

控制与决策

Control and Decision

一种新的GBPA生成方法及其在模式识别中的应用

付威, 王欣

引用本文:

付威,王欣. 一种新的GBPA生成方法及其在模式识别中的应用[J]. *控制与决策*, 2024, 39(3): 994–1002.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2022.1230>

您可能感兴趣的其他文章

Articles you may be interested in

小样本下多稀疏表示分类器的决策融合方法

Decision fusion of multiple sparse representation-based classifiers in case of small samples

控制与决策. 2021, 36(8): 1984–1990 <https://doi.org/10.13195/j.kzyjc.2019.1839>

基于模糊-两阶段超效率SBM的电网应急能力动态综合评价

Dynamic comprehensive evaluation of power grid emergency capability based on fuzzy-two-stage super efficiency SBM

控制与决策. 2021, 36(6): 1333–1341 <https://doi.org/10.13195/j.kzyjc.2019.1128>

基于广义主成分分析的重构故障子空间建模方法

Reconstructed fault subspace modelling method based on generalized principal component analysis

控制与决策. 2021, 36(4): 808–814 <https://doi.org/10.13195/j.kzyjc.2019.0818>

基于广义罚函数可行性准则的DE算法对不确定数据的处理

Application of improved DE algorithm based on generalized penalty function feasibility criteria in uncertain data processing

控制与决策. 2021, 36(2): 498–504 <https://doi.org/10.13195/j.kzyjc.2019.0728>

特征加工链选用规律的挖掘、修正及其在工艺决策中的应用

Mining and correction of selection rule of feature operation chain and their application in process design

控制与决策. 2020, 35(12): 2865–2874 <https://doi.org/10.13195/j.kzyjc.2019.0367>

一种新的 GBPA 生成方法及其在模式识别中的应用

付威¹, 王欣^{1,2†}

(1. 黑龙江大学自动化系, 哈尔滨 150080; 2. 黑龙江省信息融合估计与检测重点实验室, 哈尔滨 150080)

摘要: 广义证据理论是一种在不完备识别框架中处理多传感器信息融合问题的实用方法. 由于时代环境的影响, 人们的认知存在局限性, 难免会将不完备的识别框架认为是完备的, 经典证据理论在这种情况下并不完全适用. 因此, 根据广义证据理论提出一种新的广义基本概率赋值 (generalized basic probability assignment, GBPA) 生成方法. 该方法首先根据训练数据分别构造样本类别和测试样本的广义三角模糊数模型; 然后通过计算样本和类别间的广义三角模糊距离生成 GBPA; 最后使用广义组合规则融合所有证据并得出最终的结论. Iris 数据集的实验结果表明所提方法合理有效, 即使在样本不足的情况下仍有较高的分类精度.

关键词: 广义证据理论; 广义基本概率赋值; 广义三角模糊数; 广义三角模糊距离

中图分类号: TP274

文献标志码: A

DOI: 10.13195/j.kzyjc.2022.1230

引用格式: 付威, 王欣. 一种新的 GBPA 生成方法及其在模式识别中的应用 [J]. 控制与决策, 2024, 39(3): 994-1002.

A novel method to determine GBPA and its application in pattern recognition

FU Wei¹, WANG Xin^{1,2†}

(1. Department of Automation, Heilongjiang University, Harbin 150080, China; 2. Key Laboratory of Information Fusion Estimation and Detection in Heilongjiang Province, Harbin 150080, China)

Abstract: Generalized evidence theory (GET) is a useful method to address the problem of multi-sensor information fusion over the incomplete framework of discernment (FoD). Due to the limitations of cognition in the era, people inevitably considered the incomplete FoD as the complete FoD, and the classical evidence theory is not fully applicable in this case. Therefore, a new generalized basic probability assignment (GBPA) determination method based on the GET is proposed. According to the training data, the method first generates the generalized triangular fuzzy number (GTFN) models of the classes and the test samples, respectively. Then the GBPAs are determined by calculating the generalized triangular fuzzy distance between samples and classes. Finally, the generalized combination rule fuses all the bodies of evidence to obtain the final conclusion. The experimental results on the Iris dataset show that the proposed method is reasonably effective and has relatively high classification accuracy even in the case of insufficient samples.

Keywords: generalized evidence theory; generalized basic probability assignment; generalized triangular fuzzy number; generalized triangular fuzzy distance

0 引言

D-S 证据理论是信息融合技术中重要的研究方法之一, 其优势在于能有效地表达不确定信息, 而且能在没有先验信息的情况下融合证据^[1]. 因此, D-S 证据理论被广泛地应用于风险评估^[2]、模式识别^[3-4]和故障诊断^[5-6]等多个领域.

生成基本概率赋值 (basic probability assignment, BPA) 是应用 D-S 证据理论的第一步. BPA 通常是根

据多个传感器的识别数据以特定方法构造而成, 表示对一个或多个目标的支持程度, 如何合理地生成 BPA 仍是一个研究热点. 值得注意的是, 如果不同的 BPA 之间存在高度冲突的情况, 在组合 BPA 的过程中很可能会得出 0 信任或 1 信任等反直观结果, 使信息融合的精度降低^[7-8].

一般而言, BPA 存在高度冲突的原因有两种. 第 1 种原因是 BPA 生成方法存在固有问题. 在根据专

收稿日期: 2022-07-11; 录用日期: 2022-11-22.

基金项目: 国家自然科学基金项目 (61573132); 黑龙江省自然科学基金重点项目 (ZD2021F003); 黑龙江省自然科学基金联合引导项目 (LH2020G008); 黑龙江省省属高等学校基本科研业务费基础研究基金项目 (KJCX201809).

责任编辑: 刘宝碇.

† 通讯作者. E-mail: wangxin@hlju.edu.cn.

家经验设定BPA的方法中,由于专家的意见具有主观性,无法保证所有专家均得出一致的结论,BPA之间难免会存在高度冲突的情况^[9].还有一些BPA生成方法会假设样本数据满足特定的概率分布.例如,康兵义等^[10]假设训练样本满足均匀分布,提出了基于区间数的BPA生成方法;Xu等^[11]假设样本概率分布符合高斯模型,构造了一种嵌套结构的BPA生成函数;Chen等^[12]根据样本的概率分布拟合构造了BPA生成函数.上述方法均假设训练样本与特定分布一致,但若两者的偏差很大,就很难得到期望结果.第2种原因是经典证据理论不适用于不完备的识别框架.在现实世界中,人的认知受时代环境的限制,随着社会的发展,新事物不断涌现,识别框架中很可能会存在未知的命题,这不可避免地会导致经典证据理论中存在高度冲突的证据,从而无法得出合理的融合结果.针对这一问题,Deng^[13]在经典证据理论的基础上提出了广义证据理论(generalized evidence theory, GET),该理论放弃了经典证据理论中 $m(\emptyset) = 0$ 的严格限制条件,将空集 \emptyset 当作未知的命题可以有效地融合不完备识别框架中的不确定信息.为了解决在GET中生成广义基本概率赋值(generalized basic probability assignment, GBPA)的问题,Zhang等^[14]基于扩展三角模糊数与待测实例的相交长度生成GBPA,该方法的计算负担小,使用嵌套结构的GBPA生成方法减少了高度冲突证据的产生.但是该方法在生成GBPA的过程中,将过多的信息分配给复合焦元,使得证据的不确定性增加,信息量减小,致使算法的融合结果不够精确.与Zhang等的GBPA构造方法不同,Jiang等^[15]使用三角模糊数与待测实例之间的相交面积生成GBPA.在文献[15]的基础上,Fan等^[16]采用 k -means算法构造三角模糊数生成GBPA.与文献[14]中的GBPA生成方法相比,上述两种方法使用相交面积能直观地提取更多的信息,但是这种方法需要通过构造系数矩阵对GBPA进行修正,这在一定程度上会导致算法复杂度升高,不便于实际应用.

在GET的启发下,本文基于广义三角模糊距离提出一种新的GBPA生成方法.首先,根据训练样本的特征构造广义三角模糊数(generalized triangular fuzzy number, GTFN)模型和待测实例的三角模糊数模型.然后,计算两者之间的广义三角模糊距离,通过调整待测实例模型的置信度,为证据分配更多的信息.最后,使用广义组合规则融合证据得到相应的GBPA.该方法是一种数据驱动方法,计算负担小,仅需使用较少的先验数据即可构造三角模糊数.它根据训练样本的标准差对测试数据进行扩展,能在开放

世界中更准确地描述测试数据的特征,更合理地实现现实情况中的分类.Iris数据集的分类实验结果表明,所提方法具有较高的分类识别率,是一种合理有效的GBPA生成方法.

1 相关基础理论

1.1 广义证据理论

定义1 (广义基本概率赋值)^[13] 假设 U 为开放世界中的一个识别框架,幂集 2^U 是 U 的所有子集构成的集合.函数 $m : 2^U \rightarrow [0, 1]$ 需满足

$$\sum_{A \subseteq U} m(A) = 1, 0 \leq m(A) \leq 1, \quad (1)$$

其中 $m(A)$ 为命题 A 的GBPA.GBPA与BPA之间的区别在于GBPA规定 $m(\emptyset)$ 可以不为0.若 $m(\emptyset)$ 越大,则未知结果的可能性越大;当 $m(\emptyset) = 0$ 时,GBPA退化为BPA.

定义2 (广义组合规则)^[13] 假设识别框架 U 下存在两个证据 m_1 和 m_2 ,当 $k \neq 1$ 时,广义组合规则为

$$m(A) = \frac{(1 - m(\emptyset)) \sum_{A_i \cap B_j = A} m_1(A_i) \times m_2(B_j)}{1 - k}, \quad A \neq \emptyset; \quad (2)$$

$$m(\emptyset) = m_1(\emptyset) \times m_2(\emptyset); \quad (3)$$

$$k = \sum_{A_i \cap B_j = \emptyset} m_1(A_i) \times m_2(B_j). \quad (4)$$

当 $k = 1$ 时, $m(\emptyset) = 1$.

1.2 广义三角模糊数

为了解决不确定环境中的信息问题,Zadeh^[17]于1965年提出模糊集理论.广义三角模糊数作为模糊数的类型之一,由于其适应范围广,运算简便,且符合人类的思维习惯,而得到了广泛的应用.

定义3 (广义三角模糊数)^[18] 假设实数 R 内存在一个GTFN $\tilde{A} = (a, b, c; w)$, $w \in [0, 1]$,则从 R 到闭区间 $[0, w]$ 上的连续映射表示为 $\mu_{\tilde{A}}(x) : R \rightarrow [0, w]$.

如图1所示, $\mu_{\tilde{A}}(x)$ 表示 \tilde{A} 的隶属度函数, w 表示 \tilde{A} 的置信度.当 $w = 1$ 时,GTFN \tilde{A} 退化为普通三角模糊数. $\mu_{\tilde{A}}(x)$ 需满足以下条件:1)当 $x \in (-\infty, a] \cup [c, +\infty)$ 时, $\mu_{\tilde{A}}(x) = 0$;2)当 $x \in [a, b]$ 时, $\mu_{\tilde{A}}(x)$ 严格单调增;3)当 $x = b$ 时, $\mu_{\tilde{A}}(x) = w$;4)当 $x \in [b, c]$ 时, $\mu_{\tilde{A}}(x)$ 严格单调减.因此, $\tilde{A} = (a, b, c; w)$ 的隶属度函数可以表示为

$$\mu_{\tilde{A}}(x) = \begin{cases} 0, & x < a; \\ \frac{x - a}{b - a} w, & a \leq x \leq b; \\ \frac{c - x}{c - b} w, & b \leq x \leq c; \\ 0, & x > c. \end{cases} \quad (5)$$

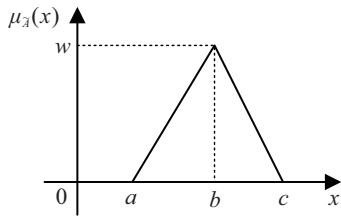
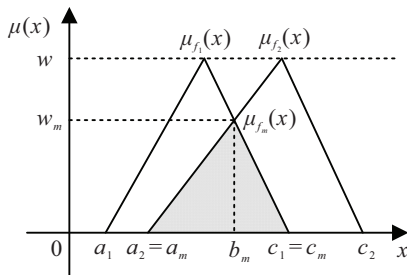
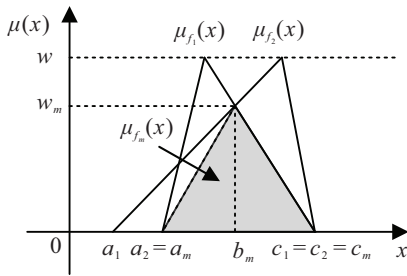


图1 GTFN \tilde{A}

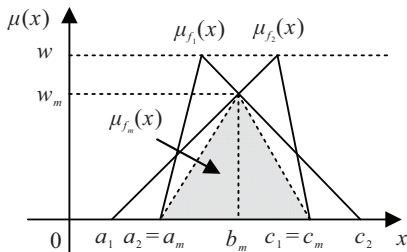
假设实数 R 中存在两个三角模糊数 $f_1 = (a_1, b_1, c_1; 1)$ 和 $f_2 = (a_2, b_2, c_2; 1)$, 且 f_1 与 f_2 之间存在交集, 则三角模糊数之间的重合部分可构造为复合命题的 GTFN, 记作 $f_m = (a_m, b_m, c_m; w_m)$. 其中, 三角模糊数之间的重合部分大致可以分为三角形、四边形和五边形 3 种情况. 如图 2(a) 所示, f_1 与 f_2 的重合部分仍然是一个三角形, 则这个三角形可以构成复合命题的 GTFN. 对于图 2 中 (b) 和 (c) 的情况, 它们的重合部分为多边形, 则将该多边形纵坐标最高点对应的横坐标以及多边形与横轴的 2 个交点的横坐标共同构成复合命题的 GTFN.



(a) 重合部分为三角形



(b) 重合部分为四边型



(c) 重合部分为五边形

图2 构造复合命题 GTFN 的 3 种情况

1.3 广义三角模糊距离

定义 4 (广义三角模糊距离)^[19] 假设已知两个 GTFN $f_1 = (a_1, b_1, c_1; w)$ 和 $f_2 = (a_2, b_2, c_2; w)$, 则 f_1 与 f_2 之间的 GTFN 距离为

$$d(f_1, f_2) = \int_0^w (\mu_{f_1}(x), \mu_{f_2}(x)) d\mu(x) = \int_0^w (\mu_{f_1}(x)^L, \mu_{f_2}(x)^L) d\mu(x) + \int_0^w (\mu_{f_1}(x)^R, \mu_{f_2}(x)^R) d\mu(x). \quad (6)$$

其中: $\int_0^w (\mu_{f_1}(x)^L, \mu_{f_2}(x)^L) d\mu(x)$ 表示 f_1 与 f_2 的左边界、 $\mu(x) = 0$ 和 $\mu(x) = w$ 共同围成的闭合区域的面积; $\int_0^w (\mu_{f_1}(x)^R, \mu_{f_2}(x)^R) d\mu(x)$ 表示 f_1 与 f_2 的右边界、 $\mu(x) = 0$ 和 $\mu(x) = w$ 共同围成的闭合区域的面积. 不妨设, 当 f_1 与 f_2 相交的重合部分为五边形时, 各部分面积和交点如图 3 所示.

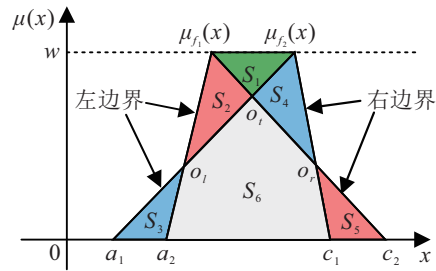


图3 相交的两个 GTFN 构成的五边形闭合区域

由图 3 得

$$\int_0^w (\mu_{f_1}(x)^L, \mu_{f_2}(x)^L) d\mu(x) = S_1 + S_2 + S_3, \quad (7)$$

$$\int_0^w (\mu_{f_1}(x)^R, \mu_{f_2}(x)^R) d\mu(x) = S_1 + S_4 + S_5. \quad (8)$$

由式 (6) ~ (8) 引出

$$d(f_1, f_2) = 2S_1 + S_{f_1} + S_{f_2} - 2S_6. \quad (9)$$

其中 S_{f_1} 和 S_{f_2} 分别为 f_1 和 f_2 与 $\mu(x) = 0$ 围成的闭合区域的面积, 即 $S_{f_1} = S_2 + S_5 + S_6, S_{f_2} = S_3 + S_4 + S_6$.

注意, 式 (9) 为 f_1 与 f_2 相交的重合部分为五边形时 $d(f_1, f_2)$ 的计算方法. 特别地, 当 f_1 与 f_2 的重合部分为四边形时, 图 3 中 o_l 或 o_r 的纵坐标退化为 0; 当 f_1 与 f_2 的重合部分为三角形时, 图 3 中 o_l 和 o_r 的纵坐标均退化为 0, 式 (9) 的计算方法同样适用. 当两个广义三角模糊数不相交时, 如图 4 所示, 此时 f_1 与 f_2 间不存在重合部分, 即 $S_6 = 0$, 代入式 (9) 可得

$$d(f_1, f_2) = 2S_1 + S_{f_1} + S_{f_2}. \quad (10)$$

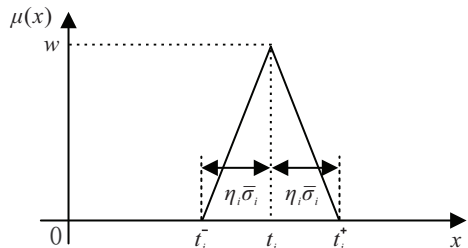


图4 由不相交的两个 GTFN 构成的闭合区域

综上所述, 式 (9) 可以处理 f_1 与 f_2 相交重合部分的所有情况. 由式 (9) 可知, S_1 越大, S_6 越小, 两个 GTFN 间

的距离越大,表明它们之间的相关性越小. 例如在图4中, $S_6 = 0$, 且 S_1 面积大于图3中 S_1 的面积, 因此不相交的两个GTFN的距离大于相交时的距离, 这表明此时的两个GTFN之间的相关性弱, 这与直观相符.

2 所提方法的GBPA生成原理

2.1 生成GTFN

假设 X 是含有 N 种属性和 M 种类别的数据集, 则 X 内属性为 i 且类别为 j 的 n 个训练数据的集合可以表示为 $X_{ij} = \{x_1^{ij}, x_2^{ij}, \dots, x_n^{ij}\}$, 其中 $i = 1, 2, \dots, N, j = 1, 2, \dots, M$. 由 X_{ij} 可以得出平均值 \bar{x}_{ij} 和标准差 σ_{ij} 分别为

$$\bar{x}_{ij} = \frac{1}{n} \sum_{k=1}^n x_k^{ij}, \quad (11)$$

$$\sigma_{ij} = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_k^{ij} - \bar{x}_{ij})^2}. \quad (12)$$

因此, 由定义3引出属性为 i 类别为 j 的三角模糊数为 $f_{ij} = (\min(X_{ij}), \bar{x}_{ij}, \max(X_{ij}); 1)$, 其中 $\min(X_{ij})$ 和 $\max(X_{ij})$ 分别代表 X_{ij} 的最小值和最大值. 对于测试集中属性为 i 的任意测试样本 t_i , 考虑到在获取样本数据的过程中可能会存在误差, 故使用GTFN描述 t_i 误差的极值. 所提方法使用先验数据 $\bar{\sigma}_i$ 反映训练集数据的离散程度, 由 $\bar{\sigma}_i$ 计算 t_i 可能性最小的值 t_i^- 和可能性最大的值 t_i^+ , 有

$$\bar{\sigma}_i = \frac{1}{M} \sum_{j=1}^M \sigma_{ij}, \quad (13)$$

$$t_i^- = t_i - \eta_i \bar{\sigma}_i, \quad (14)$$

$$t_i^+ = t_i + \eta_i \bar{\sigma}_i, \quad (15)$$

其中 η_i 为调节因子, 可以根据不同属性的样本进行调整. 如图5所示, 根据上述参数构造出的单个测试样本的GTFN为 $f_t = (t_i^-, t_i, t_i^+; w)$, 其中置信度 w 是一个变化的值, 具体的变化规则将在下一节中叙述.

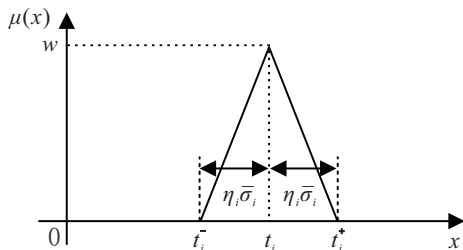


图5 单个测试集样本的GTFN

2.2 计算样本与各命题间的广义三角模糊距离

假设由数据集 X 可得出属性 i 中任意两个类别的三角模糊数 $f_{i1} = (a_1, b_1, c_1; 1)$ 和 $f_{i2} = (a_2, b_2, c_2; 1)$, 若 f_{i1} 与 f_{i2} 之间存在交集, 则由定义3可得出相应

复合命题的GTFN f_m , 设 $f_m = (a_m, b_m, c_m; w_m)$. 如果测试样本的GTFN为 $f_t = (t_i^-, t_i, t_i^+; w)$, 当计算它与任一类别的广义三角模糊距离时, 需调整测试样本GTFN的置信度 w , 使其与相应类别的置信度一致.

如图6所示, 此时 f_m 的 $a_m = a_2, c_m = c_1$. 当计算 f_t 与 f_{i1} 或 f_{i2} 的距离时, f_t 的置信度应调整为 $w = 1$; 当计算 f_t 与 f_m 的广义三角模糊距离时, f_t 的置信度应调整为 $w = w_m$. 然后, 再使用式(9)即可得出测试样本与各命题的广义三角模糊距离.

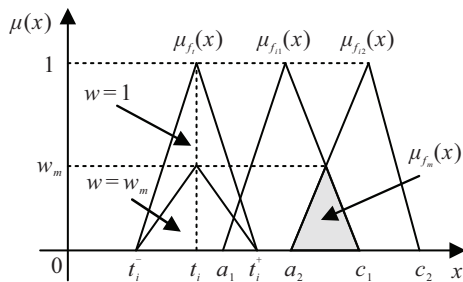


图6 随置信度变化的GTFN

注1 如果不同复合命题的GTFN完全一致, 则在计算广义三角距离时会得到重复的结果, 冗余的信息将使识别结果更倾向于模棱两可的答案, 使得最终的识别精度降低. 为了得到更为精确的结果, 本文将完全相同的GTFN归为这些命题中基数较大的复合命题的模型, 只计算它与待测样本GTFN的距离.

2.3 由广义三角模糊距离生成GBPA

在本节中, 所提方法首先根据广义三角模糊距离确定不同GTFN之间的相似度, 再生成GBPA. 假设已知两个GTFN f_1 和 f_2 之间的距离为 $d(f_1, f_2)$, 则它们的相似度为

$$S(f_1, f_2) = \frac{1}{1 + \alpha d(f_1, f_2)}, \quad (16)$$

其中调节系数 $\alpha > 0$. 由式(16)可知, 相似度与 $d(f_1, f_2)$ 成反比, 当 $d(f_1, f_2)$ 越接近于0时, f_1 与 f_2 之间的相似度越高, 反之亦然. 调节系数 α 的目的是将广义三角模糊距离离散化, 使得不同类别的相似度具有较大差异, 从而提升融合结果的精度. 式(16)求得的相似度不能直接作为GBPA, 需根据GET的要求进行处理.

假设数据集 X 的广义识别框架为 U_X , $S(A)$ 表示样本与 U_X 中命题 A 的相似度, 生成GBPA的具体策略如下:

1) 若 $\sum_{A \in U_X} S(A) \leq 1$, 则在式(16)中所求得的相似度即为GBPA, 且

$$m(\emptyset) = 1 - \sum_{A \in U_X} S(A). \quad (17)$$

2) 若 $\sum_{A \in U_X} S(A) > 1$, 则 $m(\emptyset) = 0$, U_X 中的任意命题 A_X 的 GBPA 为

$$m(A_X) = S(A_X) / \sum_{A \in U_X} S(A). \quad (18)$$

2.4 基于 GTFN 生成 GBPA 的步骤

生成 GBPA 的具体步骤如下:

step 1: 构造不同命题的 GTFN 模型. 基于训练样本中的单个特征属性, 计算各类样本的标准差和平均标准差, 并根据样本的最小值、平均值和最大值构造单子集命题的三角模糊数模型, 再由单子集命题的三角模糊数的交集构造复合命题的 GTFN 模型.

step 2: 使用平均标准差构造当前属性的测试样本的 GTFN.

step 3: 计算待识别样本与各命题的广义三角模糊距离, 并使用式(16)生成相似度.

step 4: 根据广义组合规则生成该样本在当前属性特征下的一组 GBPA 值.

为了更好地理解基于 GTFN 的 GBPA 生成方法, 算法 1 给出了该算法的计算机伪代码. 图 7 展示了该方法的流程.

算法 1 基于 GTFN 的 GBPA 生成方法.

输入: 训练集 D , 数据集属性的数量 N , 数据集类别的数量 M , 测试集 P , 测试样本的数量 T_s ;

输出: GBPA.

- 1) for $i = 1 : N$
- 2) for $j = 1 : M$
- 3) 将训练集 D 中属性为 i 且类别为 j 的数据选为 D_{ij} .
- 4) 由 D_{ij} 计算属性为 i 的三角模糊数 f_{ij} 和平均标准差 $\bar{\sigma}_i$.
- 5) end for
- 6) 由 f_{ij} 得出复合命题的广义三角模糊数 f_m .
- 7) end for
- 8) for $i = 1 : N$
- 9) 将测试集中属性为 i 的数据选为新的测试集 P_i .
- 10) for $t = 1 : T_s$
- 11) 生成并记录属性 i 中第 t 个样本的 GTFN $f_t = (t_i^-, t_i, t_i^+; w)$.
- 12) 计算样本与目标命题的 GTFN 距离 d 和相似度 S .
- 13) if $\sum_{A \in U_X} S(A) \leq 1$
- 14) $m(\emptyset) = 1 - \sum_{A \in U_X} S(A), m(A) = S(A)$

- 15) else
- 16) $m(\emptyset) = 0, m(A_X) = \frac{S(A_X)}{\sum_{A \in U_X} S(A)}$
- 17) end if
- 18) end for
- 19) end for
- 20) 使用 Dempster 组合规则融合 N 个 GBPA.
- 21) 输出最终的 GBPA.

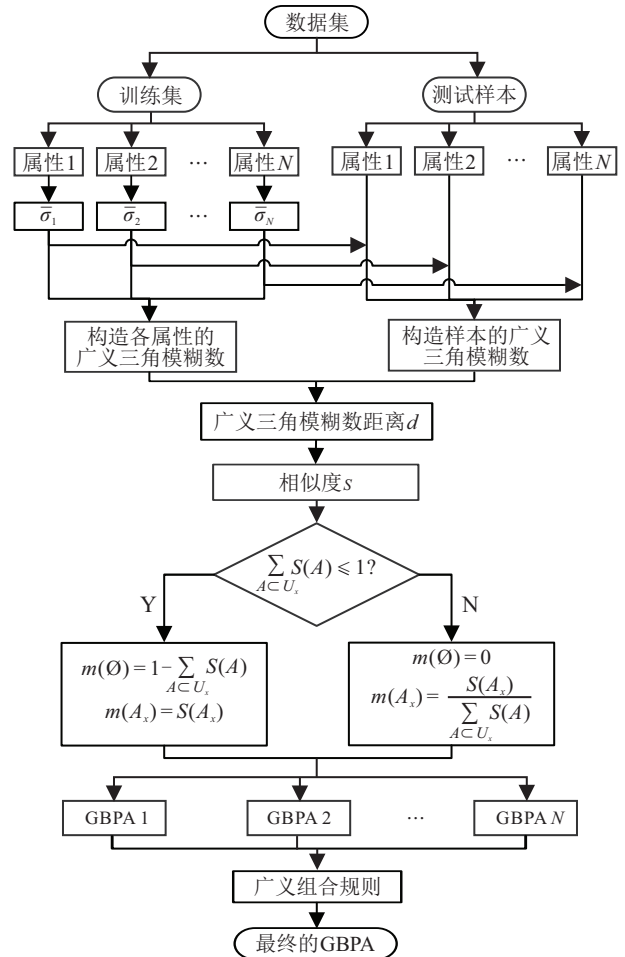


图 7 本文提出的 GBPA 生成方法的流程

3 在模式识别中典型案例的应用

3.1 使用 Iris 数据集生成 GBPA

本节使用 UC Irvine 机器学习知识库中的 Iris 数据集进行实验. Iris 数据集是模式识别中多属性决策的典型案例, 它包括 3 个类别: Setosa (Se)、Versicolour (Ve) 和 Virginica (Vi). 每个类别包含 50 个样本, 每个样本含有 4 种属性, 即萼片长度 (SL)、萼片宽度 (SW)、花瓣长度 (PL) 和花瓣宽度 (PW). 从 Iris 数据集的每一类中随机抽取 40 组样本, 共计 120 个样本作为训练集, 其余的样本作为测试集, 根据 2.4 节中提出的步骤生成 GBPA.

step 1: 构造不同命题的 GTFN 模型. 设 U_X 为 Iris 数据集在开放世界中的识别框架, 对随机抽取的训

练集数据使用式(11)~(13)计算单个测试样本GTFN的参数,具体数据如表1所示.图8展示了随机抽取的训练集中不同属性的三角模糊数模型.由图8可知,在属性SL和SW中,3种类别的三角模糊数存在较多的交集,样本不易区分.其中,属性为SL的单子集命题与复合命题的GTFN如表2所示.

表1 构造测试样本GTFN的参数

属性	σ_{Se}	σ_{Ve}	σ_{Vi}	$\bar{\sigma}$	η
SL	0.360	0.452	0.599	0.470	0.1
SW	0.396	0.306	0.323	0.342	0.1
PL	0.187	0.468	0.533	0.396	0.9
PW	0.113	0.204	0.274	0.197	1.0

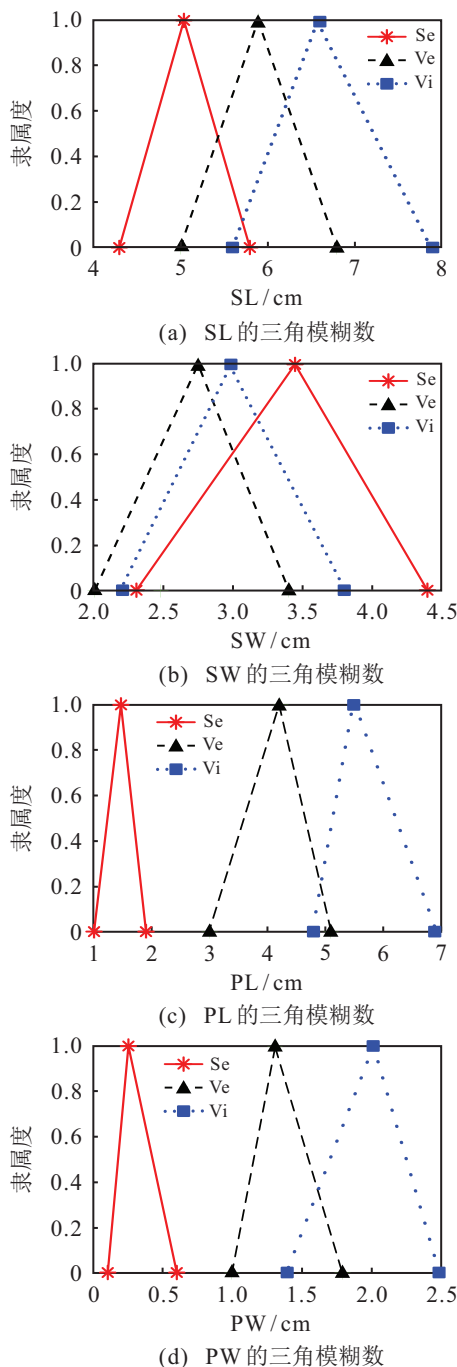


图8 各属性训练样本的三角模糊数

表2 属性为SL中各命题的GTFN

属性	广义三角模糊数
{Se}	(4.30, 5.04, 5.80; 1)
{Ve}	(5.00, 5.90, 6.80; 1)
{Vi}	(5.60, 6.59, 7.90; 1)
{Se, Ve}	(5.00, 5.43, 5.80; 0.48)
{Se, Vi}	(5.60, 5.71, 5.80; 0.11)
{Ve, Vi}	(5.60, 6.23, 6.80; 0.63)
{Se, Ve, Vi}	(5.60, 5.71, 5.80; 0.11)

step2: 构造测试样本的GTFN. 随机选取测试集中一个类别为Se的样本,数据(单位:cm)如表3所示.以SL属性为例,该样本SL属性的GTFN为(4.95, 5.05; w).

表3 随机抽取的样本的数据

属性	SL	SW	PL	PW
数值	5.0	3.4	1.5	0.2

step3: 计算广义三角模糊距离.如图9所示,当计算测试样本的GTFN与各命题的GTFN的距离时,w的值需与相应命题的GTFN模型的置信度保持一致.使用式(9)计算样本与各类别的广义三角模糊距离,实验中取调节系数 $\alpha = 42$,使用式(16)计算样本与各个命题之间的相似度,具体数值如表4所示.在表4中,由于{Se, Vi}的GTFN与{Se, Ve, Vi}的GTFN相同,由注1,只计算测试样本与{Se, Ve, Vi}的距离,因此表4中{Se, Vi}的距离和相似度为空.

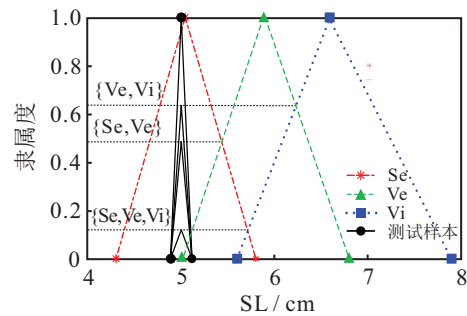


图9 属性为SL的单个测试样本的GTFN

表4 由所选样本生成的属性为SL的GBPA

命题	距离 d	相似度 S	GBPA
{Se}	0.1764	0.1189	0.1189
{Ve}	0.4487	0.0504	0.0504
{Vi}	0.8356	0.0277	0.0277
{Se, Ve}	0.2144	0.1000	0.1000
{Se, Vi}	—	—	0
{Ve, Vi}	0.6069	0.0377	0.0377
{Se, Ve, Vi}	0.3534	0.0631	0.0631
$\{\emptyset\}$	—	—	0.6022

step4: 根据开放世界规则生成GBPA. 由表4可知, $\sum_{ACU_x} S(A) < 1$, 因此 $m(\emptyset) = 1 - \sum_{ACU_x} S(A)$, 则可得该样本在属性SL下的GBPA.

重复 step 2~step 4, 可得 SW、PL 和 PW 属性的 GBPA. 最后, 使用广义组合规则可得最终的融合结果, 所有结果如表 5 所示. 由表 5 可知, 测试样本的类别分类结果为 Se, 这与该样本在 Iris 数据集中的类别一致, 表明本文提出的 GBPA 生成方法是有效的.

为了进一步测试所提方法的有效性, 本文从每个类别的测试样本中各随机抽取一组数据, 利用上述方法得到 4 种属性的 4 组 GBPA, 然后使用广义组合规则得到最终的 GBPA. 表 6 展示了随机挑选的 3 个测试样本的数据、相应的 GBPA 以及分类结果对比,

表 5 选取的样本的 GBPA 与融合结果

命题	SL	SW	PL	PW	融合结果
{Se}	0.1189	0.0856	0.4699	0.3266	0.9014
{Ve}	0.0504	0.0657	0.0177	0.0396	0.0095
{Vi}	0.0277	0.0879	0.0113	0.0260	0.0045
{Se, Ve}	0.1000	0	0	0	0
{Se, Vi}	0	0.1007	0	0	0
{Ve, Vi}	0.0377	0.0763	0.0136	0.0326	0.0006
{Se, Ve, Vi}	0.0631	0.0899	0	0	0
{∅}	0.6022	0.4939	0.4875	0.5793	0.0840

可以看出, 所提方法对于测试样本类别的判断是正确的, 是一种有效的 GBPA 生成方法.

表 6 3 组测试样本的 GBPA 与分类结果

样本	类别	SL	SW	PL	PW	GBPA	结论
1	Se	4.9 cm	3.0 cm	1.4 cm	0.2 cm	—	Se
		$m(\text{Se}) = 0.1151$	$m(\text{Se}) = 0.0739$	$m(\text{Se}) = 0.4144$	$m(\text{Se}) = 0.3226$	$m(\text{Se}) = 0.9145$	
		$m(\text{Ve}) = 0.0456$	$m(\text{Ve}) = 0.1100$	$m(\text{Ve}) = 0.0171$	$m(\text{Ve}) = 0.0396$	$m(\text{Ve}) = 0.0144$	
		$m(\text{Vi}) = 0.0262$	$m(\text{Vi}) = 0.1106$	$m(\text{Vi}) = 0.0110$	$m(\text{Vi}) = 0.0260$	$m(\text{Vi}) = 0.0058$	
		$m(\text{Se, Ve}) = 0.0844$	$m(\text{Se, Ve}) = 0$	$m(\text{Se, Ve}) = 0$	$m(\text{Se, Ve}) = 0$	$m(\text{Se, Ve}) = 0$	
		$m(\text{Se, Vi}) = 0$	$m(\text{Se, Vi}) = 0.1063$	$m(\text{Se, Vi}) = 0$	$m(\text{Se, Vi}) = 0$	$m(\text{Se, Vi}) = 0$	
		$m(\text{Ve, Vi}) = 0.0350$	$m(\text{Ve, Vi}) = 0.1302$	$m(\text{Ve, Vi}) = 0.0133$	$m(\text{Ve, Vi}) = 0.0326$	$m(\text{Ve, Vi}) = 0.0009$	
		$m(\text{Se, Ve, Vi}) = 0.0557$	$m(\text{Se, Ve, Vi}) = 0.1493$	$m(\text{Se, Ve, Vi}) = 0$	$m(\text{Se, Ve, Vi}) = 0$	$m(\text{Se, Ve, Vi}) = 0$	
$m(\emptyset) = 0.6380$	$m(\emptyset) = 0.3198$	$m(\emptyset) = 0.5442$	$m(\emptyset) = 0.5793$	$m(\emptyset) = 0.0643$			
2	Ve	5.5 cm	2.3 cm	4.0 cm	1.3 cm	—	Ve
		$m(\text{Se}) = 0.0871$	$m(\text{Se}) = 0.0416$	$m(\text{Se}) = 0.0184$	$m(\text{Se}) = 0.0455$	$m(\text{Se}) = 0.0060$	
		$m(\text{Ve}) = 0.0841$	$m(\text{Ve}) = 0.0916$	$m(\text{Ve}) = 0.1120$	$m(\text{Ve}) = 0.3180$	$m(\text{Ve}) = 0.8114$	
		$m(\text{Vi}) = 0.0391$	$m(\text{Vi}) = 0.0642$	$m(\text{Vi}) = 0.0277$	$m(\text{Vi}) = 0.0649$	$m(\text{Vi}) = 0.0843$	
		$m(\text{Se, Ve}) = 0.1771$	$m(\text{Se, Ve}) = 0$	$m(\text{Se, Ve}) = 0$	$m(\text{Se, Ve}) = 0$	$m(\text{Se, Ve}) = 0$	
		$m(\text{Se, Vi}) = 0$	$m(\text{Se, Vi}) = 0.0551$	$m(\text{Se, Vi}) = 0$	$m(\text{Se, Vi}) = 0$	$m(\text{Se, Vi}) = 0$	
		$m(\text{Ve, Vi}) = 0.0625$	$m(\text{Ve, Vi}) = 0.0810$	$m(\text{Ve, Vi}) = 0.0481$	$m(\text{Ve, Vi}) = 0.1218$	$m(\text{Ve, Vi}) = 0.0211$	
		$m(\text{Se, Ve, Vi}) = 0.1872$	$m(\text{Se, Ve, Vi}) = 0.0701$	$m(\text{Se, Ve, Vi}) = 0$	$m(\text{Se, Ve, Vi}) = 0$	$m(\text{Se, Ve, Vi}) = 0$	
$m(\emptyset) = 0.3628$	$m(\emptyset) = 0.5964$	$m(\emptyset) = 0.7938$	$m(\emptyset) = 0.4498$	$m(\emptyset) = 0.0773$			
3	Vi	6.5 cm	3.0 cm	5.8 cm	2.2 cm	—	Vi
		$m(\text{Se}) = 0.0317$	$m(\text{Se}) = 0.0739$	$m(\text{Se}) = 0.0108$	$m(\text{Se}) = 0.0244$	$m(\text{Se}) = 0.0018$	
		$m(\text{Ve}) = 0.0693$	$m(\text{Ve}) = 0.1100$	$m(\text{Ve}) = 0.0279$	$m(\text{Ve}) = 0.0534$	$m(\text{Ve}) = 0.1453$	
		$m(\text{Vi}) = 0.0789$	$m(\text{Vi}) = 0.1106$	$m(\text{Vi}) = 0.1087$	$m(\text{Vi}) = 0.1699$	$m(\text{Vi}) = 0.7184$	
		$m(\text{Se, Ve}) = 0.0421$	$m(\text{Se, Ve}) = 0$	$m(\text{Se, Ve}) = 0$	$m(\text{Se, Ve}) = 0$	$m(\text{Se, Ve}) = 0$	
		$m(\text{Se, Vi}) = 0$	$m(\text{Se, Vi}) = 0.1063$	$m(\text{Se, Vi}) = 0$	$m(\text{Se, Vi}) = 0$	$m(\text{Se, Vi}) = 0$	
		$m(\text{Ve, Vi}) = 0.1092$	$m(\text{Ve, Vi}) = 0.1302$	$m(\text{Ve, Vi}) = 0.0526$	$m(\text{Ve, Vi}) = 0.0749$	$m(\text{Ve, Vi}) = 0.0284$	
		$m(\text{Se, Ve, Vi}) = 0.0566$	$m(\text{Se, Ve, Vi}) = 0.1493$	$m(\text{Se, Ve, Vi}) = 0$	$m(\text{Se, Ve, Vi}) = 0$	$m(\text{Se, Ve, Vi}) = 0$	
$m(\emptyset) = 0.6122$	$m(\emptyset) = 0.3198$	$m(\emptyset) = 0.8000$	$m(\emptyset) = 0.6773$	$m(\emptyset) = 0.1061$			

3.2 测试不同训练规模的分类准确率

上一节根据固定训练规模的 GBPA 生成实验验证了所提方法的有效性, 本节将测试不同训练规模下所提方法的有效性.

首先将 Iris 数据集中的样本按照不同的规模随机选取为训练集, 其余的样本作为测试集, 然后使用该方法生成测试集的 GBPA, 测试并记录所提方法的分类精度. 对于不同的训练比例, 实验中均采用蒙特卡罗方法重复实验 100 次, 最终可得出 Iris 数据集中 Se、Ve、Vi 和综合数据的平均分类精度, 实验结果如表 7 和图 10 所示. 当训练比例为 6% 时, 平均准确

率仍然达到了 91.48%, 随着训练规模的提升, 测试集整体的分类精度整体呈单调上升趋势, 当训练规模为 90% 时, 平均准确率达到了 94.99%.

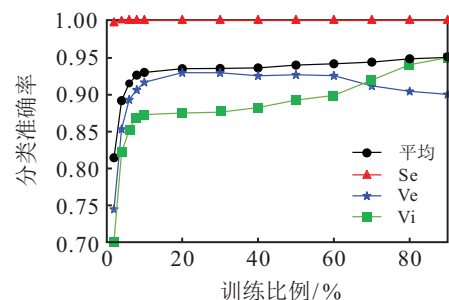


图 10 Iris 数据集的分类精度与训练比例的关系

表7 不同训练规模的分类准确率数据

比例/%	Se	Ve	Vi	平均准确率
2	0.9967	0.7448	0.7016	0.8144
4	0.9999	0.8530	0.8215	0.8915
6	1.0000	0.8928	0.8515	0.9148
8	1.0000	0.9061	0.8718	0.9260
10	1.0000	0.9163	0.8726	0.9296
20	1.0000	0.9293	0.8748	0.9347
30	1.0000	0.9291	0.8759	0.9350
40	1.0000	0.9251	0.8821	0.9357
50	1.0000	0.9264	0.8922	0.9395
60	1.0000	0.9253	0.8986	0.9413
70	1.0000	0.9115	0.9192	0.9436
80	1.0000	0.9041	0.9400	0.9480
90	1.0000	0.9000	0.9496	0.9499

为了更进一步测试所提GBPA生成方法的有效性,接下来使用该方法测试不同测试集在不同训练比例下的分类精度.随机选取Iris数据集N%的样本作为训练集,其中N的变化范围是从2~100.实验中记录了3种测试集的分类精度:1)将其余1-N%的样本作为测试集;2)将训练集作为测试集;3)将整个数据集作为测试集.每次实验均采用蒙特卡罗方法重复实验100次,最终可以得到各测试集的平均分类精度,实验结果如表8和图11所示.

表8 不同测试集在不同训练比例下的分类精度数据

比例/%	测试集	训练集	全部数据
10	0.9301	0.9711	0.9350
20	0.9352	0.9582	0.9401
30	0.9350	0.9514	0.9397
40	0.9369	0.9488	0.9409
50	0.9400	0.9492	0.9434
60	0.9418	0.9507	0.9466
70	0.9444	0.9521	0.9501
80	0.9481	0.9537	0.9521
90	0.9523	0.9548	0.9544
100	—	0.9533	0.9533

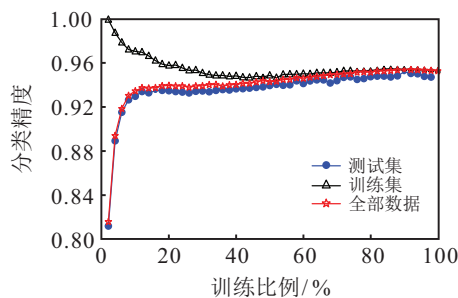


图11 不同测试集的分类精度与训练比例的关系

从图11可以看出,将剩余样本作为测试集和全部数据作为测试集的分类精度均随着训练比例的增加而升高,而将训练集作为测试集的分类精度却呈现出下降的趋势,说明本方法不易出现过拟合的现象,进一步表明了本文提出的GBPA生成方法是有效的.另外,当训练比例为10%时,分类精度达到了93.5%,这表明该方法即使小训练样本的情况下,仍可以达到较高的分类精度.

将所提方法与BPA否定方法^[4]、区间数方法^[10]和广义三角模糊数方法^[14, 20]进行比较.在本实验中,从每个类别中随机抽取的训练样本数为10、15、20、25、30、35、40和45,剩余的样本用作测试集.使用蒙特卡罗方法重复实验100次,并记录实验结果的平均值,结果如表9和图12所示,表9和图12展示了本文所提方法具有较高的分类精度.

表9 不同方法分类准确率与训练比例的关系

数量	文献[4]	文献[10]	文献[14]
10	0.8737	0.9167	0.8193
15	0.8810	0.8895	0.8649
20	0.9022	0.8822	0.8833
25	0.9027	0.8827	0.9109
30	0.9033	0.8967	0.9227
35	0.9267	0.9000	0.9236
40	0.9233	0.9133	0.9333
45	0.9267	0.9067	0.9420

数量	文献[20]-W1	文献[20]-W2	本文方法
10	0.8283	0.7900	0.9352
15	0.8533	0.8190	0.9350
20	0.8756	0.8489	0.9369
25	0.8880	0.8613	0.9400
30	0.9000	0.8867	0.9418
35	0.9111	0.8933	0.9444
40	0.9067	0.9000	0.9481
45	0.8800	0.8933	0.9523

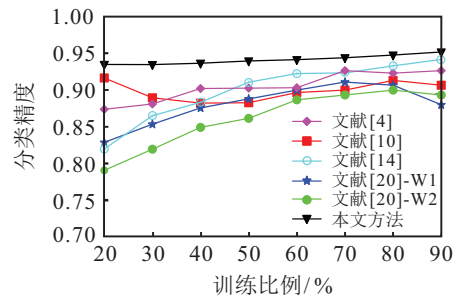


图12 5种BPA生成方法分类精度曲线

为进一步说明所提方法的有效性,仍对Iris数据集进行分类,将本文方法与5种经典分类器的分类精度进行比较,分别是径向基函数支持向量机(SVM-RBF)、决策树学习器(REPTree)、朴素贝叶斯(NB)、1最近邻分类器(1NN)以及Adaboost算法.分类精度数据如表10所示,结果表明,本文方法与经典分类器的分类精度相当,是一种有效的分类方法.

表10 不同分类方法在Iris数据集中的分类精度 %

数据集	SVM-RBF	REPTree	NB	1NN	Adaboost	本文
Iris	92.7	94.7	96.0	94.0	93.6	94.8

4 结论

受时代环境影响,人的认知存在局限性,而新事物总会随着社会的发展不断涌现,识别框架中很可能

会存在未知的命题,因此识别框架未必是完备的. 广义证据理论在处理这一类问题时更具优势,它能够在识别框架不完备的情况下融合证据. 本文基于GET提出了一种新的广义基本概率赋值生成方法. 该方法是一种数据驱动方法,它首先根据训练样本为每种属性构造一个广义三角模糊数模型,然后根据训练样本的平均标准差生成测试样本的GTFN,最后计算测试样本与模型中各命题之间的GTFN距离生成GBPA. 该方法能准确地描述测试样本的特征,并合理地现实情况下进行分类. 由Iris数据集的分类实验结果可知,所提方法是有效的,即使在训练样本不足的情况下,仍然可以达到较高的分类精度.

参考文献(References)

- [1] Denux T, Shenoy P P. An interval-valued utility theory for decision making with Dempster-Shafer belief functions[J]. *International Journal of Approximate Reasoning*, 2020, 124: 194-216.
- [2] Pan Y, Zhang L M, Wu X G, et al. Multi-classifier information fusion in risk analysis[J]. *Information Fusion*, 2020, 60: 121-136.
- [3] Boukezzoula R, Coquin D, Nguyen T L, et al. Multi-sensor information fusion: Combination of fuzzy systems and evidence theory approaches in color recognition for the NAO humanoid robot[J]. *Robotics and Autonomous Systems*, 2018, 100: 302-316.
- [4] Wu D D, Liu Z J, Tang Y C. A new classification method based on the negation of a basic probability assignment in the evidence theory[J]. *Engineering Applications of Artificial Intelligence*, 2020, 96: 103985.
- [5] 刘小峰, 舒仁杰, 柏林, 等. 小样本下多稀疏表示分类器的决策融合方法[J]. *控制与决策*, 2021, 36(8): 1984-1990.
(Liu X F, Shu R J, BO L, et al. Decision fusion of multiple sparse representation-based classifiers in case of small samples[J]. *Control and Decision*, 2021, 36(8): 1984-1990.)
- [6] Ghosh N, Paul R, Maity S, et al. Fault Matters: Sensor data fusion for detection of faults using Dempster-Shafer theory of evidence in IoT-based applications[J]. *Expert Systems with Applications*, 2020, 162: 113887.
- [7] 毛艺帆, 张多林, 王路. 基于重合度的证据冲突度量方法[J]. *控制与决策*, 2017, 32(2): 293-298.
(Mao Y F, Zhang D L, Wang L. Measurement of evidence conflict based on overlapping degree[J]. *Control and Decision*, 2017, 32(2): 293-298.)
- [8] 王欣, 付威. 基于推土机距离的证据冲突强度度量方法[J]. *通信学报*, 2022, 43(5): 204-213.
(Wang X, Fu W. Strong measurement method of evidence conflict based on earth mover's distance[J]. *Journal on Communications*, 2022, 43(5): 204-213.)
- [9] Fu W, Yu S, Wang X. A novel method to determine basic probability assignment based on adaboost and its application in classification[J]. *Entropy: Basel*, Switzerland, 2021, 23(7): 812.
- [10] 康兵义, 李娅, 邓勇, 等. 基于区间数的基本概率指派生成方法及应用[J]. *电子学报*, 2012, 40(6): 1092-1096.
(Kang B Y, Li Y, Deng Y, et al. Determination of basic probability assignment based on interval numbers and its application[J]. *Acta Electronica Sinica*, 2012, 40(6): 1092-1096.)
- [11] Xu P D, Deng Y, Su X Y, et al. A new method to determine basic probability assignment from training data[J]. *Knowledge-Based Systems*, 2013, 46: 69-80.
- [12] Chen H F, Wang X. Determination of basic probability assignment based on probability distribution[C]. 2020 39th Chinese Control Conference. Shenyang, 2020: 2941-2945.
- [13] Deng Y. Generalized evidence theory[J]. *Applied Intelligence*, 2015, 43(3): 530-543.
- [14] Zhang J F, Deng Y. A method to determine basic probability assignment in the open world and its application in data fusion and classification[J]. *Applied Intelligence*, 2017, 46(4): 934-951.
- [15] Jiang W, Zhan J, Zhou D Y, et al. A method to determine generalized basic probability assignment in the open world[J]. *Mathematical Problems in Engineering*, 2016, 2016: 1-11.
- [16] Fan Y, Ma T S, Xiao F Y. An improved approach to generate generalized basic probability assignment based on fuzzy sets in the open world and its application in multi-source information fusion[J]. *Applied Intelligence*, 2021, 51(6): 3718-3735.
- [17] Zadeh L A. Fuzzy sets[J]. *Information and Control*, 1965, 8(3): 338-353.
- [18] Wang F. Preference degree of triangular fuzzy numbers and its application to multi-attribute group decision making[J]. *Expert Systems with Applications*, 2021, 178: 114982.
- [19] 陈之宁, 王安, 周存宝. 直觉三角模糊数距离及其在多属性决策上的应用[J]. *重庆理工大学学报: 自然科学*, 2012, 26(3): 130-134.
(Chen Z N, Wang A, Zhou C B. The distance measure among intuitionistic triangular fuzzy numbers and its application on multi-attribute decision-making[J]. *Journal of Chongqing University of Technology: Natural Science*, 2012, 26(3): 130-134.)
- [20] 肖建于, 童敏明, 朱昌杰, 等. 基于广义三角模糊数的基本概率赋值构造方法[J]. *仪器仪表学报*, 2012, 33(2): 429-434.
(Xiao J Y, Tong M M, Zhu C J, et al. Basic probability assignment construction method based on generalized triangular fuzzy number[J]. *Chinese Journal of Scientific Instrument*, 2012, 33(2): 429-434.)

作者简介

付威(1995—),男,硕士生,从事信息融合决策与不确定性推理等研究, E-mail: fuwei_1995@163.com;

王欣(1978—),男,副教授,博士,从事多传感器证据理论决策与状态估计等研究, E-mail: wangxin@hlju.edu.cn.