

DOI: 10.7652/xjtuxb201509014

最短特征线段多分类器系统设计

丁建坤¹, 韩德强¹, 杨艺²

(1. 西安交通大学智能网络与网络安全教育部重点实验室, 710049, 西安;

2. 西安交通大学机械结构强度与振动国家重点实验室, 710049, 西安)

摘要: 为提高多分类器系统分类的性能,设计了一种使用最短特征线段分类器的多分类器系统。依据最短特征线段分类算法工作机理,利用特征线段长度表征样本隶属于各个类别的可能性,即模糊隶属度,对成员分类器输出形式完成由摘要级至度量级的重新建模,更多地保留输出细节以减少信息损失,进而利用基于模糊的证据融合规则实现成员分类器的度量级融合,通过隶属度到 mass 函数的转换,利用模糊-证据融合规则实现多分类器系统的构造,进一步提高了多分类器系统分类性能。采用人工数据集和 UCI 数据集设计了对比实验,实验表明,与其他邻域型分类器构造的多分类器系统相比,新多分类器系统能有效提升分类正确率。

关键词: 特征线段;隶属度;多分类器系统;证据理论

中图分类号: TP391 **文献标志码:** A **文章编号:** 0253-987X(2015)09-0077-07

Design of Multiple Classifier Systems Based on Shortest Feature Line Segment

DING Jiankun¹, HAN Deqiang¹, YANG Yi²

(1. Ministry of Education Key Lab for Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an 710049, China; 2. State Key Laboratory of Strength and Vibration, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: To improve the classification performance of multiple classifier systems, a novel multiple classifier system using shortest feature line segment (SFLS) as member classifiers is proposed. According to the SFLS's algorithmic principle on classification, the length of the shortest feature line segment is used to represent the probability, i. e., the fuzzy membership of the query sample belonging to the corresponding class. Thus, the member classifier's output is transformed from the abstract level to the measurement level. Furthermore, by transforming the fuzzy membership function into the mass function and using fuzzy-based evidential fusion rules, the classification fusion is implemented. Compared with traditional multiple classifier systems, the proposed approach can use more detailed information for implementing more effective decision-level fusion. Experimental results show that the proposed multiple classifier systems can effectively improve classification accuracy.

Keywords: feature line segment; membership function; multiple classifier systems; evidence theory

收稿日期: 2015-01-28。 作者简介: 丁建坤(1990—),男,博士生;韩德强(通信作者),男,副教授,博士生导师。 基金项目: 国家“973 计划”资助项目(2013CB329405);国家自然科学基金资助项目(61203222);教育部博士点基金资助项目(20120201120 036);中央高校基本科研业务费专项资金资助项目(xjj2012104, xjj2014122)。

网络出版时间: 2015-07-10 网络出版地址: <http://www.cnki.net/kcms/detail/61.1069.T.20150710.1032.001.html>

模式识别理论与方法的研究近年来取得了诸多进展^[1],然而在复杂环境下,单一分类器效果往往不甚理想。多分类器系统是应对复杂环境下模式识别问题的重要手段,引发了广泛的关注。目前,多分类器融合技术已广泛应用于图像识别、故障诊断、生物认证、医疗诊断等诸多领域^[2-5]。

学者们从多分类器系统的不同层面开展了研究。在多个成员分类器生成方面,设计实现了基于不同训练样本^[6]、基于不同特征子空间以及选取不同的分类算法^[7]等多种形式的融合方法。1992年,Kira等提出了Relief算法^[8],从特征评价的角度分析了成员分类器的性能及选择。在此基础上,Marko等提出了ReliefF、RReliefF等改进算法^[9]。在多分类器融合有效性方面,Dietterich最早指出,并非多个分类器的融合一定能够获取更优的分类性能,不同成员分类器间的差异性是关键所在^[10]。梁绍一等对现有的差异性度量方法进行了总结,并提出了基于几何关系的差异性度量方法^[11]。多分类器融合规则方面,Xu等在1992年针对成员分类器的输出形式进行了分析研究^[12],并总结了一系列的分类器融合方法,如投票法、Bayes法以及DS证据理论方法等。

成员分类器(个体分类器)的设计问题是多分类器系统设计的前提,其输出形式也决定了后续融合规则的选取^[12]。邻域型分类器是一个很重要的分支,最近邻(NN)分类器^[13]是一种最具代表性的简单而有效的邻域型分类算法,但是在训练样本数较小的情况下性能会受限。基于特征线(通过来自同一类的两个训练样本点的直线)的分类器,如最近特征线(NFL)^[14]是NN的一种改进算法,以测试样本点到特征线的距离为决策依据,但在扩充了训练样本空间的同时,特征线的内、外穿越情况给NFL带来了错分类问题^[15]。针对NFL存在的穿越缺陷,出现了许多特征线的改进方法,如最短特征线段(SFLS)^[15]、核最近特征线(KNFL)^[16]等方法。

本文选用SFLS作为成员分类器分类算法。SFLS作为NFL的改进算法,用特征线段代替特征线作为决策依据,可避免内、外穿越问题,且具备诸多优良的分类特性^[15]。将SFLS用于多分类器系统,有望进一步提升分类性能,然而SFLS仅输出类别标签,仅利用成员分类器输出的类别标签进行表决(投票)融合,会造成信息损失。本文利用SFLS工作机理,以模糊隶属度对输出建模更多地保留输出细节以减少信息损失,进而利用基于模糊的证据

融合规则实现成员分类器的度量级融合,提升了分类精度。实验结果验证了所提方法的合理性和有效性。

1 多分类器系统概述

多分类器系统因其能有效提高分类性能,在模式识别领域受到越来越多的关注并得到了广泛应用。多分类器系统的原理框图如图1所示。

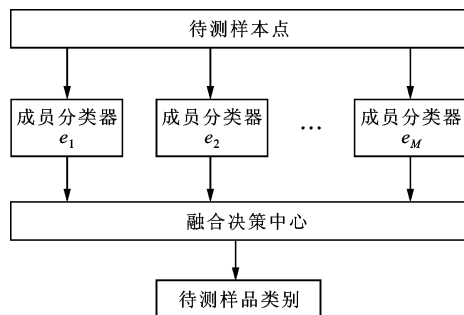


图1 多分类器系统框图

图1中 $e_i(i=1,2,\dots,M)$ 代表各成员分类器。通过选择合适的融合规则,在融合决策中心利用各成员分类器的输出结果进行融合分类。成员分类器输出结果主要有3种形式^[12]:摘要级(也称抽象级),即输出仅为类别标签;排序级,即输出属于各类别可能性大小的顺序;度量级,即输出一系列度量值,代表样本点属于各类别的可能性。

对不同的成员分类器输出形式采用不同的融合方式,对于上述3种输出形式,均可采用投票融合。对排序级输出可采用排序融合方式^[17],对于度量级的输出形式可以利用DS证据推理^[12]、模糊推理^[5,7]等方法进行融合。在多分类器系统中,成员分类器的分类性能、输出形式以及成员分类器之间的差异性都会影响到多分类器系统的性能^[10-12]。本文的工作主要关注成员分类器设计、融合规则选取及对多分类器系统性能的影响。

2 基于最短特征线段的多分类器系统

2.1 最短特征线段分类器

最短特征线段(SFLS)分类器是一种以线段特征替代点特征的邻域型分类器。设 $\mathbf{x}_i^o$ 和 $\mathbf{x}_j^o$ 均为来自训练样本中 $\omega \in \Omega, \Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ 类的两个样本点,其中 n 为类别个数,每个样本的特征维数为 D 。假设 ω 类共有 N_ω 个训练样本,则由该类训练样本可以构成 $N_\omega(N_\omega - 1)/2$ 条特征线段。对于任意一个待分类样本点 $\mathbf{x}_q$ 执行如下步骤。

(1)计算待测样本点是否被包含在由 x_i^w 和 x_j^w 构成的超球体中。如果该测试样本点与特征线段端点的两个训练样本连线夹角是锐角,表明测试样本点在超球体外,如图2中 x_{q1} 所示;若夹角为钝角表明测试样本点在超球体内,则标记线段 $\overline{x_i^w x_j^w}$ 并记录其长度,如图2中 x_{q2} 所示。

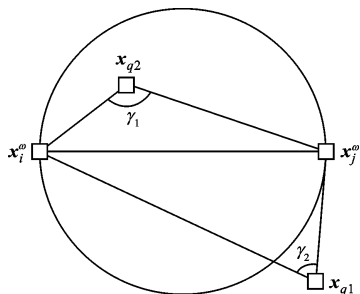


图2 特征线段和测试样本点相对位置关系

测试样本点 x_q 与特征线段端点的两个训练样本 x_i^w 和 x_j^w 的连线夹角 γ 可根据向量夹角公式求得,即

$$\cos \gamma = \frac{(x_q - x_i^w)^T (x_q - x_j^w)}{\|x_q - x_i^w\| \|x_q - x_j^w\|} \quad (1)$$

(2)在被标记的特征线段中,寻找长度最短的特征线段,则此特征线段所属的类别即为该待测样本点被判定的类别,如果没有特征线段被标记则拒绝识别该待测样本点。

SFLS 保留了 NFL 的优点,利用样本点的两两组合代替原始样本点作为判断的依据,充分地利用了样本信息。SFLS 使用的是特征线段,确定了待分类样本点与通过该特征线段的特征直线的距离上界,这样即保持了 NFL 优良的分类特性,又避免了 NFL 的内、外穿越问题^[15]。SFLS 中由特征线段构成的特征空间具有一个对分类器来说很好的性质,即聚合性。但是,SFLS 的输出仅为类别标签,丢掉了许多有用信息,不利于进一步的多分类器融合,本文在对 SFLS 内部工作机理分析研究的基础上,对其输出形式重新建模,以隶属度形式输出,便于多分类器系统做更精细的度量级融合。

2.2 SFLS 分类器输出度量化

首先在 SFLS 中所有被标记的特征线段里分别选出各个类别内的最短特征线段,其长度体现了待测样本隶属于相应类别的可能性大小(长度越长,可能性越低)。利用这些类内最短特征线段长度,可生成待测样本点对各个类别的隶属度,具体步骤为:①基于 SFLS 分类器,得到特征线段构成的特征空间中被标记的特征线段的类别 ω 和长度 d_i^w ;②在被标

记的特征线段里,分别找出属于 $\omega_i (i=1,2,\dots,n)$ 的特征线段中的最短特征线段,长度 $d_{\min}^{\omega_i} = \min_i d_i^w$,若 ω_i 类中没有被标记的特征线段,则将 $d_{\min}^{\omega_i}$ 设为 $+\infty$;③根据得到的类内最短特征线段的长度 $d_{\min}^{\omega_i}$,利用隶属度生成函数 $G(\cdot)$ 得到测试样本对类别 ω_i 的隶属度 $\theta(\omega_i)$ 。

特征线段长度越短,则待测样本属于该类别的可能性越大,即特征线段长度 d 的单调递减函数可以作为隶属度生成函数,负指数函数、反比例函数等函数形式比较简单,被广泛使用。当 d 的取值接近于 0 时,反比例函数取值会非常大,趋于无穷,并且在 $d=0$ 时反比例函数无法计算和使用(分母不能为 0),而负指数函数有映射关系:当 $d=0$ 时, $\exp(-d)=1$;当 $d \rightarrow +\infty$ 时, $\exp(-d) \rightarrow 0^+$ 。相对而言,负指数函数形式更适于生成隶属度,且鲁棒性更强,因此本文选择负指数函数作为隶属度生成函数,即待测样本点 x_q 对 ω_i 类的隶属度为

$$\theta(\omega_i) = G(d_{\min}^{\omega_i}) = \exp(-\lambda d_{\min}^{\omega_i}) \quad (2)$$

式中: $\lambda (\lambda > 0)$ 为可调参数,用于调节特征线段长度对隶属度大小的影响程度。 λ 的大小对最终融合结果影响较小,依经验设置为 $\lambda=1$ 即可。若 ω_i 类中没有被标记的特征线段,即 $d_{\min}^{\omega_i}$ 为 $+\infty$,将测试样本属于该类别的隶属度直接设置为 0,这样就完成了 SFLS 的输出度量化。由此,每一个成员分类器得到一组待测样本点,属于各个类别的隶属度,然后对隶属度进行融合。在融合决策过程中,可采用多种融合方法,证据理论是一种重要的不确定性信息融合方法,相比传统的模糊逻辑,证据理论在融合方面更有优势^[18]。本文的工作中将基于证据理论来实现模糊隶属度的融合。

2.3 模糊-证据融合方法

2.3.1 证据理论简介 设集合 Ω 为辨识框架,若 $m: 2^\Omega \rightarrow [0,1]$ 满足: $\sum_{A \subseteq \Omega} m(A) = 1, m(\emptyset) = 0$, 则称为辨识框架上的基本信度赋值(BBA),其中 2^Ω 表示 Ω 的幂集。若 $A \subseteq \Omega, m(A) > 0$, 则 A 被称为焦点^[19]。

对于辨识框架 Ω 中的某个命题 A ,计算信任函数 $\text{Bel}(A)$ 和似真函数 $\text{Pl}(A)$

$$\text{Bel}(A) = \sum_{B \subseteq A} m(B) \quad (3)$$

$$\text{Pl}(A) = \sum_{B \cap A \neq \emptyset} m(B) \quad (4)$$

区间 $[\text{Bel}(A), \text{Pl}(A)]$ 可用于表示 A 的不确定性程度。 $\forall A \subseteq \Omega, A \neq \emptyset, \Omega$ 上的两个 mass 函数 m_1, m_2

的 Dempster 组合规则为

$$m_1 \oplus m_2(A) = \frac{1}{1-K} \sum_{B \cap C=A} m_1(B)m_2(C) \quad (5)$$

式中: $K = \sum_{B \cap C=\emptyset} m_1(B)m_2(C)$ 称为冲突项。多个证据组合时, Dempster 组合规则满足交换律和结合律。

2.3.2 由隶属度至 mass 函数的生成方法 应用证据理论进行融合决策, 首先要生成 mass 函数。本文中 SFLS 的量化输出形式是隶属度, 因此这里关注由隶属度至 mass 函数的生成方法。

设辨识框架 $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$, $\theta_j(\omega_i)$ 表示第 j 个成员分类器中, 待测样本 x_q 对 ω_i 类的隶属度, 则根据隶属度生成 mass 函数有谨慎有序加权平均证据推理(COWA-ER)方法^[19]、 α -cut 方法^[20]、直接设定焦元组方法。

(1) 平均证据法。首先, 分别以悲观的方式和乐观的方式选出待测样本对各类别的隶属度, 即选出样本对各类别的最小隶属度和最大隶属度, 得到矩阵

$$\mathbf{E} = \begin{bmatrix} \min_j(\theta_j(\omega_1)) & \max_j(\theta_j(\omega_1)) \\ \min_j(\theta_j(\omega_2)) & \max_j(\theta_j(\omega_2)) \\ \vdots & \vdots \\ \min_j(\theta_j(\omega_n)) & \max_j(\theta_j(\omega_n)) \end{bmatrix} \quad (6)$$

其次, 将矩阵 $\mathbf{E}$ 中的每个元素对 E 的最大值归一化得到

$$\mathbf{E}' = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ \vdots & \vdots \\ a_n & b_n \end{bmatrix}$$

$$a_i = \frac{\min_j(\theta_j(\omega_i))}{\max(E)}; \quad b_i = \frac{\max_j(\theta_j(\omega_i))}{\max(E)} \quad (7)$$

最后, 根据 $\mathbf{E}'$ 生成 mass 函数。 $\mathbf{E}'$ 中以悲观方式得到的 a_i 用以表征样本属于类别 ω_i 的信任度, 即 $\text{Bel}(\omega_i) = a_i$; 乐观方式得到的 b_i 用以表征样本属于类别 ω_i 的似真度, 即 $\text{Pl}(\omega_i) = b_i$; 样本属于类别 ω_i 的不确定程度用区间 $[\text{Bel}(\omega_i), \text{Pl}(\omega_i)]$ 的长度表示, 即 $b_i - a_i$ 。由此可以生成 mass 函数, 即

$$\left. \begin{aligned} m_i(\omega_i) &= a_i \\ m_i(\bar{\omega}_i) &= 1 - b_i \\ m_i(\omega_i \cup \bar{\omega}_i) &= m_i(\Omega) = b_i - a_i \end{aligned} \right\} \quad (8)$$

(2) α -cut 方法。对隶属度重新排序, 得到一组顺序 $\{\pi_1, \pi_2, \dots, \pi_n\}$, 满足

$$0 \leq \theta_j(\omega_{\pi_1}) < \theta_j(\omega_{\pi_2}) < \dots < \theta_j(\omega_{\pi_n}) \leq 1$$

设 $B_k (k=1, \dots, n)$ 为转换后的 mass 函数焦元, 则有

$$\left. \begin{aligned} B_k &= \{\omega \in \Omega \mid \theta(\omega_{\pi_i}) \geq \theta(\omega_{\pi_k})\} \\ m_j(B_k) &= \frac{\theta(\omega_{\pi_k}) - \theta(\omega_{\pi_{k-1}})}{\theta(\omega_{\pi_n})} \end{aligned} \right\} \quad (9)$$

(3) 直接设定焦元组方法。首先对隶属度进行归一化, 即

$$P_j(\omega_i) = \theta_j(\omega_i) / \sum_{k=1, \dots, n} \theta_j(\omega_k) \quad (10)$$

将隶属度归一化后, 可按如下两种方式生成 mass 函数: 方式 1^[21] 是仅对单点焦元和全集赋值, β 为折扣系数, 即

$$\left. \begin{aligned} m_j(\omega_i) &= \beta P_j(\omega_i), \quad i=1, 2, \dots, n \\ m_j(\Omega) &= 1 - \sum_i P_j(\omega_i) \end{aligned} \right\} \quad (11)$$

该方法预设了焦元结构为单点焦元以及辨识框架的全集, 将概率值打折扣赋予单点焦元, 剩余部分赋给全集以表征“未知”, 当 $\beta=1$ 时简化为概率形式; 方式 2^[21] 是仅对单点焦元、其补集和全集赋值, 即

$$\left. \begin{aligned} \omega_i &= \{\omega_s \mid P_j(\omega_s) = \max_k P_j(\omega_k)\} \\ m_j(\omega_i) &= \beta P_j(\omega_i) \\ m_j(\bar{\omega}_i) &= \beta \sum_{k \neq i} P_j(\omega_k) \\ m_j(\Omega) &= 1 - \beta \end{aligned} \right\} \quad (12)$$

该方法将 mass 值赋予单点焦元及其相应补命题的焦元, 剩余部分赋予全集。根据以上方法生成 mass 函数之后, 就可以根据 Dempster 规则进行融合, 得到融合后的 mass 函数。

2.3.3 由证据至概率的转换及决策 基于多个证据函数组合之后所获取的新 mass 函数实现决策时, 往往需要借助由证据至概率的转换方法。这里使用经典的 Pignistic 概率转换^[22], 定义为

$$\text{Bet}P(A) = \sum_{B \subseteq \Omega} \frac{|A \cap B|}{|B|} \frac{m(B)}{1 - m(\emptyset)} \quad (13)$$

式中: $|B|$ 为集合 B 中元素个数。利用式(13)将融合后的 mass 函数转换为 Pignistic 概率 $\text{Bet}P(\omega_i) (i=1 \dots n)$, 由此做出最终融合决策, 即

$$P_{\text{dec}} = \text{Bet}P(\omega_k) = \max_i (\text{Bet}P(\omega_i))$$

待测样本 x_q : $\begin{cases} x_q \in \omega_k & P_{\text{dec}} > \epsilon \\ \text{拒绝识别} & P_{\text{dec}} \leq \epsilon \end{cases}$, 其中 P_{dec} 为决策概率, ϵ 为决策门限。

3 实验比较和分析

为了验证最短特征线段多分类器系统的有效性, 本文基于人工数据集和 UCI 数据集设计了对比实验。实验方法: 针对不同数据集, 将实验样本平均分成 6 份, 任意选取其中 3 份作为训练样本, 其他数

据做测试样本,每组实验包括 20 个实验。实验重复做 10 组,取平均正确率。

最短特征线段多分类器系统中隶属度生成函数式(2)的参数 λ 依经验设置为 $\lambda=1$ 。在直接设定焦元组方法的方式 1 和方式 2 中,依经验设置 $\beta=0.8$ 。实验将对构造多分类器系统时采用不同成员分类器分类算法、不同 mass 函数生成方法所获取的结果进行分析比较。采用不同 mass 函数生成方法构造的多分类器系统及其在实验中的名称对应关系如表 1 所示。

表 1 本文方法名称与 mass 函数生成方法对应表

本文方法名称	mass 函数生成方法
MCS-SFLS1	COWA-ER 方法
MCS-SFLS2	α -cut 方法
MCS-SFLS3-a	直接设定焦元组方法(方式 1)
MCS-SFLS3-b	直接设定焦元组方法(方式 2)

用 MCS-SFLS-Ave 表示直接对隶属度求平均的融合方式,即把各成员分类器得到的隶属度的平均值作为融合隶属度,取隶属度最大的类别作为决策类别。分别用 MCS-SFLS、MCS-3-NN、MCS-KNFL、MCS-NFL 表示以相同的特征子空间抽取方式,分别选用 SFLS、3-NN、KNFL、NFL 等方法作为成员分类器分类算法生成各个成员分类器,以等权值投票方式进行融合构成多分类器系统,与新多分类器系统作对比,对数据每一维特征 F_i 做归一化预处理,即

$$F'_i = \frac{F_i - \min(F_i)}{\max(F_i) - \min(F_i)}$$

(14)

3.1 人工数据集

人工数据集包含 3 个类别的样本,每个样本有 6 维互相独立的特征,每一维特征均服从高斯分布。特征的均值 μ 设置如表 2 所示,方差 σ^2 均设为 0.2。

表 2 各类别样本不同特征的均值设置

类别	μ					
	特征 1	特征 2	特征 3	特征 4	特征 5	特征 6
1	0.6	0.3	0.6	0.7	0.4	0.4
2	0.3	0.8	0.4	0.5	0.5	0.3
3	0.8	0.5	0.3	0.4	0.7	0.6

实验分别选取特征 1、2,特征 3、4,特征 5、6 作为特征子空间,生成 3 个成员分类器构成多分类器系统。不同特征子空间上对不同类别的区分程度不同,例如由特征 1、特征 2 构成的特征子空间对类别 2 的区分程度更高,如图 3 所示。

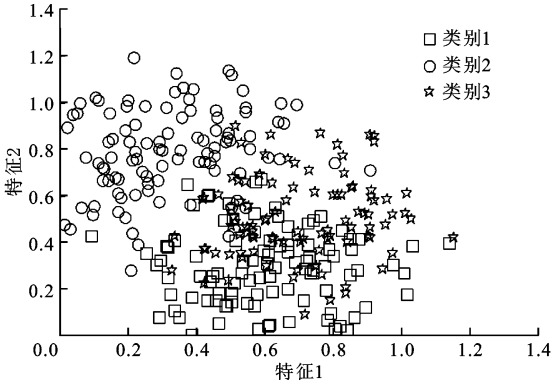


图 3 样本在特征 1、特征 2 子空间下的分布图

对多分类器系统在不同决策门限下进行测试,其中 MCS-SFLS-Ave 的融合隶属度依式(10)归一化后得到决策概率,表决融合法通过除以总票数得到决策概率,结果如表 3 所示。可以看出,在此数据集上,以模糊及模糊-证据融合规则构造的多分类器系统分类性能均优于表决融合的多分类器系统。

表 3 人工数据集下各多分类器系统分类正确率对比

多分类器系统名称	正确率/%		
	$\epsilon=0$	$\epsilon=0.4$	$\epsilon=0.5$
MCS-3-NN	73.42	73.43	73.43
MCS-KNFL	69.29	69.29	69.29
MCS-NFL	45.34	45.34	45.34
MCS-SFLS	70.31	70.31	70.31
MCS-SFLS-Ave	89.28	60.84	22.37
MCS-SFLS1	87.51	84.15	83.14
MCS-SFLS2	85.54	80.53	78.01
MCS-SFLS3-a	86.22	79.46	56.51
MCS-SFLS3-b	83.55	82.45	82.16

同时,从表 3 中可以看出,在决策门限 $\epsilon=0$ 时 MCS-SFLS1 和 MCS-SFLS-Ave 的分类性能较为突出,但是随着决策门限增大,MCS-SFLS-Ave 的分类正确率下降很快,而 MCS-SFLS1 和 MCS-SFLS2 下降并不明显。这表明在高决策门限情况下,基于模糊-证据融合规则的融合方法决策正确率高于简单隶属度平均融合方法。因而,基于模糊-证据融合规则的融合结果可靠性更高。在本实验中成员分类器较少,决策门限对投票法融合结果无实质影响。

3.2 UCI 数据集

实验使用到的数据集信息如表 4 所示。

通过选用不同的特征子空间来生成成员分类器。若样本有 D 维特征,则可构成 2^D-1 个特征子空间,以此可以得到 2^D-1 个成员分类器,从中等概率地选择 5 个成员分类器构成多分类器系统。决策

门限均设为 0.4(对二分类问题没有影响)。实验结果如表 5 所示。

表 4 实验用 UCI 数据集信息

数据集名称	特征维度	类别数目
Pima	8	2
Wdbc	30	2
Magic04	10	2
Seed	7	3
Waveform2	21	3

表 5 不同 UCI 数据集下多分类器系统分类正确率对比

多分类器系统名称	正确率/%				
	Pima	Wdbc	Magic04	Seed	Waveform2
MCS-3-NN	69.58	93.53	70.44	88.56	65.42
MCS-KNFL	67.38	92.94	69.44	87.87	63.11
MCS-NFL	66.55	91.88	68.83	81.13	61.91
MCS-SFLS	69.22	93.86	73.50	87.52	67.83
MCS-SFLS-Ave	70.41	94.36	74.89	90.89	76.73
MCS-SFLS1	70.99	95.04	75.22	91.33	77.63
MCS-SFLS2	70.22	92.28	73.33	87.66	76.34
MCS-SFLS3-a	70.23	92.42	73.39	87.82	76.77
MCS-SFLS3-b	70.23	92.09	73.33	87.66	75.26

从实验结果可以看出,最短特征线段多分类器系统的分类性能较为理想,尤其在 Seed 和 Waveform2 数据集上,分类正确率明显高于其他多分类器系统,其中基于 COWA-ER 的多分类器系统(MCS-SFLS1)的分类性能明显优于其他多分类器系统。一般意义上,对于特征线分类器来说,类别数目越多,特征维度越低,穿越错分问题越突出^[23]。在三分类问题的 Seed 和 Waveform2 数据集上,类别个数多于二分类,且特征维数属于中低水平,因此基于特征线的方法出现穿越错分问题较多,所以基于特征线段的多分类器系统优势较大。在二分类 Wdbc 数据集上,由于类别数目较少,且特征维度较高(30 维),穿越错分问题并不突出,基于特征线段的多分类器系统优势相对减弱。

总体而言,本文提出的最短特征线段多分类器系统在继承了 SFLS 分类器良好分类性能的同时,利用模糊-证据理论方法进行融合,进一步提高了多分类器系统的分类性能。

4 结 论

本文设计了利用 SFLS 作为成员分类器,以模

糊-证据融合规则构造的多分类器系统。将 SFLS 的输出改为度量级形式,进而使用证据理论实现模糊融合分类决策。实验表明,本文方法能更加充分地利用特征线段中包含的信息,相比其他基于邻域型分类器的多分类器系统而言,进一步提高了分类正确率,并且基于模糊-证据融合的识别结果比基于模糊隶属度平均方法的可靠性更高,而在各种模糊-证据融合方法中,COWA-ER 性能更优。

本文研究的重点在于成员分类器个体的设计以及融合规则的设计和选取,但对于多分类器系统来说,成员分类器之间的差异性^[19]更为关键。

参考文献:

[1] 李晶皎,赵丽红,王爱侠. 模式识别 [M]. 北京: 电子工业出版社, 2010.

[2] 杨艺, 韩崇昭, 韩德强. 利用特征子空间评价与多分类器融合的高光谱图像分类 [J]. 西安交通大学学报, 2010, 44(8): 20-24.

YANG Yi, HAN Chongzhao, HAN Deqiang. Hyperspectral image classification based on feature subspace evaluation and multiple classifier fusion [J]. Journal of Xi'an Jiaotong University, 2010, 44(8): 20-24.

[3] 张成军, 阴妍, 鲍久圣, 等. 多源信息融合故障诊断方法研究进展 [J]. 河北科技大学学报, 2014, 35(3): 213-221.

ZHANG Chengjun, YIN Yan, BAO Jiusheng, et al. Research progress in fault diagnosis methods based on multi-source information fusion [J]. Journal of Hebei University of Science and Technology, 2014, 35(3): 213-221.

[4] 王风华, 韩九强, 姚向华. 一种基于虹膜和人脸的多生物特征融合方法 [J]. 西安交通大学学报, 2008, 42(2): 133-137.

WANG Fenghua, HAN Jiuqiang, YAO Xianghua. Multimodal biometric fusion approach based on iris and face [J]. Journal of Xi'an Jiaotong University, 2008, 42(2): 133-137.

[5] 张彩坡. 模糊积分及多分类器融合在医疗诊断中的应用 [D]. 天津: 天津理工大学, 2010.

[6] BREIMAN L. Bagging predictors [J]. Machine Learning, 1996, 24(2): 123-140.

[7] 焦李成, 公茂果, 王爽, 等. 自然计算、机器学习与图像理解前沿 [M]. 西安: 西安电子科技大学出版社, 2008.

[8] KIRA K, RENDELL L. A practical approach to feature selection [C]// Proceedings of the 9th Machine Learning. Aberdeen, UK: Morgan Kaufmann Pub-

- lishers, 1992: 249-256.
- [9] MARKO R, IGOR K. Theoretical and empirical analysis of ReliefF and RReliefF [J]. Machine Learning, 2003, 53(1/2): 23-69.
- [10] DIETTERICH T G. An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting and randomization [J]. Machine Learning, 2000, 40(2): 139-157.
- [11] 梁绍一, 韩德强, 韩崇昭. 一种基于几何关系的多分类器差异性度量及其在多分类器系统构造中的应用 [J]. 自动化学报, 2014, 40(3): 449-458.
- LIANG Shaoyi, HAN Deqiang, HAN Chongzhao. A novel diversity measure based on geometric relationship and its application to design of multiple classifier systems [J]. Acta Automatica Sinica, 2014, 40(3): 449-458.
- [12] XU Lei, KRZYSAK A, SUEN C Y. Methods of combining multiple classifiers and their applications to handwriting recognition [J]. IEEE Transactions on Systems, Man and Cybernetics, 1992, 22(3): 418-435.
- [13] COVER T, HART P. Nearest neighbor pattern classification [J]. IEEE Transactions on Information Theory, 1967, 13(1): 21-27.
- [14] DU Hao, CHEN Yanqiu. Rectified nearest feature line segment for pattern classification [J]. Pattern Recognition, 2007, 40(5): 1486-1497.
- [15] HAN Deqiang, HAN Chongzhao, YANG Yi. A novel classifier based on shortest feature line segment [J]. Pattern Recognition Letters, 2011, 32(3): 485-493.
- [16] HE Yunhui. Face recognition using kernel nearest feature classifiers [C]//2006 International Conference on Computational Intelligence and Security. Piscataway, NJ, USA: IEEE, 2006: 678-683.
- [17] 杨艺, 韩德强, 韩崇昭. 基于排序融合的特征选择 [J]. 控制与决策, 2011, 26(3): 397-401.
- YANG Yi, HAN Deqiang, HAN Chongzhao. Study on feature selection based on rank-level fusion [J]. Control and Decision, 2011, 26(3): 397-401.
- [18] TACNET J M, DEZERT J. Cautious OWA and evidential reasoning for decision making under uncertainty [C]//Proceeding of the 14th International Conference on Information Fusion. Piscataway, NJ, USA: IEEE, 2011: 1-8.
- [19] SHAFER G. A mathematical theory of evidence [M]. Princeton, USA: Princeton University Press, 1967.
- [20] HAN Deqiang, DEZERT J, TACNET J M, et al. A fuzzy-cautious OWA approach with evidential reasoning [C]//Proceedings of the 15th International Conference on Information Fusion. Piscataway, NJ, USA: IEEE, 2012: 278-285.
- [21] VALENTE F, HERNANSKY H. Combination of acoustic classifiers based on Dempster-Shafer theory of evidence [C] // IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway, NJ, USA: IEEE, 2007: 1129-1132.
- [22] SMETS P. Data fusion in the transferable belief model [C]//Proceedings of the 3rd International Conference on Information Fusion. Piscataway, NJ, USA: IEEE, 2000: 21-33.
- [23] DU Hao, CHEN Yanqiu. Rectified nearest feature line segment for pattern classification [J]. Pattern Recognition, 2007, 40(5): 1486-1497.

(编辑 赵炜)

(上接第46页)

- pressures and temperatures [J]. International Journal of Hydrogen Energy, 2008, 33(23): 7274-7285.
- [14] HUANG Z H, ZHANG Y, ZENG K, et al. Measurements of laminar burning velocities for natural gas-hydrogen-air mixtures [J]. Combustion and Flame, 2006, 146(1/2): 302-311.
- [15] 蒋德明. 内燃机燃烧与排放学 [M]. 西安: 西安交通大学出版社, 2001.
- [16] GIBAUD C, SNYDER J A, SICK V, et al. Laser-induced fluorescence measurements and modeling of absolute CH concentrations in strained laminar methane/air diffusion flames [J]. Proceedings of the Combustion Institute, 2005, 30(1): 455-463.
- [17] EL-ASRAG H, LU T, LAW C K, et al. Simulation of soot formation in turbulent premixed flames [J]. Combustion and Flame, 2007, 150(1/2): 108-126.
- [18] KEE R J, GRGAR J F, SMOOKE M D, et al. A Fortran program for modeling steady laminar one-dimensional pre-mixed flames, SNAD85-8240 [R]. Albuquerque, USA: Sandia National Laboratory, 1985.
- [19] THIERRY P, DENIS V. Theoretical and numerical combustion [M]. Philadelphia, USA: R. T. Edwards, Inc., 2005.
- [20] ZHANG B, XIU G L, BAI C H. Explosion characteristics of argon/nitrogen diluted natural gas-air mixtures [J]. Fuel, 2014, 124(5): 125-132.

(编辑 荆树蓉)