

用于不平衡数据分类的 0 阶 TSK 型模糊系统

顾晓清^{1,2} 蒋亦樟¹ 王士同¹

摘要 处理不平衡数据分类时,传统模糊系统对少数类样本识别率较低.针对这一问题,首先,在前件参数学习上,提出了竞争贝叶斯模糊聚类(Bayesian fuzzy clustering based on competitive learning, BFCCL)算法, BFCCL 算法考虑不同类别样本聚类中心间的排斥作用,采用交替迭代的执行方式并通过马尔科夫蒙特卡洛方法获得模型参数最优解.其次,在后件参数学习上,基于大间隔的策略并通过参数调节使得少数类到分类面的距离大于多数类到分类面的距离,该方法能有效纠正分类面的偏移.基于上述思想以 0 阶 TSK 型模糊系统为具体研究对象构造了适用于不平衡数据分类问题的 0 阶 TSK 型模糊系统(0-TSK-IDC).人工和真实医学数据集实验结果表明,0-TSK-IDC 在不平衡数据分类问题中对少数类和多数类均具有较高的识别率,且具有良好的鲁棒性和可解释性.

关键词 不平衡数据,分类,马尔科夫蒙特卡洛, Takagi-Sugeno-Kang 型模糊系统

引用格式 顾晓清,蒋亦樟,王士同.用于不平衡数据分类的 0 阶 TSK 型模糊系统.自动化学报,2017,43(10):1773-1788

DOI 10.16383/j.aas.2017.c160200

Zero-order TSK-type Fuzzy System for Imbalanced Data Classification

GU Xiao-Qing^{1,2} JIANG Yi-Zhang¹ WANG Shi-Tong¹

Abstract When learning from imbalanced datasets, the traditional fuzzy systems have a low rate of identification over the minority class. Firstly, in the antecedent parameter learning stage, a new clustering method, called Bayesian fuzzy clustering based on competitive learning (BFCCL), is proposed to partition the input space for the antecedents of if-then rules. BFCCL considers the repulsed force of clustering prototypes between different classes, and uses an alternating iterative strategy to obtain the optimal model parameters by Markov chain Monte Carlo method. Secondly, in the consequent parameter learning stage, based on the maximum separation strategy and by keeping the distance between the minority class and the classification hyperplane larger than the distance between the majority class and the hyperplane, the method can effectively correct the skewness of the classification hyperplane. Based on the above ideas, a zero-order-Takagi-Sugeno-Kang fuzzy system for imbalanced data classification (0-TSK-IDC) is proposed. Experimental results on artificial and real-world medicine datasets illustrate the effectiveness of 0-TSK-IDC on both minority and majority classes in imbalanced data classification, as well as its good robustness and interpretability.

Key words Imbalanced data, classifying, Markov chain Monte Carlo, Takagi-Sugeno-Kang type fuzzy system

Citation Gu Xiao-Qing, Jiang Yi-Zhang, Wang Shi-Tong. Zero-order TSK-type fuzzy system for imbalanced data classification. *Acta Automatica Sinica*, 2017, 43(10): 1773-1788

模糊系统是对处理生产和实践过程中的思维、分析、推理与决策等过程构建的一种数学模型,能够将自然语言直接转译成计算机语言.由于具备不确定和模糊信息的处理能力,并具有高度的可解释

性和强大的学习能力,模糊系统在分类问题上受到广泛关注,应用领域有信号处理,医疗诊断等^[1-7]方面.模糊系统的参数学习一般可由专家经验人为赋值或基于相关数据的学习来获得,但很多情况下专家经验并不存在或不完备,而后一种方法因其强大的学习能力在实践中更具可行性.在现实生活中,不平衡数据的分类问题应用广泛,例如,医疗诊断应用中,绝大部分对象都是正常人群,只有很少一部分是疾病患者;入侵检测和钓鱼网站识别应用中,异常样本通常只占所有样本非常小的比例.然而,在许多实际应用中,与多数类样本相比,少数类样本的有效识别更具有意义,也往往是研究者关注的重点对象.

确定所需规则数和规则空间的划分以及确定模糊规则的后件参数是模糊系统建模的两大关键技

收稿日期 2016-02-29 录用日期 2016-08-02
Manuscript received February 29, 2016; accepted August 2, 2016

国家自然科学基金(61502058, 61572085, 61572236),江苏省自然科学基金资助(BK20160187),中央高校基本科研业务费专项资金资助项目(JUSRP51614A)资助

Supported by National Natural Science Foundation of China (61502058, 61572085, 61572236), Natural Science Foundation of Jiangsu Province under Grant (BK20160187), and Fundamental Research Funds for the Central Universities (JUSRP51614A)

本文责任编辑 王立威

Recommended by Associate Editor WANG Li-Wei

1. 江南大学数字媒体学院 无锡 214122 2. 常州大学信息科学与工程学院 常州 213164

1. School of Digital Media, Jiangnan University, Wuxi 214122
2. School of Information Science & Engineering, Changzhou University, Changzhou 213164

术^[8]. 对于第一项, 传统模糊系统构建分类器时常采用聚类的方法, 一种是不考虑样本的标签信息, 在整个数据集上进行聚类; 另一种是在每一个类别的样本中独立进行聚类, 然后再将各聚类结果进行整合. 但是, 这两种方法在处理不平衡数据分类问题时存在以下不足: 前者由于没有利用样本的类别标签信息, 往往会因为少数类样本的数量稀少而把少数类样本视为异常点或噪声; 后者仅是简单地将各类别样本割裂开来, 两类样本重叠区域会出现聚类中心间距过小或中心点重叠的现象. 对于模糊规则的后件参数学习, 传统模糊系统一般遵循模型误差最小化的原则, 如文献 [9] 中的递推最小二乘法, 文献 [10] 中的不对称最小二乘法, 这类方法在处理样本容量平衡的分类问题时具有较高精度. 但是在处理不平衡数据的分类问题时, 这类模糊系统往往倾向于追求多数类样本的高识别率来达到整体样本分类误差的最小化, 在这种情况下, 分类面不可避免地会向少数类样本发生偏移, 少数类样本的识别就存在较高的误判率^[11]. 因此, 研究模糊系统在不平衡数据分类上的应用是有必要的和值得关注的.

目前针对不平衡数据的分类问题, 在模糊系统领域一般通过过采样或欠采样技术来调整正负类样本的比例, 如文献 [12–13] 采用过采样技术给模糊规则设置不同权重来提高少数类样本的分类精度; 文献 [14–15] 先定位样本点的分布然后抽取不同类别的代表点实现类别间数据的平衡. 文献 [16] 提出了过采样和欠采样的结合方法来处理不平衡数据的分类问题. 但是这类算法的缺点是会改变样本的原始分布结构, 采取精确复制少数类样本的策略容易造成分类器的过拟合, 而采取欠抽样多类样本的策略容易丢失部分样本信息. 另外, 由于代价敏感学习^[17–18] 关注错分样本的代价, 其相关算法也常用来解决不平衡数据的学习问题.

针对上述模糊系统在不平衡数据分类中前/后件参数学习的不足, 本文提出了一种适用于不平衡数据分类的 0 阶 TSK 型模糊系统 (Zero-Order-Takagi-Sugeno-Kang fuzzy system for imbalanced data classification, 0-TSK-IDC), 能在较好地处理不平衡数据分类的同时, 保证所得规则的可解释性. 鉴于 0 阶 TSK 型模糊系统具有简单性和可解释性等优点^[19], 本文将其作为研究对象. 0-TSK-IDC 在不改变原有样本分布结构的基础上, 从模糊规则的前件参数学习和后件参数学习两个方面进行研究, 首先, 在模糊规则前件参数的学习上, 受文献 [20] 使用惩罚对手的竞争学习来加速聚类收敛性和文献 [21] 防止聚类中心重合而最大化聚类中心间距的启发, 本文认为在样本的类别标签已知的情况下, 不同类别样本的聚类中心在学习过程中会产生

了一种“竞争”关系, 即聚类中心受同类别样本的吸引向该类别的数据密集区域“靠近”, 同时也受到其他类别样本的“排斥”而被推离异类数据. 本文将这一思想融入贝叶斯聚类 (Bayesian fuzzy clustering, BFC)^[22] 模型中, 提出了一种新的竞争贝叶斯模糊聚类 (Bayesian fuzzy clustering based on competitive learning, BFCCL). BFCCL 在聚类过程中考虑样本的结构信息和不同类别聚类中心之间的排斥作用, 采用不同类别样本交替优化的策略, 并通过马尔科夫蒙特卡洛方法实现整个数据集的模糊划分. BFCCL 的优点在于能够在类别不平衡的空间划分中表现出准确性, 有利于后件参数的学习, 同时能有效增强所得模糊规则的可解释性.

其次, 本文设计的模糊规则后件参数学习的策略是在遵循分类面“大间隔”的同时考虑纠正分类面的偏移, 使少数类样本到分类面的距离大于多数类样本到分类面的距离. 该策略在目标函数的设计中同时考虑了结构风险和经验风险, 其训练过程可转化为二次规划问题求解, 保证结果的全局最优解, 从而 0-TSK-IDC 模糊系统具有较高的泛化性和鲁棒性.

本文结构如下: 第 1 节回顾了相关工作, 包括 TSK 型模糊系统和 BFC 算法的相关概念及原理; 在此基础上第 2 节提出了 BFCCL 算法; 第 3 节提出了用于不平衡数据分类的 0-TSK-IDC 模糊系统; 第 4 节通过在人工数据集和 4 个不平衡医学诊断数据集上的实验说明了 BFCCL 及 0-TSK-IDC 的有效性; 第 5 节总结全文.

1 相关工作

1.1 TSK 型模糊系统

经典模糊系统可分为 3 类^[23–24]: Mamdani 型模糊系统、Takagi-Sugeno-Kang (TSK) 型模糊系统和 Generalized 模糊系统. 其中,

TSK 型模糊系统的第 k 个模糊规则的表达形式如下^[25]:

$$R^k : \text{IF } x_1 \text{ is } A_1^k \text{ AND } x_2 \text{ is } A_2^k \text{ AND } \cdots \text{ AND } x_D \text{ is } A_D^k, \\ \text{Then } f_k(\mathbf{x}) = p_0^k + p_1^k x_1 + \cdots + p_d^k x_d, \\ k = 1, \cdots, K \quad (1)$$

其中, TSK 型模糊系统中每一条规则都有与之对应的输入向量 $\mathbf{x} = [x_1, x_2, \cdots, x_D]^T$, 并把输入空间的模糊子集 A^k ($A^k = [A_1^k, A_2^k, \cdots, A_D^k]^T$) 投影到输出空间的模糊集 $f_k(\mathbf{x})$. A_i^k 为输入向量 \mathbf{x} 第 i 维所对应的第 k 条规则的模糊子集, K 是模糊规则数, D 是样本维数. 当模糊规则后件部分为一个实数值,

即 $f_k(\mathbf{x}) = p_0^k$ 时, 表示该模糊系统为 0 阶 TSK 型模糊系统. 由于 0 阶 TSK 型模糊系统输出的简洁性, 以及模糊规则后件参数更具有可解释性, 而受到广泛关注. 0 阶 TSK 型模糊系统的实值输出为

$$y_0 = \sum_{k=1}^K \frac{\mu_k(\mathbf{x})}{\sum_{k'=1}^K \mu_{k'}(\mathbf{x})} f_k(\mathbf{x}) = \sum_{k=1}^K \tilde{\mu}_k(\mathbf{x}) p_0^k \quad (2)$$

其中, $\mu_k(\mathbf{x})$ 和 $\tilde{\mu}_k(\mathbf{x})$ 分别为第 k 条模糊规则对应的隶属度函数和归一化后的模糊隶属度函数. $\mu_k(\mathbf{x})$ 可由 IF 部分中各维对应的隶属度值通过合取操作获得, 即

$$\mu_k(\mathbf{x}) = \prod_{i=1}^D \mu_{A_i^k}(x_i) \quad (3)$$

若引入高斯函数作为隶属度函数, 式 (3) 中 $\mu_{A_i^k}(x_i)$ 可表示为

$$\mu_{A_i^k}(x_i) = \exp\left(\frac{-(x_i - y_i^k)^2}{2\delta_i^k}\right) \quad (4)$$

参数 y_i^k 和 δ_i^k 分别是高斯函数的中心和方差, 称为模糊规则的前件参数. 若采用 FCM (Fuzzy C-means)^[26]、BFC^[22] 等模糊聚类方法求解, $\mathbf{y}^k = [y_1^k, y_2^k, \dots, y_D^k]^T$ 为第 k 个聚类中心, $\boldsymbol{\delta}^k = [\delta_1^k, \delta_2^k, \dots, \delta_D^k]^T$ 中的元素 δ_i^k 则可通过下式计算得到:

$$\delta_i^k = \frac{h \sum_{j=1}^N u_{jk}(x_{ji} - y_i^k)^2}{\sum_{j=1}^N u_{jk}} \quad (5)$$

其中, u_{jk} 为输入向量 $\mathbf{x}_j = [x_{j1}, x_{j2}, \dots, x_{jD}]^T$ 属于第 k 类的模糊隶属度, 尺度参数 h 为一正常数.

传统的 0 阶 TSK 型模糊系统的规则后件参数学习常遵循模型误差最小化原则^[9-10], 其可表示为

$$\min E = \frac{1}{2} \sum_{i=1}^N (y_0 - y_i)^2 \quad (6)$$

其中, y_i 为期望输出. 如在式 (6) 基础上考虑结构风险最小化原则, 规则后件参数 \mathbf{p}_0 的学习可转化成线性回归的二次规划问题. 但其在处理不平衡数据分类时, 往往倾向于对多数类具有较高的识别率, 而对少数类的识别率却很低.

1.2 贝叶斯模糊聚类 (BFC)

贝叶斯模糊聚类 (BFC) 算法^[22] 从概率的角度重新阐释了 FCM 聚类, 并依据贝叶斯推理通过马尔科夫蒙特卡洛 (Markov chain Monte Carlo,

MCMC) 方法^[27] 来估计模型的最大后验概率, 得到隶属度和聚类中心的全局最优解. 对于给定的样本集 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\} \in \mathbf{R}^D$, BFC 模型可表示为:

$$p(\mathbf{X}, \mathbf{U}, \mathbf{Y}) = p(\mathbf{X}|\mathbf{U}, \mathbf{Y})p(\mathbf{U}|\mathbf{Y})p(\mathbf{Y}) \quad (7)$$

其中, $p(\mathbf{X}|\mathbf{U}, \mathbf{Y})$ 称为模糊数据似然 (Fuzzy data likelihood, FDL), $p(\mathbf{U}|\mathbf{Y})$ 称为模糊聚类先验 (Fuzzy clustering prior, FCP), $p(\mathbf{Y})$ 称为聚类中心先验.

样本集 \mathbf{X} 的模糊数据似然 $p(\mathbf{X}|\mathbf{U}, \mathbf{Y})$ 是所有样本的 FDL 的乘积, 即

$$p(\mathbf{X}|\mathbf{U}, \mathbf{Y}) = \prod_{n=1}^N \text{FDL}(\mathbf{x}_n|\mathbf{u}_n, \mathbf{Y}) = \prod_{n=1}^N \frac{1}{Z(\mathbf{u}_n, m, \mathbf{Y})} \prod_{c=1}^C \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu} = \mathbf{y}_c, \boldsymbol{\Lambda} = u_{nc}^m \mathbf{I}) \quad (8)$$

其中, N 是样本数, C 是聚类数, u_{nc} 是样本 \mathbf{x}_n 属于第 c 个聚类的隶属度, m 是模糊指数, \mathbf{y}_c 是第 c 个聚类中心, \mathbf{I} 是 D 维单位矩阵, $Z(\mathbf{u}_n, m, \mathbf{Y})$ 是归一化常数. BFC 算法假设样本 \mathbf{X} 服从正态分布 $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Lambda})$, $\boldsymbol{\mu}$ 为分布的中心, $\boldsymbol{\Lambda}$ 为分布的协方差矩阵. 可见 BFC 中每个样本 \mathbf{x}_i 具有独立的概率分布. 同一个聚类中的样本分布共享一个期望值, 但具有不同的协方差.

类似地, 样本集 \mathbf{X} 的模糊聚类先验 $p(\mathbf{U}|\mathbf{Y})$ 是所有样本的 FCP 的乘积, 即

$$p(\mathbf{U}|\mathbf{Y}) = \prod_{n=1}^N \text{FCP}(\mathbf{u}_n|\mathbf{Y}) = \prod_{n=1}^N Z(\mathbf{u}_n, m, \mathbf{Y}) \left(\prod_{c=1}^C u_{nc}^{-mD/2} \right) \text{Dirichlet}(\mathbf{u}_n|\boldsymbol{\alpha}) \quad (9)$$

$p(\mathbf{U}|\mathbf{Y})$ 由三个因子组成: $F_1 = Z(\mathbf{u}_n, m, \mathbf{Y})$, $F_2 = \prod_{c=1}^C u_{nc}^{-mD/2}$ 和 $F_3 = \text{Dirichlet}(\mathbf{u}_n|\boldsymbol{\alpha})$. 其中, F_1 和 F_2 因子的作用是分别消去 FDL 中的归一化常数 $Z(\mathbf{u}_n, m, \mathbf{Y})$ 和 $\prod_{c=1}^C u_{nc}^{-mD/2}$ 项. F_3 因子为狄利克雷 (Dirichlet) 分布, 形式为

$$\text{Dirichlet}(\mathbf{x}|\boldsymbol{\alpha}) = \frac{\Gamma\left(\sum_{c=1}^C \alpha_c\right)}{\prod_{c=1}^C \Gamma(\alpha_c)} \prod_{c=1}^C x_c^{\alpha_c - 1} \quad (10)$$

其中, $x_c \geq 0$, $c = 1, \dots, C$ 且 $\sum_{c=1}^C x_c = 1$. Dirichlet 分布打破了 FCM 算法中模糊指数 m 必

须大于 1 的限定, BFC 算法中 m 值可以取任意实数. BFC 模型假设每个聚类中心均满足高斯分布, 聚类中心先验 $p(\mathbf{Y})$ 可以写成:

$$p(\mathbf{Y}) = \prod_{c=1}^C \mathcal{N}(\mathbf{y}_c | \boldsymbol{\mu}_y, \sum_y) \quad (11)$$

此时, 高斯分布的均值和协方差分别为所有样本的平均值和协方差, 即 $\boldsymbol{\mu}_y = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n$, $\sum_y = \frac{\sigma}{N} \sum_{n=1}^N (\mathbf{x}_n - \boldsymbol{\mu}_y)(\mathbf{x}_n - \boldsymbol{\mu}_y)^T$, 其中, σ 为一常数. 将式 (8), 式 (9) 和式 (11) 三式相乘可得到 BFC 模型的样本和参数的联合概率分布, 即:

$$p(\mathbf{X}, \mathbf{U}, \mathbf{Y}) = p(\mathbf{X} | \mathbf{U}, \mathbf{Y}) p(\mathbf{U} | \mathbf{Y}) p(\mathbf{Y}) \propto \exp \left\{ -\frac{1}{2} \sum_{n=1}^N \sum_{c=1}^C u_{nc}^m \|\mathbf{x}_n - \mathbf{y}_c\|^2 \right\} \times \left[\prod_{n=1}^N \prod_{c=1}^C u_{nc}^{\alpha_c - 1} \times \exp \left\{ -\frac{1}{2} \sum_{c=1}^C (\mathbf{y}_c - \boldsymbol{\mu}_y)^T \sum_y^{-1} (\mathbf{y}_c - \boldsymbol{\mu}_y) \right\} \right] \quad (12)$$

其中, 符号 “ \propto ” 表示两边仅相差一个不依赖于任何变量的常数因子.

2 竞争贝叶斯模糊聚类 (BFCCL)

2.1 BFCCL 基本思想

为了建立适用于不平衡数据分类的模糊规则前

件参数学习方法, 同时也为了提高所得模糊规则的准确性和可解释性, 本文提出了竞争贝叶斯模糊聚类 (BFCCL). BFCCL 聚类的出发点为: 在聚类过程中, 一类样本会对另一类别的聚类中心产生排斥的影响, 特别是在不同类别样本的重叠区域, 一方面, 聚类中心受到本类别样本的吸引而靠近该区域; 另一方面, 聚类中心受到异类别样本的排斥而远离该区域. 这种排斥关系可表现为不同类别的聚类中心在样本叠加区域的一种竞争学习关系. 当样本重叠区域的重叠密度越大时, 属于不同类别的聚类中心之间的竞争力也就越大. 本文将这思想融入到 BFC 贝叶斯概率模型中, 在样本的类别标签已知的前提下, 采用在两类样本交替循环执行的策略, 两类的聚类中心分别作为已知数据, 参与到另一类样本聚类的过程中, 起到聚类中心竞争学习的作用. 两类聚类中心交替影响, 直至两类聚类中心都处于稳定状态. BFCCL 聚类的构造原理示意图如图 1 所示.

2.2 目标函数的构造和参数学习方法

由图 1 可知, BFCCL 聚类通过两类交替循环的策略实现数据集的最佳模糊划分, 本节以某一类样本为例, 重点阐述 BFCCL 聚类模型的目标函数和参数学习方法. 设给定某一类样本集 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$, 聚类数为 C_1 , 假设另一类样本的聚类中心已知, 为 $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{C_2}]^T$, 聚类数为 C_2 . 基于 BFC 模型的贝叶斯概率模型, 样本 \mathbf{X} 服从正态分布, 并且每个样本具有独立的概率分布, 因此 BFCCL 聚类中样本集 \mathbf{X} 中数据和参数的联合概率模型可表示为式 (13).

$$p(\mathbf{X}, \mathbf{Z}, \mathbf{U}, \mathbf{Y}) = p(\mathbf{X}, \mathbf{Z} | \mathbf{U}, \mathbf{Y}) p(\mathbf{U} | \mathbf{Y}) p(\mathbf{Y}) \propto \exp \left\{ -\frac{1}{2} \left(\sum_{n=1}^N \sum_{c=1}^{C_1} u_{nc}^m \|\mathbf{x}_n - \mathbf{y}_c\|^2 + \sum_{n=1}^N \sum_{c=C_1+1}^{C_1+C_2} u_{nc}^m \|\mathbf{x}_n - \mathbf{z}_c\|^2 \right) \right\} \times \left[\prod_{n=1}^N \prod_{c=1}^{C_1+C_2} u_{nc}^{\alpha_c - 1} \times \exp \left\{ -\frac{1}{2} \sum_{c=1}^{C_1} (\mathbf{y}_c - \boldsymbol{\mu}_y)^T \sum_y^{-1} (\mathbf{y}_c - \boldsymbol{\mu}_y) \right\} \right] \quad (13)$$

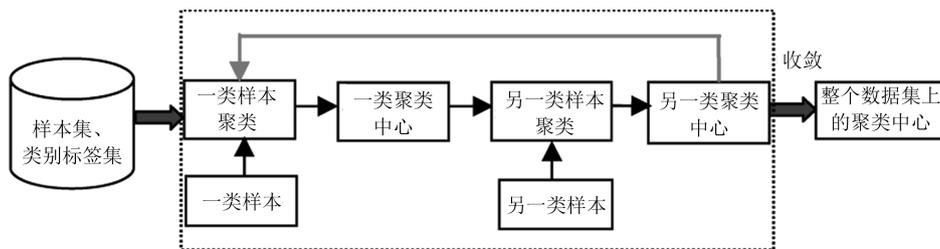


图 1 BFCCL 聚类的构造原理示意图
Fig. 1 The principle of BFCCL clustering

其中, 样本集 \mathbf{X} 和另一类聚类中心矩阵 \mathbf{Z} 为模型的已知数据, 需构建的聚类中心矩阵 $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{c_1}]^T$ 和模糊隶属度矩阵 \mathbf{U} 是模型的参数. 通过对上式取对数, 可得 BFCCL 模型的目标函数:

$$\begin{aligned}
 J(\mathbf{X}, \mathbf{Z}, \mathbf{U}, \mathbf{Y}) = & \sum_{n=1}^N \sum_{c=1}^{C_1} u_{nc}^m \|\mathbf{x}_n - \mathbf{y}_c\|^2 + \\
 & \sum_{n=1}^N \sum_{c=C_1+1}^{C_1+C_2} u_{nc}^m \|\mathbf{x}_n - \mathbf{z}_c\|^2 - \\
 & 2 \sum_{n=1}^N \sum_{c=1}^{C_1+C_2} (\alpha_c - 1) \log(u_{nc}) + \\
 & \sum_{c=1}^{C_1} (\mathbf{y}_c - \boldsymbol{\mu}_y)^T \sum_y^{-1} (\mathbf{y}_c - \boldsymbol{\mu}_y) \quad (14)
 \end{aligned}$$

结合上述联合概率模型 (13) 和目标函数式 (14), 给出如下分析和说明:

1) 不同于常规聚类方法不考虑样本所带的标签信息, BFCCL 模型目标函数中的前 2 项表示出不同类别的聚类中心间的竞争关系. 在另一类样本的聚类中心已知的前提下, 待求聚类中心必然会与这些已知聚类中心对模糊隶属度产生“竞争”关系, 这种竞争关系可理解为不同类别聚类中心间的相互排斥关系; 同时这种竞争关系能保证空间划分的清晰性和增强所得模糊规则的准确性和可解释性.

2) 由图 1 可知, 为了获得整个数据集的最佳空间划分, 需在正负类样本上采取交替循环的策略来求解式 (13) 的最大后验值. 在循环的过程中两类聚

类中心相互影响, 直至两类的聚类中心处于稳定状态.

3) 模糊指数 m 起到调节竞争力强弱的作用, m 的值越大, 聚类中心间的竞争力就越弱; 反之, m 的值越小, 聚类中心间的竞争力就越强.

为了获得最优模糊隶属度矩阵 \mathbf{U} 和聚类中心矩阵 \mathbf{Y} , BFCCL 通过先后两次使用 Metropolis-Hastings (MH) 采样^[27] 构造一条 Markov 链来获得模型参数的最大后验值, 其参数的学习过程如图 2 所示. 从图 2 可知, 该 Markov 链的第 n 次迭代的过程可主要分为 3 步:

步骤 1. 在给定样本和聚类中心的情况下, 使用 Metropolis-Hastings 方法对模糊隶属度 \mathbf{u}_n 进行采样:

a) 由 Dirichlet 分布产生一个新的模糊隶属度值 \mathbf{u}'_n , 并将 Dirichlet 分布的参数 $\boldsymbol{\alpha}$ 设置为 $\boldsymbol{\alpha} = \mathbf{1}_C$, 即

$$\mathbf{u}'_n \sim \text{Dirichlet}(\boldsymbol{\alpha} = \mathbf{1}_C) \quad (15)$$

b) 计算采样值 \mathbf{u}'_n 的接受概率: 条件分布 $p(\mathbf{U}|\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ 在样本和聚类中心已知的情况下与联合分布 $p(\mathbf{X}, \mathbf{Z}, \mathbf{U}, \mathbf{Y})$ 成比例, 同时也与条件分布 $p(\mathbf{X}, \mathbf{U}|\mathbf{Y}, \mathbf{Z})$ 成比例. 因此, 接受概率 $a_{\mathbf{u}}$ 的计算公式可写成以下形式:

$$a_U = \min \left\{ 1, \frac{p(\mathbf{X}_n, \mathbf{u}'_n | \mathbf{Y}, \mathbf{Z})}{p(\mathbf{x}_n, \mathbf{u}_n | \mathbf{Y}, \mathbf{Z})} \right\} \quad (16)$$

其中, \mathbf{u}_n 表示样本 \mathbf{x}_n 当前状态的模糊隶属度, $p(\mathbf{x}_n, \mathbf{u}_n | \mathbf{Y}, \mathbf{Z})$ 的计算式如式 (17).

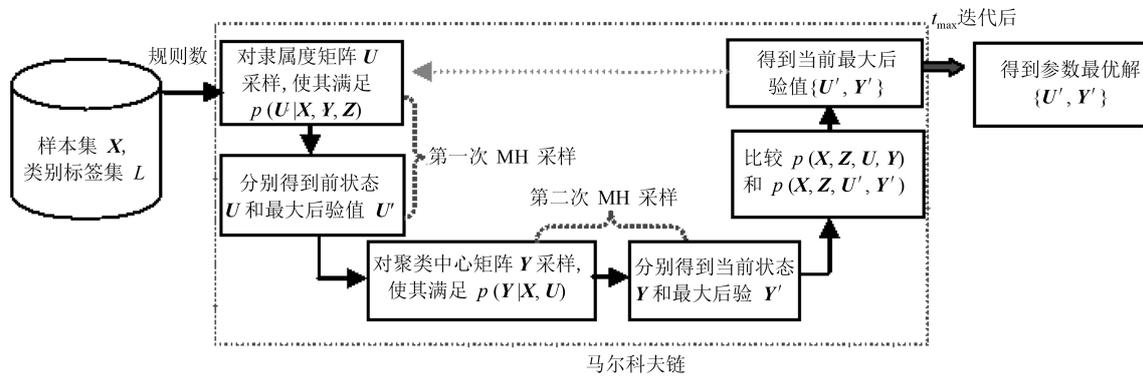


图 2 BFCCL 参数学习示意图
Fig. 2 The parameter learning strategy of BFCCL

$$\begin{aligned}
 p(\mathbf{x}_n, \mathbf{u}_n | \mathbf{Y}, \mathbf{Z}) = & p(\mathbf{x}_n | \mathbf{u}_n, \mathbf{Y}, \mathbf{Z}) p(\mathbf{u}_n | \mathbf{Y}, \mathbf{Z}) = \\
 & \prod_{c=1}^{C_1} \mathcal{N}(\mathbf{x}_n | \mathbf{y}_c, u_{nc}^m) u_{nc}^{-m \frac{D}{2}} \prod_{c=C_1+1}^{C_1+C_2} \mathcal{N}(\mathbf{x}_n | \mathbf{z}_c, u_{nc}^m) u_{nc}^{-m \frac{D}{2}} \text{Dirichlet}(\mathbf{u}_n | \boldsymbol{\alpha}) \propto
 \end{aligned}$$

$$\prod_{c=1}^{C_1} \exp\left(-\frac{1}{2}u_{nc}^m \|\mathbf{x}_n - \mathbf{y}_c\|^2\right) u_{nc}^{\alpha_c - 1} \prod_{c=C_1+1}^{C_1+C_2} \exp\left(-\frac{1}{2}u_{nc}^m \|\mathbf{x}_n - \mathbf{z}_c\|^2\right) u_{nc}^{\alpha_c - 1} \quad (17)$$

值得注意的是, 通过采样方式获得的新采样值 \mathbf{u}'_n 其分布和当前样本无关, 且 $\mathbf{u}'_n/\mathbf{u}_n$ 在不同时刻状态独立, 因此判决函数 a_u 不需要进行 Hasting 校验.

c) 比较 $p(\mathbf{x}_n, \mathbf{u}'_n | \mathbf{Y}^*, \mathbf{Z})$ 和 $p(\mathbf{x}_n, \mathbf{u}^*_n | \mathbf{Y}^*, \mathbf{Z})$, 其中 \mathbf{Y}^* 和 \mathbf{u}^*_n 分别是当前聚类中心 \mathbf{Y} 和隶属度 \mathbf{u}_n 的最优值. 如果 $p(\mathbf{x}_n, \mathbf{u}'_n | \mathbf{Y}^*, \mathbf{Z})$ 大于 $p(\mathbf{x}_n, \mathbf{u}^*_n | \mathbf{Y}^*, \mathbf{Z})$, 那么 \mathbf{u}'_n 替换 \mathbf{u}^*_n .

步骤 2. 为了获得聚类中心矩阵 \mathbf{Y} 的最优值, BFCCL 使用 Metropolis-Hastings 方法对 \mathbf{Y} 进行采样:

a) 通过以下正态分布对聚类中心进行采样, 即

$$\mathbf{y}'_c \sim \mathcal{N}\left(\mathbf{y}_c, \frac{1}{\sigma} \sum_y \mathbf{y}\right) \quad (18)$$

其中该正态分布的均值为当前聚类中心 \mathbf{y}_c , \sum_y 为样本的协方差矩阵, 参数 σ 用来控制生成的聚类中心的紧度.

b) 计算采样值 \mathbf{y}'_c 的接受概率: 对于固定的一类样本 \mathbf{X} , 隶属度矩阵 \mathbf{U} 和另一类聚类中心矩阵 \mathbf{Z} , $p(\mathbf{Y} | \mathbf{X}, \mathbf{U}, \mathbf{Z})$ 与 $p(\mathbf{Y}, \mathbf{X}, \mathbf{U}, \mathbf{Z})$ 和 $p(\mathbf{X}, \mathbf{U}, \mathbf{Y})$ 成比例, 同时也与 $p(\mathbf{X}, \mathbf{Y} | \mathbf{U})$ 成比例. 因此, 这里只需估计每一个样本 \mathbf{Y}'_c 的条件概率 $p(\mathbf{X}, \mathbf{y}'_c | \mathbf{U})$. 其接受概率 a_y 可表达式为

$$a_y = \min \left\{ 1, \frac{p(\mathbf{X}, \mathbf{y}'_c | \mathbf{U})}{p(\mathbf{X}, \mathbf{y}_c | \mathbf{U})} \right\} \quad (19)$$

由于新产生的状态独立于其他聚类中心, 因此, 条件分布 $p(\mathbf{X}, \mathbf{y}_c | \mathbf{U})$ 的计算式可以写成:

$$\begin{aligned} p(\mathbf{X}, \mathbf{y}_c | \mathbf{U}) &= p(\mathbf{X} | \mathbf{U}, \mathbf{y}_c) p(\mathbf{y}_c) \propto \\ &\exp \left\{ -\frac{1}{2} \sum_{n=1}^N u_{nc}^m \|\mathbf{x}_n - \mathbf{y}_c\|^2 \right\} \times \\ &\exp \left\{ -\frac{1}{2} (\mathbf{y}_c - \boldsymbol{\mu}_y)^\top \sum_y^{-1} (\mathbf{y}_c - \boldsymbol{\mu}_y) \right\} \end{aligned} \quad (20)$$

因为 \mathbf{y}'_c 服从高斯分布满足对称性, 式 (19) 也无需进行 Hasting 校验.

c) 比较 $p(\mathbf{X}, \mathbf{y}'_c | \mathbf{U}^*)$ 和 $p(\mathbf{X}, \mathbf{y}^*_c | \mathbf{U}^*)$, 其中 \mathbf{U}^* 和 \mathbf{y}^*_c 分别是模糊隶属度矩阵 \mathbf{U} 和聚类中心 \mathbf{y}_c 的最优值. 如果 $p(\mathbf{X}, \mathbf{y}'_c | \mathbf{U}^*)$ 大于 $p(\mathbf{X}, \mathbf{y}^*_c | \mathbf{U}^*)$, 那么 \mathbf{y}'_c 替换 \mathbf{y}^*_c .

步骤 3. 根据式 (13) 计算 $p(\mathbf{X}, \mathbf{U}^*, \mathbf{Y}^*, \mathbf{Z})$ 的值并与当前 $p(\mathbf{X}, \mathbf{U}, \mathbf{Y}, \mathbf{Z})$ 的值进行比较, 若

$p(\mathbf{X}, \mathbf{U}, \mathbf{Y}, \mathbf{Z}) > p(\mathbf{X}, \mathbf{U}^*, \mathbf{Y}^*, \mathbf{Z})$, 则用 $\{\mathbf{U}, \mathbf{Y}\}$ 替换当前的 $\{\mathbf{U}^*, \mathbf{Y}^*\}$.

2.3 BFCCL 聚类算法

本文将研究重点置于最基本的二元分类问题. 结合图 1 所示的 BFCCL 聚类构造原理示意图和第 2.2 节的内容, 本节给出 BFCCL 聚类算法在整个数据集上实施的具体步骤, 见算法 1. 设全体样本 $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ 和其对应的类别标签 $\{l_1, l_2, \dots, l_N\}$ ($l_i \in \{-1, +1\}$, $i = 1, 2, \dots, N$), 其中 $\mathbf{X}_1 = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{N_1}\}$ 和 $\mathbf{X}_2 = \{\mathbf{x}_{N_1+1}, \mathbf{x}_{N_1+2}, \dots, \mathbf{x}_N\}$ 分别表示正负两类样本. 为了获得全体样本的最佳空间划分, 需在正负类上采取交替循环的策略来执行式 (13). 两类聚类中心交替影响, 直至两类的聚类中心处于稳定状态.

BFCCL 聚类采取两类样本交替循环执行的策略来获得聚类中心和模糊隶属度的最优解, BFCCL 聚类中单次循环的时间复杂度主要在于步骤 2 的执行, 而步骤 2 又由两部分构成: 计算模糊隶属度矩阵 \mathbf{U} 和聚类中心矩阵 \mathbf{Y} , 前者的时间复杂度由式 (17) 产生, 为 $O(NCD + CD^2)$; 后者的时间复杂度由式 (20) 产生, 为 $O(CD^2)$. 因此, BFCCL 聚类算法执行的时间复杂度为 $O(v_{\max} t_{\max} (NCD + CD^2))$, 其中 t_{\max} 和 v_{\max} 分别是算法正负类交替迭代和 Metropolis-Hastings 采样方法中迭代的次数. 另外, 实验中一般可将式 (18) 中的协方差矩阵设为对角阵, 此时, 算法 1 的复杂度可减少至 $O(v_{\max} t_{\max} NCD)$. 可见, BFCCL 聚类算法运行的时间效率与样本的分布和结构有关, 为了在一定程度上提高本文方法的执行效率, 可以采用 KPCA 等降维方法对样本进行相应的预处理, 或者采用正交三角 (QR) 等分解法对式 (20) 中的协方差矩阵进行变换处理.

算法 1. BFCCL 聚类算法

输入. 正负两类数据集 \mathbf{X}_1 和 \mathbf{X}_2 , 模糊指数 m , 正、负类样本的聚类个数分别为 $C^{(1)}$ 和 $C^{(2)}$, MH 采样最大迭代次数 v_{\max} , 两类交替循环最大迭代次数 t_{\max} , 误差阈值 ε ;

输出. 模糊隶属度矩阵 $\mathbf{U}^{(1)*}, \mathbf{U}^{(2)*}$ 和聚类中心 $\mathbf{Y}^{(1)*}, \mathbf{Y}^{(2)*}$.

初始化. 在数据集 \mathbf{X}_1 和 \mathbf{X}_2 分别计算正负类样本的均值 $\mu_y^{(1)}, \mu_y^{(2)}$ 和协方差 $\sum_y^{(1)}, \sum_y^{(2)}$; 在数据集 \mathbf{X}_1 和 \mathbf{X}_2 分别进行两类模糊隶属度矩阵 $\mathbf{U}^{(1)}$ 和 $\mathbf{U}^{(2)}$ 进行初始化, 其中 $\mathbf{U}^{(1)}$ 矩阵中的元

素 $\mathbf{u}_n^{(1)} \sim \text{Dirichlet}(\boldsymbol{\alpha} = \mathbf{1}_C)$, $\mathbf{U}^{(2)}$ 矩阵中的元素 $\mathbf{u}_n^{(2)} \sim \text{Dirichlet}(\boldsymbol{\alpha} = \mathbf{1}_C)$, ($n = 1, \dots, N$, $C = C^{(1)} + C^{(2)}$); 在数据集 \mathbf{x}_1 和 \mathbf{x}_2 分别对两类聚类中心矩阵 $\mathbf{Y}^{(1)}$ 和 $\mathbf{Y}^{(2)}$ 进行初始化, 其中 $\mathbf{Y}^{(1)}$ 矩阵中的元素 $\mathbf{y}_c^{(1)}: \mathbf{y}_c^{(1)} \sim \mathcal{N}(\mu_y^{(1)}, \sum_y^{(1)})$ ($c = 1, \dots, C^{(1)}$), $\mathbf{Y}^{(2)}$ 矩阵中的元素 $\mathbf{y}_c^{(2)} \sim \mathcal{N}(\mu_y^{(2)}, \sum_y^{(2)})$ ($c = 1, \dots, C^{(2)}$); 在数据集 \mathbf{X}_1 和 \mathbf{X}_2 分别对模糊隶属度矩阵和聚类中心矩阵的最大后验值进行初始化, $\mathbf{U}^{(1)*} = \mathbf{U}(1)$, $\mathbf{U}^{(2)*} = \mathbf{U}(2)$, $\mathbf{Y}^{(1)*} = \mathbf{Y}^{(1)}$, $\mathbf{Y}^{(2)*} = \mathbf{Y}^{(2)}$;

For $v = 0$ to v_{\max}

For $t = 1$ to t_{\max}

步骤 1. 令 $\mathbf{X} = \mathbf{X}_1$, $\mathbf{Z} = \mathbf{Y}^{(2)*}(v)$, $\mathbf{U}^* = \mathbf{U}^{(1)*}$, $C_1 = C^{(1)}$, $C_2 = C^{(2)}$

步骤 2.

For $n = 1$ to $\text{size}(\mathbf{X})$

{步骤 2.1. 根据式 (15) 得到隶属度 \mathbf{u}'_n ;

步骤 2.2. 根据式 (16) 计算 $p(\mathbf{x}_n, \mathbf{u}'_n | \mathbf{Y}, \mathbf{Z})$

和 $p(\mathbf{x}_n, \mathbf{u}_n | \mathbf{Y}, \mathbf{Z})$, 并以式 (17) 的概率接受 $\mathbf{u}_n \leftarrow \mathbf{u}'_n$;

步骤 2.3. 根据式 (17), 如果 $p(\mathbf{x}_n, \mathbf{u}'_n | \mathbf{Y}^*, \mathbf{Z}) < p(\mathbf{x}_n, \mathbf{u}_n | \mathbf{Y}^*, \mathbf{Z})$, 那么 $\mathbf{u}_n^* \leftarrow \mathbf{u}'_n$;

步骤 2.4. 如果 $n > N$, 则转至步骤 2.5; 否则, 置 $n = n + 1$, 并返回步骤 2.1};

End n

For $c = 1$ to $\text{size}(C_1)$

{步骤 2.5. 根据式 (18) 采样新的聚类中心 \mathbf{y}'_c ;

步骤 2.6. 以式 (19) 的概率 a_y 接受 $\mathbf{y}_c \leftarrow \mathbf{y}'_c$;

步骤 2.7. 根据式 (20), 如果 $p(\mathbf{x}, \mathbf{y}'_c | \mathbf{U}^*) > p(\mathbf{x}, \mathbf{y}_c | \mathbf{U}^*)$, 那么 $\mathbf{y}_c^* \leftarrow \mathbf{y}'_c$;

步骤 2.8. 如果 $c > C_1$, 则转至步骤 2.9; 否则, 置 $c = c + 1$, 返回步骤 2.6;

步骤 2.9. 如果 $p(\mathbf{X}, \mathbf{U}^*, \mathbf{Y}^*, \mathbf{Z}) > p(\mathbf{X}, \mathbf{U}, \mathbf{Y}, \mathbf{Z})$, 那么 $\mathbf{U}^* \leftarrow \mathbf{U}$, $\mathbf{Y}^* \leftarrow \mathbf{Y}$;

End c

步骤 2.10. 如果 $t > t_{\max}$, 则将 \mathbf{U}^* 和 \mathbf{Y}^* 分别设置为当前样本的模糊隶属度矩阵和聚类中心矩阵的最大后验值并转至步骤 3; 否则, 转至步骤 2.1;

步骤 3. 令 $\mathbf{X} = \mathbf{X}_2$, $\mathbf{Z} = \mathbf{Y}^{(1)*}(v)$, $\mathbf{U}^* = \mathbf{U}^{(2)*}$, $C_1 = C^{(2)}$, $C_2 = C^{(1)}$

步骤 4. 转向步骤 2

End t

步骤 5. 如果 $|\mathbf{Y}^{(1)*}(v) - \mathbf{Y}^{(1)*}(v-1)| \leq \varepsilon$ 或者 $v > v_{\max}$, 则算法停止并输出 $\mathbf{U}^{(1)*}$, $\mathbf{U}^{(2)*}$ 和 $\mathbf{Y}^{(1)*}$, $\mathbf{Y}^{(2)*}$; 否则, 置 $v = v + 1$, 转至步骤 5.

End v .

3 用于不平衡数据分类的 0 阶 TSK 型模糊系统 (0-TSK-IDC)

对于给定的样本及其类别标签 $\{\mathbf{x}_i, l_i\}_{i=1}^N$ 由算法 1 完成 BFCCL 聚类后, 得到的每一个聚类对应于 0-TSK-IDC 模糊系统中的一条模糊规则. 同时分别属于两类样本的模糊隶属度矩阵 $\mathbf{U}^{(1)*}$, $\mathbf{U}^{(2)*}$ 和聚类中心矩阵 $\mathbf{Y}^{(1)*}$, $\mathbf{Y}^{(2)*}$ 为构建高斯隶属度函数 (式 (4)) 提供了依据, 高斯隶属度函数的方差 $\delta^{(1)}$ 和 $\delta^{(2)}$ 可由下式计算得到:

$$\delta_{c,i}^{(1)} = \frac{h \sum_{j=1}^{N_1} u_{jc} (x_{ji} - y_{c,i}^{(1)})^2}{\sum_{j=1}^{N_1} u_{jc}}, \quad i = 1, 2, \dots, D, \quad c = 1, 2, \dots, C^{(1)} \quad (21)$$

$$\delta_{c,i}^{(2)} = \frac{h \sum_{j=N_1+1}^N u_{jc} (x_{ji} - y_{c,i}^{(2)})^2}{\sum_{j=N_1+1}^N u_{jc}}, \quad i = 1, 2, \dots, D, \quad c = 1, 2, \dots, C^{(2)} \quad (22)$$

当模糊规则的前件参数 (矩阵 \mathbf{Y} ($\mathbf{Y} = [\mathbf{Y}^{(1)*}, \mathbf{Y}^{(2)*}]^T$) 和 $\boldsymbol{\delta}$ ($\boldsymbol{\delta} = [\boldsymbol{\delta}^{(1)}, \boldsymbol{\delta}^{(2)}]^T$) 确定后, 使用式 (3)~(5) 可将原始数据集 $\{\mathbf{x}_i, l_i\}_{i=1}^N$ 映射至特征空间形成新的数据集 $\{d(\mathbf{x}_i), l_i\}_{i=1}^N$, 其中 $d(\mathbf{x}_i)$ 可写成以下形式:

$$d(\mathbf{x}_i) = [\tilde{\mu}_1(\mathbf{x}_i), \tilde{\mu}_2(\mathbf{x}_i), \dots, \tilde{\mu}_{C^{(1)}+C^{(2)}}(\mathbf{x}_i)]^T \quad (23)$$

令 $\mathbf{p}_0 = [p_0^1, p_0^2, \dots, p_0^{C^{(1)}+C^{(2)}}]^T$, 0-TSK-IDC 模糊系统的后件参数学习转化为求解规则后件 \mathbf{p}_0 的线性问题, 即

$$f(\mathbf{x}_i) = \mathbf{p}_0^T d(\mathbf{x}_i) \begin{cases} \geq 0, & \mathbf{x}_i \in \mathbf{X}_1 \\ < 0, & \mathbf{x}_i \in \mathbf{X}_2 \end{cases} \quad (24)$$

传统分类面的大间隔的策略是达到两类间距离的最大化, 但这种分类面在处理不平衡分类问题中易向少数类发生偏移. 为纠正分类面的偏移, 本文寻找的分类面在达到两类间距离的最大化的同时, 保证少数类到分类面的距离不得小于多数类到分类面的距离, 其示意图如图 3 所示. 因此后件参数的学习问题可用以下最优化问题来表示:

$$\begin{aligned} \min_{\rho, \mathbf{p}_0, \xi} & \frac{1}{2} \mathbf{p}_0^T \mathbf{p}_0 - v \rho^2 + \lambda_1 \sum_{i=1}^{N_1} \xi_i + \lambda_2 \sum_{j=N_1+1}^N \xi_j \\ \text{s. t.} & \mathbf{p}_0^T d(\mathbf{x}_i) \geq 1 + \rho^2 - \xi_i, \quad i = 1, 2, \dots, N_1 \\ & -\mathbf{p}_0^T d(\mathbf{x}_j) \geq 1 - \xi_j, \quad j = N_1 + 1, N_1 + 2, \dots, N \end{aligned} \quad (25)$$

其中, v, λ_1 和 λ_2 为 3 个正常数, 且 $\lambda_1/\lambda_2 = N_2/N_1 = (N - N_1)/N_1$.

求解式 (25) 可以利用 Lagrange 技巧, 引入非负的 Lagrange 因子 α , $L(\mathbf{p}_0, \rho, \xi)$ 方程的各变量求偏导数, 并令各偏导方程为零, 上述最优化问题可以转换成对偶形式:

$$\begin{aligned} \min_{\alpha} \quad & -\frac{1}{2} \alpha^T \mathbf{K} \alpha \\ \text{s. t.} \quad & \sum_{i=1}^{N_1} \alpha_i = v \\ & \sum_{j=N_1+1}^N \alpha_j = v \\ & 0 \leq \alpha_i \leq \lambda_1, \quad i = 1, \dots, N_1 \\ & 0 \leq \alpha_j \leq \lambda_2, \quad j = N_1 + 1, \dots, N \end{aligned} \quad (26)$$

其中, $\mathbf{K} = [k_{ij}]_{N \times N}, k_{ij} = l_i l_j d(\mathbf{x}_i)^T d(\mathbf{x}_j)$. 根据优化理论, 当且仅当 $\mathbf{K}(\cdot)$ 为 Mercer 核时式 (26) 为二次凸规划, 求得的解是全局最优解. 这里, 证明其为 Mercer 核.

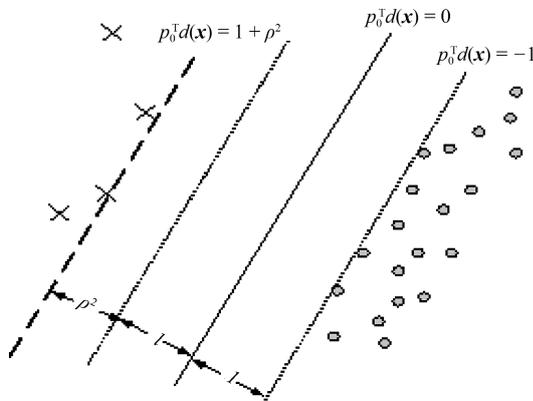


图 3 0-TSK-ICD 分类面示意图
Fig. 3 The classification hyperplane of 0-TSK-ICD

引理 1 [28]. 令 \mathbf{x} 是 \mathbf{R}^n 上的一个紧集, $\mathbf{K}(\mathbf{x}, \mathbf{x})$ 是 Mercer 核当且仅当 $\mathbf{K}(\mathbf{x}, \mathbf{x})$ 是 $\mathbf{X} \times \mathbf{X}$ 上连续对称函数且关于任意的 $\mathbf{x}_1, \dots, \mathbf{x}_l \in \mathbf{X}$ 的 Gram 矩阵半正定.

定理 1. 如果模糊系统中前件的模糊隶属函数 $\tilde{\mu}_k(\mathbf{x})$ 是连续函数, 则式 (26) 的核函数 \mathbf{K} 是 Mercer 核.

证明. 根据式 (26) 的定义, $\mathbf{K} = [k_{ij}]_{N \times N}, k_{ij} = l_i l_j d(\mathbf{x}_i)^T d(\mathbf{x}_j)$, 容易看出 \mathbf{K} 是一个实对称矩阵. 其次, 对任意的 $\alpha_1, \dots, \alpha_N \in \mathbf{R}$, 由式 (23), 有 $\sum_{i,j} \alpha_i \alpha_j k_{ij} = \sum_{i,j} \alpha_i \alpha_j l_i l_j d(\mathbf{x}_i)^T d(\mathbf{x}_j) = \left\| \sum_{i=1}^N \sum_{k=1}^{C^{(1)}+C^{(2)}} \alpha_i l_i \tilde{\mu}_k(\mathbf{x}_i) \right\|^2 \geq 0$ 且 $\tilde{\mu}_k(\mathbf{x})$ 是连续函数, 因此核函数 \mathbf{K} 是 Mercer 核. \square

因此, 设 $\alpha^* = [\alpha_1^*, \dots, \alpha_N^*]^T$ 为式 (26) 的解, 则 0-TSK-ICD 的后件参数 p_0^k 为

$$p_0^k = \sum_{i=1}^N l_i \alpha_i^* \tilde{\mu}_k(\mathbf{x}_i) \quad (27)$$

最终, 0-TSK-ICD 模糊系统的决策函数 $F(\mathbf{x})$ 可表示为

$$F(\mathbf{x}) = \text{sgn}(\mathbf{p}_0^T d(\mathbf{x})) \quad (28)$$

根据上文的分析和推导, 0-TSK-ICD 使用 BFCCL 算法得到准确性和可解释性强的规则前件参数, 规则后件参数着眼于分类面的偏移纠正. 这里总结出 0-TSK-ICD 模糊系统的具体训练步骤, 具体如下:

算法 2. 0-TSK-ICD 模糊系统

输入. 数据集 $\{\mathbf{x}_i, l_i\}_{i=1}^N$, 模糊规则数 $(C^{(1)} + C^{(2)})$, 模糊指数 m , 尺度参数 h , MH 采样最大迭代次数 v_{\max} , 两类交替循环最大迭代次数 t_{\max} , 误差阈值 ε , 调节参数 v , 正则化参数 λ_1 和 λ_2 ;

输出. 分类决策函数 $F(\mathbf{x})$.

步骤 1. 根据算法 1 得到隶属度矩阵 $\mathbf{U}^{(1)*}$, $\mathbf{U}^{(2)*}$ 和聚类中心矩阵 $\mathbf{Y}^{(1)*}, \mathbf{Y}^{(2)*}$;

步骤 2. 由式 (22) 计算高斯隶属度函数的方差矩阵 $\delta^{(1)}$ 和 $\delta^{(2)}$;

步骤 3. 由式 (3)~(5) 0-TSK-ICD 将原始数据集 $\{\mathbf{x}_i, l_i\}_{i=1}^N$ 映射至特征空间形成新的数据集 $\{d(\mathbf{x}_i), l_i\}_{i=1}^N$;

步骤 4. 由式 (26) 求解得到 Lagrange 因子 α ;

步骤 5. 由式 (27) 得到 0-TSK-ICD 的后件参数 \mathbf{p}_0 ;

步骤 6. 由式 (28) 得到 0-TSK-ICD 的决策函数.

0-TSK-ICD 模糊系统的时间复杂度主要集中在算法 1 和二次规划求极值问题上. 二次规划问题的时间复杂度为 $O(N^3)$, 因此, 0-TSK-ICD 模糊系统的时间复杂度为 $O(v_{\max} t_{\max} (NCD + CD^2) + N^3)$. 本文使用 SMO (Sequential minimal optimization)^[20] 等分解方法处理二次规划问题, 0-TSK-ICD 模糊系统的时间复杂度降低至 $O(v_{\max} t_{\max} (NCD + CD^2) + N^2)$.

需要指出的是, 0-TSK-ICD 模糊系统后件参数的学习策略与支持向量机 (Support vector machine, SVM)^[29] 较为相似, 但这两种方法之间存在着本质差异, 0-TSK-ICD 通过模糊规则完成特征空间的映射, 而 SVM 通过核函数直接将原始数据映射到特征空间上; 其次, 0-TSK-ICD 分类决策函数中的参数同时也是模糊系统中模糊规则的参数, 其具有高度的语义性和可解释性, 这一特性是模糊系统特有的.

4 实验结果与分析

4.1 实验设置

一般认为当少数类与多数类的样本比例低于 1:2 时, 数据集具有不平衡特征. 而现实生活中的医学诊断数据往往呈现出非平衡性, 本文的实验部分将利用人工 Banana 数据集^[30] 和 4 个 UCI 医学数据集^[31] (基本信息见表 1) 所示对 BFCCL 算法和 0-TSK-IDC 模糊系统进行验证和评价. 依照不平衡数据分类问题中常用的设定方法, 实验中将少数类指定为正类, 将多数类指定为负类.

本文的实验分成两个部分: 1) 0-TSK-IDC 中规则前件参数学习分别使用 BFCCL 算法和 BFC 算法来进行性能的比较; 2) 0-TSK-IDC 模糊系统与 FS-FCSVM^[32]、L2-TSK-FS^[25]、Adaboost^[33] 和 CS-SVM^[34] 算法进行性能的比较, 其中 FS-FCSVM 和 L2-TSK-FS 是基于模糊聚类的模糊系统; Adaboost 和 CS-SVM 是处理不平衡数据分类的算法. 为了验证本文所提的模糊规则的后件参数学习方法在不平衡分类问题上的有效性, 实验中还将在 0-TSK-IDC 模糊系统与 BFCCL-TSK-FS 模糊系统进行了对比, BFCCL-TSK-FS 模糊系统中规则前件参数的学习使用 BFCCL 聚类方法; 规则后件参数的学习使用与 FS-FCSVM 模糊系统相同的方法, 即标准的 SVM 算法.

实验参数设置如下: 0-TSK-IDC 中最大迭代次数 $t_{\max} = 10^3$, $v_{\max} = 10^3$, 式 (19) 中的参数 $\sigma = 3$, 误差阈值 $\varepsilon = 10^{-5}$, 根据文献^[35] 模糊指数 m 取值为 2, 此时物理意义最明确, 其他可人工设定的参数通过网格搜索的方式来确定, 尺度参数 h 取值为 $\{0.2^2, 0.4^2, \dots, 2^2\}$, 调节参数 v 取值为 $\{1, 5, \dots, 30\}$, 正则化参数 λ_2 取值为 $\{10^{-3}, \dots, 10^3\}$, λ_1 的值根据 $\lambda_1 = \lambda_2 N_2 / N_1$ 来确定, 模糊规则数 C_1 和 $C^{(2)}$ 取值均为 $\{1, 2, \dots, 10\}$. 实验中, FS-FCSVM 和 L2-TSK-FS 的规则前件由 FCM 聚类获得, 模糊指数 m 取值为 2, 规则后件学习的正则化参数的设置与 λ_2 相同. CS-SVM 算法中的核函数采用高斯核, 核参数取值为 $\{10^{-3}, \dots, 10^3\}$, 正则化参数的设置与 λ_1 和

λ_2 相同. Adaboost 算法中设置弱分类器的个数为 10. BFCCL-TSK-FS 算法中 BFCCL 聚类的参数设置参照 0-TSK-IDC, 其他参数的设置参照 FS-FCSVM.

为了更好地呈现出不同程度的非平衡性对算法性能产生的影响, 本文采用 G -mean 和 F -measure^[36-37] 两种评价准则来评价算法的分类性能:

$$G\text{-mean} = \sqrt{\text{Positive Accuracy} \times \text{Negative Accuracy}} \quad (29)$$

$$F\text{-measure} = \frac{(1 + \beta^2) \times \text{Recall} \times \text{Precision}}{\beta^2 \times \text{Recall} + \text{Precision}} \quad (30)$$

其中, G -mean 是关注数据集整体分类性能的评价指标, $\text{Positive Accuracy} = \text{Recall} = TP / (TP + FN)$, $\text{Negative Accuracy} = TN / (TN + FP)$ 为真负率. F -measure 是查全率和查准率的调和均值, $\text{Recall} = TP / (TP + FN)$ 为查全率, $\text{Precision} = TP / (TP + FP)$ 为查准率, β 通常设置为 1. 这 2 个评价准则中用到的 TP 指标是指被正确分类的正类样本的数目, FN 指标是指被错误分类的正类样本的数目, FP 指标是指被错误分类的负类样本的数目, TN 指标是指被正确分类的负类样本的数目. 可见, G -mean 准则可以合理地评价在保持正、负类分类精度平衡下最大化两类的精度, 而 F -measure 准则可以合理地评价分类器对于少数类的分类性能. 本文实验中每个数据集进行 5 折交叉验证, 运行 5 次以平均后的 G -mean 和 F -measure 作为最终分类精度. 本文的实验在 2.53 GHz quad-core CPU, 8 GB RAM, Windows 7 系统下执行, 所有算法均在 Matlab 2009b 环境下实现.

4.2 BFCCL 与 BFC 在 0-TSK-IDC 模糊分类器中的性能比较

4.2.1 人工数据集

实验中构造了 3 组正负类比例不同的 Banana

表 1 数据集的基本信息

Table 1 The basic information of datasets

数据集	正类样本数	负类样本数	正负类比例	属性个数
Banana dataset	600, 200, 100	1 500	2:5, 2:15, 1:15	2
Heart statlog	120, 60, 30, 20, 12	170	12:17, 6:17, 3:17, 2:17, 6:85	13
Breast wisconsin	241, 200, 150, 100, 40	458	241:458, 100:229, 75:229, 50:229, 20:229	10
Liver disorders	145, 100, 50, 20	200	29:40, 1:2, 1:4, 1:10	7
Haberman	81, 40, 25	225	9:25, 8:45, 1:9	3

数据集, 即: 选取全部 1500 个负类样本, 并分别随机选取 600, 200, 100 个正类样本. 图 4~7 是分别使用 BFC 和 BFCCL 算法正负类聚类数均为 3 和 4 时, 交叉验证中某一折的聚类结果示意图, 两图中正类样本分别是 480, 160 和 80 个 (用黑点表示), 负类样本均为 1200 个 (用灰点表示), 正类和负类的聚类中心分别用黑色和白色矩形表示. 可以看出:

1) 当在正负两类样本中分别使用 BFC 算法聚类时, 图 4 中 (a)~(c) 中负类样本的聚类效果完全相同, 图 6 亦如此. 同时在样本重叠区域出现了聚类中心间的距离狭小的现象, 自然由此聚类效果得到的模糊规则的模糊集会存在交叉和重叠的现象. 这是因为 BFC 聚类不考虑不同类别聚类中心间的排斥关系, 运用到模糊系统前件学习中时, 得到的模糊

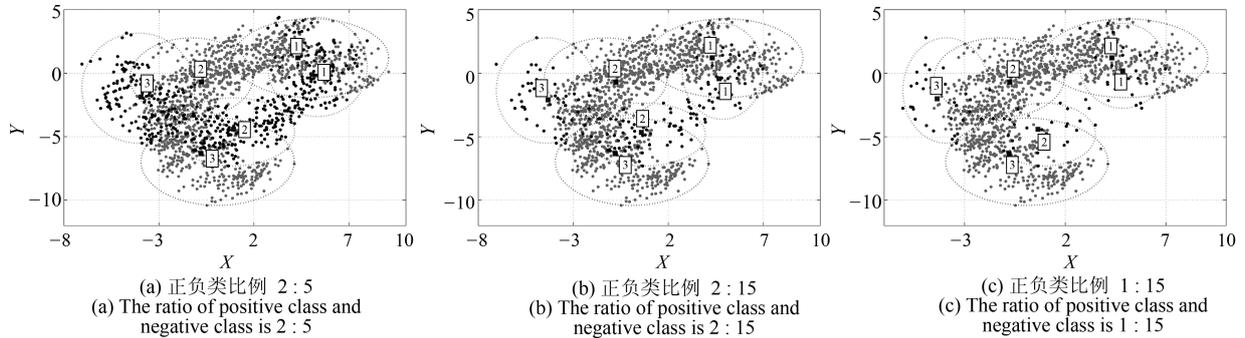


图 4 BFC 在 Banana 集上正负类聚类数均为 3 时的聚类效果
 Fig. 4 The clustering results on the Banana dataset in BFC with three clustering on the positive and negative classes, respectively

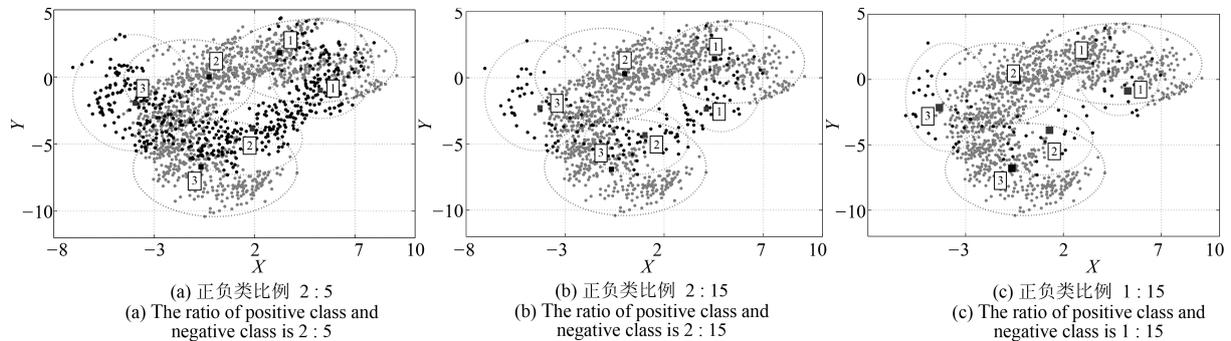


图 5 BFCCL 在 Banana 集上正负类聚类数均为 3 时的聚类效果
 Fig. 5 The clustering results on the Banana dataset in BFCCL with three clustering on the positive and negative classes, respectively

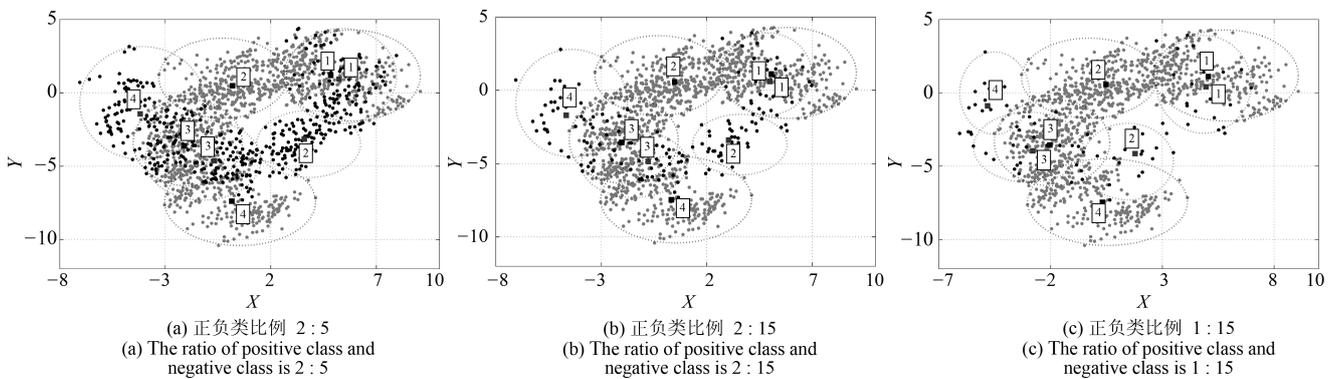


图 6 BFC 在 Banana 集上正负类聚类数均为 4 时的聚类效果
 Fig. 6 The clustering results on the Banana dataset in BFC with four clustering on the positive and negative classes, respectively

规则其清晰性得不到保证.

2) BFCCL 算法考虑了在样本重叠区不同类别聚类中心的竞争关系, 聚类中心由于受到另一个类别样本的排斥而尽量地远离该区域, 这种排斥力最终促使正负类样本的聚类中心尽可能地远离对方样本区域, 从图 5 和图 7 的聚类效果可知, 随着正类样本数量和分布的变化, 其对负类样本的聚类中心的排斥力的强弱也发生了变化, 负类样本的聚类效果也随之受到了影响. 因此在样本重叠区域不会发生聚类中心间的距离狭小的情况.

图 8 给出了 0-TSK-IDC 模糊系统基于图 7 (b) 聚类结果所获得的模糊集示意图. 我们知道, 聚类中心 (即模糊规则中心) 的分布是决定所获得的模糊规则清晰性和可解释性的决定性因素之一. 由图 8 可知, BFCCL 聚类中的不同类别间聚类中心竞争学习的策略可以保证获得的模糊规则具有良好的可解释性. 表 3 为 0-TSK-IDC 模糊系统中规则前件分别使用 BFCCL 与 BFC 算法的性能比较. 根据表 3 结果可知, 在正负类样本具有不同比例不平衡性的情

况下, 与 BFC 聚类相比, 使用 BFCCL 聚类获得的解释性强的规则前件参数有助于提高 0-TSK-IDC 模糊系统的分类性能, 此时 G -mean 和 F -measure 指标均有相应的提高.

4.2.2 真实数据集

为了对 BFCCL 聚类算法性能作进一步地探讨和分析, 本节使用 4 个不平衡 UCI 医学诊断集对 BFCCL 与 BFC 算法应用到模糊规则前件参数学习时, 所得的 0-TSK-IDC 模糊系统的分类性能进行比较. G -mean 和 F -measure 值及其方差的比较结果如表 3 所示. 从表 3 结果可知, 真实数据集上的实验结果与上一节人工数据集实验上观察到的结果基本保持一致, 在 0-TSK-IDC 模糊系统中使用 BFCCL 算法方法后 F -measure 和 G -mean 评价指标均高于使用 BFC 算法的情况, 说明在使用聚类算法获取模糊规则前件参数的学习中, 充分考虑不同类别聚类中心之间的竞争关系, 有助于提升模糊空间划分的准确性和清晰性.

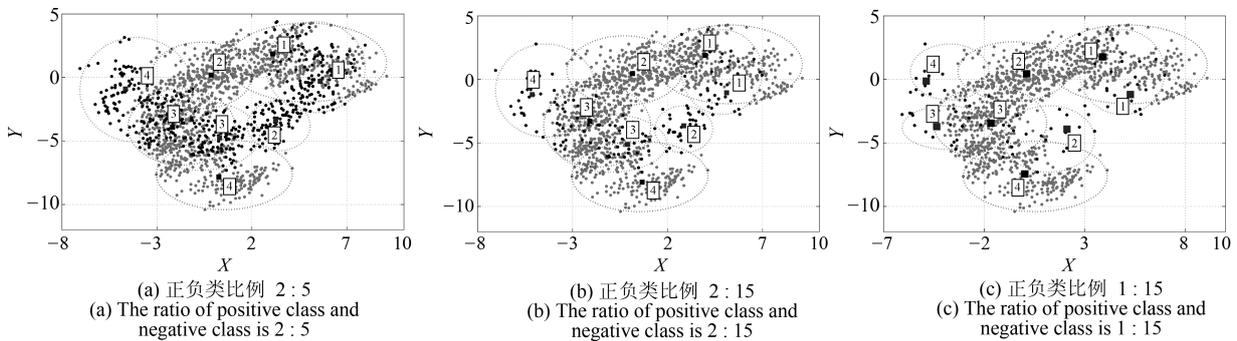


图 7 BFCCL 在 Banana 集上正负类聚类数均为 4 时的聚类效果

Fig. 7 The clustering results on the Banana dataset in BFCCL with four clustering on the positive and negative classes, respectively

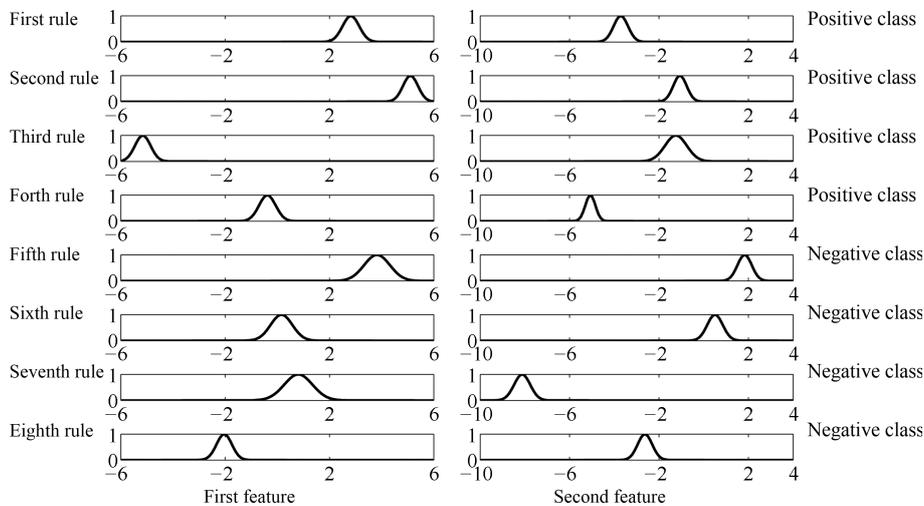


图 8 0-TSK-IDC 基于图 7 (b) 聚类结果的所获得模糊集示意图

Fig. 8 A plot of rulebase of 0-TSK-IDC from the clustering result in Fig. 7 (b)

表 2 Banana 集上基于 BFC 与 BFCCL 图 4~7 聚类结果的 0-TSK-IDC 模糊系统中的 G -mean 和 F -measure 及其方差的比较

Table 2 G -mean, F -measure and their standard deviations comparison of 0-TSK-IDC with the clustering results in Fig. 4~7 by using the BFC and BFCCL on the Banana dataset

规则数	正负类比例	BFC		BFCCL	
		G -mean (%)	F -measure (%)	G -mean (%)	F -measure (%)
6	2:5	96.48(0.60)	96.2(0.64)	96.97 (0.53)	96.44 (0.54)
	2:15	95.89(0.52)	95.77(0.49)	96.20 (0.31)	96.22 (0.36)
	1:15	94.14(0.47)	93.79(0.41)	95.45 (0.55)	94.99 (0.48)
8	2:5	97.98(0.31)	97.23(0.34)	99.75 (0.27)	99.74 (0.23)
	2:15	97.03(0.29)	96.92(0.29)	99.32 (0.34)	99.32 (0.35)
	1:15	96.76(0.36)	96.65(0.32)	98.68 (0.30)	98.63 (0.33)

表 3 UCI 医学集上分别使用 BFCCL 与 BFC 得到规则前件时 0-TSK-IDC 模糊系统中的 G -mean 和 F -measure 值及其方差的比较

Table 3 G -mean, F -measure and their standard deviations comparison of 0-TSK-IDC with BFC and BFCCL on UCI datasets

数据集	正负类比例	BFC		BFCCL	
		G -mean (%)	F -measure (%)	G -mean (%)	F -measure (%)
Heart	12:17	87.01(1.69)	86.87(1.78)	89.56 (1.91)	89.36 (1.90)
	6:17	86.24(2.00)	85.48(1.88)	88.14 (1.89)	87.88 (1.92)
	3:17	85.41(1.97)	84.08(2.00)	87.29 (2.01)	86.40 (2.00)
	2:17	82.50(2.30)	80.02(2.31)	85.71 (2.24)	83.63 (2.23)
	6:85	81.05(2.17)	75.37(2.19)	84.65 (2.11)	78.50 (2.04)
Breast	241:458	93.62(2.60)	91.46(2.57)	96.56 (2.34)	95.03 (2.20)
	100:229	91.14(2.05)	90.02(2.55)	95.59 (1.97)	94.24 (2.01)
	75:229	90.37(2.00)	89.14(2.04)	93.75 (1.88)	91.22 (1.89)
	50:229	87.99(1.90)	85.73(1.89)	91.59 (2.03)	89.29 (2.10)
	20:229	84.21(2.23)	81.26(2.21)	87.56 (2.00)	86.05 (1.99)
Liver	29:40	70.38(0.80)	66.28(0.82)	72.50 (0.77)	68.51 (0.76)
	1:2	69.77(0.75)	61.27(0.75)	71.15 (0.69)	62.50 (0.60)
	1:4	67.82(0.79)	52.98(0.78)	70.24 (0.73)	55.22 (0.79)
	1:10	65.08(0.81)	47.31(0.83)	67.18 (0.75)	50.65 (0.72)
Haberman	9:25	76.05(1.73)	52.75(1.73)	76.56 (1.60)	53.61 (1.60)
	8:45	68.02(1.86)	51.09(1.85)	68.97 (1.85)	52.60 (1.87)
	1:9	64.21(1.69)	48.22(1.73)	65.42 (1.74)	50.01 (1.69)

4.3 0-TSK-IDC 模糊分类器与相关分类算法性能比较

为了对 0-TSK-IDC 模糊分类器的性能进行评估, 本节对 0-TSK-IDC 与 FS-FCSVM、L2-TSK-FS、BFCCL-TSK-FS、Adaboost 和 CS-SVM 五种分类算法进行比较.

4.3.1 人工数据集

本节依旧采用与第 4.2.1 节相同的 Banana 数

据集进行训练和测试, 表 4 列出了 6 种对比算法的 G -mean 和 F -measure 指标的比较. 从实验结果可以看出, 随着 Banana 数据集中正负类不平衡比例的提高, 6 种算法的 F -measure 和 G -mean 指标都呈现出一定的下降趋势, 可见, 数据的不平衡性将严重影响分类的效果. 特别是 FS-FCSVM、L2-TSK-FS 和 BFCCL-TSK-FS 算法, 由于未考虑类别不平衡性, 当正类样本太少而不能充分学习会导致分类面向正类样本发生了较大的偏移, 其结果是

表 4 Banana 数据集上 0-TSK-IDC 模糊分类器与其他算法的 G -mean 和 F -measure 值及其方差的比较
Table 4 G -mean, F -measure and their standard deviations comparison of 0-TSK-IDC and other algorithms on the Banana dataset

正负类样本数	算法	G -mean (%)	F -measure (%)
2:5	FS-FCSVM	95.90(0.89)	95.31(0.84)
	L2-TSK-FS	96.26(0.47)	95.48(0.45)
	BFCCL-TSK-FS	96.79(0.50)	96.14(0.41)
	Adaboost	98.77(0.87)	98.71(0.90)
	CS-SVM	98.98(0.33)	98.91(0.30)
	0-TSK-IDC	99.75(0.27)	99.74(0.23)
2:15	FS-FCSVM	90.53(0.65)	89.46(0.60)
	L2-TSK-FS	89.23(0.71)	88.47(0.72)
	BFCCL-TSK-FS	92.70(0.58)	92.26(0.62)
	Adaboost	97.92(0.64)	97.75(0.68)
	CS-SVM	98.22(0.37)	98.05(0.36)
	0-TSK-IDC	99.32(0.34)	99.32(0.35)
1:15	FS-FCSVM	86.06(0.81)	82.83(0.84)
	L2-TSK-FS	87.95(0.55)	84.67(0.54)
	BFCCL-TSK-FS	88.84(0.43)	86.33(0.49)
	Adaboost	97.46(0.58)	97.28(0.52)
	CS-SVM	97.79(0.74)	97.61(0.70)
	0-TSK-IDC	98.68(0.30)	98.63(0.33)

正类样本的分类准确率迅速降低, 因而 G -mean 和 F -measure 指标值也迅速降低. 由于充分考虑了不同类别的聚类中心间的竞争学习及数据的不平衡性, 本文提出的 0-TSK-IDC 模糊系统在分类性能上相比其他 5 种算法都有提高. CS-SVM 算法更多地考虑少数类即正类样本的分类代价, 对正类样本的识别率较高而负类样本的识别率较低; 而 Adaboost 算法通过过采样技术来增加少数类样本的数量, 由于改变了样本的分布结构容易造成分类器过拟合的情况. 因此, 这两种算法获得的 G -mean 和 F -measure 指标值在绝大多数情况下均落后于 0-TSK-IDC.

4.3.2 真实数据集

本节给出 0-TSK-IDC、FS-FCSVM、BFCCL-TSK-FS、L2-TSK-FS、Adaboost 和 CS-SVM 算法在 4 个 UCI 医学诊断数据集上的性能比较, G -mean 和 F -measure 变化曲线如图 9 和图 10 所示. 从实验结果可以看出,

1) UCI 数据集上获得的实验结果与人工数据集的实验结果是一致的. 本文所提的 0-TSK-IDC 在不同程度的类别不平衡场合均表现出优良的分类性能; 特别对不平衡性较大的数据集, 0-TSK-IDC 更能体现其较强的鲁棒性. 其原因在于 0-TSK-IDC 模糊系统前件参数的学习使用了 BFCCL 聚类, BFCCL 聚类通过类别间竞争学习的策略能合理获取不平衡

两类上的空间划分, 能保证获得的模糊规则中的模糊集的准确性和可解释性; 0-TSK-IDC 模糊系统后件参数的学习从纠正分类面的偏移入手, 最终得到的分类面在达到两类间距离的最大化的同时, 保证少数类到分类面的距离不小于多数类到分类面的距离.

2) 真实数据集上类别间样本数量的不平衡性严重影响算法的分类性能, 如图 9 和图 10 所示, 随着实验中类别不平衡性的加剧, 三种模糊系统 FS-FCSVM、L2-TSK-FS 和 BFCCL-TSK-FS 得到的 G -mean 和 F -measure 值也急剧下降. 但由于 BFCCL-TSK-FS 能够借助 BFCCL 聚类算法得到准确的空间划分结果, 所以其分类效果要优于 FS-FCSVM 和 L2-TSK-FS, 具体表现为 BFCCL-TSK-FS 在 4 个 UCI 数据集上获得的 G -mean 和 F -measure 值均高于 FS-FCSVM 和 L2-TSK-FS 获得的值.

3) Adaboost 和 CS-SVM 算法在 G -mean 和 F -measure 值上得到的结果较为相近, 与 FS-FCSVM、L2-TSK-FS 和 BFCCL-TSK-FS 这三种模糊系统相比, 两者具有明显的优势; 但与 0-TSK-IDC 相比两者的性能略差. 值得注意的是, 0-TSK-IDC 模糊系统另一个突出优点是其构建的模糊规则具有高度的可解释性, 而这一特性是 Adaboost 和 CS-SVM 算法所不具备的.

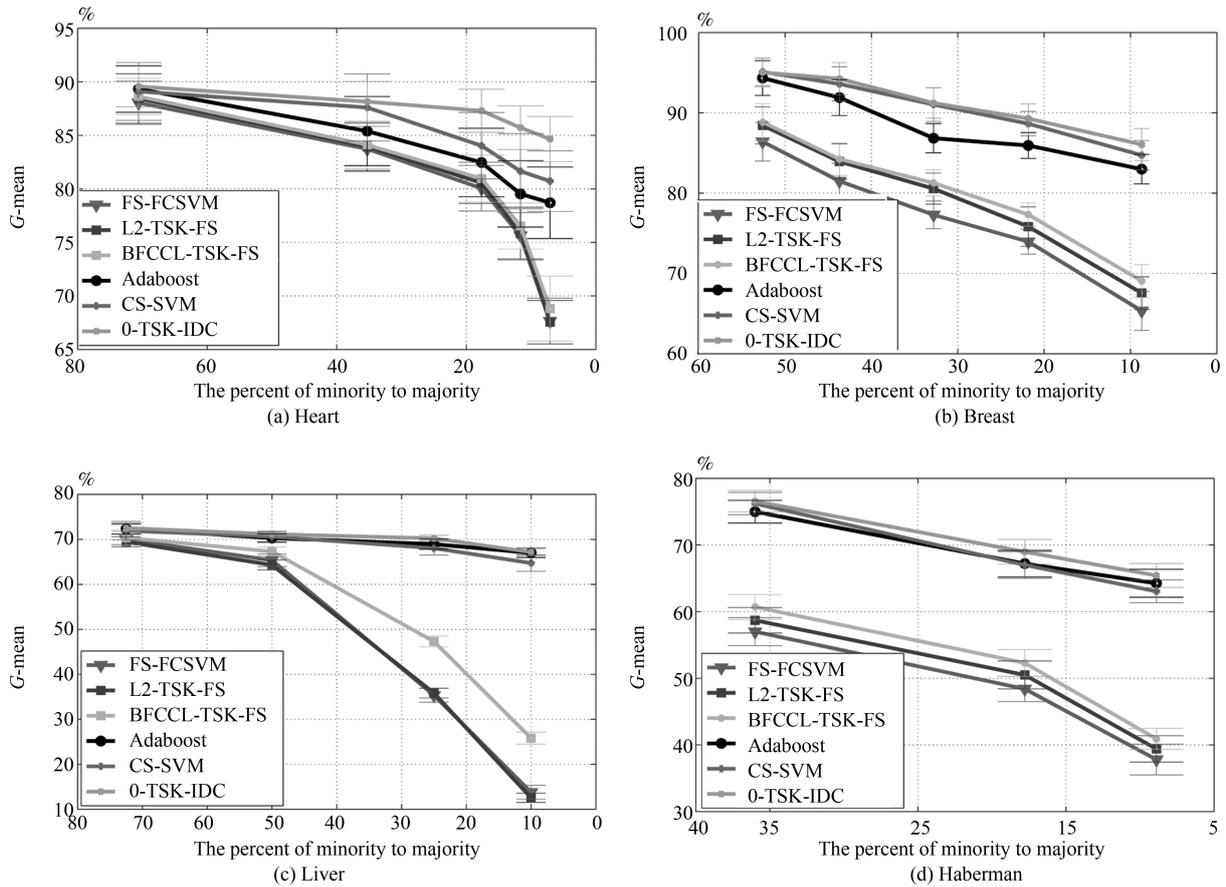


图 9 UCI 医学集上不同算法的 G -mean 比较

Fig. 9 G -mean and its standard deviation comparison of 0-TSK-IDC and other algorithms on UCI dataset

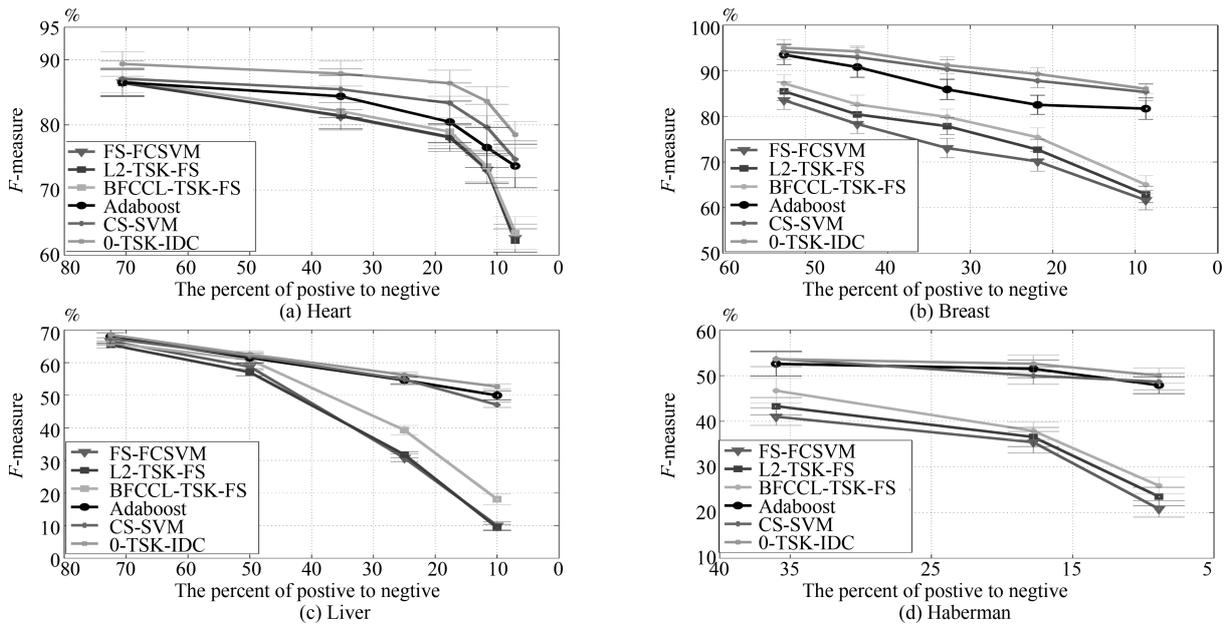


图 10 UCI 医学集上不同算法的 F -measure 比较

Fig. 10 F -measure and its standard deviation comparison of 0-TSK-IDC and other algorithms on UCI dataset

为了进一步考察在真实数据集上 0-TSK-IDC 构建的模糊规则的可解释性, 图 11 给出了 Heart 集正负类样本数 6:17、Breast 集正负类样本数 75:229、Liver 正负类样本数 1:4 和 Haberman 集正负类样本数 8:45 时, G -mean 指标随规则数变化的曲线. 可以看出, 图 9 中 0-TSK-IDC 模糊系统识别率最高时所需的模糊规则数分别是 7、12、7 和 6. 模糊系统的特点是建立分类性能和模糊规则数之间的平衡, 通常情况下规则数越少的模糊系统其可解释性越强, 图 11 的实验结果可以说明从规则数角度看 0-TSK-IDC 模糊系统具有高度的可解释性.

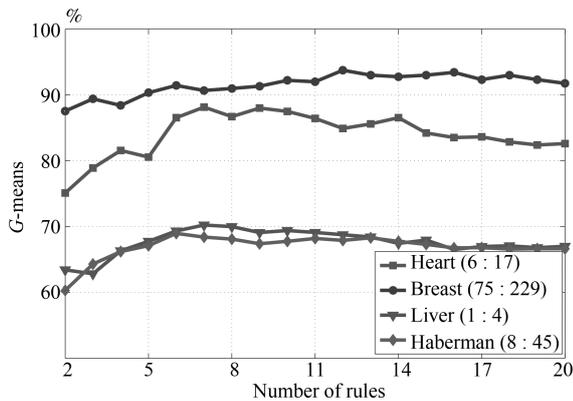


图 11 UCI 医学集上 G -mean 随规则数变化的示意图
Fig. 11 G -mean with the different fuzzy rules on UCI databases

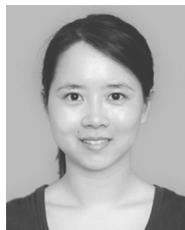
5 结论

本文提出的 0-TSK-IDC 模糊系统利用 BFCCL 算法进行模糊规则前件参数的学习, 使基于聚类中心竞争机制和概率模型的聚类算法在类别不平衡的空间划分中表现出了清晰性和可解释性; 0-TSK-IDC 模糊系统在模糊规则后件参数的学习中, 使用了“大间隔”的机制并设置少数类到分类面的距离大于多数类到分类面的距离, 从而使得 0-TSK-IDC 具有较强的泛化能力. 本文的主要贡献可以归纳为: 1) 建立一种利用概率模型改进模糊系统的框架; 2) 提出了一种利用聚类模糊系统解决不平衡分类问题的方法. 另外, 0-TSK-IDC 模糊系统亦可处理类别平衡的分类问题, 只要设置式 (23) 和 (24) 中参数 v 为 0 即可. 对于多分类问题, 0-TSK-IDC 可以采用类似于 SVM 方法, 用类别组合的方式来实现. 应当指出, 本文对 0-TSK-IDC 模糊系统能否有效解决大样本等问题没有进行深入探讨, 当样本容量极大时, 若从聚类速度和二次规划求解角度而言, 0-TSK-IDC 仍面临进一步提高实用性的挑战, 这将作为我们近期的研究重点.

References

- Richardson J, Korniak J, Reiner P D, Wilamowski B M. Nearest-neighbor spline approximation (NNSA) improvement to TSK fuzzy systems. *IEEE Transactions on Industrial Informatics*, 2016, **12**(1): 169–178
- Deng Z H, Cao L B, Jiang Y Z, Wang S T. Minimax probability TSK fuzzy system classifier: a more transparent and highly interpretable classification model. *IEEE Transactions on Fuzzy Systems*, 2015, **23**(4): 813–826
- Jia Li, Yang Ai-Hua, Qiu Ming-Sen. Research on multi-signal based neuro-fuzzy Hammerstein-Wiener model. *Acta Automatica Sinica*, 2013, **39**(5): 690–696 (贾立, 杨爱华, 邱铭森. 基于多信号源的神经模糊 Hammerstein-Wiener 模型研究. *自动化学报*, 2013, **39**(5): 690–696)
- Liu Y J, Tong S C, Chen C L P, Li D J. Neural controller design-based adaptive control for nonlinear MIMO systems with unknown hysteresis inputs. *IEEE Transactions on Cybernetics*, 2016, **46**(1): 9–19
- Cheng W Y, Juang C F. A fuzzy model with online incremental SVM and margin-selective gradient descent learning for classification problems. *IEEE Transactions on Fuzzy Systems*, 2014, **22**(2): 324–337
- Jiang Y Z, Chung F L, Ishibuchi H, Deng Z H, Wang S T. Multitask TSK fuzzy system modeling by mining intertask common hidden structure. *IEEE Transactions on Cybernetics*, 2015, **45**(3): 534–547
- Liu Y J, Tong S C. Adaptive fuzzy control for a class of unknown nonlinear dynamical systems. *Fuzzy Sets and Systems*, 2015, **263**: 49–70
- Wong S Y, Yap K S, Yap H J, Tan S C, Chang S W. On equivalence of FIS and ELM for interpretable rule-based knowledge representation. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, **26**(7): 1417–1430
- Leski J M. TSK-fuzzy modeling based on ϵ -insensitive learning. *IEEE Transactions on Fuzzy Systems*, 2005, **13**(2): 181–193
- Leski J M. Fuzzy $(c+p)$ -means clustering and its application to a fuzzy rule-based classifier: toward good generalization and good interpretability. *IEEE Transactions on Fuzzy Systems*, 2015, **23**(4): 802–812
- Fernández A, del Jesus M J, Herrera F. On the 2-tuples based genetic tuning performance for fuzzy rule based classification systems in imbalanced data-sets. *Information Sciences*, 2010, **180**(8): 1268–1291
- Fernández A, del Jesus M, Herrera F. Hierarchical fuzzy rule based classification systems with genetic rule selection for imbalanced data-sets. *International Journal of Approximate Reasoning*, 2009, **50**(3): 561–577
- Ramentol E, Caballero Y, Bello R, Herrera F. SMOTE-RSB*: a hybrid preprocessing approach based on oversampling and undersampling for high imbalanced data-sets using SMOTE and rough sets theory. *Knowledge and Information Systems*, 2012, **33**(2): 245–265
- López V, Fernández A, del Jesus M, Herrera F. A hierarchical genetic fuzzy system based on genetic programming for addressing classification with highly imbalanced and borderline datasets. *Knowledge Based Systems*, 2013, **38**: 85–104

- 15 Galar M, Fernández A, Barrenechea E, Herrera F. EUS-Boost: enhancing ensembles for highly imbalanced data-sets by evolutionary undersampling. *Pattern Recognition*, 2013, **46**(12): 3460–3471
- 16 Chawla N V, Bowyer K W, Hall L O, Kegelmeyer W P. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 2002, **16**(1): 321–357
- 17 Sun Y M, Kamel M S, Wong A K C, Wang Y. Cost-sensitive boosting for classification of imbalanced data. *Pattern Recognition*, 2007, **40**(12): 3358–3378
- 18 Tang Y C, Zhang Y Q, Chawla N V, Krasser S. SVMs modeling for highly imbalanced classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2009, **39**(1) 281–288
- 19 Deng Z H, Jiang Y Z, Chung F L, Ishibuchi H, Wang S T. Knowledge-leverage-based fuzzy system and its modeling. *IEEE Transactions on Fuzzy Systems*, 2013, **21**(4): 597–609
- 20 Zhu L, Chung F L, Wang S T. Generalized fuzzy C-means clustering algorithm with improved fuzzy partitions. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2009, **39**(3): 578–591
- 21 Deng Z H, Choi K S, Chung F L, Wang S T. Enhanced soft subspace clustering integrating within-cluster and between-cluster information. *Pattern Recognition*, 2010, **43**(3): 767–781
- 22 Glenn T C, Zare A, Gader P D. Bayesian fuzzy clustering. *IEEE Transactions on Fuzzy Systems*, 2015, **23**(5): 1545–1561
- 23 Jiang Yi-Zhang, Deng Zhao-Hong, Wang Shi-Tong. Mamdani-Larsen type transfer learning fuzzy system. *Acta Automatica Sinica*, 2012, **38**(9): 1393–1409 (蒋亦樟, 邓赵红, 王士同. ML 型迁移学习模糊系统. *自动化学报*, 2012, **38**(9): 1393–1409)
- 24 Azeem M F, Hanmandlu M, Ahmad N. Generalization of adaptive neuro-fuzzy inference systems. *IEEE Transactions on Neural Networks*, 2000, **11**(6): 1332–1346
- 25 Deng Z H, Choi K S, Chung F L, Wang S T. Scalable TSK fuzzy modeling for very large datasets using minimal-enclosing-ball approximation. *IEEE Transactions on Fuzzy Systems*, 2011, **19**(2): 210–226
- 26 Hall L O, Goldgof D B. Convergence of the single-pass and online fuzzy C-means algorithms. *IEEE Transactions on Fuzzy Systems*, 2011, **19**(4): 792–794
- 27 Meyn S P, Tweedie R L. *Markov Chains and Stochastic Stability*. London: Springer, 1993.
- 28 Nesterov Y. *Introductory Lectures on Convex Optimization: A Basic Course*. US: Springer, 2004.
- 29 Vapnik V N. *Statistical Learning Theory*. New York: Wiley, 1998.
- 30 Ni T G, Chung F L, Wang S T. Support vector machine with manifold regularization and partially labeling privacy protection. *Information Sciences*, 2015, **294**: 390–407
- 31 UCI database [Online], available: <http://www.ics.uci.edu/>.
- 32 Juang C F, Chiu S H, Shiu S J. Fuzzy system learned through fuzzy clustering and support vector machine for human skin color segmentation. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 2007, **37**(6): 1077–1087
- 33 Wang S, Yao X. Multiclass imbalance problems: analysis and potential solutions. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2012, **42**(4): 1119–1130
- 34 Masnadi-Shirazi H, Vasconcelos N, Iranmehr A. Cost-sensitive support vector machines. *Journal of Machine Learning Research*, 2012, arXiv: 1212.0975
- 35 Bezdek J C. A physical interpretation of fuzzy ISODATA. *IEEE Transactions on Systems, Man, and Cybernetics*, 1976, **SMC-6**(5): 387–389
- 36 Sun Z B, Song Q B, Zhu X Y, Sun H L, Xu B W, Zhou Y M. A novel ensemble method for classifying imbalanced data. *Pattern Recognition*, 2015, **48**(5): 1623–1637
- 37 Parambath S A P, Usunier N, Grandvalet Y. Optimizing F-measures by cost-sensitive classification. In: *Proceedings of Advances in Neural Information Processing Systems 27*. Montreal, Canada: Curran Associates, Inc., 2014. 2123–2131



顾晓清 常州大学信息科学与工程学院讲师。江南大学数字媒体学院博士研究生。主要研究方向为模式识别, 机器学习。本文通信作者。

E-mail: czxqgu@163.com

(GU Xiao-Qing Lecturer at the School of Information Science and Engineering, Changzhou University. Ph. D. candidate at the School of Digital Media, Jiangnan University. Her research interest covers pattern recognition and machine learning. Corresponding author of this paper.)



蒋亦樟 江南大学数字媒体学院讲师。2016 年获得江南大学数字媒体学院博士学位。主要研究方向为人工智能, 模式识别, 模糊系统。

E-mail: s101914015@vip.jiangnan.edu.cn

(JIANG Yi-Zhang Lecturer at the School of Information Science and Engineering. He received his Ph. D. degree from the School of Digital Media, Jiangnan University in 2016. His research interest covers artificial intelligence, pattern recognition, and fuzzy system.)



王士同 江南大学数字媒体学院教授。主要研究方向为人工智能, 神经网络, 模式识别。

E-mail: wxwangst@aliyun.com

(WANG Shi-Tong Professor at the School of Digital Media, Jiangnan University. His research interest covers artificial intelligence, neuro-fuzzy systems, and pattern recognition.)