:

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

或:

$$A = \{[t_A(x_1), 1-f_A(x_1)], [t_A(x_2), 1-f_A(x_2)], \cdots, [t_A(x_m), 1-f_A(x_m)]\}$$

三维表示为:

$$A = \sum_{i=1}^{m}(t_A(x_i), f_A(x_i), \pi_A(x_i))/x_i;\ x_i\in X$$

或:

$$A = \{(t_A(x_1), f_A(x_1), \pi_A(x_1)), (t_A(x_2), f_A(x_2), \pi_A(x_2)), \cdots, (t_A(x_m), f_A(x_m), \pi_A(x_m))\}$$

当 X 为连续时,其上 Vague 集 A 的二维表示为:

$$A = \int_X [t_A(x), 1-f_A(x)]/x;\ x\in X$$

三维表示为:

$$A = \int_X (t_A(x), f_A(x), \pi_A(x))/x;\ x\in X$$

或:

$$A = \int_X (t_x, f_x, \pi_x)/x;\ x\in X$$

收稿日期:2007 - 12 - 20;**修回日期**:2008 - 02 - 11。

基金项目:海南省自然科学基金资助项目(80473);海南省教育厅资助项目(HJKJ200732;HJKJ2008-46)。

作者简介:符晓芳(1973 -),女,海南昌江人,讲师,硕士研究生,主要研究方向:信息处理、数据挖掘;　张福金(1956 -),男,山东济宁人,副教授,主要研究方向:电工技术、制冷设备、工业控制;　王鸿绪(1946 -),男,辽宁本溪人,教授,主要研究方向:模糊控制和模糊信息处理。

2 Vague 值的扩展

定义1 对于论域 X 上的三维表示的 Vague 值为 $x=(t_x,f_x,\pi_x)$，定义：

$t_x^{(0)}=t_x,\ f_x^{(0)}=f_x,\pi_x^{(0)}=\pi_x;t_x^{(n+1)}=t_x^{(n)}+\pi_x^{(n)}t_x,$
$f_x^{(n+1)}=f_x^{(n)}+\pi_x^{(n)}f_x,\pi_x^{(n+1)}=\pi_x^{(n)}\pi_x(n=0,1,2,3,\cdots)$ (1)

定理1 对 $n=0,1,2,3,\cdots$，有：

1) $[t_x^{(n)},1-f_x^{(n)}]$ 是论域 X 上的 Vague 值；

2) $x^{(n)}$ 的三维表示为 $x^{(n)}=(t_x^{(n)},f_x^{(n)},\pi_x^{(n)})$。

说明：记 $x^{(n)}=[t_x^{(n)},1-f_x^{(n)}]$，其三维表示为 $x^{(n)}=(t_x^{(n)},f_x^{(n)},\pi_x^{(n)})$，称为Vague值$x$的$(t_x,f_x)$扩展的$n$次Vague值。用投票模型可解释为：当第一轮投票没有胜出者时，需要进行第二轮投票。在第一轮投弃权票的人中，可能在第二轮投票时仍按与第一轮相同的赞成与反对的比例进行投票。如此进行下去便可得到 $x^{(n)}$。这种扩展的思路来自于文献[3]，在文献[3]中定义为：

$$t_x^{(n)}=t_x+t_x(1+\pi_x+\pi_x^2+\cdots+\pi_x^n)$$
$$f_x^{(n)}=f_x+f_x(1+\pi_x+\pi_x^2+\cdots+\pi_x^n)$$
$$\pi_x^{(n)}=\pi_x^{n+1}$$

与本文的(t_x,f_x)扩展式(1)相比较可见，式(1)这种递推形式的定义能更直观地反映出三维表示的思想。$t_x^{(n+1)}$、$f_x^{(n+1)}$、$\pi_x^{(n+1)}$ 分别由 $t_x^{(n)}$、$f_x^{(n)}$、$\pi_x^{(n)}$ 递推而来的。还应指出，式(1)也是对数据 $x=(t_x,f_x,\pi_x)$ 的一种“挖掘”，挖掘的过程即是式(1)，而$x^{(n)}$的二维表示组成的闭区间有如下的包含关系：

$$x^{(+\infty)}\subseteq\cdots\subseteq x^{(n+1)}\subseteq x^{(n)}\subseteq\cdots\subseteq x^{(1)}\subseteq x^{(0)}\subseteq[0,1]$$

其中 $x^{(+\infty)}=[t_x^{(+\infty)},1-f_x^{(+\infty)}]=[t_x/(t_x+f_x),1-f_x/(t_x+f_x)]$。挖掘的结果为 $x^{(+\infty)}$，它是 Vague 值 x 转化成的模糊值。

3 Vague 集间的相似度量

定义2 设有 Vague 集(三维表示)为：

$A=\sum_{i=1}^{m}(t_A(x_i),\ f_A(x_i),\pi_A(x_i))/x_i=\sum_{i=1}^{m}(t_{x_i},\ f_{x_i},\pi_{x_i})/x_i;\ x_i\in X$

$B=\sum_{i=1}^{m}(t_B(x_i),\ f_B(x_i),\pi_B(x_i))/x_i=\sum_{i=1}^{m}(t_{y_i},\ f_{y_i},\pi_{y_i})/x_i;\ x_i\in X$

它们是论域$X=\{x_1,x_2,\cdots,x_m\}$上的Vague集。对$n=0,1,2,\cdots$，定义A与B之间的函数为：

$$R_{k,n}(A,B)=\frac{1}{m}\sum_{i=1}^{m}R_{k,n}(A(x_i),B(x_i));\ k=1,2,3$$

其中：

$$R_{1,n}(x_i,y_i)=(t_{xi}^{(n)}\wedge t_{y_i}^{(n)})+(f_{x_i}^{(n)}\wedge f_{y_i}^{(n)})+(\pi_{x_i}^{(n)}\wedge\pi_{y_i}^{(n)})\qquad(2)$$

$$R_{2,n}(x_i,y_i)=\frac{(t_{x_i}^{(n)}\wedge t_{y_i}^{(n)})+(f_{x_i}^{(n)}\wedge f_{y_i}^{(n)})}{(t_{x_i}^{(n)}\vee t_{y_i}^{(n)})+(f_{x_i}^{(n)}\vee f_{y_i}^{(n)})}\qquad(3)$$

$$R_{3,n}(x_i,y_i)=\frac{(t_{xi}^{(n)}\wedge t_{y_i}^{(n)})+(f_{xi}^{(n)}\wedge f_{y_i}^{(n)})+(\pi_{xi}^{(n)}\wedge\pi_{y_i}^{(n)})}{(t_{xi}^{(n)}\vee t_{y_i}^{(n)})+(f_{xi}^{(n)}\vee f_{y_i}^{(n)})+(\pi_{xi}^{(n)}\vee\pi_{y_i}^{(n)})}\qquad(4)$$

其中，$\wedge$ 和 $\vee$ 是模糊集中的取大、取小运算。即对 $a,b\in[0,1]$，则定义 $a\vee b=\max\{a,b\}$，$a\wedge b=\min\{a,b\}$。

若元素 $x_i(i=1,2,\cdots,m)$ 的权重为 $\omega_i(\in(0,1])(n=0,1,2,\cdots)$，还可定义加权函数：

$$RW_{k,n}(A,B)=\Big[\sum_{i=1}^{m}\omega_iR_{k,n}(A(x_i),B(x_i))\Big]\Big/\Big[\sum_{i=1}^{m}\omega_i\Big];$$
$k=1,2,3$

定理2 对于 $i=1,2,\cdots,m;n=0,1,2,\cdots;k=1,2,3$，则：

1) $R_{k,n}(x_i,y_i)\in[0,1]$；

2) $R_{k,n}(x_i,y_i)=R_{k,n}(y_i,x_i)$；

3) 当 $x_i=y_i$ 时，$R_{k,n}(x_i,y_i)=1$；

4) 当 $x_i=(0,1,0)$，$y_i=(1,0,0)$ 或者 $x_i=(1,0,0)$，$y_i=(0,1,0)(1\leqslant i\leqslant m)$ 时，$R_{k,n}(x_i,y_i)=0$。

证明 仅证明 $R_{1,n}(x_i,y_i)$ 满足这四条性质，其他结论类似可证。

1) 因$0\leqslant(t_{x_i}^{(n)}\wedge t_{y_i}^{(n)})+(f_{xi}^{(n)}\wedge f_{y_i}^{(n)})+(\pi_{xi}^{(n)}\wedge\pi_{y_i}^{(n)})\leqslant t_{x_i}^{(n)}+f_{xi}^{(n)}+\pi_{xi}^{(n)}=1$，则 $R_{1,n}(x_i,y_i)\in[0,1]$。

2) 显然有 $R_{1,n}(x_i,y_i)=R_{1,n}(y_i,x_i)$ 成立。

3) 当 $x_i=y_i$ 时，$t_{x_i}^{(n)}=t_{y_i}^{(n)}$，$f_{xi}^{(n)}=f_{y_i}^{(n)}$，$\pi_{xi}^{(n)}=\pi_{y_i}^{(n)}$，所以 $R_{1,n}(x_i,y_i)=(t_{xi}^{(n)}\wedge t_{y_i}^{(n)})+(f_{xi}^{(n)}\wedge f_{y_i}^{(n)})+(\pi_{xi}^{(n)}\wedge\pi_{y_i}^{(n)})=t_{x_i}^{<n>}+f_{x_i}^{<n>}+\pi_{x_i}^{<n>}=1$。

4) 当 $x_i=(0,1,0)$，$y_i=(1,0,0)(1\leqslant i\leqslant m)$ 时，则：$R_{1,n}(x_i,y_i)=(0\wedge1)+(1\wedge0)+(0\wedge0)=0+0+0=0$，同理可证 $y_i=(0,1,0)(1\leqslant i\leqslant m)$ 时，$R_{1,n}(x_i,y_i)=0$。

定义3 称下列函数为Vague集A与B间的相似度量，如果 $R(A,B)$ 满足下列公理：

$$M(A,B)=\begin{cases}\text{任意}, & A=B=\{(0,0,1),(0,0,1),\cdots,(0,0,1)\}\\ R(A,B), & \text{其他}\end{cases}$$

公理1 有界性，$R(A,B)\in[0,1]$；

公理2 对称性，$R(A,B)=R(B,A)$；

公理3 边界条件1。当$A=B$且$\neq\{(0,0,1),(0,0,1),\cdots,(0,0,1)\}$时，有 $R(A,B)=1$；

公理4 边界条件2。当$A(x_i)=(0,1,0)$，$B(x_i)=(1,0,0)$ 或者 $A(x_i)=(1,0,0)$，$B(x_i)=(0,1,0)(i=1,2,\cdots,m)$ 时，$R(A,B)=0$。

则称函数 $R(A,B)$ 为 Vague 集 A 和 B 之间的相似度量的伴随函数。

定理3 当$n=1,2,\cdots$时，下列函数是Vague集A与B之间的相似度量和加权相似度量：

$$M_{k,n}(A,B)=\begin{cases}\text{任意}, & A=B=\{(0,0,1),(0,0,1),\cdots,(0,0,1)\}\\ R_{k,n}(A,B), & \text{其他}\end{cases};$$
$k=1,2,3$

$$W_{k,n}(A,B)=\begin{cases}\text{任意}, & A=B=\{(0,0,1),(0,0,1),\cdots,(0,0,1)\}\\ RW_{k,n}(A,B), & \text{其他}\end{cases};$$
$k=1,2,3$

而函数 $R_{1,n}(A,B)$、$R_{2,n}(A,B)$、$R_{3,n}(A,B)$ 是 Vague 集 A 和 B 之间的相似度量的伴随函数，$RW_{1,n}(A,B)$、$RW_{2,n}(A,B)$、$RW_{3,n}(A,B)$ 是 Vague 集 A 和 B 间的加权相似度量的伴随函数。

式(2)、(3)、(4)分别给出三个系列各无穷多个伴随函数，从而得到相应的相似度量公式。需要说明：1)应用取大取小运算是本文公式的特点之一；2)紧密依赖于 Vague 值的三维表示也是本文公式的特点之一；3)定义3采用分段函数来定义 Vague 集间的相似度量，原因详见文献[4]，这里简单说明如下。早在文献[2]中就有论述，用$1-t_x-f_x$表示不确定特征知识的精确度，如果它很小，我们的关于x的知识是相对精确的；如果它很大，我们知道相应的很少。那么当$1-t_x-f_x$达到最大值1时，人们对x的知识知道得是最少的，此时

$t_x=f_x=0$。也就是 Vague 值 $x=[0,1]$(三维表示为 $x=(0,0,1)$) 时,我们对 x 的知识知道得是最少的。尽管当 $x=[0,1]$, $y=[0,1]$ 时,有 $x=y$,也不能定义此时 x 与 y 之间的相似度量为 1。因为对两个完全不了解的 Vague 值,谈不上它们是最相似的。此时,合乎情理的作法是定义它们的相似度量的值为"不确定" 或者"任意"。所以我们沿用文献[4] 的方法,用分段函数来定义 Vague 集之间的相似度量;4) 文献[5] 中的函数 $M_Q(x,y)$ 是本文伴随函数 $R_{3,n}(A,B)$ 当取 $n=0, m=1$(即 Vague 集 A 和 B 分别退化为 Vague 值 x 和 y) 时的特例 $R_{3,0}(A,B)$;5) 表1是本文公式(皆取 $n=2$) 与文献[6 -9] 所给公式:

$$M_C(A,B)=1-\frac{1}{m}\sum_{i=1}^{m}\left|(t_{x_i}-f_{x_i})-(t_{y_i}-f_{y_i})\right|/2$$

$$M_H(A,B)=1-\frac{1}{m}\sum_{i=1}^{m}[\,|t_{x_i}-t_{y_i}|+|f_{x_i}-f_{y_i}|\,]/2$$

$$M_L(A,B)=1-\frac{1}{m}\sum_{i=1}^{m}\{[\,|(t_{x_I}-f_{x_i})-(t_{y_I}-f_{y_i})|/4]+[\,|t_{x_i}-t_{y_i}|+|f_{x_i}-f_{y_i}|\,]/4\}$$

$$M_O(A,B)=1-\frac{1}{m}\sum_{i=1}^{m}\{[\,|t_{x_i}-t_{y_i}|2+|f_{x_i}-f_{y_i}|^2]/2\}^{1/2}$$

用文献[10]例 1 中的部分 Vague 集 A 和 B 进行 Vague 集间计算结果的比较。由表 1 明显可见本文公式皆是分辨力较高的。

$A=\{[0.4,0.8],[0.4,0.8],[0.4,0.8],[0.4,0.8]\}$

$B=\{[0.3,0.8],[0.4,0.7],[0.4,0.9],[0.5,0.8]\}$

表 1　Vague 值间相似度量值比较

相似度量公式	Vague 值			
	[0.4,0.8] [0.3,0.8]	[0.4,0.8] [0.4,0.7]	[0.4,0.8] [0.4,0.9]	[0.4,0.8] [0.5,0.8]
M_C	0.950	0.950	0.950	0.950
M_H	0.950	0.950	0.950	0.950
M_L	0.950	0.950	0.950	0.950
M_O	0.930	0.930	0.930	0.930
$M_{1,2}$	0.901	0.895	0.863	0.929
$M_{2,2}$	0.859	0.834	0.790	0.896
$M_{3,2}$	0.820	0.810	0.759	0.867

4　相似度量在网络信息过滤中的应用

文献[11]提出基于模糊语义关联度的正文检索方法。因为 Vague 集特别是三维表示较模糊集能更直接全面地描述模糊信息,那么用 Vague 集对网络信息的描述将会更加方便和符合实际。因此可以把该检索方法推广为 Vague 集相似度量的网络过滤检索法,步骤如下。

1)编制不良信息库,设为 $B_1, B_2, \cdots, B_k$。存放在不良信息库中的正文,按文献[11]提出的方法抽取 Vague 集关键字集,建立 Vague 集关键字集与原文间的指针联系。目的是从 Vague 集关键字集能沿指针方便地找到它的相应的原正文。

2)用户查询正文 T, 以确定它是否是不良信息。同样按文献[11]的方法抽取正文 T 的 Vague 集关键字集 A,用 A 进行检索。即计算 A 与库中 B_i 之间的相似度量 $M(A,B_i)$。若对事先确定的阈值 $\lambda(\in(0,1])$,有 $M(A,B_i)>\lambda$,则 A 所代表的正文 T 属于 B_i 所代表的正文 T_i 类不良信息。若对所有 $1\le i\le k$, $M(A,B_i)\le\lambda$,则 A 所代表的正文不是不良信息。

例 1:设用 Vague 集的三维形式表示的不良信息正文 T_1、T_2、T_3 的关键字集库为 B_1、B_2、B_3 和待检索的正文 T 的关键字集 A 的数据如下:

$B_1=\{(1.0,0.0,0.0),(0.7,0.0,0.3),(0.6,0.2,0.2)\}$

$B_2=\{(0.7,0.2,0.1),(1.0,0.0,0.0),(0.9,0.1,0.0)\}$

$B_3=\{(0.7,0.0,0.3),(0.8,0.1,0.1),(1.0,0.0,0.0)\}$

$A=\{(0.6,0.2,0.2),(0.7,0.2,0.1),(0.6,0.1,0.3)\}$

表 2　检索计算表

相似度量公式	i		
	1	2	3
$M_{1,3}(A,B_i)$	0.7955	0.8806	0.8427
$M_{2,3}(A,B_i)$	0.6813	0.7973	0.7343
$M_{3,3}(A,B_i)$	0.6631	0.7974	0.7535

由表2可见,若取 $\lambda=0.9$,则 A 所代表的正文 T 不是不良信息。若取 $\lambda=0.85$,由 $M_{1,3}(A,B_2)$ 的数据显示,A 所代表的正文 T 是 B_2 所代表的正文 T_2 类不良信息。但其他数据显示仍然不是不良信息。

由此例可见,阈值 λ 的选取直接决定了检索结果。λ 的选取与采用哪个系列的相似度量公式、参数 n 的具体取值、不良信息库中的 $B_1, B_2, \cdots, B_k$ 及待检索的 A 都有关系,须经大量的实验归纳得到。具体阈值 λ 的选取原则是要使实验数据的错判率在允许的误差(例如错判率低于百分之一) 范围内。

5　结语

Vague 值用三维表示可更直接地体现出 Vague 集理论的思想。本文应用 Vague 值的三维表示,把 Vague 值 $A(x)$ 进行 (t_x, f_x) 扩展,并在此基础上,应用模糊集中最基本的取大取小运算,给出 Vague 集间三个系列相似度量公式,文献[5]给出的公式之一是本文公式的特例。这些公式为 Vague 集在工程和理论研究中更广泛的应用提供更多的相似度量公式的选择余地。和其他相似度量公式一样,虽然我们采用分段函数的定义,也并不能证明这些公式满足所有的反例。

参考文献:

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

[2] GAU W L, BUEHRER D J. Vague sets [J]. IEEE Transactions on Systems, Man and Cybernetics, 1993, 23(2): 610 -614.

[3] 刘华文, 王凤英. Vague 集的转化与相似度量[J]. 计算机工程与应用, 2004, 40(32): 79 -81, 84.

[4] 张福金, 王鸿绪. 再论 Vague 集间的相似度量[J]. 计算机科学, 2006, 33(5): 197 -199.

[5] 邱卫根. Vague 集及其相似度量[J]. 计算机科学, 2007, 34(1): 156 -158.

[6] CHEN S M. Similarity measures between Vague sets and between elements [J]. IEEE Transactions on System, Man and Cybernetics, 1997, 27(1): 153 -158.

[7] HONG D H, KIM C.　A note on similarity measures between Vague sets and between elements [J]. Information Science, 1999(115): 83 -96.

[8] 李凡, 徐章艳. Vague 集间的相似度量[J]. 软件学报, 2001, 12(6): 922 -927.

[9] 李艳红, 迟忠先, 阎德勤. Vague 集间的相似度量与 Vague 熵[J]. 计算机科学, 2002, 29(12): 129 -132.

[10] 刘华文. 模糊模式识别的基础——相似度量[J]. 模式识别与人工智能, 2004, 17(2): 141 -145.

[11] 何新贵. 模糊集间的语义关联度及其应用[J]. 软件学报, 1994, 5(6): 19 -24.