

基于深度森林与异质集成的标记分布学习方法*

王艺霏, 祝继华, 刘新媛, 周熠扬

(西安交通大学 软件学院, 陕西 西安 710049)

通信作者: 祝继华, E-mail: zhujh@xjtu.edu.cn



摘要: 作为一种解决标签模糊性问题的新学习范式, 标记分布学习 (LDL) 近年来受到了广泛的关注. 为了进一步提升标记分布学习的预测性能, 提出一种联合深度森林与异质集成的标记分布学习方法 (LDLDF). 所提方法采用深度森林的级联结构模拟具有多层处理结构的深度学习模型, 在级联层中组合多个异质分类器增加集成的多样性. 相较于其他现有 LDL 方法, LDLDF 能够逐层处理信息, 学习更好的特征表示, 挖掘数据中丰富的语义信息, 具有强大的表示学习能力和泛化能力. 此外, 考虑到深层模型可能出现的模型退化问题, LDLDF 采用一种层特征重用机制 (layer feature reuse) 降低模型的训练误差, 有效利用深层模型每一层的预测能力. 大量的实验结果表明, 所提方法优于近期的同类方法.

关键词: 标记分布学习; 深度森林; 深度集成模型; 异质集成学习; 特征重用

中图法分类号: TP18

中文引用格式: 王艺霏, 祝继华, 刘新媛, 周熠扬. 基于深度森林与异质集成的标记分布学习方法. 软件学报, 2024, 35(7): 3410–3427. <http://www.jos.org.cn/1000-9825/6936.htm>

英文引用格式: Wang YF, Zhu JH, Liu XY, Zhou YY. Label Distribution Learning Method Based on Deep Forest and Heterogeneous Ensemble. Ruan Jian Xue Bao/Journal of Software, 2024, 35(7): 3410–3427 (in Chinese). <http://www.jos.org.cn/1000-9825/6936.htm>

Label Distribution Learning Method Based on Deep Forest and Heterogeneous Ensemble

WANG Yi-Fei, ZHU Ji-Hua, LIU Xin-Yuan, ZHOU Yi-Yang

(School of Software Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: As a new learning paradigm to solve the problem of label ambiguity, label distribution learning (LDL) has received much attention in recent years. To further improve the prediction performance of LDL, this study proposes an LDL based on deep forest and heterogeneous ensemble (LDLDF), which uses the cascade structure of deep forest to simulate deep learning models with multi-layer processing structure and combines multiple heterogeneous classifiers in the cascade layer to increase the diversity of ensemble. Compared with other existing LDL methods, LDLDF can process information layer by layer and learn better feature representations to mine rich semantic information in data, and it has better representation learning ability and generalization ability. In addition, by considering the degradation problem of deep models, LDLDF adopts a layer feature reuse mechanism to reduce the training error of the model, which effectively utilizes the prediction ability of each layer in the deep model. Sufficient experimental results show that LDLDF is superior to other methods.

Key words: label distribution learning (LDL); deep forest; deep ensemble model; heterogeneous ensemble learning; feature reuse

传统的单标记学习 (single-label learning, SLL) 和多标记学习 (multi-label learning, MLL)^[1,2] 分别假设每个实例与一个或多个标签相关联. 而在现实世界中, 不同标签对同一实例的描述可能存在程度上的差别^[3]. 为了更好地解决这种标签模糊性问题, Geng 等人^[3] 提出的标记分布学习 (label distribution learning, LDL) 考虑到不同标签的相对重要性, 为每个实例分配一个代表标签描述程度的向量. 这种新型的学习范式广泛适用于许多具有标签模糊性

* 基金项目: 国家重点基础研究发展计划 (2020AAA0109602)

收稿时间: 2022-11-22; 修改时间: 2023-01-04, 2023-03-08; 采用时间: 2023-03-15; jos 在线出版时间: 2023-08-23

CNKI 网络首发时间: 2023-08-25

的场景.

近年来, 一些学者对标记分布学习进行研究并提出了许多高效算法, 有效地将该范式应用于机器学习的多个领域^[4-6], 这些算法主要分为3类: 第1类是基于问题转换的算法, 将标记分布学习问题转换成单标记或多标记问题进行求解, 例如 PT-Bayes^[3], 该算法假设每类标记服从 Gaussian 分布, 通过 Bayes 公式计算每个类的后验概率作为对应标记的描述度. 第2类是自适应算法, 通过扩展已有算法解决标记分布学习问题, 例如 LDLogitBoost^[7], 该算法采用基于二阶导数的 Hessian 矩阵, 在函数空间上应用坐标下降优化的方法, 结合 Boosting 和逻辑回归, 提出了两种扩展了传统最大熵模型的 LDL 模型. 第3类是专门用于处理标记分布学习问题的算法, 例如 IIS-LLD^[4], 该算法基于最大熵模型进行迭代优化, 用于解决面部年龄估计问题.

考虑到深度神经网络 (DNNs) 在当前机器学习的各个领域都取得了良好效果^[8,9], 也有一些标记分布学习算法采用了 DNNs 模型, 例如 Gao 等人^[10]的研究提出了深度标记分布学习算法 (deep label distribution learning, DLDL), 主要应用于年龄估计和头部姿态估计. 深度神经网络有着强大的表示学习能力, 同时也存在着很多弊端, 如超参数多、可解释性弱、数据规模大等. 这样的矛盾激发了很多非神经网络的深层模型快速发展, 值得一提的是 Zhou 等人^[11]提出的深度森林模型. 深度森林是一种基于树集成的深层模型, 可以像深度神经网络模型一样进行表示学习, 具有超参数少、预测精度高、不依赖反向传播的优势. 由此本文提出一种联合深度森林与异质集成的标记分布学习方法 (label distribution learning based on deep forest and heterogeneous ensemble, LDLDF). 该方法采用深度森林的级联结构模拟具有多层处理结构的深度学习模型, 并在级联层中组合多个异质分类器增加集成的多样性. 级联结构赋予模型强大的表示学习能力, 让模型可以像深度神经网络一样逐层处理和转换特征, 挖掘数据中丰富的语义信息. 考虑到具有一定的复杂性的模型层可能会产生更好的拟合结果, 我们在 LDLDF 中融入异质集成学习 (heterogeneous ensemble learning) 的思想, 将级联结构中每层的分类器设置为多种异质分类器的组合, 如随机森林 (random forest, RF)^[12]、极端梯度提升树 (XGBoost, XGB)^[13]、线性回归 (linear regression, LR)、 k 近邻 (k -nearest neighbour algorithm, KNN), 通过集成不同算法的优势来提升模型性能. 同时, 考虑到 LDLDF 作为深层模型可能出现的模型退化问题, 受到残差网络^[14]和置信度筛选^[15]的启发, 本文设计一种层特征重用机制 (layer feature reuse), 通过比较模型在当前层和上一层的预测性能, 重用模型对部分实例更优的历史预测结果.

本文主要贡献如下.

(1) 提出一种联合深度森林与异质集成的标记分布学习方法 (LDLDF). 该方法采用深度森林的级联结构, 通过对原始特征的逐层处理和转换, 挖掘数据中丰富的语义信息, 赋予模型良好的表示学习能力. 同时在级联层中组合多个异质分类器增加集成的多样性, 实现良好的泛化性能和较高的预测精度.

(2) 提出一种解决深层模型退化问题的层特征重用机制. 该机制通过比较模型在当前层和上一层的预测性能, 重用模型对部分实例更优的历史预测结果, 降低模型的训练误差, 更有效地利用深层模型每一层的预测能力.

(3) 在 14 个标记分布学习领域广泛使用的数据集上对 LDLDF 算法进行验证. 实验结果表明, 本文提出的方法在多个标记分布评价指标上均取得较优效果, 显著优于同类对比算法.

1 相关工作

1.1 集成方法和深度集成方法

集成学习^[16]是一种重要的机器学习方法, 通过集成多个独立的弱学习器得到一个强学习器, 提高模型的预测精度和泛化能力. 根据基学习器种类的不同, 可以将集成学习分为异质集成学习和同质集成学习^[17]. 目前同质集成学习的应用已十分广泛, 包括 Bagging 算法、Boosting 算法等^[16]. 在 Bagging 算法中, 每个基学习器有放回地选取一部分训练数据进行训练, 通过多数投票机制产生最终的预测结果. 基学习器之间没有强依赖关系, 可以并行构建. 随机森林算法是 Bagging 算法的代表, 它使用决策树作为基学习器, 先随机选取一部分特征属性, 再从中选择最优划分属性. 在 Boosting 算法中, 基学习器之间存在强依赖关系, 需要串行生成. AdaBoost (adaptive Boosting) 算法^[18]是 Boosting 算法的代表, 它能够根据前一个基学习器的表现, 自适应地调整样本的权值分布, 使预测错误

的样本在后续训练中受到更多关注. 迭代训练多个基学习器后, AdaBoost 算法根据分类误差将所有基学习器加权组合成一个强学习器. GBDT (gradient boosted decision trees) 算法^[19]是 Boosting 算法的扩展, 它使用决策树作为基学习器, 沿着损失函数梯度下降的方向迭代产生新的决策树, 累加所有决策树的预测作为最终结果. 异质集成学习能够综合多种基学习器的优势, 相比同态集成具有更好的性能, 目前已有许多方法将其应用于各个领域^[20,21]. 异质集成学习中比较有代表性的就是 Stacking 算法^[22], 该算法分为两层结构, 第 1 层使用多种异质分类器构成基分类器, 输入训练集进行训练, 第 2 层使用一种分类器, 以第 1 层基学习器的输出作为特征和训练集共同输入模型训练.

由于结合了深度学习和集成学习的优势, 深度集成模型受到越来越多的关注^[17]. Zhou 等人^[11]提出了一个有代表性的深度集成模型——深度森林 (gcForest). 该模型有着深度神经网络中 layer-by-layer 的结构, 模型每层包含两个完全随机森林和两个普通随机森林^[12]. 完全随机森林是在完整的特征空间中随机选取特征分裂, 普通随机森林是在一个随机特征子空间内通过基尼系数选取分裂节点. 深度森林将每层所有随机森林的预测和原始输入特征串联后作为下一层的输入特征. 在模型的最终层, 将所有随机森林预测的平均值作为模型的最终输出结果.

深度森林综合了传统集成算法和深度神经网络的特点, 有着超参数少、鲁棒性好、可解释性强的优势, 在许多任务上都取得优越的性能. 如: Pang 等人^[15]提出了一种改进的随机森林 gcForest_{cs}, 通过置信度筛选将高置信度的实例直接传递到模型最后一层, 在保证模型预测精度的同时降低模型复杂度, 提升运行速度; Feng 等人^[23]提出了一种基于树集成的自动编码任务的方法 eForest, 证明了森林可以携带与深度神经网络一样多的信息; Liang 等人^[24]提出了 MLDF 方法, 通过设计多层级联结构来学习标签之间的相关性, 是深度森林模型在多标签学习领域的首次实践; Guo 等人^[25]提出的 BCDForest 方法将深度森林应用于癌症亚型分类任务. Wang 等人^[26]提出了一种基于树模型的弱标签深度学习 LCForest, 结合级联森林结构和标签补充结构提升利用弱标签数据的能力.

1.2 标记分布学习

现有的大多数数据集只具有简单的逻辑标签, 无法适配标记分布学习任务. 由于人工标注标记分布的成本较高, 且标签的描述度通常没有客观的衡量标准, 在实际场景中难以直接获得标记分布数据集. 为了解决该问题, Xu 等人^[27]提出了标记增强, 利用原始数据集的特征和标签信息探究标签对实例的相对重要程度, 从而恢复可靠的标记分布. 目前已有许多学者致力于标记增强算法的研究^[28-30], 为标记分布学习提供数据支持. 基于此, 标记分布学习范式被更加广泛应用于解决各类实际问题, 在计算机视觉^[4,31,32]、自然语言处理^[33]、医学诊断^[34]等多个领域内取得了优异的成果.

Jia 等人^[35]提出了一种利用局部低秩标签相关性学习面部情绪分布的方法, 提高了面部表情识别的精确度; Yin 等人^[36]提出了年龄标记分布的概念, 设计了 SA-IIS 和条件概率神经网络 (conditional probability neural network, CPNN) 两种算法用于解决人脸年龄估计任务; Gao 等人^[10]将深度神经网络与标记分布学习相结合, 提出了深度标记分布学习 (DLDF) 算法, 该方法主要用于年龄估计和头部姿态估计, 将每幅图像的标签转换为离散的标签分布, 利用 KL 散度拟合标签分布. 该方法在特征学习和分类器学习中有效地利用了标签的模糊性, 避免了模型的过拟合; Qi 等人^[6]提出了一种新的标签分布学习方法来解决多源人 Re-ID 任务, 该方法旨在探讨不同类别之间的关系, 减轻不同域之间的域偏移, 提高模型判别能力的同时学习域不变特性; Li 等人^[37]提出了一种考虑标记间协作的标记分布学习算法 (LDLCL), 通过标记空间中的稀疏重构学习标记相关性; 考虑到不同分类器在不同评价指标上可能具有各自的优势, Wang 等人^[38]提出了一种基于标记分布的异态集成学习算法, 采用两层分类器模型结构, 第 1 层使用多个异质分类器组合学习, 第 2 层使用单一分类器; Geng 等人^[39]提出了一种标签分布支持向量回归器 (LDSVR) 的方法, 通过多输出支持向量机对标签分布的每个分量拟合一个 Sigmoid 函数, 这种方法有别于使用最大熵模型表示标签分布的标记分布学习方法; Xing 等人^[7]提出了 LDLogitBoost 和 AOSO-LDLogitBoost 方法, 在模型中分别选择加权回归树和向量树作为基分类器; 考虑到多数 LDL 方法在表示学习方面具有局限性, 受可微决策树的启发, Shen 等人^[40]提出了一种端到端策略标签分布学习森林 (LDLFs), 通过融合叶节点的预测得到一般形式的标签分布, 在面部年龄估计问题中取得了很好的效果. 然而, LDLFs 本质上是利用深度神经网络 (DNNs)

进行表示学习, 同时将网络的最后一层设置为一个决策森林层输出最终的预测分布, 该方法对特征的逐层处理能力依赖于选择的神经网络.

由此本文提出一个联合深度森林与异质集成的标记分布学习方法 (LDLDF). 该方法采用深度森林的级联结构, 让模型像深度神经网络一样逐层处理和转换特征, 相比于深度神经网络, 它具有超参数少、预测精度高、不依赖反向传播的优势. 并且, LDLDF 在级联层中组合多个异质分类器, 从基分类器集成设计的角度增加深度集成模型内部的多样性, 挖掘数据中丰富的语义信息, 具有强大的表示学习能力和泛化能力. 此外, 我们设计一种层特征重用机制 (layer feature reuse) 解决 LDLDF 作为深层模型可能出现的模型退化问题.

2 联合深度森林与异质集成的标记分布学习方法

2.1 符号说明

首先表 1 对下文中要使用的符号进行说明. 我们使用 $X = [x_1, x_2, \dots, x_n] \in R^{n \times d}$ 表示输入特征空间, x_i 表示第 i 个样本. $Y = \{y_1, y_2, \dots, y_c\}$ 表示有限标签集, 其中 c 表示标签的数量. $D = [D_1, D_2, \dots, D_n] \in R^{n \times c}$ 表示输入空间对应的标记分布空间. $D_i = (d_{x_i}^1, d_{x_i}^2, \dots, d_{x_i}^c)^T \in [0, 1]^c$ 表示样本 x_i 对应的标记分布, 其中 $d_{x_i}^j \in [0, 1]$ 称为标签 y_j 对样本 x_i 的描述程度, 并且该描述度满足 $\sum_j d_{x_i}^j = 1$.

表 1 符号说明

编号	符号	含义
1	$X = [x_1, x_2, \dots, x_n] \in R^{n \times d}$	输入特征空间
2	x_i	第 i 个样本
3	$Y = \{y_1, y_2, \dots, y_c\}$	有限标签集
4	c	标签数量
5	$D = [D_1, D_2, \dots, D_n] \in R^{n \times c}$	输入空间对应的标记分布空间
6	$D_i = (d_{x_i}^1, d_{x_i}^2, \dots, d_{x_i}^c)^T \in [0, 1]^c$	样本 x_i 对应的标记分布
7	$d_{x_i}^j \in [0, 1]$	标签 y_j 对样本 x_i 的描述程度
8	$layer_k$	LDLDF 模型的第 k 层
9	$L^{(k,j)}$	第 k 层中第 j 个基分类器的预测结果
10	$L^{(k)}$	第 k 层中所有分类器的预测结果
11	$C^{(k,j)}$	$L^{(k,j)}$ 通过层特征重用模块后得到的增强特征
12	$C^{(k)}$	$L^{(k)}$ 通过层特征重用模块后得到的增强特征
13	$F^{(k)}$	第 k 层向第 $k+1$ 层的输入特征
14	S	每层中分类器的个数
15	m	模型层数

2.2 LDLDF 总体框架

深度神经网络中的表示学习主要依赖于对原始特征的逐层处理. 受此启发, LDLDF 方法采用深度森林的级联结构, 每个级联层接收其上一层处理的特征信息, 并将本层的预测结果输出到下一层. 图 1 中描述了 LDLDF 方法的整体框架, 其中 $layer_k$ 表示模型的第 k 层, $L^{(k)}$ 表示第 k 层中所有分类器的预测结果, $C^{(k)}$ 表示 $L^{(k)}$ 通过层特征重用模块后得到的增强特征, 将 $C^{(k)}$ 与原始特征向量串联后得到向第 $k+1$ 层的输入特征 $F^{(k)}$.

LDLDF 在构建级联层时集成了多种异质分类器 (例如图 1 中的随机森林和极端梯度提升树), 所有分类器预测的结果形成一个总的概率分布向量, 将其通过层特征重用模块后的输出作为增强特征, 与原始特征向量串联后输入到下一层级. 例如在 $layer_2$, 我们将所有分类器输出的结果 $L^{(2)} = \{L^{(2,1)}, L^{(2,2)}, \dots, L^{(2,S)}\}$ 输入层特征重用模块, 比较模型在 $layer_2$ 和 $layer_1$ 对每个实例的预测性能. 对 $layer_1$ 中预测更优的实例, 层特征重用模块会重用它们在

$layer_1$ 中产生的结果, 在 $layer_2$ 中预测更优的实例, 模块会保留它们在 $layer_2$ 中的结果, 得到增强特征 $C^{(2)}$. $C^{(2)}$ 与原始输入特征 X 串联后作为下一层的输入特征 $F^{(2)}$. 如果一个数据集的标签数为 4, 模型每层有 4 个分类器, 则每个分类器将产生一个 4 维的概率分布向量, 下一层接收 16 (=4×4) 维的增强特征. 此外, LDLDF 通过 5 倍交叉验证生成每个分类器预测的概率分布. 具体来说, 每个实例将被作为训练数据使用 4 次, 得到 4 个概率分布, 该实例最终的预测向量为 4 个预测分布的平均值. 当产生新的级联层时, 模型可以在验证集上估计整个级联的性能, 如果最近的 3 层中性能没有得到改善或改善的幅度低于阈值, 整个训练过程将终止, 从而自动确定级联层的数量.

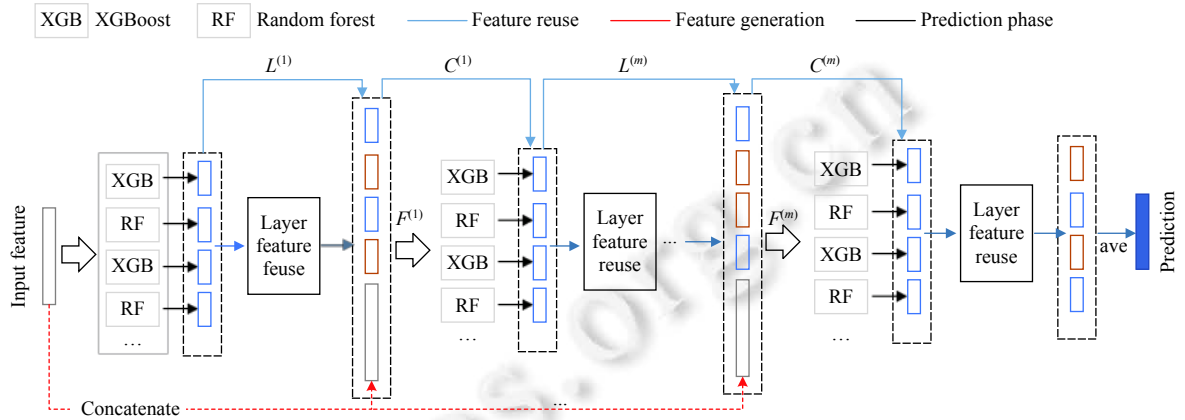


图 1 LDLDF 算法框架

2.3 层内异质集成学习

考虑到具有一定的复杂性的模型层可能会产生更好的拟合结果^[11], 我们在级联层中组合多个异质分类器, 从基分类器集成设计的角度增加深度集成模型内部的多样性. 异质集成算法可以充分利用不同算法的性能优势, 如何选择合适的基分类器就成为重点关注的问题. 一般而言, 基分类器的多样性和互补性是良好集成的关键因素^[41]. 标记分布的形式类似于概率分布, 可以采用回归函数拟合实例的标记分布. 因此本文选择了随机森林 (RF)、极端梯度提升树 (XGBoost)、线性回归 (LR)、 k 近邻 (KNN) 这 4 种分类器算法组成 LDLDF 的层内异质集成学习模块, 这些算法都能很好地解决各种回归问题, 下文将具体介绍这 4 种分类器算法.

随机森林 (RF) 以决策树为基本单元, 集成多棵决策树构成森林. 该算法可以处理大量的输入变量, 对数据集适应能力强, 模型抗干扰能力强, 通常比单个决策树具有更好的泛化能力. 在 LDL 中规模较大、特征较多的数据集上, RF 会取得优秀的结果. 假设森林中有 P 棵决策树, 随机森林会采用 Bagging 的思想, 随机抽样出一部分数据建立 P 个决策树模型, 每棵树学习到的模型为 $h(x; \theta_p)$. 对于一个测试样本 x , 它的预测模型 $\widehat{h}(x; \theta)$ 由公式 (1) 表示.

$$\widehat{h}(x; \theta) = \frac{1}{P} \sum_{p=1}^P h(x; \theta_p) \tag{1}$$

极端梯度提升树 (XGBoost) 使用了集成学习 Boosting 算法的思想, 是梯度提升树 (GBDT) 的高效实现. 在处理 LDL 问题上, XGBoost 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, 通过其自动处理缺失值特征的策略, 有效地利用输入的弱标签数据, 输出最终的预测分布. 它的基本思想是让第 t 个分类器拟合前面 $t-1$ 个分类器预测结果与真实值的残差. 对比 GBDT, XGBoost 从算法的实现方式、效率、健壮性方面都进行了优化. 假设有 K 棵树, 第 i 个样本的预测值 \widehat{D}_i 由 K 棵树集成得到, 输入模型的训练样本为 $I = \{(x_1, D_1), \dots, (x_n, D_n)\}$, 整体的目标函数可以表示为公式 (2).

$$L = \sum_{i=1}^n l(D_i, \widehat{D}_i) + \sum_{k=1}^K \Omega(f_k) \tag{2}$$

由于模型通过叠加式训练得到, 令 $\widehat{D}_i^{(t)}$ 为第 i 个样本在第 t 次迭代的预测值, 可以将目标函数写成梯度提升的

形式, 如公式 (3) 所示.

$$L^{(t)} = \sum_{i=1}^n l(D_i, \widehat{D}_i^{(t-1)} + f_i(x_i)) + \Omega(f_i) \quad (3)$$

其中, $l(D_i, \widehat{D}_i^{(t-1)})$ 计算了前 $t-1$ 个分类器预测结果的损失累加值. 对目标函数中 $f_i(x_i)$ 进行二阶泰勒展开, 如公式 (4) 所示.

$$L^{(t)} \approx \sum_{i=1}^n [l(D_i, \widehat{D}_i^{(t-1)}) + g_i f_i(x_i) + \frac{1}{2} h_i f_i^2(x_i)] + \Omega(f_i) \quad (4)$$

其中, g_i 为 $f_i(x_i)$ 的一阶导数, h_i 为 $f_i(x_i)$ 的二阶导数. 由于 $l(D_i, \widehat{D}_i^{(t-1)})$, g_i 和 h_i 都是已知常数项, 所以可以将需要优化的目标函数简化为公式 (5).

$$L^{(t)} \approx \sum_{i=1}^n [g_i f_i(x_i) + \frac{1}{2} h_i f_i^2(x_i)] + \Omega(f_i) \quad (5)$$

在处理 LDL 问题时, 线性回归 (LR) 模型思想简单、容易实现, 具有较高的可解释性. LR 算法利用已知数据训练模型, w 为模型参数, 对于一个测试样本 x , 它的预测标签分布将会通过公式 (6) 得到:

$$\widehat{D}(x) = w^T x \quad (6)$$

k 近邻算法可以用于预测连续的数据标签, 在处理 LDL 问题时, 该算法理论简单、容易实现、准确性高, 对于一个测试样本 x , 它的预测标签分布是通过计算该样本 K 个邻居的平均分布而获得的, 如公式 (7) 所示.

$$\widehat{D}(x) = \frac{1}{K} \sum_{k=1}^K D(x_k) \quad (7)$$

如算法 1 所示. 在训练时, 每一层所有分类器都会对训练实例进行预测, 预测的概率分布形成一个总的分布向量. 该向量通过层特征重用模块后得到增强特征, 将其与原始特征向量串联后作为下一层级的输入特征. 这样的特征生成方式促使算法关注到层中所有分类器的预测结果, 充分利用异质分类器的优势, 提高模型的预测精度.

算法 1. 层内异质集成学习.

输入: $layer_t$ 的输入矩阵 $F^{(t-1)}$, $layer_t$ 的基分类器集合 $H_t = \{h_1, h_2, \dots, h_s\}$, 原始输入特征 X , 基分类器的最优输出矩阵 $Z = \{Z_1, Z_2, \dots, Z_n\}$, X 的真实标签分布 D ;

输出: $layer_{t+1}$ 的输入矩阵 $F^{(t)}$, 预测性能 $p[t]$.

1. 初始化参数: $L^{(0)} = []$
 2. **for** $s = 1$ to S **do**
 3. 训练基分类器 s 并得到其预测结果: $L^{(s)} = h_s(F^{(0)})$
 4. $L^{(0)} = L^{(0)}.concat(L^{(s)})$ /*拼接所有基分类器的预测分布*/
 5. $C^{(0)} =$ 层特征重用($L^{(0)}, Z, D$) /*将 $L^{(0)}$ 通过层特征重用模块后得到增强特征 $C^{(0)}$ */
 6. 更新 Z 为 $C^{(0)}$
 7. 利用 $C^{(0)}$ 计算预测性能 $p[t]$
 8. $F^{(0)} = C^{(0)}.concat(X)$ /*将增强特征 $C^{(0)}$ 和原始样本特征 X 串联得到 $layer_{t+1}$ 的输入矩阵 $F^{(0)}$ */
 9. **return** $F_t, p[t], Z$
-

2.4 层特征重用

深层模型中深度的增加并不意味着更好的模型性能, 给适当深度的模型添加更多的层有可能带来更大的训练误差^[14]. LDLDF 方法采用级联结构模拟多层处理结构的深度学习模型, 通过多层的异质集成得到最终的结果. 原始的级联森林结构将当前层所有实例的预测结果作为增强特征, 与原始特征串联后输入下一层. 例如对于第 t 层,

深度森林将每个实例在 $layer_t$ 的预测和原始特征串联后输入到第 $t+1$ 层, 模型第 $t+1$ 层的预测结果会替换其在第 t 层的结果作为下一层的输入, 直到输出最终层的预测结果. 然而, 随着层数的增长, 模型对某些实例在前一层的预测会优于后一层, 使用传统的特征更新方式会增加模型训练的误差, 出现模型退化问题. 由此, LDLDF 提出了层特征重用机制 (算法 2) 来解决这样的模型退化问题, 它的主要思想是通过比较模型对每个实例在当前层和上一层的预测性能, 重用模型对部分实例更优的历史预测结果, 更有效地利用模型每一层的预测能力.

算法 2. 层特征重用.

输入: $layer_t$ 中基分类器的输出矩阵 $L^{(t)} = \{L_1^{(t)}, L_2^{(t)}, \dots, L_n^{(t)}\}$, 基分类器的最优输出矩阵 $Z = \{Z_1, Z_2, \dots, Z_n\}$, X 的真实标签分布 D ;

输出: 通过层特征重用模块后得到的增强特征 $C^{(t)}$.

1. 初始化参数: $C^{(t)} = L^{(t)}$
2. **for** $i = 1$ **to** n **do**
3. 比较 $L_i^{(t)}$ 、 Z_i 和真实标签分布 D_i 的距离, 根据公式 (9) 计算 $output_{x_i}$
4. 将 $C_i^{(t)}$ 更新为 $output_{x_i}$
5. **return** $C^{(t)}$

该模块的关键在于选择合适的特征重用标准. 考虑到 LDL 问题的目标是得到与真实标签分布相近的预测标签分布, 本文选取了 KL 散度 (Kullback-Leibler divergence, 也称相对熵) 作为特征重用标准, 这是一个用来衡量两个概率分布相似性的度量指标. 例如分布 Q 和分布 P 的差异可以通过公式 (8) 表示.

$$D_{KL}(P||Q) = H(P, Q) - H(P) = \sum_i P(x_i) \log \frac{1}{Q(x_i)} - \sum_i P(x_i) \log \frac{1}{P(x_i)} = \sum_i P(x_i) \log \frac{P(x_i)}{Q(x_i)} \quad (8)$$

其中, $H(P)$ 表示 P 分布的信息熵, $H(P, Q)$ 表示 P 、 Q 分布的交叉熵. 分布 P 、 Q 的相似度越高, KL 散度的值越小. 层特征重用模块的工作原理通过公式 (9) 表示.

$$output_{x_i} = \begin{cases} L_i, & D_{KL}(L_i||D_i) < D_{KL}(Z_i||D_i) \\ Z_i, & D_{KL}(L_i||D_i) > D_{KL}(Z_i||D_i) \end{cases} \quad (9)$$

其中, L_i 表示当前层分类器对实例 x_i 预测的分布, Z_i 表示之前层对实例 x_i 最好的预测分布, D_i 表示 x_i 的真实标签分布. $output_{x_i}$ 表示样本 x_i 通过层特征重用模块后的结果, $output_{x_i}$ 将作为增强特征和原始特征向量串联后输入下一层.

2.5 算法实现

本文提出的 LDLDF 方法采用级联结构模拟多层处理结构的深度学习模型, 在级联层中组合多个异质分类器, 层中所有分类器将预测结果通过层特征重用模块后得到增强特征, 将其与原始特征向量共同输入模型下一层. 如果模型在最近的 3 层中性能没有得到改善或改善的幅度低于阈值, 则模型停止增长, 经过所有层后可以得到最终的预测.

LDLDF 算法的伪代码见算法 3.

算法 3. LDLDF.

输入: 最大深度 T , 训练数据 $\{(X, D)\} = \{(x_1, D_1), (x_2, D_2), \dots, (x_n, D_n)\}$, 更新阈值 θ ;

输出: 模型集合 M , 最大层数 m .

1. 初始化参数: 模型集合 $M = \emptyset$, 基分类器的最优输出矩阵 $Z = []$, 每层的预测性能 $p[1:T]$, 训练集上最好预测性能 p_{best} , 模型停止更新计数器 $a = 0$.
2. **for** $t = 1$ **to** T **do**
3. $F^{(t)}, p[t], Z =$ 层内异质集成学习 ($F^{(t-1)}, H_t, X, Z, D$) /*调用 LDLDF 层内异质集成学习方法*/

```

4. if  $p[t] - p_{\text{best}} \geq \theta$  then
5.      $p_{\text{best}} = p[t]$ 
6.      $m = t$ 
7. else if  $p[t] - p_{\text{best}} < \theta$ 
8.      $a+ = 1$ 
9.     if  $a == 3$  then
10.        break
11.  $M = M \cup \text{layer}_t$ 
12. return  $M, m$ 

```

3 实验

3.1 数据集和评价指标

本文在 14 个标记分布学习数据集上进行了实验. S-JAFFE^[42]来自面部表情数据库 JAFFE, 它包含了 213 张日本女性的灰度表情图片, 每张图片都由 60 个人对 6 个基本的情绪标签进行评分, 每种情绪归一化后的平均得分被用来表示情绪标签的分布. SBU_3DFE^[43]来自面部表情数据集 BU_3DFE, 里面包含 2500 张图像, 处理方式与 SJAFFE 数据集相同. Flickr_LDL 和 Twitter_LDL 是两个情感图像数据集, 每张图片由 8 种不同的情绪 (快乐、愤怒、敬畏、满足、厌恶、兴奋、恐惧、悲伤) 标记. SCUT-FBP5500^[44]是一个与面容相关的数据集, 包括 5500 个具有不同属性和不同标签的正面面孔, 所有的面部图片都进行了得分标注. Yeast 系列的 10 个数据集来自对酿酒酵母不同时间点基因表达水平的生物学实验^[45], 总共有 2465 个酵母基因, 每个基因都由一个长度为 24 的载体向量表示, 标签对应于不同生物实验中的离散时间点, 所有时间点的标准化基因表达水平构成了一个特定酵母基因的标签分布. Movie 数据集来自 Netflix 上用户对电影的评分, 评分分为 5 个等级, 每个电影的标签分布由每个等级水平的百分比计算得到. Natural Scene 数据集包含 2000 张自然场景图像, 依据 9 种与这些图像相关的标签 (植物、天空、云、雪、建筑、沙漠、山、水、太阳) 进行图像标注. Human Gene 数据集来自人类基因与疾病之间关系的研究, 标签对应于 68 种疾病, 将每种疾病的基因表达水平归一化后作为相应标签的描述程度. 实验中选取的图像数据集均已预先从图片中提取到特征向量. 后文表 2 中总结了这 13 个数据集的简要信息.

标记分布学习算法的性能通常由预测标记分布和真实标记分布之间的距离或相似度计算. 本文选取了 Geng 等人^[3]提出的 6 个评价 LDL 算法性能的指标, 其中包含 Chebychev 距离、Clark 距离、Canberra 距离、Kullback-Leibler 散度这 4 个距离度量以及 Cosine 余弦相关性系数、Intersection 交叉相似度这两个相似性度量. 这些性能指标的定义在后文表 3 中, 其中 d_j 表示了分布 D 中的第 j 个元素.

3.2 实验设置

本文中的所有实验都通过随机采样选取 80% 的样本数据构成训练集, 20% 的样本数据构成测试集. LDLDF 的超参数设置如下: LDLDF 的最大层数设置为 10, 每层基分类器个数设置为 6, 在每一层中都使用五折交叉验证以避免过拟合. 当选取随机森林 (RF) 或极端梯度提升树 (XGB) 为基分类器时, 在第 1 层中每个森林中取 40 棵树, 后一层比前一层多取 20 棵树, 直到树的数量达到 100 棵. 设置第 1 层的森林最大深度为 3, 后一层比前一层最大深度增加 3. 当选取 k 近邻 (KNN) 为基分类器时取近邻数 k 为 3.

3.3 异质集成的模型选择

首先对 LDLDF 中选取的基分类器进行说明. 为了提高算法模型的精度, 本文依据标记分布问题的性质和对分类器多样性、准确性的要求, 选取了 4 种分类器, 分别为: 随机森林 (RF)、极端梯度提升树 (XGB)、线性回归 (LR)、 k 近邻 (KNN). 这几种分类器算法都可以很好地解决各种回归问题. 我们针对选取的 14 个数据集进行了实验, 表 4、表 5 以 S-JAFFE、SBU_3DFE、Yeast-elu 这 3 个数据集为例, 分别展示采用 2 种和 3 种分类器组合的结果.

表 2 实验数据集的统计数据

编号	数据集	样本个数	特征维度	标记个数
1	S-JAFFE	213	243	6
2	SBU_3DFE	2500	243	6
3	Natural scene	2000	294	9
4	Movie	7755	1869	5
5	Human gene	17892	36	68
6	Yeast-alpha	2465	24	18
7	Yeast-cdc	2465	24	15
8	Yeast-elu	2465	24	14
9	Yeast-diau	2465	24	7
10	Yeast-heat	2465	24	6
11	Yeast-spo	2465	24	4
12	Flickr_LDL	11150	256	8
13	SCUT_FBP5500	5500	512	5
14	Twitter_LDL	10045	128	8

表 3 标记分布学习算法的评价指标

名称	公式
Kullback-Leibler (KL)↓	$Dis_1(D, \hat{D}) = \sum_{j=1}^c d_j \ln \frac{d_j}{\hat{d}_j}$
Chebyshev (Cheby)↓	$Dis_2(D, \hat{D}) = \max_j d_j - \hat{d}_j $
Clark↓	$Dis_3(D, \hat{D}) = \sqrt{\sum_{j=1}^c \frac{(d_j - \hat{d}_j)^2}{(d_j + \hat{d}_j)^2}}$
Canberra (Can)↓	$Dis_4(D, \hat{D}) = \sum_{j=1}^c \frac{ d_j - \hat{d}_j ^2}{d_j + \hat{d}_j}$
Cosine (Cos)↑	$Sim_1(D, \hat{D}) = \frac{\sum_{j=1}^c d_j \hat{d}_j}{\sqrt{\sum_{j=1}^c d_j^2} \sqrt{\sum_{j=1}^c \hat{d}_j^2}}$
Intersection (Inter)↑	$Sim_2(D, \hat{D}) = \sum_{j=1}^c \min(d_j, \hat{d}_j)$

表 4 两种分类器组合在 S-JAFFE、SBU_3DFE、Yeast-elu 上的实验结果

数据集	分类器	KL_div↓	Cheby↓	Clark↓	Canberra↓	Cosine↑	Intersection↑	运行时间 (s)
S-JAFFE	RF+XGB	0.0399±0.0048	0.0915±0.0042	0.3235±0.0078	0.6615±0.0235	0.9572±0.0050	0.8864±0.0069	6.23
	RF+KNN	0.0501±0.0019	0.0963±0.0030	0.3406±0.0025	0.7086±0.0090	0.9526±0.0017	0.8790±0.0017	1.97
	RF+LR	0.0456±0.0048	0.0935±0.0059	0.3457±0.0092	0.7044±0.0220	0.9571±0.0046	0.8818±0.0049	3.20
	XGB+KNN	0.0412±0.0135	0.0942±0.0097	0.3379±0.0178	0.6921±0.0429	0.9546±0.0075	0.8822±0.0088	5.00
	XGB+LR	0.0472±0.0080	0.0941±0.0050	0.3394±0.0081	0.6894±0.0201	0.9557±0.0032	0.8831±0.0034	8.01
	KNN+LR	0.0683±0.0362	0.0936±0.0084	0.3664±0.0307	0.7449±0.0620	0.9498±0.0064	0.8773±0.0105	1.55
	RF+RF	0.0527±0.0022	0.0992±0.0014	0.3410±0.0073	0.7138±0.0185	0.9503±0.0021	0.8773±0.0027	3.43
	XGB+XGB	0.0486±0.0120	0.1065±0.0059	0.3769±0.0171	0.7605±0.0302	0.9484±0.0052	0.8684±0.0061	9.88
	KNN+KNN	0.0622±0.0061	0.1031±0.0063	0.3717±0.0234	0.7654±0.0468	0.9413±0.0057	0.8675±0.0094	1.29
	RF+XGB	0.0431±0.0042	0.0974±0.0036	0.3123±0.0079	0.6458±0.0124	0.9500±0.0038	0.8833±0.0035	73.49
RF+KNN	0.0658±0.0013	0.1172±0.0017	0.3633±0.0027	0.7697±0.0084	0.9349±0.0012	0.8611±0.0015	38.71	
RF+LR	0.0519±0.0015	0.0996±0.0018	0.3249±0.0032	0.6732±0.0082	0.9480±0.0016	0.8792±0.0017	28.00	
XGB+KNN	0.0447±0.0009	0.1030±0.0008	0.3272±0.0009	0.6782±0.0046	0.9455±0.0011	0.8772±0.0008	77.49	
XGB+LR	0.0515±0.0012	0.0995±0.0027	0.3272±0.0078	0.6692±0.0146	0.9496±0.0023	0.8802±0.0031	60.23	
KNN+LR	0.0566±0.0016	0.1061±0.0006	0.3537±0.0042	0.7335±0.0115	0.9439±0.0016	0.8694±0.0020	9.19	
RF+RF	0.0651±0.0029	0.1184±0.0025	0.3644±0.0063	0.7830±0.0161	0.9358±0.0027	0.8595±0.0030	43.19	
XGB+XGB	0.0458±0.0017	0.1043±0.0014	0.3331±0.0034	0.6839±0.0075	0.9489±0.0017	0.8755±0.0017	174.31	
KNN+KNN	0.0851±0.0022	0.1315±0.0017	0.4148±0.0083	0.8616±0.0170	0.9168±0.0020	0.8430±0.0029	6.57	
RF+XGB	0.0058±0.0001	0.0158±0.0002	0.1949±0.0023	0.5723±0.0031	0.9943±0.0003	0.9594±0.0009	49.81	
RF+KNN	0.0062±0.0002	0.0162±0.0003	0.2010±0.0041	0.5919±0.0101	0.9939±0.0001	0.9582±0.0006	23.93	
RF+LR	0.0061±0.0002	0.0163±0.0003	0.1980±0.0029	0.5801±0.0042	0.9940±0.0001	0.9590±0.0002	22.08	
XGB+KNN	0.0060±0.0002	0.0161±0.0001	0.1989±0.0002	0.5847±0.0025	0.9940±0.0001	0.9587±0.0002	45.47	
XGB+LR	0.0059±0.0004	0.0162±0.0005	0.1994±0.0066	0.5839±0.0166	0.9940±0.0003	0.9588±0.0010	39.35	
KNN+LR	0.0061±0.0002	0.0160±0.0003	0.1990±0.0040	0.5874±0.0129	0.9940±0.0002	0.9585±0.0009	11.14	
RF+RF	0.0060±0.0001	0.0161±0.0002	0.1986±0.0032	0.5844±0.0069	0.9940±0.0002	0.9585±0.0009	34.10	
XGB+XGB	0.0061±0.0001	0.0164±0.0001	0.2016±0.0015	0.5844±0.0069	0.9940±0.0002	0.9585±0.0009	80.69	
KNN+KNN	0.0069±0.0001	0.0172±0.0001	0.2117±0.0019	0.6221±0.0073	0.9932±0.0001	0.9561±0.0005	4.87	

表5 3种分类器组合在 S-JAFFE、SBU_3DFE、Yeast-elu 上的实验结果

数据集	分类器	KL_div↓	Cheby↓	Clark↓	Canberra↓	Cosine↑	Intersection↑	运行时间 (s)
S-JAFFE	RF+XGB+KNN	0.0361±0.0125	0.0837±0.0087	0.3141±0.0195	0.6413±0.0441	0.9635±0.0078	0.8925±0.0083	7.98
	RF+XGB+LR	0.0413±0.0072	0.0890±0.0077	0.3246±0.0175	0.6607±0.0415	0.9588±0.0065	0.8877±0.0078	10.34
	XGB+KNN+LR	0.0410±0.0064	0.0877±0.0083	0.3155±0.0256	0.6515±0.0556	0.9610±0.0060	0.8892±0.0093	5.76
	RF+KNN+LR	0.0424±0.0091	0.0909±0.0104	0.3293±0.0223	0.6732±0.0480	0.9576±0.0065	0.8853±0.0090	3.24
SBU_3DFE	RF+XGB+KNN	0.0478±0.0032	0.1025±0.0016	0.3233±0.0046	0.6730±0.0117	0.9458±0.0021	0.8781±0.0021	72.00
	RF+XGB+LR	0.0464±0.0031	0.0959±0.0022	0.3070±0.0046	0.6308±0.0090	0.9524±0.0019	0.8862±0.0018	81.89
	XGB+KNN+LR	0.0467±0.0022	0.0966±0.0020	0.3097±0.0062	0.6347±0.0127	0.9524±0.0021	0.8855±0.0025	87.51
	RF+KNN+LR	0.0523±0.0021	0.0991±0.0020	0.3142±0.0052	0.6513±0.0108	0.9477±0.0023	0.8823±0.0022	23.14
Yeast-elu	RF+XGB+KNN	0.0061±0.0002	0.0162±0.0004	0.2009±0.0035	0.5892±0.0076	0.9938±0.0002	0.9584±0.0005	31.40
	RF+XGB+LR	0.0058±0.0001	0.0159±0.0001	0.1943±0.0010	0.5709±0.0019	0.9943±0.0001	0.9596±0.0010	26.92
	XGB+KNN+LR	0.0062±0.0001	0.0163±0.0001	0.2006±0.0018	0.5895±0.0042	0.9938±0.0001	0.9584±0.0002	23.06
	RF+KNN+LR	0.0059±0.0003	0.0159±0.0003	0.1962±0.0050	0.5773±0.0142	0.9942±0.0003	0.9592±0.0009	5.00

通过分析比较可知: 异质分类器集成的深层模型构造方式可以利用不同分类器的优势, 相比同态分类器集成有着更优的性能. 相较于两种分类器集成, 3种分类器集成的结果在所选指标上表现更优. 在 S-JAFFE 数据集上, 采用 XGB、RF、KNN 这 3 种分类器的组合能够取得更加准确的结果. 在 SBU_3DFE 和 Yeast-elu 数据集上, 采用 XGB、RF、LR 这 3 种分类器的组合能够取得更加准确的结果.

此外, 表 4、表 5 中对比了不同分类器组合的运行时间. 实验表明, 加入 XGB 和 RF 分类器可以有效地提高模型精度, 算法整体耗时也在合理范围之内. 对于不同的数据集, 加入适合于数据集的第 3 种分类器可以优化模型表现. 考虑到实验参数的统一, 本文在对比实验中采用了 XGB、RF 分类器的组合.

3.4 对比算法和实验结果

本文将提出的 LDLDF 算法与 9 种算法进行了比较: 分别是 PT-SVM 算法^[3]、AA-KNN 算法^[3]、CPNN 算法^[4]、SA-IIS 算法^[3]、SA-BFGS 算法^[3]、StructRF 算法^[46]、LDL-HR 算法^[47]、LDL-LDM^[48]算法、LDLCL^[37]算法.

(1) PT-SVM: 该算法在重采样后的单标签数据集上训练 SVM 分类器, SVM 分类器通过预测每个标签的概率得到标签分布.

(2) AA-KNN: 该算法将 KNN 算法用于解决标记分布问题.

(3) CPNN: 该算法基于一个 3 层神经网络, 以目标变量 (如年龄) 和条件特征向量 (如面部特征) 作为输入, 输出给定特征向量的目标变量的条件概率. CPNN 可以在学习真实年龄的同时有效地利用相邻的年龄.

(4) SA-IIS、SA-BFGS: 这两种算法均使用最大熵模型作为输出模型, 使用 KL 散度作为目标函数, 分别使用 IIS 算法和 BFGS 算法作为优化方式.

(5) StructRF: 该算法是一种处理标记分布学习问题的结构化随机森林算法, 考虑了标签相关性.

(6) LDL-HR: 该算法通过共同学习 top 标签和其他标签描述度来提高分类性能, 解决标记分布学习中目标不匹配问题.

(7) LDL-LDM: 该算法提出了标签分布的底层流形结构可以编码标签之间的相关性, 以数据驱动的方式有效地利用全局和局部标签相关性来处理 LDL 和不完全 LDL 问题.

(8) LDLCL: 该算法通过标记空间中的稀疏重构学习标记相关性, 提出了一种考虑标记间协作的标记分布学习算法.

为了保证实验的公平性, 所有对比算法均采用原文中的推荐设置和超参数, 并全部使用五折交叉验证. 本文对每种方法都进行 10 次实验, 记录 10 次实验的平均值和标准差用于比较研究. 表 6 记录了每个方法在 6 个评价指标的结果, 实验结果以“平均值±标准差 (mean±std.)”的形式给出, 表中用粗体标注出每个度量的最佳结果. 分析表 6 的实验结果发现: LDLDF 方法在每个评价指标上的平均排名都可以达到最优, 在较大规模的数据集 SBU_3DFE、SCUT-FBP5500、Movie、Natural sene、Human gene、Flickr_LDL、Twitter_LDL 和较小规模的数据集 S-JAFFE 上都能取得优异的性能. 在 Yeast 系列数据集上的优异表现体现出该方法对标记分布数据集较为有效.

表 6 对比算法实验结果

数据集	算法	KL_div↓	Cheby↓	Clark↓	Canberra↓	Cosine↑	Intersection↑
Yeast-alpha	PT-SVM	0.0060±0.0003	0.0139±0.0004	0.2217±0.0062	0.7251±0.0205	0.9941±0.0003	0.9600±0.0011
	AA-KNN	0.0065±0.0002	0.0146±0.0004	0.2305±0.0041	0.7532±0.0140	0.9936±0.0002	0.9584±0.0008
	CPNN	0.0056±0.0021	0.0135±0.0014	0.2316±0.0301	0.7252±0.0409	0.9936±0.0031	0.9596±0.0035
	SA-IIS	0.0067±0.0012	0.0148±0.0010	0.2334±0.0121	0.7630±0.0421	0.9934±0.0011	0.9577±0.0021
	SA-BFGS	0.0062±0.0007	0.0135±0.0009	0.2108±0.0133	0.6848±0.0455	0.9945±0.0007	0.9620±0.0025
	StructRF	0.0054±0.0007	0.0133±0.0008	0.2079±0.0134	0.6745±0.0461	0.9947±0.0007	0.9628±0.0025
	LDL-HR	0.0156±0.0134	0.0216±0.0052	0.3692±0.0151	1.2341±0.0259	0.9856±0.0020	0.9345±0.0100
	LDL-LDM	0.0054±0.0010	0.0136±0.0006	0.2123±0.0021	0.6918±0.0104	0.9944±0.0004	0.9618±0.0012
	LDLCL	0.0054±0.0001	0.0134±0.0001	0.2094±0.0001	0.6797±0.0001	0.9946±0.0001	0.9624±0.0001
Ours	0.0052±0.0001	0.0131±0.0002	0.2072±0.0018	0.6740±0.0128	0.9947±0.0001	0.9628±0.0007	
Yeast-cdc	PT-SVM	0.0075±0.0064	0.0170±0.0083	0.2252±0.1006	0.6749±0.2806	0.9926±0.0059	0.9555±0.0181
	AA-KNN	0.0082±0.0064	0.0175±0.0080	0.2366±0.0978	0.7156±0.2763	0.9920±0.0059	0.9528±0.0178
	CPNN	0.0075±0.0014	0.0215±0.0015	0.2495±0.0187	0.6697±0.0575	0.9929±0.0011	0.9567±0.0039
	SA-IIS	0.0082±0.0010	0.0178±0.0010	0.2352±0.0120	0.7088±0.0360	0.9921±0.0010	0.9531±0.0020
	SA-BFGS	0.0071±0.0009	0.0163±0.0009	0.2161±0.0138	0.6492±0.0416	0.9931±0.0008	0.9572±0.0027
	StructRF	0.0068±0.0009	0.0161±0.0010	0.2139±0.0139	0.6403±0.0427	0.9934±0.0008	0.9579±0.0028
	LDL-HR	0.0146±0.0007	0.0233±0.0007	0.3292±0.0089	1.0019±0.0204	0.9863±0.0005	0.9347±0.0011
	LDL-LDM	0.0069±0.0002	0.0163±0.0004	0.2162±0.0035	0.6472±0.0103	0.9932±0.0002	0.9574±0.0007
	LDLCL	0.0069±0.0001	0.0161±0.0001	0.2153±0.0001	0.6455±0.0001	0.9933±0.0001	0.9575±0.0001
Ours	0.0066±0.0004	0.0160±0.0003	0.2130±0.0074	0.6379±0.0120	0.9936±0.0004	0.9580±0.0005	
Yeast-elu	PT-SVM	0.0069±0.0003	0.0171±0.0003	0.2115±0.0043	0.6242±0.0134	0.9934±0.0002	0.9559±0.0009
	AA-KNN	0.0073±0.0004	0.0176±0.0004	0.0176±0.0004	0.6419±0.0179	0.9929±0.0004	0.9547±0.0013
	CPNN	0.0062±0.0029	0.0163±0.0031	0.2215±0.0291	0.5910±0.2451	0.9939±0.0019	0.9551±0.0076
	SA-IIS	0.0073±0.0005	0.0178±0.0010	0.2160±0.0070	0.6392±0.0190	0.9929±0.0005	0.9547±0.0010
	SA-BFGS	0.0063±0.0004	0.0164±0.0006	0.1992±0.0058	0.5838±0.0172	0.9939±0.0004	0.9588±0.0012
	StructRF	0.0061±0.0004	0.0160±0.0006	0.1961±0.0058	0.5756±0.0171	0.9941±0.0004	0.9593±0.0012
	LDL-HR	0.0103±0.0002	0.0213±0.0002	0.2619±0.0030	0.7757±0.0100	0.9899±0.0002	0.9451±0.0007
	LDL-LDM	0.0061±0.0001	0.0162±0.0002	0.1992±0.0031	0.5834±0.0076	0.9940±0.0002	0.9588±0.0005
	LDLCL	0.0061±0.0001	0.0162±0.0001	0.1985±0.0001	0.5816±0.0001	0.9941±0.0001	0.9589±0.0001
Ours	0.0058±0.0001	0.0158±0.0002	0.1949±0.0023	0.5723±0.0031	0.9943±0.0003	0.9594±0.0009	
Yeast-heat	PT-SVM	0.0138±0.0006	0.0441±0.0011	0.1902±0.0047	0.3798±0.0110	0.9868±0.0006	0.9376±0.0019
	AA-KNN	0.0144±0.0010	0.0447±0.0017	0.1939±0.0070	0.3903±0.0135	0.9863±0.0009	0.9359±0.0022
	CPNN	0.0147±0.0011	0.0563±0.0033	0.2235±0.0085	0.3906±0.0008	0.9863±0.0011	0.9376±0.0009
	SA-IIS	0.0133±0.0004	0.0430±0.0010	0.1881±0.0030	0.3772±0.0050	0.9870±0.0004	0.9380±0.0010
	SA-BFGS	0.0127±0.0005	0.0423±0.0009	0.1828±0.0032	0.3647±0.0067	0.9880±0.0005	0.9401±0.0011
	StructRF	0.0118±0.0005	0.0406±0.0009	0.1764±0.0032	0.3526±0.0068	0.9887±0.0005	0.9422±0.0011
	LDL-HR	0.0131±0.0002	0.0430±0.0005	0.1872±0.0014	0.3737±0.0029	0.9875±0.0002	0.9386±0.0005
	LDL-LDM	0.0129±0.0005	0.0427±0.0010	0.1850±0.0035	0.3674±0.0064	0.9877±0.0005	0.9397±0.0010
	LDLCL	0.0123±0.0001	0.0417±0.0001	0.1807±0.0001	0.3605±0.0001	0.9882±0.0001	0.9408±0.0001
Ours	0.0117±0.0003	0.0406±0.0006	0.1761±0.0018	0.3522±0.0040	0.9887±0.0003	0.9422±0.0006	
Yeast-diau	PT-SVM	0.0177±0.0019	0.0443±0.0034	0.2371±0.0141	0.5096±0.0333	0.9836±0.0018	0.9289±0.0048
	AA-KNN	0.0149±0.0009	0.0391±0.0013	0.2110±0.0062	0.4539±0.0151	0.9863±0.0009	0.9370±0.0022
	CPNN	0.0133±0.0021	0.0375±0.0047	0.2072±0.0275	0.4673±0.0611	0.9876±0.0025	0.9342±0.0094
	SA-IIS	0.0140±0.0010	0.0386±0.0010	0.2090±0.0070	0.4490±0.0170	0.9870±0.0010	0.9380±0.0020
	SA-BFGS	0.0131±0.0011	0.0370±0.0015	0.2008±0.0082	0.4309±0.0193	0.9879±0.0011	0.9401±0.0028
	StructRF	0.0124±0.0012	0.0358±0.0018	0.1941±0.0085	0.4164±0.0201	0.9884±0.0011	0.9421±0.0029
	LDL-HR	0.0147±0.0005	0.0392±0.0008	0.2127±0.0044	0.4570±0.0097	0.9864±0.0005	0.9365±0.0013
	LDL-LDM	0.0133±0.0004	0.0375±0.0007	0.2025±0.0039	0.4339±0.0070	0.9877±0.0003	0.9397±0.0008
	LDLCL	0.0128±0.0001	0.0365±0.0001	0.1983±0.0001	0.4257±0.0001	0.9881±0.0001	0.9409±0.0001
Ours	0.0123±0.0001	0.0357±0.0009	0.1939±0.0054	0.4152±0.0046	0.9886±0.0006	0.9422±0.0013	

表6 对比算法实验结果(续)

数据集	算法	KL_div↓	Cheby↓	Clark↓	Canberra↓	Cosine↑	Intersection↑
Yeast-spo	PT-SVM	0.0292±0.0028	0.0633±0.0032	0.2675±0.0123	0.5504±0.0253	0.9725±0.0027	0.9089±0.0043
	AA-KNN	0.0291±0.0023	0.0629±0.0027	0.2668±0.0103	0.5490±0.0220	0.9727±0.0022	0.9095±0.0037
	CPNN	0.0203±0.0059	0.0576±0.0030	0.2403±0.0052	0.5228±0.0251	0.9767±0.0041	0.9115±0.0079
	SA-IIS	0.0254±0.0030	0.0600±0.0040	0.2550±0.0170	0.5230±0.0340	0.9760±0.0030	0.9140±0.0050
	SA-BFGS	0.0246±0.0031	0.0584±0.0039	0.2504±0.0175	0.5134±0.0355	0.9769±0.0027	0.9154±0.0056
	StructRF	0.0240±0.0027	0.0575±0.0036	0.2461±0.0154	0.5044±0.0298	0.9774±0.0023	0.9170±0.0047
	LDL-HR	0.0253±0.0011	0.0588±0.0020	0.2508±0.0055	0.5156±0.0175	0.9762±0.0014	0.9149±0.0016
	LDL-LDM	0.0247±0.0010	0.0578±0.0013	0.2488±0.0060	0.5137±0.0137	0.9768±0.0010	0.9154±0.0022
	LDLCL	0.0244±0.0001	0.0581±0.0001	0.2489±0.0001	0.5117±0.0001	0.9771±0.0001	0.9157±0.0001
Ours	0.0238±0.0014	0.0573±0.0006	0.2455±0.0029	0.5039±0.0072	0.9777±0.0037	0.9172±0.0012	
S-JAFFE	PT-SVM	0.0790±0.0058	0.1224±0.0080	0.4470±0.0161	0.9335±0.0314	0.9258±0.0056	0.8410±0.0059
	AA-KNN	0.0538±0.0107	0.0988±0.0139	0.3485±0.0306	0.7142±0.0640	0.9484±0.0109	0.8763±0.0133
	CPNN	0.1243±0.0156	0.1822±0.0143	0.6159±0.0259	1.0896±0.0689	0.8749±0.0291	0.8165±0.0098
	SA-IIS	0.0700±0.0120	0.1175±0.0150	0.4190±0.0340	0.8750±0.0860	0.9340±0.0121	0.8511±0.0162
	SA-BFGS	0.0861±0.0231	0.1184±0.0145	0.4657±0.0633	0.9565±0.1269	0.9277±0.0113	0.8426±0.0152
	StructRF	0.0544±0.0117	0.1047±0.0137	0.3709±0.0399	0.7726±0.0920	0.9483±0.0117	0.8690±0.0165
	LDL-HR	0.0648±0.0053	0.1095±0.0060	0.4090±0.0104	0.8458±0.0035	0.9390±0.0029	0.8565±0.0085
	LDL-LDM	0.0451±0.0084	0.0930±0.0052	0.3443±0.0102	0.7621±0.0201	0.9428±0.0035	0.8631±0.0120
	LDLCL	0.0417±0.0003	0.1061±0.0002	0.3361±0.0002	0.7023±0.0002	0.9421±0.0013	0.8775±0.0006
Ours	0.0399±0.0048	0.0915±0.0042	0.3235±0.0078	0.6615±0.0235	0.9572±0.0050	0.8864±0.0069	
SBU_3DEF	PT-SVM	0.0909±0.0048	0.1415±0.0058	0.4236±0.0151	0.9154±0.0315	0.9128±0.0045	0.8356±0.0058
	AA-KNN	0.0818±0.0039	0.1276±0.0035	0.4031±0.0087	0.8315±0.0193	0.9202±0.0035	0.8479±0.0038
	CPNN	0.0671±0.0230	0.1265±0.0194	0.3465±0.0291	0.7523±0.0879	0.9319±0.0181	0.8677±0.0091
	SA-IIS	0.0680±0.0045	0.1113±0.0040	0.4161±0.0092	0.9341±0.0224	0.9355±0.0041	0.8624±0.0053
	SA-BFGS	0.0603±0.0043	0.1089±0.0051	0.3676±0.0121	0.7631±0.0239	0.9413±0.0043	0.8647±0.0046
	StructRF	0.0581±0.0052	0.1123±0.0056	0.3476±0.0166	0.7451±0.0337	0.9425±0.0052	0.8661±0.0062
	LDL-HR	0.0726±0.0043	0.1198±0.0025	0.3752±0.0102	0.8648±0.0085	0.9352±0.0060	0.8558±0.0038
	LDL-LDM	0.0537±0.0008	0.1047±0.0015	0.3466±0.0022	0.7256±0.0052	0.9472±0.0009	0.8710±0.0011
	LDLCL	0.0611±0.0001	0.1151±0.0001	0.3602±0.0001	0.7639±0.0001	0.9394±0.0001	0.8626±0.0001
Ours	0.0431±0.0042	0.0974±0.0036	0.3123±0.0079	0.6458±0.0124	0.9500±0.0038	0.8833±0.0035	
Natural scene	PT-SVM	1.5258±0.2236	0.4273±0.0328	2.5662±0.0330	7.2686±0.1413	0.4567±0.0610	0.3483±0.0394
	AA-KNN	1.1154±0.0596	0.3142±0.0131	1.9384±0.0377	4.6736±0.1066	0.7041±0.0129	0.5571±0.0125
	CPNN	1.4778±0.0140	0.7732±0.0162	2.8829±0.0063	8.5877±0.0211	0.5704±0.0334	0.2268±0.0040
	SA-IIS	0.8700±0.0260	0.3411±0.0170	2.4610±0.0250	6.7650±0.1040	0.6980±0.0080	0.4870±0.0120
	SA-BFGS	0.8540±0.0620	0.3220±0.0170	2.4110±0.0230	6.6200±0.0970	0.7100±0.0170	0.5480±0.0170
	StructRF	0.6142±0.0292	0.2651±0.0148	2.4080±0.0251	6.5025±0.1081	0.8026±0.0121	0.6045±0.0131
	LDL-HR	1.0712±0.0135	0.3363±0.242	2.2880±0.1253	6.7837±0.1445	0.6123±0.0087	0.4019±0.0259
	LDL-LDM	0.7624±0.0131	0.3116±0.0052	2.3754±0.0092	6.5499±0.0382	0.7438±0.0045	0.5523±0.0054
	LDLCL	0.8506±0.0104	0.2811±0.0003	2.4458±0.0011	6.6211±0.0050	0.7789±0.0007	0.6003±0.0008
Ours	0.5865±0.0217	0.2452±0.0162	2.3621±0.0136	6.4739±0.1023	0.8098±0.0312	0.6382±0.0265	
Movie	PT-SVM	0.2938±0.0697	0.2292±0.0191	0.8323±0.0800	1.6277±0.1939	0.7810±0.0495	0.6870±0.0440
	AA-KNN	0.1177±0.0051	0.1240±0.0026	0.5488±0.0102	1.0553±0.0210	0.9224±0.0032	0.7976±0.0051
	CPNN	0.1272±0.0214	0.1404±0.0333	0.8542±0.0139	1.4706±0.0255	0.9104±0.0314	0.7821±0.0435
	SA-IIS	0.1372±0.0131	0.1502±0.0080	0.5912±0.0283	1.1370±0.0574	0.9050±0.0082	0.8004±0.0100
	SA-BFGS	0.1273±0.0109	0.1317±0.0062	0.5665±0.0246	1.0948±0.0511	0.9183±0.0063	0.8155±0.0083
	StructRF	0.0921±0.0066	0.1108±0.0051	0.5042±0.0235	0.9629±0.0452	0.9393±0.0042	0.8421±0.0062
	LDL-HR	0.1292±0.102	0.1410±0.0053	0.5832±0.0141	1.1056±0.0367	0.9205±0.0024	0.8202±0.0068
	LDL-LDM	0.0984±0.0012	0.1151±0.0007	0.5171±0.0024	0.9912±0.0052	0.9350±0.0007	0.8359±0.0009
	LDLCL	0.1114±0.0008	0.1122±0.0001	0.5354±0.0013	1.0085±0.0002	0.9365±0.0002	0.8385±0.0002
Ours	0.0863±0.0016	0.1071±0.0012	0.4900±0.0075	0.9313±0.0140	0.9435±0.0010	0.8481±0.0019	

表 6 对比算法实验结果 (续)

数据集	算法	KL_div↓	Cheby↓	Clark↓	Canberra↓	Cosine↑	Intersection↑
Human gene	PT-SVM	—	—	—	—	—	—
	AA-KNN	0.2980±0.3799	0.0644±0.0767	2.3767±1.1127	16.1936±8.8100	0.7697±0.2303	0.7433±0.1384
	SA-IIS	0.2380±0.0190	0.0534±0.0040	2.1230±0.0880	14.5410±0.6530	0.8330±0.0110	0.7830±0.0100
	SA-BFGS	0.2365±0.0194	0.0533±0.0038	2.1111±0.0864	14.4532±0.6455	0.8343±0.0107	0.7842±0.0097
	StructRF	0.2336±0.0196	0.0532±0.0039	2.1026±0.0912	14.3870±0.6685	0.8359±0.0116	0.7858±0.0103
	LDL-HR	0.2825±0.0248	0.0556±0.0073	2.4330±0.1541	15.3091±1.1115	0.7802±0.0503	0.7242±0.0419
	LDL-LDM	0.2428±0.0145	0.0542±0.0063	2.1175±0.0235	14.5113±0.1499	0.8324±0.0035	0.7829±0.0056
	LDLCL	0.2353±0.0001	0.0532±0.0001	2.1065±0.0001	14.4195±0.0001	0.8352±0.0001	0.7849±0.0001
Ours	0.2315±0.0042	0.0531±0.0033	2.1022±0.0094	14.3971±0.0712	0.8362±0.0013	0.7859±0.0041	
FBP5500	PT-SVM	—	—	—	—	—	—
	AA-KNN	0.1637±0.0029	0.1560±0.0050	1.0129±0.0032	1.6936±0.0087	0.9375±0.0019	0.8450±0.0014
	CPNN	0.1406±0.0042	0.1614±0.0069	1.3033±0.0040	2.2741±0.0102	0.9397±0.0029	0.8213±0.0020
	SA-IIS	0.2262±0.0023	0.2118±0.0037	1.3315±0.0022	2.4097±0.0085	0.9086±0.0022	0.7511±0.0012
	SA-BFGS	0.1459±0.0021	0.1723±0.0019	1.2910±0.0018	2.2511±0.0092	0.9273±0.0031	0.8019±0.0009
	LDL-HR	0.4468±0.0008	0.3214±0.0036	1.4353±0.0039	2.7331±0.0131	0.8114±0.0005	0.6162±0.0013
	LDL-LDM	0.1146±0.0008	0.1489±0.0032	1.2836±0.0005	2.2002±0.0074	0.9513±0.0020	0.8359±0.0015
	LDLCL	0.1425±0.0001	0.1394±0.0001	1.3256±0.0001	2.2695±0.0014	0.9548±0.0001	0.8448±0.0001
Ours	0.1058±0.0030	0.1351±0.0004	1.2755±0.0077	2.1686±0.0133	0.9565±0.0009	0.8512±0.0006	
Flickr_LDL	PT-SVM	—	—	—	—	—	—
	AA-KNN	0.9558±0.2158	0.2558±0.0519	1.8745±0.1904	4.1554±0.5981	0.8190±0.0157	0.6590±0.0029
	CPNN	0.7632±0.2011	0.3545±0.0425	1.9590±0.2001	5.1726±0.4214	0.6029±0.0364	0.5733±0.0466
	SA-IIS	0.7569±0.0124	0.3411±0.0132	2.2122±0.0095	5.4952±0.0072	0.7338±0.0141	0.5558±0.0109
	SA-BFGS	0.5449±0.0235	0.3085±0.0051	1.8974±0.0012	4.3199±0.0021	0.6833±0.0049	0.6528±0.0051
	LDL-HR	0.9522±0.0142	0.4301±0.0053	2.1788±0.0061	5.4485±0.0241	0.6125±0.0014	0.4388±0.0030
	LDL-LDM	0.5388±0.0052	0.2932±0.0004	2.1600±0.0001	5.2994±0.0315	0.8238±0.0004	0.6143±0.0018
	LDLCL	0.5935±0.0001	0.2532±0.0001	2.2063±0.0001	5.4747±0.0001	0.8317±0.0001	0.6582±0.0001
Ours	0.4898±0.0034	0.2500±0.0012	2.1983±0.0025	5.4744±0.0071	0.8357±0.0018	0.6616±0.0026	
Twitter_LDL	PT-SVM	—	—	—	—	—	—
	AA-KNN	0.7951±0.2335	0.4375±0.1419	2.4126±0.2402	6.0847±0.2374	0.6892±0.1329	0.5633±0.1249
	CPNN	0.5905±0.0018	0.4216±0.0024	2.4779±0.0066	6.4501±0.0049	0.6995±0.0030	0.5491±0.0012
	SA-IIS	0.3721±0.0052	0.3196±0.0044	2.4459±0.0121	6.1998±0.0089	0.7725±0.0045	0.6038±0.0039
	SA-BFGS	0.3639±0.0031	0.2954±0.0028	2.3967±0.0078	6.1405±0.0057	0.7521±0.0030	0.5915±0.0027
	LDL-HR	1.2344±0.0131	0.5288±0.0093	2.3891±0.0079	6.1680±0.0109	0.5377±0.0067	0.3573±0.0031
	LDL-LDM	0.7501±0.0025	0.3857±0.0028	2.3677±0.0045	6.1319±0.0060	0.7818±0.0054	0.5198±0.0022
	LDLCL	0.6801±0.0001	0.3065±0.0001	2.3836±0.0001	6.1903±0.0012	0.6137±0.0001	0.8223±0.0001
Ours	0.3525±0.0002	0.2863±0.0004	2.3377±0.0012	5.9338±0.0027	0.8186±0.0002	0.6134±0.0004	

同时, 考虑到对比实验中部分方法的实验数值较为相近, 本文采用基于 Wilcoxon-Holm 方法^[49]计算的 CD 图 (critical difference diagram) 进行方法的显著性检验. 图 2 展示了不同的 LDL 方法在 6 种指标 (KL_div, Cheby, Clark, Canberra, Cosine, Intersection) 上的 CD 图. 图中横坐标为 CD 值, CD 值越小代表对应方法越好, 粗线连接的方法代表它们在该指标上的表现上没有明显的差异. 显著性实验结果表明, 本文提出的 LDLDF 方法的性能明显优于其他同类 LDL 方法.

3.5 层特征重用机制影响

为了解决深层模型的模型退化问题, 本文在所提方法中加入层特征重用机制, 该机制选取每个实例在当前层的预测分布和之前层的最优预测分布, 利用它们和真实分布之间的相似性进行特征重用. 根据算法 2, 以第 i 个实例为例, 如果第 t 层的分类器对其的预测 $L_t^{(i)}$, 性能劣于之前层对其的最优预测 Z_t 时, 我们在第 t 层重用 Z_t 作为对第 i 个实例的预测结果.

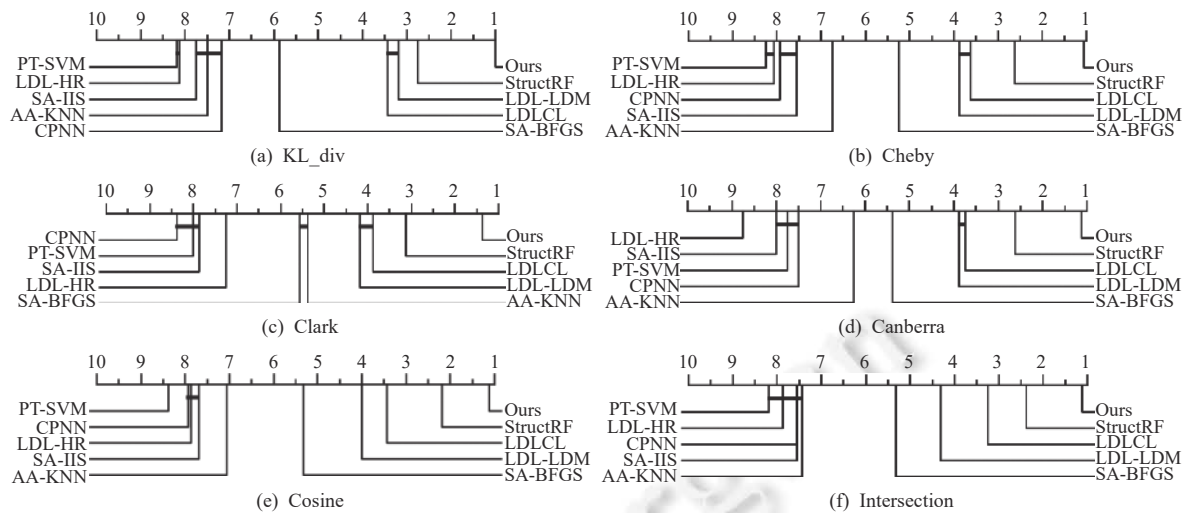


图2 不同的 LDL 方法在 6 种指标上的 CD 图

为了验证在 LDLDF 算法中加入层特征重用机制的必要性, 我们去掉该机制, 其他实验参数设置保持不变进行了消融实验. 层特征重用机制内置于层内异质集成学习模块, 如果跳过算法 1 中第 5、6 步, 直接将当前层的预测 $L^{(i)}$ 作为增强特征与原始特征样本串联, 即为不考虑层特征重用机制的 LDLDF 算法. 表 7 展示了在 14 个数据集上的实验结果, 数据集对应的第 1 行代表加入层特征重用机制, 第 2 行代表去掉层特征重用机制, 粗体标记出了每个度量的最佳结果. 由实验结果分析可知: 层特征重用机制在 LDLDF 方法中起着至关重要的作用, 加入该机制可以显著提高算法的预测性能.

表 7 是否考虑层特征重用机制影响的实验结果

数据集	KL_div↓	Cheby↓	Clark↓	Canberra↓	Cosine↑	Intersection↑
Yeast-alpha	0.0052	0.0131	0.2072	0.6740	0.9947	0.9628
	0.0065	0.0134	0.2099	0.6819	0.9945	0.9623
Yeast-cdc	0.0066	0.0160	0.2130	0.6379	0.9936	0.9580
	0.0102	0.0161	0.2148	0.6449	0.9933	0.9575
Yeast-elu	0.0058	0.0158	0.1949	0.5723	0.9943	0.9594
	0.0068	0.0160	0.1960	0.5749	0.9941	0.9594
Yeast-heat	0.0117	0.0406	0.1761	0.3522	0.9887	0.9422
	0.0136	0.0412	0.1788	0.3589	0.9883	0.9411
Yeast-diau	0.0123	0.0357	0.1939	0.4152	0.9886	0.9422
	0.0150	0.0404	0.2171	0.4660	0.9860	0.9351
Yeast-cold	0.0114	0.0496	0.1355	0.2320	0.9895	0.9430
	0.0129	0.0512	0.1397	0.2402	0.9883	0.9409
Yeast-spo	0.0238	0.0573	0.2455	0.5039	0.9777	0.9168
	0.0284	0.0590	0.2532	0.5212	0.9761	0.9143
S-JAFFE	0.0399	0.0915	0.3235	0.6615	0.9572	0.8864
	0.0458	0.0946	0.3329	0.6795	0.9550	0.8841
SBU_3DFE	0.0431	0.0974	0.3123	0.6458	0.9500	0.8833
	0.0452	0.0997	0.3182	0.6635	0.9498	0.8819
Natural scene	0.5865	0.2452	2.3621	6.4739	0.8098	0.6382
	0.6315	0.2650	2.3249	6.5387	0.7843	0.6375
Movie	0.0863	0.1071	0.4900	0.9313	0.9435	0.8481
	0.0916	0.1107	0.5106	0.9712	0.9398	0.8413

表 7 是否考虑层特征重用机制影响的实验结果 (续)

数据集	KL_div↓	Cheby↓	Clark↓	Canberra↓	Cosine↑	Intersection↑
Human gene	0.2315	0.0531	2.1022	14.3870	0.8362	0.7859
	0.2414	0.0542	2.1185	14.4940	0.8330	0.7836
Flickr_LDL	0.4898	0.2500	2.1983	5.4744	0.8367	0.6616
	0.4935	0.2549	2.1833	5.4983	0.8364	0.6557
Twitter_LDL	0.3525	0.2863	2.3377	5.9338	0.8186	0.6134
	0.3564	0.2867	2.3412	5.9462	0.8099	0.5951

3.6 参数敏感性实验

为了探究 LDLDF 层中分类器数量 S 和模型更新阈值 θ 对实验结果的影响, 本文进行了参数敏感性实验, S 在 $[2, 10]$ 范围内以 2 为步长取值 ($S=[2, 4, 6, 8, 10]$), 更新阈值 θ 同样选择了 4 种不同的取值 $\theta = [0.001, 0.005, 0.01, 0.05, 0.1]$, 在所有数据集上运行 LDLDF 方法. 图 3 中以 S-JAFFE 数据集和 Yeast-cdc 数据集为例, 展示出参数敏感性实验的结果, 在其他数据集上可以得到相同的观察结果. 可以看出, LDLDF 对分类器个数 S 和模型更新阈值 θ 这两个参数是具有鲁棒性的, 参数在一定范围内的变化不会引起预测结果较大的波动, 这样有利于将我们的方法推广到更多的数据集上. 综合在 14 个数据集上的实验结果, 我们发现在 $S \in [4, 6]$ 、 $\theta = 0.005$ 时模型可以取得稳定的优异性能.

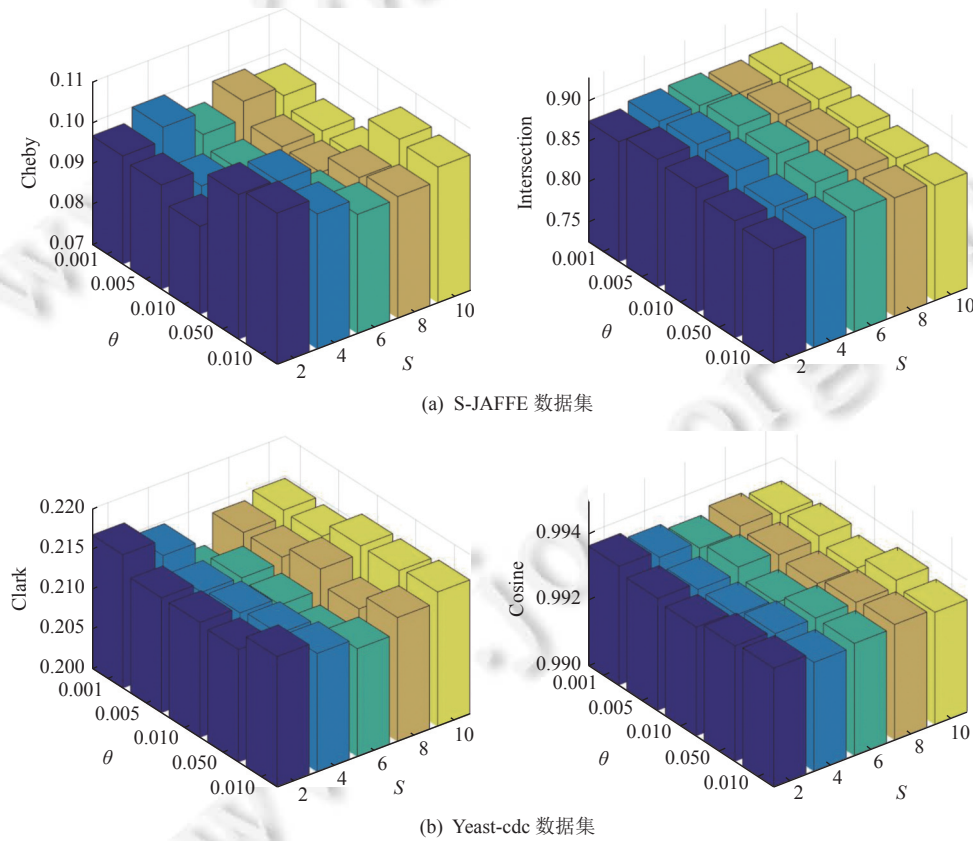


图 3 在 S-JAFFE 和 Yeast-cdc 数据集上的参数敏感性分析

4 结 论

本文提出了一个联合深度森林与异质集成的标记分布学习方法 (LDLDF). 该方法采用深度森林的级联结构模拟多层处理结构的深度学习模型, 在每个级联层中组合多个异质分类器增加集成的多样性, 赋予模型强大的表示学习能力和泛化能力. 在模型更新过程中, 考虑到深层模型可能出现的模型退化问题, 我们引入层特征重用机制, 降低模型的训练误差, 有效利用深层模型每层的预测能力. 通过大量实验验证了本文这一方法的有效性. 在后续的工作中, 我们也将继续探讨和证明如何在 LDLDF 方法中引入标签的全局相关性和局部相关性, 并探究更有效的基分类器算法来提高模型的预测性能.

References:

- [1] Gibaja E, Ventura S. Multi-label learning: A review of the state of the art and ongoing research. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2014, 4(6): 411–444. [doi: [10.1002/widm.1139](https://doi.org/10.1002/widm.1139)]
- [2] Geng X, Luo LR. Multilabel ranking with inconsistent rankers. In: *Proc. of the 2014 IEEE Conf. on Computer Vision and Pattern Recognition*. Columbus: IEEE, 2014. 3742–3747. [doi: [10.1109/CVPR.2014.478](https://doi.org/10.1109/CVPR.2014.478)]
- [3] Geng X. Label distribution learning. *IEEE Trans. on Knowledge and Data Engineering*, 2016, 28(7): 1734–1748. [doi: [10.1109/TKDE.2016.2545658](https://doi.org/10.1109/TKDE.2016.2545658)]
- [4] Geng X, Yin C, Zhou ZH. Facial age estimation by learning from label distributions. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2013, 35(10): 2401–2412. [doi: [10.1109/TPAMI.2013.51](https://doi.org/10.1109/TPAMI.2013.51)]
- [5] Zhang ZX, Wang M, Geng X. Crowd counting in public video surveillance by label distribution learning. *Neurocomputing*, 2015, 166: 151–163. [doi: [10.1016/j.neucom.2015.03.083](https://doi.org/10.1016/j.neucom.2015.03.083)]
- [6] Qi L, Shen JY, Liu JQ, Shi YH, Geng X. Label distribution learning for generalizable multi-source person re-identification. *IEEE Trans. on Information Forensics and Security*, 2022, 17: 3139–3150. [doi: [10.1109/TIFS.2022.3204219](https://doi.org/10.1109/TIFS.2022.3204219)]
- [7] Xing C, Geng X, Xue H. Logistic boosting regression for label distribution learning. In: *Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016. 4489–4497. [doi: [10.1109/CVPR.2016.486](https://doi.org/10.1109/CVPR.2016.486)]
- [8] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 2017, 60(6): 84–90. [doi: [10.1145/3065386](https://doi.org/10.1145/3065386)]
- [9] Hinton G, Deng L, Yu D, Dahl GE, Mohamed AR, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath TN, Kingsbury B. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 2012, 29(6): 82–97. [doi: [10.1109/MSP.2012.2205597](https://doi.org/10.1109/MSP.2012.2205597)]
- [10] Gao BB, Xing C, Xie CW, Wu JX, Geng X. Deep label distribution learning with label ambiguity. *IEEE Trans. on Image Processing*, 2017, 26(6): 2825–2838. [doi: [10.1109/TIP.2017.2689998](https://doi.org/10.1109/TIP.2017.2689998)]
- [11] Zhou ZH, Feng J. Deep forest: Towards an alternative to deep neural networks. In: *Proc. of the 26th Int'l Joint Conf. on Artificial Intelligence*. Melbourne: AAAI Press, 2017. 3553–3559. [doi: [10.5555/3172077.3172386](https://doi.org/10.5555/3172077.3172386)]
- [12] Breiman L. Random forests. *Machine Learning*, 2001, 45(1): 5–32. [doi: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324)]
- [13] Chen TQ, Guestrin C. XGBoost: A scalable tree boosting system. In: *Proc. of the 22nd ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*. San Francisco: ACM, 2016. 785–794. [doi: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785)]
- [14] He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. In: *Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016. 770–778. [doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90)]
- [15] Pang M, Ting KM, Zhao P, Zhou ZH. Improving deep forest by confidence screening. In: *Proc. of the 2018 IEEE Int'l Conf. on Data Mining*. Singapore: IEEE, 2018. 1194–1199. [doi: [10.1109/ICDM.2018.00158](https://doi.org/10.1109/ICDM.2018.00158)]
- [16] Zhou ZH. Ensemble learning. In: Zhou ZH, ed. *Machine Learning*. Singapore: Springer, 2021. 181–210.
- [17] Ganaie MA, Hu MH, Malik AK, Tanveer M, Suganthan PN. Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, 2022, 115: 105151. [doi: [10.1016/j.engappai.2022.105151](https://doi.org/10.1016/j.engappai.2022.105151)]
- [18] Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 1997, 55(1): 119–139. [doi: [10.1006/jcss.1997.1504](https://doi.org/10.1006/jcss.1997.1504)]
- [19] Friedman JH. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 2001, 29(5): 1189–1232. [doi: [10.1214/aos/1013203451](https://doi.org/10.1214/aos/1013203451)]
- [20] van Rijn JN, Holmes G, Pfahringer B, Vanschoren J. The online performance estimation framework: Heterogeneous ensemble learning

- for data streams. *Machine Learning*, 2018, 107(1): 149–176. [doi: [10.1007/s10994-017-5686-9](https://doi.org/10.1007/s10994-017-5686-9)]
- [21] Fang ZC, Wang Y, Peng L, Hong HY. A comparative study of heterogeneous ensemble-learning techniques for landslide susceptibility mapping. *Int'l Journal of Geographical Information Science*, 2021, 35(2): 321–347. [doi: [10.1080/13658816.2020.1808897](https://doi.org/10.1080/13658816.2020.1808897)]
- [22] Wolpert DH. Stacked generalization. *Neural Networks*, 1992, 5(2): 241–259. [doi: [10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1)]
- [23] Feng J, Zhou ZH. AutoEncoder by forest. In: *Proc. of the 32nd AAAI Conf. on Artificial Intelligence*. New Orleans: AAAI, 2018. 2967–2973. [doi: [10.5555/3504035.3504397](https://doi.org/10.5555/3504035.3504397)]
- [24] Yang L, Wu XZ, Jiang Y, Zhou ZH. Multi-label learning with deep forest. In: *Proc. of the 24th European Conf. on Artificial Intelligence*. Santiago de Compostela: IOS Press, 2020. 1634–1641. [doi: [10.3233/FAIA200274](https://doi.org/10.3233/FAIA200274)]
- [25] Guo Y, Liu SH, Li ZH, Shang XQ. BCDForest: A boosting cascade deep forest model towards the classification of cancer subtypes based on gene expression data. *BMC Bioinformatics*, 2018, 19(5): 118. [doi: [10.1186/s12859-018-2095-4](https://doi.org/10.1186/s12859-018-2095-4)]
- [26] Wang QW, Yang L, Li YF. Learning from weak-label data: A deep forest expedition. In: *Proc. of the 34th AAAI Conf. on Artificial Intelligence*. New York: AAAI, 2020. 6251–6258. [doi: [10.1609/aaai.v34i04.6092](https://doi.org/10.1609/aaai.v34i04.6092)]
- [27] Xu N, Liu YP, Geng X. Label enhancement for label distribution learning. *IEEE Trans. on Knowledge and Data Engineering*, 2021, 33(4): 1632–1643. [doi: [10.1109/TKDE.2019.2947040](https://doi.org/10.1109/TKDE.2019.2947040)]
- [28] Zheng QH, Zhu JH, Tang HY, Liu XY, Li ZY, Lu HM. Generalized label enhancement with sample correlations. *IEEE Trans. on Knowledge and Data Engineering*, 2023, 35(1): 482–495. [doi: [10.1109/TKDE.2021.3073157](https://doi.org/10.1109/TKDE.2021.3073157)]
- [29] Xu N, Shu J, Zheng RY, Geng X, Meng DY, Zhang ML. Variational label enhancement. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2023, 45(5): 6537–6551. [doi: [10.1109/TPAMI.2022.3203678](https://doi.org/10.1109/TPAMI.2022.3203678)]
- [30] Gao YB, Zhang Y, Geng X. Label enhancement for label distribution learning via prior knowledge. In: *Proc. of the 29th Int'l Joint Conf. on Artificial Intelligence*. Yokohama: IJCAI.org, 2021. 3223–3229. [doi: [10.5555/3491440.3491886](https://doi.org/10.5555/3491440.3491886)]
- [31] Gao BB, Zhou HY, Wu JX, Geng X. Age estimation using expectation of label distribution learning. In: *Proc. of the 27th Int'l Joint Conf. on Artificial Intelligence*. Stockholm: AAAI, 2018. 712–718. [doi: [10.5555/3304415.3304517](https://doi.org/10.5555/3304415.3304517)]
- [32] Su K, Geng X. Soft facial landmark detection by label distribution learning. In: *Proc. of the 33rd AAAI Conf. on Artificial Intelligence*. Honolulu: AAAI, 2019. 5008–5015. [doi: [10.1609/aaai.v33i01.33015008](https://doi.org/10.1609/aaai.v33i01.33015008)]
- [33] Zhao ZJ, Ma XJ. Text emotion distribution learning from small sample: A meta-learning approach. In: *Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing*. Hong Kong: Association for Computational Linguistics, 2019. 3957–3967. [doi: [10.18653/v1/D19-1408](https://doi.org/10.18653/v1/D19-1408)]
- [34] Wu XP, Wen N, Liang J, Lai YK, She DY, Cheng MM, Yang JF. Joint acne image grading and counting via label distribution learning. In: *Proc. of the 2019 IEEE/CVF Int'l Conf. on Computer Vision*. Seoul: IEEE, 2019. 10641–10650. [doi: [10.1109/ICCV.2019.01074](https://doi.org/10.1109/ICCV.2019.01074)]
- [35] Jia XY, Zheng X, Li WW, Zhang XQ, Li ZC. Facial emotion distribution learning by exploiting low-rank label correlations locally. In: *Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019. 9833–9842. [doi: [10.1109/CVPR.2019.01007](https://doi.org/10.1109/CVPR.2019.01007)]
- [36] Yin C, Geng X. Facial age estimation by conditional probability neural network. In: *Proc. of the 2012 Chinese Conf. on Pattern Recognition*. Beijing: Springer, 2012. 243–250. [doi: [10.1007/978-3-642-33506-8_31](https://doi.org/10.1007/978-3-642-33506-8_31)]
- [37] Li RY, Zhu JH, Liu XY. Label distribution learning with collaboration among labels. *Ruan Jian Xue Bao/Journal of Software*, 2022, 33(2): 539–554. (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/6139.htm> [doi: [10.13328/j.cnki.jos.006139](https://doi.org/10.13328/j.cnki.jos.006139)]
- [38] Wang YB, Tian WQ, Cheng YS. Heterogeneous ensemble learning algorithm based on label distribution learning. *Pattern Recognition and Artificial Intelligence*, 2019, 32(10): 945–954 (in Chinese with English abstract). [doi: [10.16451/j.cnki.issn1003-6059.201910009](https://doi.org/10.16451/j.cnki.issn1003-6059.201910009)]
- [39] Geng X, Hou P. Pre-release prediction of crowd opinion on movies by label distribution learning. In: *Proc. of the 24th Int'l Conf. on Artificial Intelligence*. Buenos: AAAI, 2015. 3511–3517. [doi: [10.5555/2832581.2832738](https://doi.org/10.5555/2832581.2832738)]
- [40] Shen W, Zhao K, Guo YL, Yuille A. Label distribution learning forests. In: *Proc. of the 31st Int'l Conf. on Neural Information Processing Systems*. Long Beach: Curran Associates Inc., 2017. 834–843. [doi: [10.5555/3294771.3294851](https://doi.org/10.5555/3294771.3294851)]
- [41] Zhao CS, Xin Y, Li XF, Yang YX, Chen YL. A heterogeneous ensemble learning framework for spam detection in social networks with imbalanced data. *Applied Sciences*, 2020, 10(3): 936. [doi: [10.3390/app10030936](https://doi.org/10.3390/app10030936)]
- [42] Lyons M, Akamatsu S, Kamachi M, Gyoba J. Coding facial expressions with Gabor wavelets. In: *Proc. of the 3rd IEEE Int'l Conf. on Automatic Face and Gesture Recognition*. Nara: IEEE, 1998. 200–205. [doi: [10.1109/AFGR.1998.670949](https://doi.org/10.1109/AFGR.1998.670949)]
- [43] Yin LJ, Wei XZ, Sun Y, Wang J, Rosato MJ. A 3D facial expression database for facial behavior research. In: *Proc. of the 7th Int'l Conf. on Automatic Face and Gesture Recognition*. Southampton: IEEE, 2006. 211–216. [doi: [10.1109/FGR.2006.6](https://doi.org/10.1109/FGR.2006.6)]
- [44] Liang LY, Lin LJ, Jin LW, Xie DR, Li MR. SCUT-FBP5500: A diverse benchmark dataset for multi-paradigm facial beauty prediction. In: *Proc. of the 24th Int'l Conf. on Pattern Recognition*. Beijing: IEEE, 2018. 1598–1603. [doi: [10.1109/ICPR.2018.8546038](https://doi.org/10.1109/ICPR.2018.8546038)]

- [45] Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc. of the National Academy of Sciences of the United States of America, 1998, 95(25): 14863–14868. [doi: 10.1073/pnas.95.25.14863]
- [46] Chen MT, Wang XG, Feng B, Liu WY. Structured random forest for label distribution learning. Neurocomputing, 2018, 320: 171–182. [doi: 10.1016/j.neucom.2018.09.002]
- [47] Wang J, Geng X. Learn the highest label and rest label description degrees. In: Proc. of the 30th Int'l Joint Conf. on Artificial Intelligence. Montreal: ijcai.org, 2021. 3097–3103. [doi: 10.24963/ijcai.2021/426]
- [48] Wang J, Geng X. Label distribution learning by exploiting label distribution manifold. IEEE Trans. on Neural Networks and Learning Systems, 2023, 34(2): 839–852. [doi: 10.1109/TNNLS.2021.3103178]
- [49] Ismail Fawaz H, Forestier G, Weber J, Idoumghar L, Muller PA. Deep learning for time series classification: A review. Data Mining and Knowledge Discovery, 2019, 33(4): 917–963. [doi: 10.1007/s10618-019-00619-1]

附中文参考文献:

- [37] 李睿钰, 祝继华, 刘新媛. 考虑标记间协作的标记分布学习. 软件学报, 2022, 33(2): 539–554. <http://www.jos.org.cn/1000-9825/6139.htm> [doi: 10.13328/j.cnki.jos.006139]
- [38] 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]



王艺霏(1999—), 女, 硕士生, 主要研究领域为机器学习.



刘新媛(1996—), 女, 博士生, 主要研究领域为机器学习.



祝继华(1982—), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为计算机视觉, 机器学习.



周熠炀(1998—), 男, 硕士生, 主要研究领域为机器学习.