

Vague 集(值)间的距离测度^{*})

卢正鼎 王天江

(华中科技大学计算机学院 武汉430074)

Distance Measures between Vague Sets (VALUES)

LU Zheng-Ding WANG Tian-Jiang

(Computer College Huazhong University of Science and Technology, Wuhan 430074)

Abstract At present in the area of intelligence it is more and more widely and deeply to research vague set. For measuring vague set, similarity measures between vague sets have been proposed in [1]. But in many cases we need the measuring method with distance characteristics, so we propose a series of distance measures for vague sets (values) and elements of vague sets. And we prove all characteristics of vague sets proposed by us, and then an example is also presented to illustrate the application of the proposed distance measures in handling behavior analysis problems. The method proposed in the example can provide a useful way in handling the behavior analysis problems. It provides a useful tool for intelligence system.

Keywords Vague set, Distance measures, Distance, Fuzzy set, Similarity measure

1 引言

自从 Zadeh 提出 fuzzy 集理论^[2]以来, fuzzy 集理论就在不断地发展和完善,并在许多领域里得到了成功的应用。但在实际应用中我们也发现它存在一些问题,例如, fuzzy 集的隶属度是一个单值,该单值既包含了支持 $u \in U$ (U 为论域)的证据,也包含了反对 $u \in U$ 的证据,它不可能表示其中的一个,更不可能同时表示支持和反对的证据。为此, Gau 等在文 [3] 中提出了 vague 集的概念,用一个真隶属函数 t_A 和一个假隶属函数 f_A 来描述 vague 集,真假隶属函数形成了隶属函数 $\mu_A(u)$ 的界,即 $t_A(u) \leq \mu_A(u) \leq 1 - f_A(u)$ 。vague 集的主要特征是它同时给出了支持和反对的证据。例如 $[t_A(u), 1 - f_A(u)] = [0.5, 0.8]$, 则有 $t_A(u) = 0.5, 1 - f_A(u) = 0.8, f_A(u) = 0.2$ 。在投票模型中这可解释为在 10 人中,有 5 人赞成,2 人反对,3 人弃权。目前在智能领域中 vague 集得到了广泛深入的研究与应用。为了度量 vague 集,有学者给出了 vague 集上的相似度量。但许多时候却需要具有距离性质的度量方法。为此,本文提出了一套 vague 集(值)间以及 vague 集中元素间的距离测度,并证明了距离测度的各种性质,然后还给出了一个实例说明所给出的距离测度在处理行为分析问题中的应用。

2 Vague 集(值)

我们对 vague 集(值)进行一个简要的分析。

令 U 是一个点(对象)的空间,其中的任意一个元素用 u 表示, U 中的一个 vague 集 A 用一个真隶属函数 t_A 和一个假隶属函数 f_A 表示, $t_A(u)$ 是从支持 u 的证据所导出的 u 的隶属度下界, $f_A(u)$ 则是从反对 u 的证据所导出的 u 的否定隶属度下界, $t_A(u)$ 和 $f_A(u)$ 将区间 $[0, 1]$ 中的一个实数与 U 中的每一个点联系起来,即

$$t_A: U \rightarrow [0, 1], f_A: U \rightarrow [0, 1]$$

当 U 是连续的时候, vague 集 A 可表示为:

$$A = \int_U [t_A(u), 1 - f_A(u)]/u \quad u \in U$$

当 U 为离散的时候, vague 集 A 可表示为:

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

其中, $t_A(u) + f_A(u) \leq 1$

定义1 假定 $0 \leq t_x \leq 1 - f_x \leq 1$, 则称 $x = [t_x, 1 - f_x]$ 是一个 vague 值。vague 值 x 可被分成三部分:真隶属度部分(即 t_x),假隶属度部分(即 f_x),和未知部分(即 $1 - t_x - f_x$)。

定义2 假定 $x = [t_x, 1 - f_x]$ 是 vague 值,若 $t_x = 1$ 且 $f_x = 0$ (即 $x = [1, 1]$),则 x 被称为单位 vague 值。

定义3 假定 $x = [t_x, 1 - f_x]$ 是 vague 值,若 $t_x = 0$ 且 $f_x = 1$ (即 $x = [0, 0]$),则 x 被称为零 vague 值。

定义4 假定 $x = [t_x, 1 - f_x], y = [t_y, 1 - f_y]$ 是两个 vague 值,若 $t_x = t_y$ 且 $f_x = f_y$,则称 x 和 y 相等。

定义5 假定 $x = [t_x, 1 - f_x], y = [t_y, 1 - f_y]$ 是两个 vague 值,若 $t_y = f_x$ 且 $f_y = t_x$,则称 y 是 x 的补。

定义6 假定 A 和 B 是论域 U 上的两个 vague 集, $U = \{u_1, u_2, \dots, u_n\}$, 其中 A 和 B 分别表示为:

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

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

若 $\forall u_i \in U, [t_A(u_i), 1 - f_A(u_i)] = [t_B(u_i), 1 - f_B(u_i)]$, 则称 vague 集 A 和 B 相等。

定义7 假定 A 是论域 U 上的 vague 集, $U = \{u_1, u_2, \dots, u_n\}$, 其中 A 表示为:

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

若 $\forall u_i \in U, t_A(u_i) = 1$, 且 $f_A(u_i) = 0$, 则称 vague 集 A 为单位

^{*}) 本课题得到国家高性能计算基金(00303)和华中科技大学科学研究基金(M99015)项目资助。卢正鼎 教授,博士生导师,主要研究领域为计算机辅助软件工程,智能信息系统。王天江 博士研究生,主要研究领域为人工生命,面向信息网络的智能应用,模糊推理,遗传算法。

vague 集。

定义8 假定 A 是论域 U 上的 vague 集, $U = \{u_1, u_2, \dots, u_n\}$, 其中 A 表示为:

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

若 $\forall u_i \in U, t_A(u_i) = 0$, 且 $f_A(u_i) = 1$, 则称 vague 集 A 为零 vague 集。

定义9 假定 A 是论域 U 上的 vague 集, $U = \{u_1, u_2, \dots, u_n\}$, 其中 A 表示为:

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

若 $\forall u_i \in U, t_A(u_i) = 0$, 且 $f_A(u_i) = 0$, 则称 vague 集 A 为空 vague 集。

定义10 假定 A 和 B 是论域 U 上的两个 vague 集, $U = \{u_1, u_2, \dots, u_n\}$, 其中 A 和 B 分别表示为:

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

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

若 $\forall u_i \in U, t_B(u_i) = f_A(u_i), f_B(u_i) = t_A(u_i)$, 则称 B 是 A 的补。

定义11 设 T_1 是一 t -范数, 其相应的 t -余范数为 $\perp_1$, 两个 vague 集 A 和 B 的交集是 vague 集 C , 即 $C = A \cap B$, 其真/假隶属函数分别为:

$$t_C = t_A T_1 t_B,$$

$$1 - f_C = (1 - f_A) \wedge (1 - f_B)$$

定义12 设 T_1 是一 t -范数, 其相应的 t -余范数为 $\perp_1$, 两个 vague 集 A 和 B 的并集是 vague 集 C , 即 $C = A \cup B$, 其真/假隶属函数分别为:

$$t_C = t_A \vee t_B,$$

$$1 - f_C = (1 - f_A) \perp_1 (1 - f_B)$$

其中, “ $\vee$ ” 和 “ $\wedge$ ” 分别为取大和取小算子。

3 Vague 集上的距离

在利用 vague 集研究不确定性问题时, 需要对 vague 集(值)间的关系及 vague 集中元素间的关系进行度量, 而且许多时候都需要具有距离性质的度量方法。为此, 我们提出了一套距离测度, 并证明了距离测度的各种性质。

定义13 假定 $x = [t_x, 1 - f_x]$, $y = [t_y, 1 - f_y]$ 是两个 vague 值, $S(x)$ 和 $S(y)$ 的定义与文[4]中的定义相同, x 和 y 之间的距离可定义为如下形式:

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

定理1 $L(x, x) = 0$

证明: 由 $L(x, y)$ 的定义即可证明。

定理2 $1 \geq L(x, y) > 0$ ($x \neq y$)

证明: $\because x \neq y \therefore$ 至少有 $t_x \neq t_y$ 或者 $f_x \neq f_y$ 成立, 假定 $t_x \neq t_y$,

$$\therefore L(x, y) = \frac{|S(x) - S(y)|}{4} + \frac{|t_x - t_y| + |f_x - f_y|}{4} > 0$$

又 $\because -1 \leq S(x) \leq 1, -1 \leq S(y) \leq 1, -1 \leq t_x - t_y \leq 1, -1 \leq f_x - f_y \leq 1$

$$\therefore L(x, y) \leq \frac{2}{4} + \frac{1+1}{4} = 1 \quad \text{证毕。}$$

定理3 $L(x, y) = L(y, x)$

证明: 由 $L(x, y)$ 的定义即可证明。

定理4 $L(x_1, x_2) + L(x_1, x_3) \geq L(x_2, x_3)$

证明:

$$\begin{aligned} L(x_1, x_2) + L(x_1, x_3) &= \frac{|S(x_1) - S(x_2)|}{4} + \frac{|t_{x_1} - t_{x_2}| + |f_{x_1} - f_{x_2}|}{4} + \frac{|S(x_1) - S(x_3)|}{4} + \frac{|t_{x_1} - t_{x_3}| + |f_{x_1} - f_{x_3}|}{4} \\ &\geq \frac{|S(x_1) - S(x_2) + S(x_3) - S(x_1)|}{4} + \frac{|t_{x_1} - t_{x_2} + t_{x_3} - t_{x_1}| + |f_{x_1} - f_{x_2} + f_{x_3} - f_{x_1}|}{4} \\ &= \frac{|S(x_3) - S(x_2)|}{4} + \frac{|t_{x_3} - t_{x_2}| + |f_{x_3} - f_{x_2}|}{4} = L(x_3, x_2) = L(x_2, x_3) \quad \text{证毕。} \end{aligned}$$

定理5 $\forall \epsilon > 0$ 如果 $L(x_1, x_2) \leq \frac{\epsilon}{2}, L(x_1, x_3) \leq \frac{\epsilon}{2}$, 则 $L(x_2, x_3) \leq \epsilon$

证明: 由定理4有 $L(x_2, x_3) \leq L(x_1, x_2) + L(x_1, x_3) \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$ 证毕。

定义14 假定 A 和 B 是两个 vague 集, A 和 B 之间的距离可定义为如下形式:

$$\begin{aligned} R(A, B) &= \frac{1}{n} \sum_{i=1}^n L(V_A(u_i), V_B(u_i)) \\ &= \frac{1}{n} \sum_{i=1}^n \left(\frac{|S(V_A(u_i)) - S(V_B(u_i))|}{4} + \frac{|t_A(u_i) - t_B(u_i)| + |f_A(u_i) - f_B(u_i)|}{4} \right) \quad (2) \end{aligned}$$

其中, $V_A(u_i) = [t_A(u_i), 1 - f_A(u_i)], V_B(u_i) = [t_B(u_i), 1 - f_B(u_i)], u_i \in U, 1 \leq i \leq n$

定理6 $R(A, A) = 0$

证明: 由定理1即可证明。

定理7 $1 \geq R(A, B) > 0$ ($A \neq B$)

证明: $R(A, B) = \frac{1}{n} \sum_{i=1}^n L(V_A(u_i), V_B(u_i)) > \frac{1}{n} \sum_{i=1}^n 0 = 0$ (由定理2得)

另一方面, $R(A, B) = \frac{1}{n} \sum_{i=1}^n L(V_A(u_i), V_B(u_i)) \leq \frac{1}{n} \sum_{i=1}^n 1 = 1$ (由定理2得) 证毕。

定理8 $R(A, B) = R(B, A)$

证明: $R(A, B) = \frac{1}{n} \sum_{i=1}^n L(V_A(u_i), V_B(u_i)) = \frac{1}{n} \sum_{i=1}^n L(V_B(u_i), V_A(u_i)) = R(B, A)$ (由定理3得) 证毕。

定理9 $R(A, B) + R(A, C) \geq R(B, C)$

证明: $R(A, B) + R(A, C) = \frac{1}{n} \sum_{i=1}^n L(V_A(u_i), V_B(u_i)) + \frac{1}{n} \sum_{i=1}^n L(V_A(u_i), V_C(u_i)) = \frac{1}{n} \sum_{i=1}^n L(V_A(u_i), V_B(u_i)) + L(V_A(u_i), V_C(u_i)) \geq \frac{1}{n} \sum_{i=1}^n L(V_B(u_i), V_C(u_i)) = R(B, C)$ (由定理4得) 证毕。

定理10 $\forall \epsilon > 0$ 如果 $R(A, B) \leq \frac{\epsilon}{2}, R(A, C) \leq \frac{\epsilon}{2}$, 则 $R(B, C) \leq \epsilon$

证明: 由定理9有 $R(B, C) \leq R(A, B) + R(A, C) \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$ 证毕。

定义15 假定 $A_1, A_2, \dots, A_m$ 是论域 U 上的 vague 集, $U = \{u_1, u_2, \dots, u_n\}$, 它们分别表示为:

$$A_1 = \sum_{i=1}^n [t_{A_1}(u_i), 1 - f_{A_1}(u_i)] / u_i, \quad u_i \in U$$

$$A_2 = \sum_{i=1}^n [t_{A_2}(u_i), 1 - f_{A_2}(u_i)] / u_i, \quad u_i \in U$$

.....

$$A_m = \sum_{i=1}^n [t_{A_m}(u_i), 1 - f_{A_m}(u_i)] / u_i, \quad u_i \in U$$

则论域 U 中任意两元素 u_i, u_j 在 vague 集 $A_1, A_2, \dots, A_m$ 下的距离定义为:

$$\begin{aligned} R_e(u_i, u_j) &= \frac{1}{m} \sum_{k=1}^m L(V_{A_k}(u_i), V_{A_k}(u_j)) \\ &= \frac{1}{m} \sum_{k=1}^m \left(\frac{|S(V_{A_k}(u_i)) - S(V_{A_k}(u_j))|}{4} + \right. \\ &\quad \left. \frac{|t_{A_k}(u_i) - t_{A_k}(u_j)| + |f_{A_k}(u_i) - f_{A_k}(u_j)|}{4} \right) \end{aligned} \quad (3)$$

定理11 $R_e(u_i, u_i) = 0$

证明:由定理1即可证明。

定理12 $1 \geq R_e(u_i, u_j) > 0 (i \neq j)$

证明: $R_e(u_i, u_j) = \frac{1}{m} \sum_{k=1}^m L(V_{A_k}(u_i), V_{A_k}(u_j)) > \frac{1}{m} \sum_{k=1}^m 0 = 0$ (由定理2得)

另一方面, $R_e(u_i, u_j) = \frac{1}{m} \sum_{k=1}^m L(V_{A_k}(u_i), V_{A_k}(u_j)) \leq \frac{1}{m} \sum_{k=1}^m 1 = 1$ (由定理2得) 证毕。

定理13 $R_e(u_i, u_i) = R_e(u_j, u_j)$

证明: $R_e(u_i, u_j) = \frac{1}{m} \sum_{k=1}^m L(V_{A_k}(u_i), V_{A_k}(u_j)) = \frac{1}{m} \sum_{k=1}^m L(V_{A_k}(u_j), V_{A_k}(u_i)) = R_e(u_j, u_i)$ (由定理3得) 证毕。

定理14 $R_e(u_i, u_j) + R_e(u_i, u_p) \geq R_e(u_j, u_p)$

证明: $R_e(u_i, u_j) + R_e(u_i, u_p) = \frac{1}{m} \sum_{k=1}^m L(V_{A_k}(u_i), V_{A_k}(u_j)) + \frac{1}{m} \sum_{k=1}^m L(V_{A_k}(u_i), V_{A_k}(u_p)) = \frac{1}{m} \sum_{k=1}^m L(V_{A_k}(u_i), V_{A_k}(u_j)) + L(V_{A_k}(u_i), V_{A_k}(u_p)) \geq \frac{1}{m} \sum_{k=1}^m L(V_{A_k}(u_j), V_{A_k}(u_p)) = R_e(u_j, u_p)$ (由定理4得) 证毕。

定理15 $\forall \epsilon > 0$ 如果 $R_e(u_i, u_j) \leq \frac{\epsilon}{2}, R_e(u_i, u_p) \leq \frac{\epsilon}{2}$, 则 $R_e(u_j, u_p) \leq \epsilon$

证明:由定理14有 $R_e(u_j, u_p) \leq R_e(u_i, u_j) + R_e(u_i, u_p) \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$ 证毕。

4 实例分析

下面我们给出一个距离测度在行为分析中的应用例子来说明距离测度。这个例子中的问题来自文[1]。假定有4个成员 (u_1, u_2, u_3, u_4) 和三个组 (A, B, C), 每个成员属于每个组的可能性用 vague 集的隶属度表示如下:

$$A = [0.4, 0.6] / u_1 + [0.8, 0.9] / u_2 + [0.9, 0.9] / u_3 + [0, 0] / u_4$$

$$B = [0.6, 0.7] / u_1 + [0.3, 0.5] / u_2 + [0, 0] / u_3 + [0.5, 0.8] / u_4$$

$$C = [0, 0] / u_1 + [0.4, 0.6] / u_2 + [0.8, 0.9] / u_3 + [1, 1] / u_4$$

根据距离测度我们可以回答下面两类问题^[5]:

第一类:在多大程度上,组 A 和组 B 可能合作?

第二类:在多大程度上,成员 u_2 和 u_3 可能在同一个组中?

为了回答第一类问题,我们考虑两个组的距离越小,它们可能合作的程度就越大。首先利用式(2)计算 A 和 B 之间的距离如下:

$$\begin{aligned} R(A, B) &= \frac{1}{4} \left[\frac{|(0.4-0.4)-(0.6-0.3)|}{4} + \frac{|0.4-0.6|+|0.4-0.3|}{4} \right] + \left[\frac{|(0.8-0.1)-(0.3-0.5)|}{4} + \frac{|0.8-0.3|+|0.1-0.5|}{4} \right] + \\ &\quad \left[\frac{|(0.9-0.1)-(0-1)|}{4} + \frac{|0.9-0|+|0.1-1|}{4} \right] + \left[\frac{|(0-1)-(0.5-0.2)|}{4} + \frac{|0-0.5|+|1-0.2|}{4} \right] = 0.5375 \end{aligned}$$

即 A 与 B 之间的距离是 0.5375, 由于距离越小表示二者之间越靠近, 而可能性是希望用数值越大表示可能性越大, 因此我们用 (1-距离值) 来表示可能性。也就是说组 A 和组 B 可能合作的可能性是 $1-0.5375$, 即 0.4625。

为了回答第二类问题,我们认为成员 u_2 和 u_3 之间的距离越小,就越可能是在同一个组中,所以利用式(3)计算 u_2 和 u_3 之间的距离如下:

$$\begin{aligned} R_e(u_2, u_3) &= \frac{1}{3} \left[\frac{|(0.8-0.1)-(0.9-0.1)|}{4} + \frac{|0.8-0.9|+|0.1-0.1|}{4} \right] + \left[\frac{|(0.3-0.5)-(0-1)|}{4} + \frac{|0.3-0|+|0.5-1|}{4} \right] + \\ &\quad \left[\frac{|(0.4-0.4)-(0.8-0.1)|}{4} + \frac{|0.4-0.8|+|0.4-0.1|}{4} \right] = 0.2667 \end{aligned}$$

即 u_2 与 u_3 之间的距离大约是 0.2667, 也就是说成员 u_2 与 u_3 在同一组的可能性大约是 $1-0.2667$, 即 0.7333。在本例中,利用我们的距离测度得到的结果与文[1]中用相似度量方法得到的结果相吻合。

结论 本文提出了一套 vague 集(值)间的距离测度及 vague 集中元素之间的距离测度,证明了距离测度的各种性质,然后还给出了一个实例说明所给出的距离测度在处理行为分析问题中的应用。在实例中,利用我们的距离测度得到的结果与文[1]中利用相似度量方法得到的结果相吻合。比文[5]中利用 Fuzzy 相似度量的方法更灵活。实例中应用距离测度的方法为处理行为分析问题提供了一个有用的方法。距离测度的提出为智能系统提供了一个有用的工具。

参考文献

- 1 Chen S M. Similarity Measures Between Vague Sets and Between Elements. IEEE Trans. Systems Man Cybernetic-Part B, 1997, 27:153~158
- 2 Zadeh L A. Fuzzy Sets and Their Applications to Cognitive and Decision Processes [M]. In: L. A. Zadeh, et al. eds. Academic Press, New York, 1975. 1~39
- 3 Gau W L, Buehrer D J. Vague sets IEEE Trans. Systems Man Cybernetic, 1993, 23:610~614
- 4 Chen S M. Measures of similarity between vague sets. Fuzzy set and System, 1995, 74:217~223
- 5 Lee-Kwang H, Song Y S, Lee K M. Similarity measure between fuzzy sets and between elements. Fuzzy Sets and Systems, 1994, 62:291~293
- 6 李凡. 模糊信息处理系统[M]. 北京:北京大学科学出版社, 1998