

基于标记分布学习的异态集成学习算法

王一宾^{1,2} 田文泉¹ 程玉胜^{1,2}

摘 要 为了提高预测的准确性,文中结合机器学习中堆积(Stacking)集成框架,组合多个分类器对标记分布进行学习,提出基于标记分布学习的异态集成学习算法(HELA-LDL).算法构造两层模型框架,通过第一层结构将样本数据采用组合方式进行异态集成学习,融合各分类器的学习结果,将融合结果输入到第二层分类器,预测结果是带有置信度的标记分布.在专用数据集上的对比实验表明,HELA-LDL 可以发挥各种算法在不同场景下的性能较优,稳定性分析进一步说明算法的有效性.

关键词 标记分布学习, 异态集成学习, 回归拟合, 堆积

引用格式 王一宾,田文泉,程玉胜.基于标记分布学习的异态集成学习算法.模式识别与人工智能,2019,32(10):945-954.

DOI 10.16451/j.cnki.issn1003-6059.201910009

中图法分类号 TP 18

Heterogeneous Ensemble Learning Algorithm Based on Label Distribution Learning

WANG Yibin^{1,2}, TIAN Wenquan¹, CHENG Yusheng^{1,2}

ABSTRACT To improve prediction accuracy, a stacking integration framework in machine learning is employed to learn label distribution through multiple classifiers, and a heterogeneous ensemble learning algorithm based on label distribution learning(HELA-LDL) is proposed. A two-layer model framework is constructed, and the sample data are combined through the first layer structure to integrate the learning results of each classifier. Finally, the fusion results are input to the second layer classifier as the original feature, and the labels are predicted to be a label distribution with confidence. Comparative experiments on specialized datasets show that HELA-LDL is superior to other algorithms in various scenes. The stability analysis further illustrates the effectiveness of HELA-LDL.

Key Words Label Distribution Learning, Heterogeneous Ensemble Learning, Regression Fitting, Stacking

Citation WANG Y B, TIAN W Q, CHENG Y S. Heterogeneous Ensemble Learning Algorithm Based on Label Distribution Learning. Pattern Recognition and Artificial Intelligence, 2019, 32(10):945-954.

收稿日期:2018-12-28;录用日期:2019-07-08

Manuscript received December 28, 2018;

accepted July 8, 2019

安徽省高校重点科研项目(No. KJ2017A352)、安徽省高校重点实验室基金项目(No. ACAIM160102)资助

Supported by Natural Science Research Funds of Education Department of Anhui Province (No. KJ2017A352), Foundation Project of Key Laboratories of Anhui Province(No. ACAIM160102)

本文责任编辑 陈松灿

Recommended by Associate Editor CHEN Songcan

1. 安庆师范大学 计算机与信息学院 安庆 246133

在标记学习中,一个示例可能具有多个标记^[1],并且不同标记对于同一个实例的描述程度并不都相同,有一定的差别.针对标记存在主次等问题,耿新等^[2-3]提出标记分布学习算法,使用类似概率的形式对传统的-1和1标记进行描述,对示例的

2. 安庆师范大学 安徽省高校智能感知与计算重点实验室 安庆 246133

1. School of Computer and Information, Anqing Normal University, Anqing 246133

2. University Key Laboratory of Intelligent Perception and Computing of Anhui Province, Anqing Normal University, Anqing 246133

标记分布情况进行准确直观表示.

标记分布学习在各领域发挥广泛作用^[4-8], 目前学者们已经提出问题转化贝叶斯 (Problem Transformation Bayes, PT-Bayes)、问题转化支持向量机 (Problem Transformation Support Vector Machine, PT-SVM)、算法适应 K 近邻 (Algorithm Adaptation K Nearest Neighbors, AA-KNN)、算法适应反向传播 (Algorithm Adaptation Backpropagation, AA-BP) 等应用算法. 同时针对标记分布学习提出改进迭代尺度专用算法 (Specialized Algorithms Improved Iterative Scaling, SA-IIS), 利用迭代求解直接解决标记分布问题^[9]. 上述具有代表性的 5 种学习算法都是通过单个模型完成标记学习. 由于算法处理策略的不同, 在不同数据集上预测精度存在较大差异.

为了提高准确率, 现有机器学习大都通过选择合适的分类器以获得较优效果, 但不同的分类器由于方法各异导致分类效果不尽相同. 如何利用各分类器在不同评价指标上的各种优势, 保持分类器多样性、提高算法精度及泛化能力是研究重点.

学者们考虑到不同算法的各种优势, 研究基分类器融合的方法, 组合多个分类器以获得较优性能. 崔颖等^[10] 基于集成学习引入分类器加权组合思想, 提高分类器泛化能力. Tsoumakas 等^[11] 提出多标记的堆积 (Stacking) 分类方法, 研究对基分类器的融合方法, 提高模型整体性能. 张笑铭等^[12] 提出适用于卷积神经网络的 Stacking 算法, 将多个卷积神经网络作为基分类器, 提高分类精度. 周星等^[13] 以三种常用的集成方法, 从多角度研究分类器算法. 李巧等^[14] 根据集成学习的思想, 从模型融合角度, 研究提高准确率和鲁棒性的方法. 傅艺绮等^[15] 提出组合机器学习算法, 有效提高软件缺陷预测性能. 但是, 标记分布的标记值是连续型变量, 目前还无相关的异态集成研究算法.

基于此种情况, 本文组合多个分类器, 重点研究在不同分类器异态集成下提高标记分布学习准确率的方法, 提出基于标记分布的异态集成学习算法 (Heterogeneous Ensemble Learning Algorithm Based on Label Distribution Learning, HELA-LDL). 首先结合 Stacking 思想构造两层模型框架, 选择 K 近邻 (K Nearest Neighbors, KNN)、线性回归 (Linear Regression, LR)、随机森林 (Random Forest, RF) 和梯度提升树 (Gradient Boosting Decision Tree, GBDT) 作为第一层学习器, RF 作为第二层分类器. 然后将样本数据集分为训练集和测试集, 通过第一层多种学习算法在相同训练集上进行训练预测, 预测结果作

为中间特征融合输出. 最后将中间特征输入到第二层分类器, 得出最终结果. 采用不同分类器组合的方式可以提高算法整体的精确度, 预测结果为类概率形式, 使输出数据是带有置信度的预测值, 符合标记分布范式. 为了验证算法, 先对比异态集成算法与组成它的基分类器, 再通过专用化数据集与多种实验算法进行对比, 结果表明, 本文算法在多个标记分布评价指标上均取得较优效果.

1 标记分布学习

标记分布学习是以概率的形式将传统标记-1 和 1 进行转化的学习范式. 对于一个样本标记集合, 集合中的每个分量表示标记对样本示例描述的程度大小.

对于一个示例集合 x , 标记集合 y , 描述度 d_x^y 表示例 x 中标记分量 y 占有所有示例标记的比重, 值介于 0 和 1 之间. 标记分布学习通过训练集 S 建立 $P(y|x)$ 函数模型, d_x^y 满足的条件和概率分布一致, 即

$$d_x^y \in [0, 1], \sum_y d_x^y = 1.$$

标记分布学习可以完成标记数值的相互转化. 相比传统学习模型, 泛化程度更高, 在标记学习中应用更广泛.

2 集成学习

2.1 集成学习理论

集成学习最初来自于组合预测理论, 通过合理组合多个基分类器, 集成为一个整体学习系统, 达到提高模型预测精度的目的. 异态集成学习综合各种优点, 具有比单分类器更高的精确性、鲁棒性及泛化性能. 常见的集成算法有平均法、投票法、学习法等.

在二分类研究中, 假设有 N 个相互独立的基分类器, 每个分类器 C_m 的错误率为 ε , m 为第 m 个分类器, 则分类问题与真实函数 f 之间的关系表示为

$$p(C_m(x) \neq f(x)) = \varepsilon.$$

如果采用投票法集成多个模型, 集成模型的错误率为

$$p(C(x) \neq f(x)) = \sum_{m=0}^{N/2} \binom{N}{m} (1 - \varepsilon)^m \varepsilon^{N-m}.$$

当每个独立的分类器错误率 ε 都小于 0.5 时, 集成分类器的错误率随着 N 的增大而降低, 最终趋

于 0.

2.2 基分类器选择

为了提高算法模型的精确度,基分类器的选择具有 2 个原则:1)分类器具有多样性,保证异态集成分类器之间存在一定的差异性;2) 保证分类器的准确性,分类效果优于随机预测.

标记分布是一种类似概率分布的形式,对标记分布的预测可以采用回归函数拟合概率分布模型.集成学习可以通过对数据集的共同学习建立逼近样本真实分布的模型,因此本文采用 KNN、RF、LR 和 GBDT 分类器进行异态集成学习.通过 4 种不同策略在原始数据集的共同学习,将原始特征转化为高级特征.四种分类器算法的说明如下.

KNN 为基于示例之间距离的学习,在标记分布学习中,数据标签是连续的,可以使用基于邻居的回归,即对于示例 x ,最近 k 个邻居的平均标记分布作为分布的预测值.

2 个示例 $\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{in})$, $\mathbf{x}_j = (x_{j1}, x_{j2}, \cdots, x_{jn})$ 的欧几里得距离为

$$dis(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2}.$$

计算示例 \mathbf{x} 邻居的平均标记分布数值:

$$p(y_j | \mathbf{x}) = \frac{1}{k} \sum d_{x_i}^{y_j}, j = 1, 2, \cdots, c,$$

其中 c 为标记数目.

LR 为分析各变量之间关系的分析方法,函数关系如下.在标记分布学习中,利用特征及标记之间的相关性,建立线性回归模型:

$$y(\boldsymbol{\omega}, \mathbf{x}) = \omega_0 + \omega_1 x_1 + \cdots + \omega_n x_n,$$

$$y(\boldsymbol{\omega}, \mathbf{x}) = \boldsymbol{\omega}^T \mathbf{x} = \mathbf{W}^T \mathbf{X},$$

$$\mathbf{X} = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}, \mathbf{W} = \begin{bmatrix} \omega_0 \\ \omega_1 \\ \vdots \\ \omega_n \end{bmatrix}.$$

线性回归求解系数为 $\boldsymbol{\omega} = (\omega_1, \omega_2, \cdots, \omega_n)^T$ 的线性模型,就是解决 $\min \|\mathbf{X}\boldsymbol{\omega} - \mathbf{y}\|^2$ 问题.求解得

$$\boldsymbol{\omega} = (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}^T \mathbf{y}.$$

在标记分布中设数据集 $S = \{(\mathbf{x}_i, \mathbf{D}_i)\}_{i=1}^N$,其中 \mathbf{x}_i 为 n 维特征向量, \mathbf{D}_i 为分布式标记,通过已知分布求解 $\boldsymbol{\omega}$,预测输出标记分布,则标记分布的最小化目标函数:

$$f = \sum_i^N \|f(\mathbf{x}_i) - \mathbf{D}_i\|^2.$$

LR 用于建立分析特征与标记之间关系,针对标记分布的特点,LR 具有较快的速度、较高的准确度

和鲁棒性.

RF 在处理标记分布数据时,首先采用随机可重复取样的方式,抽取子样本.然后在子样本数据中,随机抽取部分特征建立决策树模型.假设随机森林中有 M 棵决策树,每棵树学习得到的模型为 f_m ,则在标记分布学习中,给定样本数据集 $S = \{(\mathbf{x}_i, \mathbf{D}_i)\}_{i=1}^N$,对于未知样本 \mathbf{x}' 的预测模型:

$$f = \frac{1}{M} \sum_{i=1}^M f_m(\mathbf{x}').$$

随着学习器数目的增加,RF 可以取得较小的泛化误差,产生较高准确度的分类器,并且能够处理多分类问题,在处理标记分布数据集,尤其当特征数目较大时具有较优结果.

GBDT 由多个分类器经过迭代形成强分类器的组合决策树模型,对所有决策树的输出进行累加,在当前模型基础上拟合回归树.

设含有 N 个样本的数据集 $S = \{(\mathbf{x}_i, \mathbf{D}_i)\}_{i=1}^N$,其中 \mathbf{x}_i 为 n 维特征向量, \mathbf{D}_i 为与示例 \mathbf{x}_i 的分布式标记,损失函数 $L(\mathbf{y}, F(\mathbf{x}))$ 可微,决策树的个数为 M .

使用常量 γ_0 初始化模型:

$$F_0(\mathbf{x}) = \arg \min \sum_{i=1}^n L(y_i, \gamma_0),$$

对于 $m = 1, 2, \cdots, M$,残差

$$\gamma_{im} = - \left[\frac{\partial L(y_i, F(\mathbf{x}_i))}{\partial F(\mathbf{x}_i)} \right]_{F(\mathbf{x}) = F_{m-1}(\mathbf{x})}, i = 1, 2, \cdots, n.$$

根据训练集构建弱分类器 $h_m(\mathbf{x})$,计算乘子

$$\gamma_m = \arg \min \sum_{i=1}^n L(y_i, F_{m-1}(\mathbf{x}_i) + \gamma h_m(\mathbf{x}_i)).$$

引入学习率 $v(0 < v < 1)$ 提高算法的泛化能力,修改后的更新规则为

$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + v \gamma_m h_m(\mathbf{x}).$$

GBDT 将弱分类器组合成强分类器,沿着梯度下降的方向构建新模型,不断更新,使模型最终趋于稳定.在处理标记分布学习时,可以有效处理数据中连续的特征,并充分分析考虑特征间的相互影响,最终输出预测结果.

3 基于标记分布的异态集成学习算法

在标记学习中,不同算法在各评价指标下具有不同优势,构建一个多分类器系统模型能够适应更广泛的场景.异态集成算法就是充分利用不同算法的优势,形成一个强大的框架.

利用集成技术构造标记学习模型能够改善模型的预测精度和泛化能力,本文利用 Stacking 思想构建标记分布学习模型.学习框架利用不同的基分类器组合学习,提高模型预测精度.

分类器集成的构建主要有 2 部分:生成基分类器和合并结果.一方面保证异态集成学习分类器的多样性,选择 KNN、LR、RF、GBDT 分类器方法训练样本,获得一组基分类器,另一方面组合信息的表示.由于标记分布具有类似概率的特点,将多个分类器的结果融合训练,使输出的标记为类概率形式,可得出更符合示例描述程度的分布,结果更准确.具体流程图如图 1 所示.

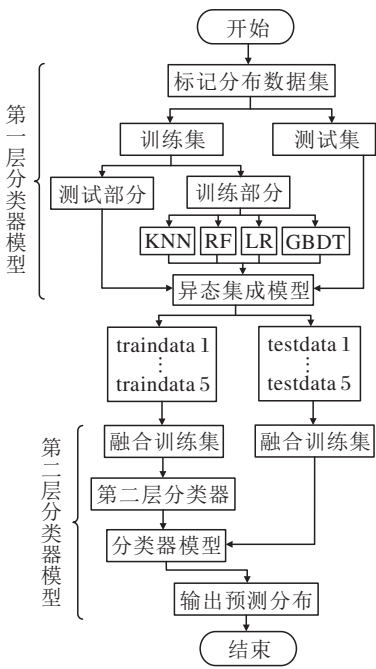


图 1 HELA-LDL 流程图

Fig. 1 Flow chart of HELA-LDL

算法分为两层分类器模型结构;第一层结构中各分类器对数据进行学习,各算法共同学习的结果是原始数据经过基分类器学习的高级特征,决定最终实验预测结果的精度;第二层结构学习高级特征,预测输出分布.

对于第一层模型,首先将原始标记分布数据集分为训练集和测试集,对训练集随机分成五份,一份作为测试部分,其余四份为训练部分.然后使用 KNN、LR、RF 和 GBDT 分类器在相同数据集上进行训练,生成多个基分类器.

使用基分类器模型对测试部分进行预测,生成数据集 traindata 1,完成第一份数据的预测工作.同

时预测测试集生成 testdata 1.重复 5 次分别生成 traindata 和对应的 testdata.合并 traindata 数据,得出训练集经过特征学习后的融合数据,记为融合训练集.将 testdata 数据集合并取平均,得出测试集经过特征学习后的融合数据,记为融合测试集.第一层结构的特征学习工作完成.

第一层输出作为中间特征输入到第二层结构,第二层结构选择 RF.融合训练集并建立随机森林回归模型,预测时将随机森林中各个树取平均值即为最终的预测标记值.

Stacking 可以保证集成后模型整体性能提高.采用类概率分布作为新数据的属性,使输出数据具有置信度的预测值.异态集成学习能综合各分类器的优点,使模型更符合标记分布范式.

基于标记分布的异态集成学习算法 (HELA-LDL) 步骤如下.

算法 HELA-LDL

输入 标记分布 train 为

$$S = \{(x_1, D_1), (x_2, D_2), \dots, (x_N, D_N)\};$$

标记分布 test 为

$$S^* = \{(x_1^*), (x_2^*), \dots, (x_M^*)\};$$

第一层学习算法 C_1, C_2, \dots, C_n ;

第二层学习算法 C

输出 预测分布 P

将标记分布 train 分为训练集和验证集

for 标记分布训练集 $x_i \in S$

$$h_n = C_n(S); \quad /* \text{每个基分类器训练模型} */$$

end

融合第一层实验数据集 S' ,输入到第二层训练

$$H = C(S') \quad /* \text{第二层分类器训练模型} */$$

for 标记分布 test $x_j^* \in S^*$

$$\text{测试集预测输出 } f(x^*) = H(x^*);$$

$/* \text{测试集预测输出标记分布} */$

end

$$P \leftarrow f(x^*)$$

4 实验及结果分析

4.1 实验数据集和评价指标

Yeast 数据集系列是关于酿酒酵母菌不同时间点基因表达水平的记录.s-JAFFE 数据集和 s-BU_3DFE 数据集为不同表情数据.Natural scene 数据集为自然场景图像.Movie 数据集为关于电影的 5 个评分级别.

表 1 实验数据集

Table 1 Experimental datasets

名称	样本数	特征数	标记数
Yeast-alpha	2465	24	18
Yeast-elu	2465	24	14
Yeast-diau	2465	24	7
Yeast-heat	2465	24	6
Yeast-spo	2465	24	6
Yeast-cold	2465	24	4
Yeast-dtt	2465	24	4
Yeast-spo5	2465	24	3
s-JAFFE	213	243	6
s-BU_3DFE	2500	243	6
Natural Scene	2000	294	9
Movie	7755	1869	5

标记分布算法预测结果是概率分布形式,对预测结果可以使用距离指标和相似度指标进行度量.选择 6 种典型评价指标,分别是距离指标(Chebyshev、Clark、Canberra、Kullback-Leibler)及相似度指标(Cosine、Intersection).预测分布与真实分布的距离越小、相似度越大,说明算法性能越优.表 2 中箭头↑表示表中数据越大越好,↓表示表中数据越小越好.

表 2 评价指标

Table 2 Evaluation indexes

评价指标	计算公式
Chebyshev ↓	$Dis_1(D, \hat{D}) = \max_j d_j - \hat{d}_j $
Clark ↓	$Dis_2(D, \hat{D}) = \sqrt{\sum_{j=1}^c \frac{(d_j - \hat{d}_j)^2}{(d_j + \hat{d}_j)^2}}$
Canberra ↓	$Dis_3(D, \hat{D}) = \sum_{j=1}^c \frac{ d_j - \hat{d}_j }{d_j + \hat{d}_j}$
Kullback-Leibler ↓	$Dis_4(D, \hat{D}) = \sum_{j=1}^c d_j \ln \left(\frac{d_j}{\hat{d}_j} \right)$
Cosine ↑	$Sim_1(D, \hat{D}) = \frac{\sum_{j=1}^c d_j \hat{d}_j}{\sqrt{\sum_{j=1}^c d_j^2} \sqrt{\sum_{j=1}^c \hat{d}_j^2}}$
Intersection ↑	$Sim_2(D, \hat{D}) = \sum_{j=1}^c \min(d_j, \hat{d}_j)$

4.2 异态集成算法实验结果

表 3 为在 s-JAFFE 数据集上采用两个分类器进行组合的实验结果对比.表 4 为在 s-JAFFE 数据集上采用三个分类器进行组合的实验结果对比.采用不同分类器的组合算法在各评价指标上最优值使用黑体数字表示.

表 3 两个分类器组合在 s-JAFFE 数据集上的测试结果

Table 3 Test results of 2 classifiers on s-JAFFE dataset

评价指标	RF+LR	RF+GBDT	RF+KNN	LR+GBDT	LR+KNN	GBDT+KNN
Chebyshev ↓	0.1349	0.1211	0.1200	0.1124	0.1217	0.1204
Clark ↓	0.4856	0.4473	0.4247	0.4112	0.4304	0.3968
Canberra ↓	0.9353	0.8580	0.8262	0.8137	0.8602	0.7996
Cosine ↑	0.9408	0.9383	0.9414	0.9569	0.9422	0.9537
Intersection ↑	0.8607	0.8538	0.8376	0.8578	0.8548	0.8649

表 4 三个分类器组合在 s-JAFFE 数据集上的测试结果

Table 4 Test results of 3 classifiers on s-JAFFE dataset

评价指标	RF+LR+GBDT	RF+LR+KNN	RF+GBDT+KNN	KNN+LR+GBDT
Chebyshev ↓	0.1061	0.1203	0.1122	0.1119
Clark ↓	0.4119	0.4135	0.4111	0.4065
Canberra ↓	0.7847	0.8108	0.8153	0.7908
Cosine ↑	0.9471	0.9388	0.9486	0.9498
Intersection ↑	0.8638	0.8570	0.8782	0.8675

通过表 3 和表 4 的分析对比可知,相比两个分类器组合,采用三个分类器组合的方式在各指标上具有一定优势.由表 3 和表 4 可知,在 Chebyshev、Canberra 和 Cosine 指标上,以 LR 和 GBDT 组合的分类器实验效果较好.

在 Clark 和 Intersection 指标上采用 GBDT 和 KNN 进行组合,效果较好.实验表明在 s-JAFFE 数据集上,加入 GBDT 的组合算法可以有效提高模型的预测精度,GBDT 与不同算法的组合可以在各评价指标上发挥良好效果.

4.3 对比算法实验结果

在 12 个数据集上对比 5 种算法,其中 SA-IIS 迭代次数设置为 50. 在每个数据集上采用十折交叉验证方式,记录各算法在不同评价指标上的平均值和标准差,并以统计形式给出排序. 实验结果如表 5 ~ 表 10 所示,黑体数字为最优值.

分析表 5 ~ 表 10 的实验结果,结论如下.
本文算法在 6 种评价指标上排名优于其它算法,各算法排名如下:

$$\text{HELA-LDL} > \text{AA-KNN} > \text{SA-IIS} > \text{PT-SVM} > \text{AA-BP} > \text{PT-Bayes}.$$

SA-IIS 在各评价指标上性能接近 AA-KNN. HELA-LDL 在 Yeast 数据集上数值优于各对比算法,说明算法针对标记分布数据集较有效. 在 Natural Scene 数据集,HELA-LDL 在 Clark、Canberra 指标上

性能略低于 AA-KNN,其它 4 个指标均最优,综合排名仍为第一.

在 s-JAFFE、s-BU_3DFE 数据集上,相比对比算法,HELA-LDL 在精度上有明显提高. 在 Movie 数据集上,样本及特征数目较多,HELA-LDL 预测结果仍然最好.

对比现有多个算法,HELA-LDL 在标记分布框架下具有较优效果,说明 HELA-LDL 的有效性. 异态集成学习能够融合各模型优势,有利于标记分布学习.

表 11 为 6 种算法的耗时对比. 采用集成学习能够提高整体精度,但是随着基分类器数目的增多,算法整体耗时增大. 通过对比发现 HELA-LDL 耗时高于 PT-Bayes 和 AA-KNN,但仍在合理接受范围,其耗时优于专用算法.

表 5 六种算法在 Chebyshev 距离上的测试结果
Table 5 Experimental results of 6 algorithms measured by Chebyshev distance

数据集	PT-Bayes	PT-SVM	AA-BP	AA-KNN	SA-IIS	HELA-LDL
Yeast-alpha	0.0996±0.0051	0.0139±0.0004	0.0367±0.0013	0.0146±0.0004	0.0170±0.0005	0.0131±0.0004
Yeast-elu	0.1121±0.0061	0.0171±0.0003	0.0389±0.0018	0.0176±0.0004	0.0203±0.0004	0.0161±0.0003
Yeast-diau	0.1577±0.0069	0.0443±0.0034	0.0501±0.0020	0.0391±0.0013	0.0412±0.0009	0.0354±0.0012
Yeast-heat	0.1741±0.0105	0.0441±0.0011	0.0530±0.0027	0.0447±0.0017	0.0466±0.0011	0.0401±0.0016
Yeast-spo	0.1754±0.0090	0.0633±0.0032	0.0671±0.0035	0.0629±0.0027	0.0613±0.0015	0.0567±0.0012
Yeast-cold	0.1831±0.0133	0.0571±0.0033	0.0591±0.0026	0.0550±0.0017	0.0566±0.0016	0.0503±0.0013
Yeast-dtt	0.1818±0.0134	0.0380±0.0016	0.0433±0.0015	0.0397±0.0017	0.0436±0.0013	0.0346±0.0019
Yeast-spo5	0.2013±0.0133	0.0925±0.0034	0.0942±0.0035	0.0970±0.0054	0.0950±0.0023	0.0888±0.0034
Natural Scene	0.4053±0.0250	0.4273±0.0328	0.3944±0.0172	0.3142±0.0131	0.3557±0.0139	0.2673±0.0131
s-JAFFE	0.1232±0.0083	0.1224±0.0080	0.1391±0.0129	0.0988±0.0139	0.1160±0.0141	0.0891±0.0085
s-BU_3DFE	0.1386±0.0037	0.1415±0.0058	0.1429±0.0064	0.1276±0.0035	0.1344±0.0050	0.1010±0.0030
Movie	0.2014±0.0028	0.2292±0.0191	0.1391±0.0034	0.1240±0.0026	0.1473±0.0020	0.1131±0.0020
算法排位	1.42	3.42	2.33	4.33	3.5	6

表 6 六种算法在 Clark 距离上的测试结果
Table 6 Experimental results of 6 algorithms measured by Clark distance

数据集	PT-Bayes	PT-SVM	AA-BP	AA-KNN	SA-IIS	HELA-LDL
Yeast-alpha	1.1729±0.0395	0.2217±0.0062	0.7349±0.0311	0.2305±0.0041	0.2614±0.0071	0.2052±0.0062
Yeast-elu	1.0343±0.0496	0.2115±0.0043	0.5449±0.0279	0.2171±0.0057	0.2404±0.0035	0.1972±0.0031
Yeast-diau	0.7531±0.0333	0.2371±0.0141	0.2753±0.0104	0.2110±0.0062	0.2215±0.0052	0.1927±0.0062
Yeast-heat	0.6838±0.0386	0.1902±0.0047	0.2321±0.0132	0.1939±0.0070	0.2010±0.0047	0.1743±0.0070
Yeast-spo	0.6843±0.0302	0.2675±0.0123	0.2907±0.0164	0.2668±0.0103	0.2624±0.0074	0.2431±0.0048
Yeast-cold	0.4954±0.0351	0.1542±0.0085	0.1609±0.0069	0.1497±0.0047	0.1532±0.0049	0.1370±0.0041
Yeast-dtt	0.4986±0.0358	0.1034±0.0044	0.1183±0.0042	0.1075±0.0051	0.1172±0.0039	0.0939±0.0057
Yeast-spo5	0.4180±0.0278	0.1864±0.0080	0.1895±0.0073	0.1958±0.0113	0.1914±0.0051	0.1799±0.0081
Natural Scene	2.5206±0.0275	2.5662±0.0330	2.4994±0.0212	1.9384±0.0377	2.4715±0.0169	1.9736±0.0182
s-JAFFE	0.4397±0.0152	0.4470±0.0161	0.5229±0.0487	0.3485±0.0306	0.4155±0.0256	0.3321±0.0208
s-BU_3DFE	0.4125±0.0070	0.4236±0.0151	0.4654±0.0199	0.4031±0.0087	0.4138±0.0072	0.3234±0.0086
Movie	0.8065±0.0086	0.8323±0.0800	0.6405±0.0165	0.5488±0.0102	0.5824±0.0076	0.5095±0.0067
算法排位	1.58	3.33	2.42	4.17	3.58	5.92

表 7 六种算法在 Canberra 标准上的测试结果

Table 7 Experimental results of 6 algorithms measured by Canberra metric

数据集	PT-Bayes	PT-SVM	AA-BP	AA-KNN	SA-IIS	HELA-LDL
Yeast-alpha	4.1937 ±0.1492	0.7251 ±0.0205	2.4208 ±0.0956	0.7532 ±0.0140	0.8618 ±0.0232	0.6679 ±0.0205
Yeast-elu	3.2778 ±0.1712	0.6242 ±0.0134	1.5983 ±0.0830	0.6419 ±0.0179	0.7125 ±0.0093	0.5779 ±0.0097
Yeast-diau	1.7086 ±0.0840	0.5096 ±0.0333	0.5944 ±0.0207	0.4539 ±0.0151	0.4787 ±0.0101	0.4128 ±0.0126
Yeast-heat	1.4440 ±0.0827	0.3798 ±0.0110	0.4671 ±0.0236	0.3903 ±0.0135	0.4042 ±0.0098	0.3465 ±0.0124
Yeast-spo	1.4431 ±0.0689	0.5504 ±0.0253	0.5951 ±0.0337	0.5490 ±0.0220	0.5394 ±0.0159	0.4983 ±0.0101
Yeast-cold	0.8669 ±0.0620	0.2660 ±0.0150	0.2773 ±0.0123	0.2593 ±0.0072	0.2649 ±0.0085	0.2357 ±0.0076
Yeast-dtt	0.8719 ±0.0649	0.1779 ±0.0083	0.2041 ±0.0076	0.1847 ±0.0081	0.2027 ±0.0072	0.1613 ±0.0058
Yeast-spo5	0.6488 ±0.0450	0.2862 ±0.0113	0.2911 ±0.0112	0.3003 ±0.0170	0.2939 ±0.0077	0.2759 ±0.0116
Natural Scene	7.1395 ±0.1106	7.2686 ±0.1413	6.9622 ±0.0905	4.6736 ±0.1066	6.8127 ±0.0721	4.8120 ±0.0656
s-JAFFE	0.9225 ±0.0391	0.9335 ±0.0314	1.0716 ±0.1133	0.7142 ±0.0640	0.8620 ±0.0580	0.6715 ±0.0409
s-BU_3DFE	0.9025 ±0.0160	0.9154 ±0.0315	0.9823 ±0.0385	0.8315 ±0.0193	0.8969 ±0.0180	0.6698 ±0.0166
Movie	1.5635 ±0.0188	1.6277 ±0.1939	1.2238 ±0.0305	1.0553 ±0.021	1.1199 ±0.0164	0.9720 ±0.0153
算法排位	1.5	3.42	2.16	4.42	3.58	5.92

表 8 六种算法在 Kullback-Leibler 散度上的测试结果

Table 8 Experimental results of 6 algorithms measured by Kullback-Leibler divergence

数据集	PT-Bayes	PT-SVM	AA-BP	AA-KNN	SA-IIS	HELA-LDL
Yeast-alpha	0.2776 ±0.0238	0.0060 ±0.0003	0.0872 ±0.9475	0.0065 ±0.0002	0.0085 ±0.0004	0.0055 ±0.0005
Yeast-elu	0.2825 ±0.0328	0.0069 ±0.0003	0.0608 ±0.0092	0.0073 ±0.0004	0.0091 ±0.0003	0.0061 ±0.0009
Yeast-diau	0.2698 ±0.0276	0.0177 ±0.0019	0.0264 ±0.0024	0.0149 ±0.0009	0.0157 ±0.0006	0.0124 ±0.0013
Yeast-heat	0.2684 ±0.0345	0.0138 ±0.0006	0.0216 ±0.0029	0.0144 ±0.0010	0.0152 ±0.0007	0.0115 ±0.0010
Yeast-spo	0.2788 ±0.0398	0.0292 ±0.0028	0.0332 ±0.0039	0.0291 ±0.0023	0.0268 ±0.0015	0.0244 ±0.0012
Yeast-cold	0.2174 ±0.0356	0.0147 ±0.0013	0.0162 ±0.0016	0.0139 ±0.0011	0.0146 ±0.0010	0.0117 ±0.0013
Yeast-dtt	0.2264 ±0.0393	0.0068 ±0.0007	0.0089 ±0.0006	0.0074 ±0.0008	0.0080 ±0.0006	0.0058 ±0.0017
Yeast-spo5	0.2066 ±0.0427	0.0299 ±0.0023	0.0312 ±0.0023	0.0347 ±0.0039	0.0314 ±0.0015	0.0280 ±0.0020
Natural Scene	2.2634 ±0.4388	1.5258 ±0.2236	1.1608 ±0.0725	1.1154 ±0.0596	0.9396 ±0.0272	0.7923 ±0.0540
s-JAFFE	0.0763 ±0.0063	0.0790 ±0.0058	0.1173 ±0.0245	0.0538 ±0.0107	0.0693 ±0.0116	0.0464 ±0.0078
s-BU_3DFE	0.0849 ±0.0029	0.0909 ±0.0048	0.1023 ±0.0096	0.0818 ±0.0039	0.0819 ±0.0040	0.0528 ±0.0028
Movie	0.7297 ±0.0590	0.2938 ±0.0697	0.1664 ±0.0107	0.1177 ±0.0051	0.1317 ±0.0046	0.0955 ±0.0040
算法排位	1.33	3.58	2.42	3.92	3.75	6

表 9 六种算法在 Cosine 相关性上的测试结果

Table 9 Experimental results of 6 algorithms measured by Cosine coefficient

数据集	PT-Bayes	PT-SVM	AA-BP	AA-KNN	SA-IIS	HELA-LDL
Yeast-alpha	0.8485 ±0.0070	0.9941 ±0.0003	0.9475 ±0.0032	0.9936 ±0.0002	0.9914 ±0.0004	0.9948 ±0.0003
Yeast-elu	0.8531 ±0.0092	0.9934 ±0.0002	0.9597 ±0.0039	0.9929 ±0.0004	0.9910 ±0.0003	0.9941 ±0.0002
Yeast-diau	0.8638 ±0.0070	0.9836 ±0.0018	0.9773 ±0.0017	0.9863 ±0.0009	0.9853 ±0.0005	0.9887 ±0.0007
Yeast-heat	0.8668 ±0.0103	0.9868 ±0.0006	0.9803 ±0.0020	0.9863 ±0.0009	0.9854 ±0.0006	0.9891 ±0.0008
Yeast-spo	0.8611 ±0.0093	0.9725 ±0.0027	0.9695 ±0.0032	0.9727 ±0.0022	0.9747 ±0.0013	0.9779 ±0.0006
Yeast-cold	0.8936 ±0.0095	0.9860 ±0.0013	0.9848 ±0.0014	0.9868 ±0.0008	0.9861 ±0.0008	0.9889 ±0.0007
Yeast-dtt	0.8950 ±0.0098	0.9935 ±0.0005	0.9916 ±0.0005	0.9929 ±0.0006	0.9915 ±0.0005	0.9945 ±0.0007
Yeast-spo5	0.8980 ±0.0106	0.9735 ±0.0018	0.9723 ±0.0018	0.9694 ±0.0033	0.9721 ±0.0012	0.9757 ±0.0017
Natural Scene	0.5589 ±0.0133	0.4567 ±0.0610	0.5925 ±0.0196	0.7041 ±0.0129	0.6659 ±0.0107	0.7736 ±0.0146
s-JAFFE	0.9281 ±0.0063	0.9258 ±0.0056	0.8991 ±0.0189	0.9484 ±0.0109	0.9347 ±0.0112	0.9560 ±0.0077
s-BU_3DFE	0.9179 ±0.0028	0.9128 ±0.0045	0.9037 ±0.0069	0.9202 ±0.0035	0.9203 ±0.0036	0.9473 ±0.0029
Movie	0.8495 ±0.0022	0.7810 ±0.0495	0.9028 ±0.0045	0.9224 ±0.0032	0.9081 ±0.0030	0.9369 ±0.0025
算法排位	1.5	3.33	2.25	4.25	3.67	6

表 10 六种算法在 Intersection 相似性上的测试结果

数据集	PT-Bayes	PT-SVM	AA-BP	AA-KNN	SA-IIS	HELA-LDL
Yeast-alpha	0.7725 ±0.0078	0.9600 ±0.0011	0.8740 ±0.0043	0.9584 ±0.0008	0.9518 ±0.0013	0.9630 ±0.0011
Yeast-elu	0.7727 ±0.0109	0.9559 ±0.0009	0.8917 ±0.0052	0.9547 ±0.0013	0.9491 ±0.0007	0.9593 ±0.0010
Yeast-diau	0.7690 ±0.0108	0.9289 ±0.0048	0.9177 ±0.0028	0.9370 ±0.0022	0.9331 ±0.0013	0.9428 ±0.0018
Yeast-heat	0.7708 ±0.0122	0.9376 ±0.0019	0.9235 ±0.0035	0.9359 ±0.0022	0.9332 ±0.0016	0.9433 ±0.0020
Yeast-spo	0.7691 ±0.0108	0.9089 ±0.0043	0.9022 ±0.0053	0.9095 ±0.0037	0.9109 ±0.0026	0.9178 ±0.0014
Yeast-cold	0.7965 ±0.0139	0.9341 ±0.0038	0.9317 ±0.0031	0.9360 ±0.0016	0.9344 ±0.0020	0.9420 ±0.0018
Yeast-dtt	0.7964 ±0.0147	0.9561 ±0.0021	0.9496 ±0.0019	0.9544 ±0.0018	0.9496 ±0.0018	0.9601 ±0.0022
Yeast-spo5	0.7987 ±0.0133	0.9075 ±0.0034	0.9058 ±0.0035	0.9030 ±0.0054	0.9050 ±0.0023	0.9110 ±0.0033
Natural Scene	0.3500 ±0.0129	0.3483 ±0.0394	0.4278 ±0.0160	0.5571 ±0.0125	0.4631 ±0.0102	0.6181 ±0.0134
s-JAFFE	0.8430 ±0.0077	0.8410 ±0.0059	0.8181 ±0.0195	0.8763 ±0.0133	0.8536 ±0.0119	0.8860 ±0.0070
s-BU_3DFE	0.8388 ±0.0032	0.8356 ±0.0058	0.8236 ±0.0067	0.8479 ±0.0038	0.8394 ±0.0037	0.8791 ±0.0032
Movie	0.7226 ±0.0028	0.6870 ±0.0440	0.7976 ±0.0051	0.8224 ±0.0038	0.8032 ±0.0033	0.8397 ±0.0030
算法排位	1.5	3.33	2.25	4.33	3.59	6

表 11 六种算法在各数据集上耗时对比

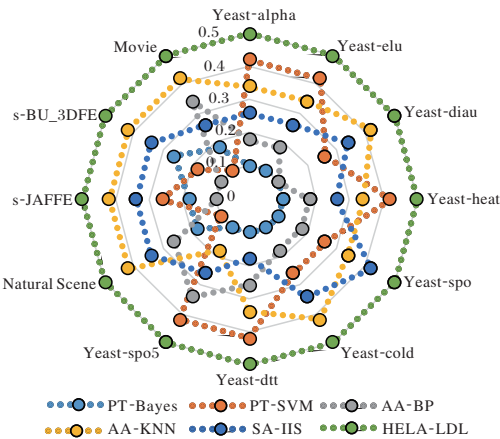
Table 11 Time consuming comparison of 6 algorithms on different datasets

数据集	PT-Bayes	PT-SVM	AA-KNN	AA-BP	SA-IIS	HELA-LDL
Yeast-alpha	11.91	864.87	16.86	740.25	220.17	164.07
Yeast-elu	11.77	515.76	17.29	720.92	184.77	124.88
Yeast-diau	11.71	105.26	17.09	877.36	99.25	57.45
Yeast-heat	10.96	74.69	17.19	941.35	91.91	51.64
Yeast-spo	11.05	74.13	17.01	652.81	57.26	41.64
Yeast-cold	10.76	302.17	16.96	598.26	59.82	27.46
Yeast-dtt	10.61	30.01	17.05	600.34	52.26	28.29
Yeast-spo5	10.61	15.74	17.20	998.17	34.47	20.52
s-JAFFE	1.29	74.81	1.04	156.97	544.60	20.47
s-BU_3DFE	13.05	824.04	23.63	893.96	430.11	383.61
Movie	146.26	1312.61	775.14	8081.5	4698.3	221.24
平均消耗	22.73	381.28	85.13	1387.45	588.44	103.75

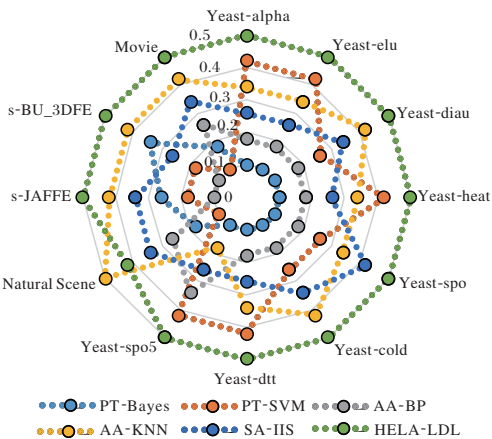
4.4 稳定性分析

为了验证标记分布中不同学习算法的稳定性, 本文利用蜘蛛网图^[17] 对各算法进行表示. 将不同评

价指标的结果标准化在[0.1,0.5] 内作为通用标准. 通过归一化后的值表示稳定指数. 图2显示 12 个数据集上 6 种算法的稳定性.



(a) Chebyshev



(b) Clark

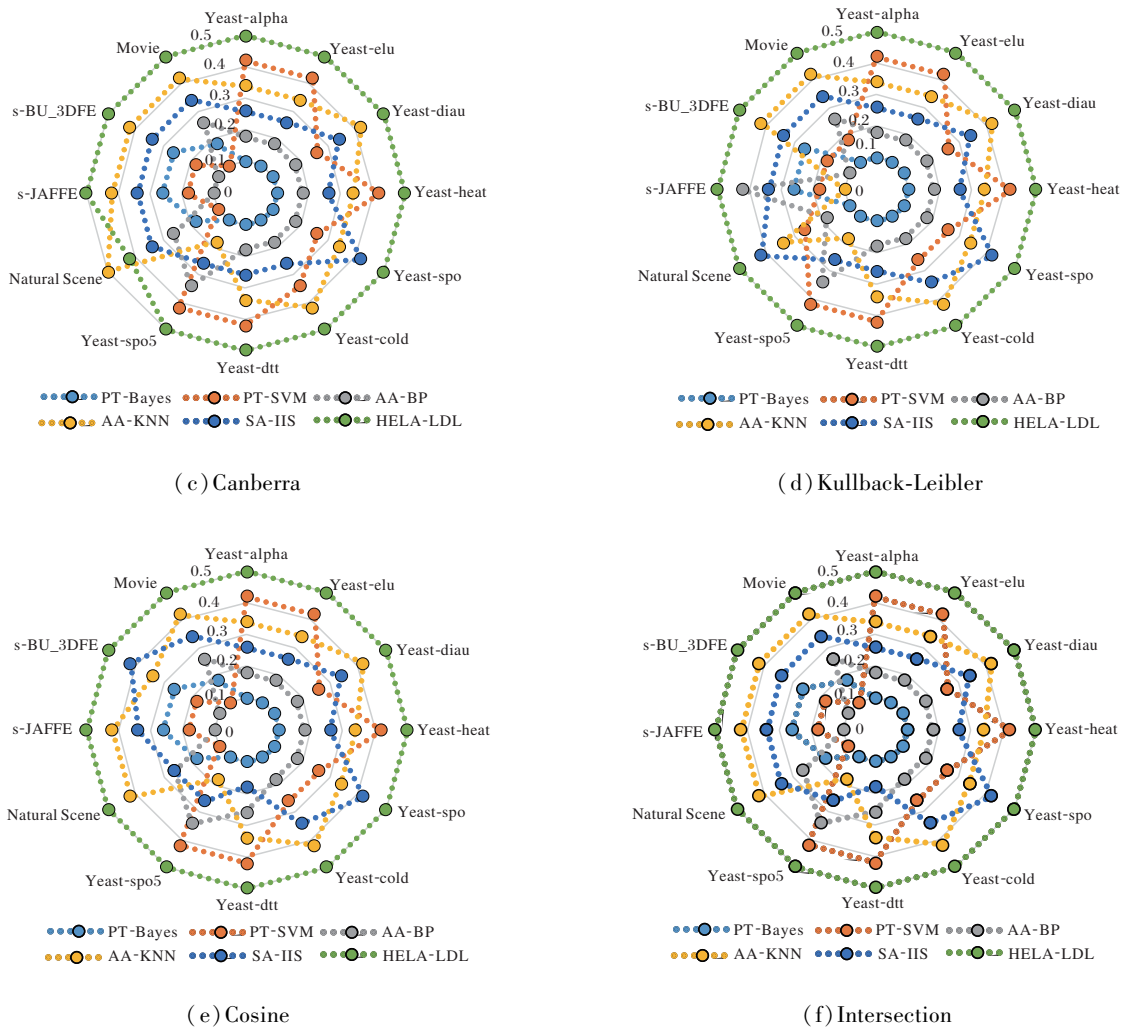


图 2 六种算法在 6 个评价指标上的稳定性指数值

Fig. 2 Values of stability index of 6 algorithms with 6 evaluation metrics

由图 2(a) 可知,在 Chebyshev 评价指标上, HELA-LDL 在 12 个数据集上平均精度的稳定指数值是 0.5,获得相当稳定的效果.由(b)、(c) 可知,在 Clark、Canberra 评价指标上, HELA-LDL 在自然场景数据集上性能略低于 AA-KNN,仍稳定在 $[0.4,0.5]$ 内,在其它数据集上稳定性指标均优于其它算法.由 (d) 可知,在 Kullback-Leibler 评价指标上, HELA-LDL 在 12 个数据集上平均精度的稳定指数值均为 0.5.由(e)、(f) 可知,在 Cosine、Intersection 评价指标上, HELA-LDL 在 12 个数据集上平均精度的稳定指数值均为 0.5,在各数据集上表现较稳定,稳定性指标均较优.

图 2 表明,相比现有多种算法, HELA-LDL 具有更好的稳定性,虽然在个别数据集上略低于 AA-KNN,但是稳定指数值波动不大,整体较优,说明 HELA-LDL 适合于标记分布学习,更有利于提高

整体性能.

5 结 束 语

在标记分布学习中,针对单个算法对预测模型性能产生影响的问题,本文提出基于标记分布的异态集成学习算法(HELA-LDL). 首先使用 KNN、LR、RF、GBDT 对原始数据特征进行学习预测,对结果融合输出作为中间特征,再将新特征学习通过 RF 构建标记分布预测模型. 在专用数据集上对比现有算法,结果表明, HELA-LDL 在不同规模的标记分布数据集上均能产生良好效果,尤其在人脸表情等真实数据集上均取得较优精度. 对算法的稳定性分析进一步说明算法的有效性.

由于基分类器学习的结果对预测模型的影响较

大,针对分布数据集选择准确度更高的学习算法可以有效提高模型整体的预测性能,下一步将研究组合更有效的学习算法.

参 考 文 献

[1] ZHANG M L, ZHOU Z H. A Review on Multi-label Learning Algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 2014, 26(8): 1819–1837.

[2] 耿 新,徐 宁,邵瑞枫. 面向标记分布学习的标记增强. *计算机研究与发展*, 2017, 54(6): 1171–1184.
(GENG X, XU N, SHAO R F. Label Enhancement for Label Distribution Learning. *Journal of Computer Research and Development*, 2017, 54(6): 1171–1184.)

[3] GENG X. Label Distribution Learning. *IEEE Transactions on Knowledge and Data Engineering*, 2016, 28(7): 1734–1748.

[4] GENG X, YIN C, ZHOU Z H. Facial Age Estimation by Learning from Label Distributions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(10): 2401–2412.

[5] HE Z Z, LI X, ZHANG Z F, *et al.* Data-Dependent Label Distribution Learning for Age Estimation. *IEEE Transactions on Image Processing*, 2017, 26(8): 3846–3858.

[6] ZHANG Z X, WANG M, GENG X. Crowd Counting in Public Video Surveillance by Label Distribution Learning. *Neurocomputing*, 2015, 166: 151–163.

[7] GAO B B, XING C, XIE C W, *et al.* Deep Label Distribution Learning with Label Ambiguity. *IEEE Transactions on Image Processing*, 2017, 26(6): 2825–2838.

[8] ZHOU Y, XUE H, GENG X. Emotion Distribution Recognition from Facial Expressions // *Proc of the 23rd ACM Conference on Multimedia*. New York, USA: ACM, 2015: 1247–1250.

[9] 赵 权,耿 新. 标记分布学习中目标函数的选择. *计算机科学与探索*, 2017, 11(5): 708–719.
(ZHAO Q, GENG X. Selection of Target Function in Label Distribution Learning. *Journal of Frontiers of Computer Science and Technology*, 2017, 11(5): 708–719.)

[10] 崔 颖,徐 凯,陆忠军,等. 主动学习策略融合算法在高光谱图像分类中的应用. *通信学报*, 2018, 39(4): 91–99.
(CUI Y, XU K, LU Z J, *et al.* Combination Strategy of Active Learning for Hyperspectral Images Classification. *Journal on Communications*, 2018, 39(4): 91–99.)

[11] TSOU MAKAS G, DIMOU A, SPYROMITROS E, *et al.* Correlation-Based Pruning of Stacked Binary Relevance Models for Multi-label Learning // *Proc of the 1st International Workshop on Learning from Multi-label Data*. Berlin, Germany: Springer, 2009: 101–116.

[12] 张笑铭,王志君,梁利平. 一种适用于卷积神经网络的 Stacking 算法. *计算机工程*, 2018, 44(4): 243–247.

(ZHANG X M, WANG Z J, LIANG L P. A Stacking Algorithm for Convolution Neural Network. *Computer Engineering*, 2018, 44(4): 243–247.)

[13] 周 星,丁立新,万润泽,等. 分类器集成算法研究. *武汉大学学报(理学版)*, 2015, 61(6): 503–508.
(ZHOU X, DING L X, WAN R Z, *et al.* Research on Classifier Ensemble Algorithms. *Journal of Wuhan University (Natural Science Edition)*, 2015, 61(6): 503–508.)

[14] 李 巧,周双娥,杨 晶. 模型融合在用户续购行为分析中的应用. *小型微型计算机系统*, 2017, 38(10): 2231–2235.
(LI Q, ZHOU S E, YANG J. Application of Model Blending in User Renewal Behavior Analysis. *Journal of Chinese Computer Systems*, 2017, 38(10): 2231–2235.)

[15] 傅艺绮,董 威,尹良泽,等. 基于组合机器学习算法的软件缺陷预测模型. *计算机研究与发展*, 2017, 54(3): 633–641.
(FU Y Q, DONG W, YIN L Z, *et al.* Software Defect Prediction Model Based on the Combination of Machine Learning Algorithms. *Journal of Computer Research and Development*, 2017, 54(3): 633–641.)

[16] WOLPERT D H. Stacked Generalization. *Neural Networks*, 1992, 5(2): 241–259.

[17] LIN Y J, LI Y W, WANG C X, *et al.* Attribute Reduction for Multi-label Learning with Fuzzy Rough Set. *Knowledge-Based Systems*, 2018, 152: 51–61.

作者简介



王一宾,硕士,教授,主要研究方向为多标记学习、机器学习、软件安全. E-mail: wangyb07@mail.ustc.edu.cn.
(WANG Yibin, master, professor. His research interests include multi-label learning, machine learning and software security.)



田文泉,硕士研究生,主要研究方向为机器学习、大数据、数据统计. E-mail: tianwq16@126.com.
(TIAN Wenquan, master student. His research interests include machine learning, big data and data statistics.)



程玉胜(通讯作者),博士,教授,主要研究方向为大数据、粗糙集、特征选择的机器学习. E-mail: Chengyshaq@163.com.
(CHENG Yusheng(Corresponding author), Ph. D. , professor. His research interests include big data, rough sets and machine learning for feature selection.)