文章编号:1000-2367(2023)03-0111-12 DOI:10.16366/j.cnki.1000-2367.2023.03.014

基于 Stacking 特征增强多粒度联级 Logistic 的个人信用评估

侯天宝,王爱银

(新疆财经大学 统计与数据科学学院,乌鲁木齐 830012)

摘 要:主要针对广受关注的 P2P 网贷信用评估问题,利用机器学习方法提高申请人网贷违约预测准确率,研究出基于 Stacking 特征增强多粒度联级 Logistic 方法及其应用.所提分类器是一种混合模型,结合了 Stacking 集成学习和联级 Logistic 学习的思想.首先,通过网格搜索技术分别建立 XGBoost,Catboost,LightGBM,AdaBoost 以及 Gradient Boosting 模型,并筛选出适合的基评估器作为 Stacking 集成的初级学习器,logistic 模型作为次级学习器,构建基于 Stacking 的多粒度扫描器,生成预测结果作为元特征,拼接成新特征数据.其次,通过新特征数据以及元特征在每级 Logistic 上的特征增强建立联级 Logistic Regression 模型,并且与现有的单一集成学习器和各基评估器在 3 个不同的 P2P 网贷信用评估数据集上进行对比.实验结果表明,通过 AUC、准确率等指标对其进行评价,相比于各基评估器以及其他单一集成分类器,基于 Stacking 增强多粒度联级 Logistic 模型有较高的准确率,预测效果更优.

关键词:个人信用;特征增强;Stacking 集成;多粒度扫描;联级 Logistic 模型

中图分类号:TP393 **文献标志码**:A

在“互联网金融”的快速发展背景下,P2P 网贷近几年来发展规模日益壮大,其业务涉及领域也层出不穷.该模式下借贷方式是个人对个人进行信贷,P2P 平台作为中介实现借款人想要借款和贷款人放贷进行投资的需求.其典型的运行模式是借款人和贷款人在平台上自由竞价,每次借贷双方达成交易,平台收取相应的中介费.P2P 网贷平台具有投资门槛低、成本代价小并且交易更加便捷等优势.但万物具有两面性,P2P 不断发展的同时潜在的信用风险问题也日益凸显,不少借款人因违约未按期还款给平台以及行业造成了不利的影响.据有利网在官方微信公众号发布严重失信借款人信息,每周都有失信人信息公布,且人数都在 2 000 人左右.由此可见,个人信用评估在 P2P 平台的重要性.

针对个人信用评估问题,提出一种基于 Stacking 特征增强多粒度联级 Logistic 模型,该模型的构建主要包括两部分:第一部分,基于 Stacking 集成的多粒度扫描器的构建,利用网格搜索技术进行调优,以准确率为指标,选出合适的基评估器作为 Stacking 框架的初级学习器,并用 Logistic 模型作为次级学习器.第二部分,多层联级 Logistic 模型的构建,为防止整个模型的过拟合现象,在每一层中的 Logistic 模型进行交叉检验,同时在训练过程中用模型的惩罚系数去自适应调节优化整个联级 Logistic 模型的预测性能,以此提升模型准确率.

1 相关工作

信用评估是指评估机构利用相关权威专家的判断或者数学建模方式,结合融资者所提供的财务信息、经营相关信息以及历史还款信息等各类指标综合进行分析,对融资者是否如期按照约定偿还债务本息的能力以及意愿进行评估,并按照其违约概率的不同以划分等级或分数的形式给出评估结论的行为.

收稿日期:2022-04-24;**修回日期**:2022-05-11.

基金项目:国家社科基金(18BJL072).

作者简介:侯天宝(1998—),男,新疆乌鲁木齐人,新疆财经大学硕士研究生,研究方向为机器学习在金融风险中的应用,E-mail:2228316609@qq.com.

通信作者:王爱银(1978—),女,新疆乌鲁木齐人,新疆财经大学副教授,博士,研究方向为信用风险、投资组合研究,E-mail:way8848@sina.com.

在个人信用评估研究领域,评估的方式有三大类:一类是基于统计学方法,运用统计方面的性质来分析数据的分布及规律并进行建模,常见的方法有逻辑回归(Logistic Regression)和线性判别分析(Linear Discriminant Analysis).如 OHLSON[1] 首次将逻辑回归模型应用到信用评估问题中,WIGINTON[2] 在研究模型信用评估问题上,将逻辑回归与线性判别分析法加以对比,实验结果表明逻辑回归评估效果优于线性判别分析.STEENACKERS 等[3] 结合逻辑回归模型与极大似然估计建立了信用评估模型,将数据划分为好贷款、坏贷款以及拒绝贷款等三大类,用极大似然估计法去迭代逻辑回归模型,从而找出最接近真实值的最优参数.虽然统计学方法可以建立各种分类问题模型,但由于需要较为严格的假设条件,并且缺少对数据特征的学习,导致预测准确率不高.第二类,继统计学方法之后,随着机器学习的兴起,不少学者将机器学习方法引入到信用评估研究领域,无须刻意找寻数据中存在的规律,重点在于通过对数据特征的学习来提高预测精度,打破了统计学方法中建立模型的局限性.如 YEH 等[4] 建立 K 近邻(K-nearest neighbor)、逻辑回归、神经网络(Neural networks)、朴素贝叶斯(Naive Bayesian)和判别分析等模型来进行信用评估,发现 K 近邻、神经网络以及朴素贝叶斯模型都优于逻辑回归与判别分析.随着研究的不断深入,影响信用评估的因素不断被探索出来,数据的维度也呈现爆炸式增长,单一的机器学习器评估性能受限,HANSEN 等[5] 研究发现构建多个模型并按一定的规则结合起来,能显著提高整个模型的泛化性能,准确率有所提高,说明集成模型比单一模型预测性能更好.TWALA[6] 通过实验发现,集成算法能有效提高信用评估的准确度.集成学习在信用评估中的应用主要分为装袋法(Bagging)、提升法(Boosting)和堆叠法(Stacking),装袋法的主要思想是从原始数据集中抽取训练集,每轮从原始样本中使用自助采样法抽取多个训练样本,并用多个基评估器进行训练,对于分类问题,采取投票的方式获得最终结果;对于回归问题,则以均值的方式获得最终结果,如经典的随机森林[7](Random Forest,RF)算法.提升法的主要思想是每次使用全部的样本,每轮训练减小上一轮的预测正确的样本权重,增大预测错误样本的权重,通过自动调节权重来提升模型的预测性能,如类别型提升[8](Categorical Boosting,CatBoost)、自适应提升[9](Adaptive Boosting,AdaBoost)、梯度提升树[10](Gradient Boosting Decision Tree,GBDT)、极端梯度提升[11](eXtreme Gradient Boosting,XGBoost)、轻量级梯度提升机[12](Light Gradient Boosting Machine,LightGBM),其中 XGBoost 和 LightGBM 运算速度快、准确度高常在 Kaggle 竞赛中被使用,各类集成算法被广泛应用于信用评估研究领域.Stacking 的主要思想是通过组合多种评估器来提升模型泛化性,分为初级学习器和次级学习器两层,将初级学习器的预测结果作为元特征来构建次级学习器模型,以提高预测性能.丁岚等[13] 以 logistic 回归、决策树(Decision Tree)、支持向量机(SVM)作为初级学习器,以 SVM 作为次级学习器来搭建 Stacking 集成框架违约风险评估模型,与单一学习器进行比较,发现集成 Stacking 框架预测性能更好.随着对深度学习的不断深入研究,发现神经网络结构具有良好的特征学习能力以及优化性能,如刘伟江等[14] 通过数据成像技术,将个人违约数据转化成图像,应用 LeNet-5 模型进行信用评估识别,有着较高的识别精确度;王重仁等[15] 将原始数据集进行编码形成包含时间维度和行为维度的灰度图像,并构建具有注意力机制的长短期神经网络(Long Short-term Memory,LSTM)与卷积神经网络(Convolutional Neural Network,CNN)融合模型进行信用评估,并与集成模型进行对比,具有较优的预测性能.第三类,近年来不少学者将机器学习方法与统计学方法相结合应用于信用评估领域,比起单一的学习器预测性能更好,如陈倩等[16] 构建基于随机森林的弹性网络回归模型(RF-Elastic-Net-Logistic),并与 RF-Lasso 回归进行对比,实验结果表明 RF-ElasticNet-Logistic 模型具有较强的预测性能;李佳欣[17] 通过应用逐步 Logistic 回归依据 AIC 准则对数据进行特征选择,并分别与决策树、条件推荐树、随机森林和支持向量机模型进行对比,发现选择较少特征的模型在验证集上的误差比全变量模型要低,并且基于逐步 Logistic 回归的随机森林预测准确率最优;曹再辉等[18] 以支持向量机、随机森林、人工神经网络(ANN)及梯度提升树作为初级学习器,逻辑回归作为次级学习器来搭建 Stacking 集成框架违约风险评估模型,与单一学习器、投票集成方法以及人工神经网络进行对比,预测性能最优.

综上所述,本文提出基于 Stacking 特征增强多粒度联级 Logistic 模型(Deep_Logistic),该结构类似于深度神经网络(Deep Neural Network,DNN),由于 Logistic 模型缺少对于数据的特征学习,造成预测准确度下降,通过用 Stacking 集成获取元特征来增强每一级 Logistic 进行训练前的特征.除此之外,在整个信用评估建模过程中,包括数据预处理、建立模型、调参、评估.之后,通过 AUC、准确率等评价指标对模型进行论证,

与此同时与构建 Stacking 集成的基评估器、常见现有的单一集成分类器在 3 个不同的个人信用评估数据集上进行对比分析.

2　模型构建理论基础

2.1　相关模型的概述

通过对相关文献的梳理，发现集成学习的方法可以融合多个单一学习器的预测效果，可提升预测的准确度，并且 Stacking 集成相对 Bagging 和 Boosting，往往预测精度更高，而且过拟合的风险会更低[19]，据此选用 Stacking 作为元特征提取器.为了生成有效的元特征对后续联级 Logistic 模型进行特征增强，需要建立有效的 Stacking 集成模型和初级学习器.据此，本文选取信用风险预警问题中应用最广泛且常见集成学习方法作为 Stacking 中的预选初级学习器，包括 XGBoost，Catboost，LightGBM，Adaboost 以及 GBDT 模型.

极端梯度提升(XGBoost)是一种改进的梯度提升决策树模型，仅以决策树为基分类器，属于 Boosting 集成学习，并且可进行多线程并行运算，故而运算速度较快，在数据挖掘领域受到广泛应用.XGBoost 的实质是由众多决策树集成而来，模型：

$$\hat{y}=\sum_{k=1}^{K} f_k(X_i), f_k \in F, \tag{1}$$

其中，K 表示模型中决策树的个数，X_i 表示输入模型中第 i 个样本，$f_k(X_i)$ 表示第 k 树在样本为 X_i 下的预测值，$\hat{y}$ 表示经过 K 次迭代后模型的预测值，$F=\{f(x)=u_{q(x)}\}(q:\mathbf{R}^m \rightarrow T, u \in \mathbf{R}^{\mathrm{T}})$，$F$ 表示决策树的集合，q 表示树的结构，T 表示每个决策树的叶子节点数，u 表示每棵树的叶子结点对应分数的集合.

LightGBM 与 XGBoost 的原理相似，该算法由微软提出，属于 Boosting 集成算法的一种.LightGBM 是一种基于直方图算法的决策树，直方图将浮点型连续特征进行离散化，在直方图遍历数据后，根据直方图的离散值，通过累计必要的统计信息，寻找出最优分界点.

AdaBoost 是 Boosting 中经典算法之一，该算法通过不断更新迭代正确、错误样本权重，能降低数据随机性波动导致的泛化错误率，使其具有较好的泛化能力[20].AdaBoost 首先给样本 $(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$ 赋予同等权重 $1/n$，然后引入分类算法构建分类模型，模型在不断迭代学习的过程中，不断更新样本的权重，每轮提高被错判样本的权重，减少正确判断的权重，以提升对分类问题中样本量较少的类别的敏感度.经过重复迭代 T 次后得到预测函数 $f_1,f_2,\cdots,f_T$，最后经过加权投票决定，预测效果越好权重越大，反之则越小.在实际分类问题中，这就使得即使在数据集不均衡的条件下 AdaBoost 更偏向错判的样本，但由于其优点，也容易导致过拟合，不太稳定.

GBDT[21−22]，又称 MART(Multiple Additive Regression Tree)，是一种基于迭代的决策树算法，属于 Boosting 集成的一种，该算法由多个决策树组成.和 AdaBoost 类似，GBDT 也是重复选择一个模型并且每次基于先前的模型表现进行调整，所不同的是，AdaBoost 是通过提升错分数点的权重来定位模型的不足，而 GBDT 是通过计算梯度来定位模型的不足，这也使得在数据集不均衡的条件下更偏向错判的样本.GBDT 主要思想是通过不断迭代的方式去减少残差，并且沿着梯度方向进行不断优化来形成多个分类回归决策树，最后将得到的所有决策树的结论进行累加起来得到最终的模型，通过这种不断迭代学习的过程，从而得到较为精准的预测结果.

CatBoost，由 2017 年俄罗斯搜索巨头 Yandex 所研究，属于 Boosting 集成.CatBoost 是基于一种对称决策树的 GBDT 框架，具有支持分类、精度高等优点.CatBoost 主要的优点是合理进行处理类别特征，Categorical 和 Boosting 组成了 CatBoost.其次，CatBoost 还解决了梯度偏差(Gradient Bias)和预测偏移(Prediction Shift)的问题，减少了过拟合现象的发生，同时使用整个数据集进行训练，对数据信息进行高效提取.

对于二分类问题.首先，CatBoost 对数值特征进行二值化，即通过 oblivious 树为基础预测器，将浮点特征、统计信息及独热编码进行二值化.其次，将类别型特征转化为数值型特征，主要是对观测集进行随机排序生成多个随机序列、给定的某个序列利用训练集的平均标签值代替类别(对于出现次数较少的类别进行平滑处理)、利用先验的思想将分类数据转化为数值型.之后，采用“贪婪策略”进行特征组合.其中在处理梯度偏差问题方面，在确定树结构基础上计算叶子节点值，根据不同分割方式下获得的叶节点值，对获得的树进行

评分,从而获得最佳分割方式,之后采用梯度或牛顿步长来逼近叶子节点值.

CatBoost 实现了训练数据集与处理类特征的同步,极大提高了处理特征的效率;通过计算叶节点的算法来避免"过拟合",提高模型的泛化性.

2.2 模型的构建

本文提出的个人信用模型构建的基本流程图,如图 1 所示.

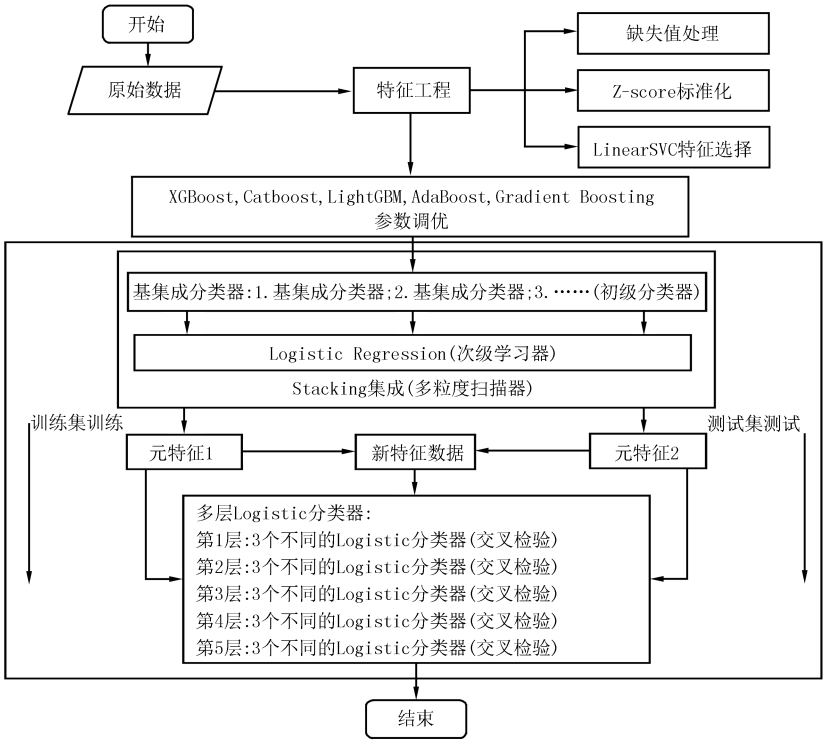

图1 信用评估建模流程

Fig. 1 The Procession of credit evaluation modeling

2.2.1 数据预处理概述

为了降低数据集本身对于分类器的影响,对原始数据集采取预处理,包括缺失值处理、规范化处理和特征选择,其中规范化处理采用 Z 分数标准化(Z-Score Standardization),特征选择使用线性分类支持向量机[23](LinearSVC)方法.

2.2.2 分类器设计

(1)分类器设计思路及结构说明

分层训练的训练过程采取逐层训练的方式.联级结构属于一种多层联级的训练框架,通过算法将不同层的训练结果进行结合.在前向传播设计方面,基于联级结构中后续层的分类器通过前一层的反馈进行训练,由于每层的 Logistic 模型自身缺乏对数据特征的学习,在进行前向传播训练中需要进行特征增强处理.受深度森林[24]构造原理启发,在联级结构外部设计基于 Stacking 集成的多粒度扫描器进行训练预测生成多个元特征,并组合生成新特征向量用于联级结构的起始输入训练以及元特征用于每级训练的特征增强,为避免过拟合现象的发生,每级 Logistic 模型均使用交叉检验.整体的分类器主要涉及两方面的结构,一是多粒度扫描器的构建,二是联级 Logistic 结构的构建.

(2)多粒度扫描器的构建

受卷积神经网络和深度森林[24]构造原理的启发,对原始输入特征使用基于不同随机模式 Stacking 的多粒度扫描方式产生联级 Logistic 的输入特征向量.其中,Stacking 集成中初级基集成学习器选用预测性能较高的学习器,即基集成学习器作为单一学习器经过网格搜索调优后,训练预测正确率之间有着明显差异中较高得分的学习器作为初级学习器.

例如图 2 中所示,对于维度为的输入 v 数据,若采用 d 维滑动窗口对输入的特征进行处理,且分类标签类别数为 c,通过 Stacking 的训练预测,最终可以得到 $c(v-d+1)$ 维度特征向量.在实验中,通过对滑动窗口维度的设置,从而最终可以得到不同粒度的特征向量.

(3)联级 Logistic 结构的构建

受多层感知器神经网络[25](MLP)的构造原理启发,类似于神经网络中对输入特征经过每层以 sigmoid 激活函数的多个神经元进行训练,从而构造联级 Logistic.在联级 Logistic 中,每一级都包含多个不同随机模式的 Logistic,由多粒度扫描器产生的元特征对每级训练前进行特征增强,使得每经过一级的训练,预测准确率都有明显提升.为了防止过拟合现象的发生,每级中的不同 Logistic 均进行交叉检验.整个构造原理如

图 3所示，联级 Logistic 的级数可以进行自定义，每一级 Logistic 学习输入特征向量的信息，经过处理后传输到下一层，每一级中 Logistic 均采取自适应惩罚系数进行调节，同时对于起初某些不太均衡的数据集可由 Logistic 自身 class_weight 参数设置进行微调，并且每级结束后均进行预测估计.与此同时，设定了早停机制，即每级相比上一级预测效果没有显著提升，不必增加层级的深度，终止训练过程.

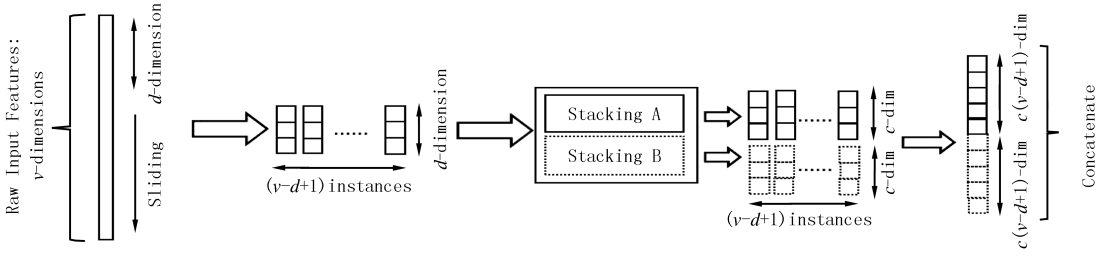

图2　基于Stacking多粒度扫描器结构

Fig.2　Structure of multi-grained scanner based on Stacking

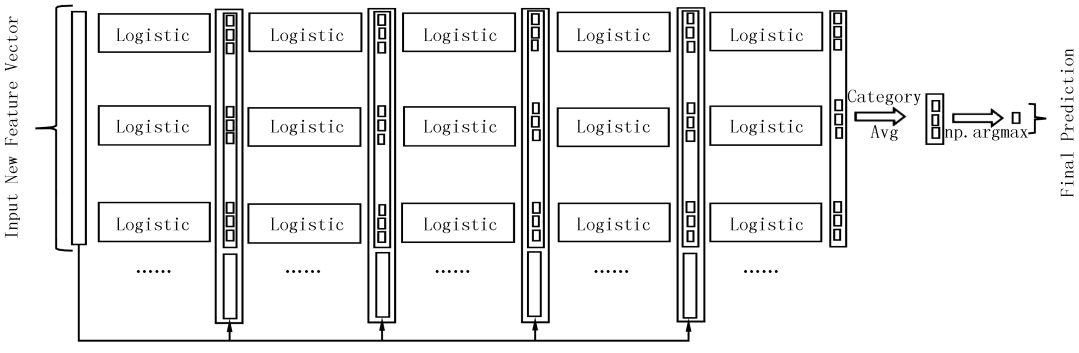

图3　联级Logistic结构

Fig.3　Structure of cascade Logistic

(4)总体结构

图 4 是基于 Stacking 特征增强多粒度联级 Logistic 的总体框架，若输入特征为 v 维，分类标签种数为 c，多粒度扫描器模块具有 2 个滑动窗口，经过第一个单位滑动维度为 d_1 的多粒度扫描器和第二个单位滑动维度为 d_2 的多粒度扫描器，最终可以得到维度为 $2c(2v-(d_1+d_2)+2)$的特征向量作为联级 Logistic 的第一级输入.每一个多粒度扫描器除了生成联级 Logistic 的开始输入，还交替用于每级 Logistic 训练前的特征增强，经过每级的训练，不断重复这一过程，直到预测性能达到最优.

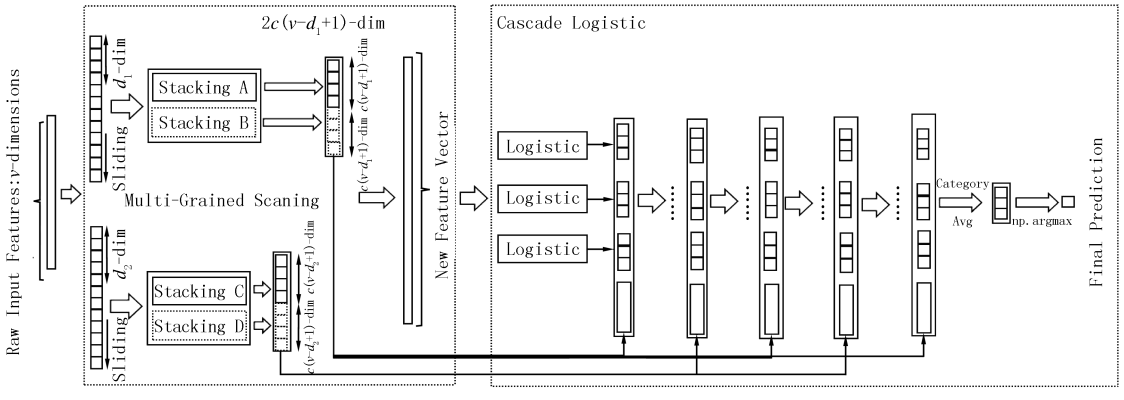

图4　基于Stacking多粒度扫描器联级Logistic总体框架

Fig.4　Overall framework of cascade Logistic based on multi-grained scanner of Stacking

3 信用评估实验

3.1 评价指标构建

在分类算法中,常见的分类指标有准确率(Accuracy)、精准率(Precision)、召回率(Recall)、真正例率 TPR(True Positive Rate)、假正率 FPR(False Positive Rate)、ROC(Receiver Operating Characteristic)曲线和 AUC(Area Under Curve)[26-28]等,这些指标都是由混淆矩阵(Confuse Matrix)中的真正类 TP(True Positive,T_P)、假负类 FN(False Negative,F_N)、假正类 FP(False Positive,F_P)和真负类 TN(True Negative,T_N)计算得来.

准确率(Accuracy,简称为 A)表示预测正确的概率:

$$A=\frac{T_P+T_N}{T_P+T_N+F_P+F_N}. \tag{2}$$

精确率(Precision,简称为 P)表示正确预测为正样本(T_P)占所有预测为正样本(T_P+F_P)的比重:

$$P=\frac{T_P}{T_P+F_P}. \tag{3}$$

召回率(Recall,简称为 R),也称灵敏度(Sensitivity,简称为 S),表示正确预测为正样本(T_P)占所有真实正样本(T_P+F_N)的比重:

$$R=S=\frac{T_P}{T_P+F_N}. \tag{4}$$

由式(3)和(4),可以发现精确率和召回率之间是具有相互的影响,为了同时兼顾二者,并且在尽可能提高精确率和召回率的同时,减少二者之间的差异,故此引入 F_1-score:

$$F_1\text{-score}=\frac{2\times P\times R}{P+R}. \tag{5}$$

真正率与召回率公式相同,都表示正确预测为正样本占所有预测为正样本的比重:

$$T_{PR}=\frac{T_P}{T_P+F_N}. \tag{6}$$

假正率指的是错误预测为正样本占所有真实为负样本的比重:

$$F_{PR}=\frac{F_P}{F_P+T_N}. \tag{7}$$

ROC 曲线是对整个分类器性能测试的一种可视化,通过曲线的形式来描述出真正例率与假正例率之间的变化关系.一个好的预测分类器往往尽可能靠近 ROC 曲线的左上角,越靠近左上角,表示分类器的性能越好.另外,为了更好量化 ROC 曲线所体现出的测试性能,可用 ROC 曲线下方的面积即 AUC 值来表示整个模型的性能表现.

3.2 数据预处理

在信用评估模型中,融资者(借款人)的个人信息一般包括贷款金额、期限、年龄、职业、贷款用途以及还款记录等数据.由于数据集中包含许多缺失数据、类别型数据等,同时为降低后期训练预测模型的复杂度,提升其正确率,因此需要在建模前进行对数据预处理.在填补缺失值方面,对于连续性数据采取均值进行填补,类别型数据用众数进行填补,并对其进行编码转化成数值型.在特征提取方面,采取线性支持向量机(LinearSVC)进行特征选择.

实验中所使用的数据集来自机器学习开放数据集网站[29](UCI Machine Learning Repository)中的德国信用数据集,Kaggle 网站上的公开 P2P 平台 Lending Club 的贷款数据集[30]以及信用卡数据集[31](Credit Card Approval Prediction).其中,为了更加符合现实业务需要,并且由于数据量大,故将 Lending Club 数据类别标签进行规整分类,对 Fully Paid 和 Current 归结为正常还款类,对 Late(31~120 d),In Grace Period 以及 Late(16~30 d)归结为延期贷款类,对 Default 和 Charged Off 归结为糟糕贷款类,并对数据进行下采样处理.表 1 所示为各样本数据分别经预处理后得到的实验数据及其相关数据属性.

3.3　实验与结果分析

为了证实所构建模型的有效性，本文根据前期预选的初级学习器，在不同数据集上进行信用评估实验，构建基于各预选初级学习器的模型和 Deep_Logistic 模型.其中，Deep_Logistic 模型的构建（由图 4 知），首先针对每种数据，Stacking 多粒度扫描器滑动窗口均可设置成原始数据特征 v 维和 $v-1$ 维，联级 Logistic 级数可均设置为 5 级；其次，根据基于各预选初级学习器的模型训练预测情况，确定 Stacking 多粒度扫描器的基学习器，再来构建 Deep_Logistic 模型.最后以 Lending Club 数据集为例，与近几年文献中在此数据集上所提信用评估模型上的表现进行对比.每个数据集上，在进行联级 Logistic 过程中交叉检验均采取 5 折，以确保实验结果的稳定性、可靠性.

表 1　实验中的样本数据集

Tab. 1　Data sets used in experiments

数据集	数据维度	样本大小	样本类别比(1∶0 或 2∶1∶0)
德国信用	35	1 000	700∶300
信用卡	20	25 128	25 007∶121
Lending Club	81	36 609	12 203∶12 203∶12 203

(1)德国信用数据集实验

表 2～6 所示是各单一集成学习器在德国信用数据集上的参数调优结果.

表 2　德国信用数据集上 XGBoost 参数调优结果

Tab. 2　XGBoost parameter tuning results on German credit data sets

参数名称	参数含义	搜索空间	调优结果	测试集准确率
scale_pos_weight	控制正负样本比例	0.10 至 1.00 之间，以 0.05 步长搜索	0.90	0.780 00
learning_rate	更新学习过程中的收缩步长	[0.010,0.050,0.070,0.075,0.080,0.085,0.090,0.100]	0.010	
n_estimators	控制弱学习器的数量	[50,100,150,200,250,300,350,400,450,500]	500	
max_depth	树的最大深度	[3,5,6,7,8,9,10]	3	
subsample	控制每棵树，随机采样的比例	0.15 至 0.95 之间，以 0.10 步长搜索	0.85	
colsample_bytree	建立树时对特征随机采样的比例	[0.60,0.65,0.70,0.75,0.80,0.85,0.90,0.95,1.00,1.50,2.00]	0.90	

表 3　德国信用数据集上 CatBoost 参数调优结果

Tab. 3　CatBoost parameter tuning results on German credit data sets

参数名称	参数含义	搜索空间	调优结果	测试集准确率
class_weights	控制正负样本比例	根据正负样本比例设置	[3,7]	0.783 33
iterations	最大树数	[400,450,500,550,600,650,700,750,800]	450	
depth	树的最大深度	[3,4,5,6,7,8,9,10]	10	

表 4　德国信用数据集上 LightGBM 参数调优结果

Tab. 4　LightGBM parameter tuning results on German credit data sets

参数名称	参数含义	搜索空间	调优结果	测试集准确率
learning_rate	更新学习过程中的收缩步长(学习率)	0.01 至 1.00 之间，以步长 0.05 搜索	0.81	0.716 67
n_estimators	boosting 的迭代次数	[50,100,150,200,250,300,350,400,450,500]	350	
max_depth	树的最大深度	[3,4,5,6,7,8,9]	3	
subsample	控制每棵树，随机采样的比例	[0.15,0.25,0.35,0.45,0.55,0.65,0.75,0.85,0.95]	0.15	
colsample_bytree	建立树时对特征随机采样的比例	[0.60,0.65,0.70,0.75,0.80,0.85,0.90,0.95,1.00,1.50]	0.90	
min_child_samples	一个叶子上的最小数据量	[2,3,4,5,6,7,8]	2	
min_split_gain	执行节点分裂的最小增益	[2,3,4,5,6,7,8]	5	
class_weight	控制正负样本比例		balanced	

由表 2～6 可知每一学习器在测试集上的预测准确率，确定 Stacking 多粒度扫描器的构建，其涉及初级学习器有 XGBoost，CatBoost.构建 Stacking，如表 7 所示.

表 5 德国信用数据集上 AdaBoost 参数调优结果

Tab. 5 AdaBoost parameter tuning results on German credit data sets

参数名称	参数含义	搜索空间	调优结果	测试集准确率
learning_rate	更新学习过程中的收缩步长(学习率)	0.01 至 1.00 之间,以步长 0.01 搜索	0.01	0.773 33
n_estimators	控制弱学习器的数量	[50,100,150,200,250,300,350,400]	50	
algorithm	控制分类学习的算法	["SAMME","SAMME.R"]	SAMME	
base_estimator	基分类器		RandonForest	

表 6 德国信用数据集上 GBDT 参数调优结果

Tab. 6 GBDT parameter tuning results on German credit data sets

参数名称	参数含义	搜索空间	调优结果	测试集准确率
loss	控制学习过程中损失函数	["deviance","exponential"]	exponential	
learning_rate	更新学习过程中的收缩步长(学习率)	[0.100,0.010,0.050,0.001,0.200,0.300,0.500,0.700,0.800,0.600]	0.800	
n_estimators	控制弱学习器的数量	[50,100,150,200,250,300,350,400,450,500]	100	
max_depth	树的最大深度	[1,2,3,4,5,6,7,8,9,10,11]	3	0.750 00
min_samples_split	控制分裂所需的最少样本量	[1,2,3,4,5,6,7,8,9,10,11]	2	
min_samples_leaf	控制每个叶子的最少样本量	[1,2,3,4,5,6,7,8,9,10,11]	1	
max_leaf_nodes	最大叶子节点	[1,2,3,4,5,6,7,8,9,10,11]	8	

表 7 德国信用数据集上 Stacking 多粒度扫描器构建

Tab. 7 Construction of Stacking on German credit data sets

Stacking A	初级学习器	XGBClassifier(learning_rate=0.01,n_estimators=500,max_depth=3,subsample=0.85,colsample_bytree=0.9,n_jobs=-1,objective="binary:logistic",scale_pos_weight=0.9,random_state=0)
		CatBoostClassifier(iterations=450,class_weights=[sample_weiht_0,sample_weiht_1],depth=10,random_state=0)
	次级学习器	LogisticRegression(C,class_weight,random_state=50),C=[0.01,1](Interval=0.001),class_weight='balanced'
Stacking B	初级学习器	XGBClassifier(learning_rate=0.01,n_estimators=500,max_depth=3,subsample=0.85,colsample_bytree=0.9,n_jobs=-1,objective="binary:logistic",scale_pos_weight=0.9,random_state=420)
		CatBoostClassifier(iterations=450,class_weights=[sample_weiht_0,sample_weiht_1],depth=10,random_state=420)
	次级学习器	LogisticRegression(C,class_weight,random_state=420),C=[0.01,1.00](Interval=0.001),class_weight='balanced'

表 8 和图 5 是分别使用预选初级学习器 XGBoost, Catboost, LightGBM, Adaboost 以及 GBDT 模型与该文所提 Deep_Logistic 模型在德国信用数据集上的比较,表 9 是 Deep_Logistic 模型中联级 Logistic 每级训练最优参数表.

表 8 各模型在德国信用数据集上的结果

Tab. 8 Results of different models on German credit data sets

分类模型	AUC	Accuracy	Precision	Recall	F_1-score
XGBoost	0.79	0.78	0.68	0.44	0.54
Catboost	0.80	0.78	0.63	0.58	0.61
LightGBM	0.75	0.72	0.50	0.66	0.57
Adaboost	0.68	0.77	0.65	0.47	0.54
GBDT	0.74	0.75	0.56	0.51	0.54
Deep_Logistic	0.80	0.79	0.63	0.60	0.62

(2)信用卡数据集实验和 Lending Club 数据集实验

为了更好地进行有效对比实验,设置在信用卡和 Lending Club 数据集上的各分类预测器参数配置、寻优方法、Deep_Logistic 模型构建方法均与在德国信用数据集上处理方式相同,并且采取一样的参数搜索范围.最终确定信用卡数据集上 Stacking 多粒度扫描器的构建,其涉及初级学习器有 XGBoost,CatBoost,Lending Club 数据集上 Stacking 多粒度扫描器的构建,其涉及初级学习器有 XGBoost,LightGBM 和 CatBoost.由此可得各模型在信用卡、Lending Club 数据集上的预测性能对比,如表 10 所示.表 11 是在信用卡数据集和 Lending Club 数据集上联级 Logistic 每级训练最优参数表.

由表 8 和图 5 可知，在德国信用数据集中，Deep_Logistic 的 AUC 值为 0.80，与 CatBoost 同排名第一，比排名第二的 XGBoost 高出 1%，比第三的 LightGBM 高出 5%；Accuracy 值为 0.79，也为最高，超过排名第二的 XGBoost 和 CatBoost 模型 1%，比排名第三的 AdaBoost 高 2%；F_1-score 值为 0.62，也为最高，比排名第二模型 CatBoost 高 1%，比排名第三模型 LightGBM 高 5%；相比于构建 Stacking 初级学习器中 XGBoost 和 CatBoost 模型，灵敏度达到 60%，有着明显提升.因此，在德国信用数据集中，Deep_Logistic 效果最好.

表 9 表示的是，在德国信用数据集上训练 Deep_Logistic 模型时，每级 Logistic 的最优参数.根据早停机制，最优级数为 1，并且最优参数自适应惩罚系数 C 为 0.01.

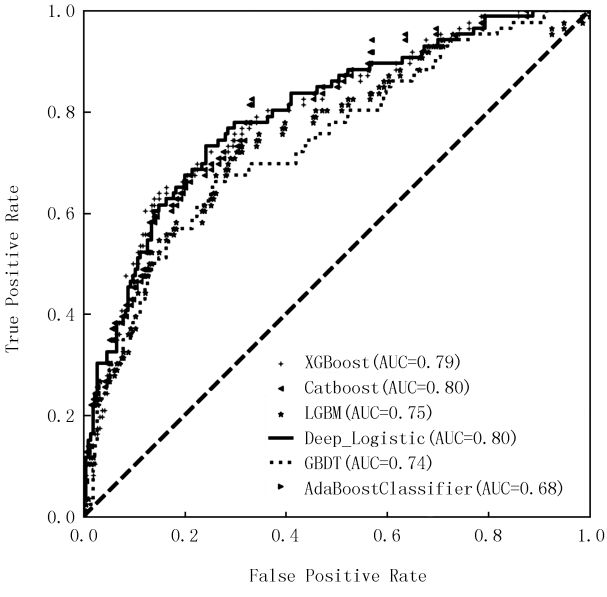

图5　各模型在德国数据集上的ROC曲线

Fig. 5　ROC curves of different models on German dataset

表 9　德国信用数据集，每级 Logistic 最优参数

Tab. 9　The optimal parameters of each layer Logistic model on German credit data sets

层级	Logistic 最优参数	测试集上准确率
1	惩罚系数 C=0.01，权重 class_weight='balanced'，random_state 采取随机值.	0.79

表 10　各模型在信用卡、Lending Club 数据集上的结果

Tab. 10　Results of different models on credit card and Lending Club data sets

分类模型	平均方式	AUC	Accuracy	Precision	Recall	F_1-score
Results of different models on credit card data sets						
XGBoost		1.000 000	0.999 337	0.999 600	0.999 734	0.999 667
Catboost		1.000 000	0.999 204	1.000 000	0.999 201	0.999 600
LightGBM		1.000 000	0.999 071	0.999 733	0.999 334	0.999 534
Adaboost		0.682 000	0.997 214	0.997 210	1.000 000	0.998 603
GBDT		1.000 000	0.999 204	0.999 467	0.999 734	0.999 600
Deep_Logistic		1.000 000	0.999 735	1.000 000	0.999 734	0.999 867
Results of different models on Lending Club data sets						
XGBoost	macro	0.893 320	0.859 621	0.858 372	0.858 576	0.857 143
Catboost		0.893 458	0.859 638	0.858 551	0.858 672	0.857 325
LightGBM		0.892 371	0.858 253	0.857 109	0.857 242	0.855 868
Adaboost		0.888 107	0.852 922	0.851 453	0.851 378	0.850 132
GBDT		0.891 629	0.857 401	0.856 127	0.856 207	0.854 867
Deep_Logistic		0.956 724	0.861 221	0.860 335	0.860 610	0.858 873

由表 10 可知，在信用卡数据集中，Deep_Logistic 的 AUC 值为 1，处于较好水平；Accuracy 值为 0.999 735，也为最高，超过排名第二的 XGBoost 模型 0.000 398，比排名第三的 CatBoost 和 GBDT 高 0.000 531；F_1-score 值为 0.999 867，也为最高，相比排名第二模型 XGBoost 高 0.000 2，比排名第三模型 CatBoost 和 GBDT 高 0.000 267；相比于构建 Stacking 初级学习器中 XGBoost 和 CatBoost 模型，灵敏度达到 0.999 734，处于最优；Precision 值为 1，与 CatBoost 同排名第一，处于最优.因此，在信用卡数据集上，Deep_Logistic 效果最好.由表 11 可知，训练 Deep_Logistic 模型，得最优级数为 1，最优参数自适应惩罚系数 C 为 0.01.

表 11 在信用卡、Lending Club 数据集上每级 Logistic 最优参数

Tab. 11 The optimal parameters of each layer Logistic model on credit card and Lending Club data sets

层级	Logistic 最优参数	测试集上准确率
The optimal parameters of each layer Logistic model on credit card data sets		
1	惩罚系数 C=0.01,权重 class_weight='balanced',random_state 采取随机值.	0.999 735
The optimal parameters of each layer Logistic model on Lending Club data sets		
1	惩罚系数 C=0.1,random_state 采取随机值.	0.856 96
2	惩罚系数 C=0.1,random_state 采取随机值.	0.858 87

Lending Club 数据集实验,由上表 10 可知,在 Lending Club 数据集中.Deep_Logistic 的 AUC 值为 0.956 724,排名第一,超过排名第二的 CatBoost 模型 0.063 266,比排名第三的 XGBoost 模型高 0.063 404;Accuracy 值为 0.861 221,排名第一,超过排名第二的 CatBoost 模型 0.001 583,比排名第三的 XGBoost 高 0.001 6;Precision 值为 0.860 335,排名第一,超过排名第二的 CatBoost 模型 0.001 784,比排名第三的 XGBoost 高 0.001 963;Recall 值为 0.860 610,排名第一,超过排名第二的 CatBoost 模型 0.001 938,比排名第三的 XGBoost 高 0.002 034;F_1-score 值为 0.858 873,排名第一,超过排名第二的 CatBoost 模型 0.001 548,比排名第三的 XGBoost 模型高 0.001 73.因此,在 Lending Club 数据集上,Deep_Logistic 效果最佳.由表 11 可知,训练 Deep_Logistic 模型,得最优级数为 2,且每级的最优参数自适应惩罚系数为 0.1.

(4)与相关文献方法对比实验

在近两年信用评估领域中同样有在基于现有机器学习模型进行改进的模型的应用研究,在表 12 中对文献[14]中引入到信用评估领域的经典"手写数字"识别模型 LeNet-5、文献[24]中引入到借款人信用评估研究的深度森林模型、本文所提 Deep_Logistic 模型,使用信用评分模型中常用评价指标进行综合对比分析.由于在两篇文献中没有涉及德国信用数据集与信用卡数据集上的实验,因此只在 Lending Club 数据集上进行对比实验.

表 12 Deep_Logistic 模型与相关文献模型效果对比

Tab. 12 Effect comparison of Deep_Logistic and andother models in literatures

分类模型	平均方式	AUC	Precision	Recall	F_1-score	Accuracy
LeNet-5		0.820 732	0.831 471	0.713 982	0.764 306	0.714 377
深度森林	macro	0.892 165	0.857 895	0.856 837	0.856 987	0.855 595
Deep_Logistic		0.956 724	0.861 221	0.860 335	0.860 610	0.858 873

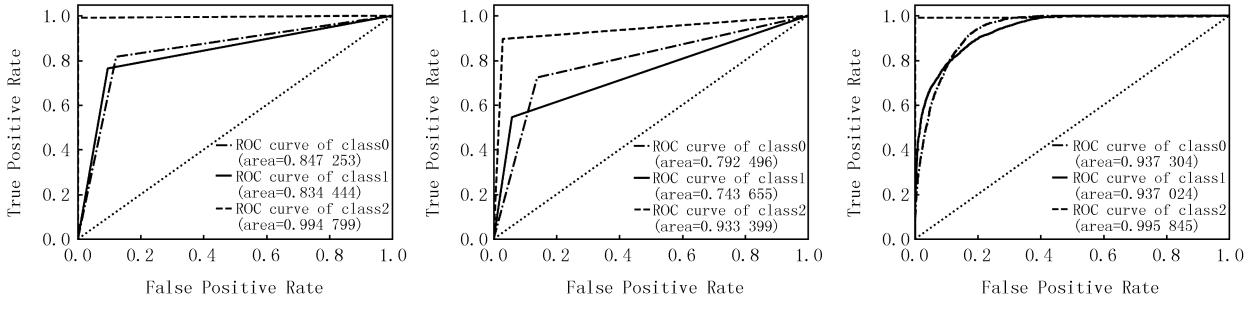

图6 深度森林,LeNet-5,Deep_Logistic模型在Lending Club数据集上的ROC曲线

Fig. 6 ROC curves of DeepForest,LeNet-5 and Deep_Logistic models on Lending Club data sets

从表 12 可以看出,Deep_Logistic 模型的 AUC 值在 Lending Club 数据集中是最高的,Accuracy 值也都表现不错,明显优于 LeNet-5 和深度森林模型,在精准度和敏感度方面也比较突出.图 6 是深度森林,LeNet-5,Deep_Logistic 模型在 Lending Club 数据集上的 ROC 曲线,综合分析可知,明显可以发现 Deep_Logistic 模型下 ROC 曲线更接近左上角(0,1),一定程度上可说明模型性能较好.

结合表 8、表 10 中的各项评价指标可看出:

(1)在常用信用评估方法中,XGBoost 和 CatBoost 进行信用风险评估的效果较好;

(2)在各数据集上,Deep_Logistic 相比于构成其 Stacking 多粒度扫描器的初级学习器中各基评估器,表

现更加稳定，信用评估效果要更好，可有效地融合各基评估器的预测性能，提高模型的整体信用评估效果.

4　结束语

本文的目的在于设计一种基于集成学习，拥有联级结构，能融合多个单一集成学习器的分类器，从而有效提出进行贷款和不贷款的策略性建议，辅助 P2P 平台及相关贷款人提供一种解决方案.本文基于 Stacking 特征增强多粒度联级 Logistic 模型对借款人进行信用风险评估，在德国信用数据集、信用卡数据集和 Lending Club 数据集基础上进行分类预测，同时将其与 XGBoost，CatBoost，LightGBM，GBDT 以及 AdaBoost 模型分类方法进行对比，并且用该 Lending Club 数据与文献中 LeNet-5，深度森林方法也进行对比.实验表明，所提方法在 3 个数据集上表现出的性能均优于其他对比方法，并且也较优于文献中方法.从以上实验中可以发现本文分类器可应用于二分类、三分类问题，理论上也可应用于多分类任务中.但是，由于本文方法中联级部分仅使用 Logistic 模型，同时在构建 Stacking 多粒度扫描器方面，初级学习器中的基评估器种类数量和生成多粒度的滑动窗口维度的增加，会增加模型的整体复杂性，在之后的研究中可以尝试使用更多的分类器作为联级结构的组成部分，如 SVC，KNN 等，并且探索出更优的方法对于 Stacking 初级学习器中的基评估器种类数量的选择，以及对最佳的滑动窗口维度的选择.为了进一步尝试使用到其他数据集进行实验应用，从而将本文方法应用于信用评估领域的相关其他方面，并尝试探索联级结构中学习器的数量与预测结果之间的关系以及相互影响，以确保整体的预测模型取得更佳效果.

参　考　文　献

[1] OHLSON J A.Financial Ratios and the Probabilistic Prediction of Bankruptcy[J].Journal of Accounting Research，1980，18(1)：109-131.

[2] WIGINTON J C.A Note on the Comparison of Logit and Discriminant Models of Consumer Credit Behavior[J].Journal of Financial Quantitative Analysis，1980，15(3)：757-770.

[3] STEENACKERS A，GOOVAERTS M J.A credit scoring model for personal loans[J].Insurance：Mathematics and Economics，1989，8(1)：31-34.

[4] YEH I C，LIEN C H.The Comparisons of Data Mining Techniques for the Predictive Accuracy of Probability of Default of Credit Card Clients[J].Expert Systems with Applications，2009，36(2)：2473-2480.

[5] HANSEN L K，SALAMON P.Neural Network Ensembles[J].IEEE Transactions on Pattern Analysis and Machine Intelligence，1990，12(10)：993-1001.

[6] TWALA B.Multiple Classifier Application to Credit Risk Assessment[J].Expert Systems with Applications，2010，37(4)：3326-3336.

[7] BREIMAN L.Random forests[J].Machine Learning，2001，45(1)：5-32.

[8] 马晓君，宋嫣琦，常百舒，等.基于 CatBoost 算法的 P2P 违约预测模型应用研究[J].统计与信息论坛，2020，35(7)：11-12.
MA X J，SONG Y Q，CHANG B S，et al.Research on Application of P2P Default Prediction Model Based on CatBoost Algorithm[J].Statistics and Information Forum，2020，35(7)：11-12.

[9] FREUND Y，SCHAPIRE R E.A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting[J].Journal of Computer and System Sciences，1997，55(1)：119-139.

[10] FRIEDMAN J H.Greedy function approximation：a gradient boosting machine[J].The Annals of Statistics，2001，29(5)：1189-1232.

[11] CHEN T Q，GUESTRIN C.XGBoost：a scalable tree boosting system[C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.New York：ACM，2016：785-794.

[12] KE G L，MENG Q，FINLEY T，et al.LightGBM：a highly efficient gradient boosting decision tree[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems.New York：Curran Associates Inc，2017：3149-3157.

[13] 丁岚，骆品亮.基于 Stacking 集成策略的 P2P 网贷违约风险预警研究[J].投资研究，2017，36(4)：41-52.
DING L，LUO P L.Research on Early Warning of P2P Network Loan Default Risk Based on Stacking Integration Strategy[J].Investment research，2017，36(4)：41-52.

[14] 刘伟江，魏海，云天鹤.基于卷积神经网络的客户信用评估模型研究[J].数据分析与知识发现，2020，4(6)：80-90.
LIU W J，WEI H，YUN T H.Research on customer credit evaluation model based on convolutional neural network[J].Data Analysis and Knowledge Discovery，2020，4(6)：80-90.

[15] 王重仁，王雯，佘杰，等.融合深度神经网络的个人信用评估方法[J].计算机工程，2020，46(10)：308-314.
WANG C R，WANG W，SHE J，et al.Personal credit evaluation method based on deep neural network[J].Computer Engineering，2020，46(10)：308-314.

[16] 陈倩，贺兴时，杨新社.基于 RF 的 Elastic Net-Logistic 个人信用违约风险评估[J].西安工程大学学报，2021，35(3)：116-122.
CHEN Q，HE X S，YANG X S.RF based Elastic Net Logistic Personal Credit Default Risk Assessment[J].Journal of Xi'an University of

Engineering,2021,35(3):116-122.
[17] 李佳欣.基于逐步 Logistic 回归下分类算法的个人信用评估分析[J].湖南文理学院学报(自然科学版),2021,33(1):5-9.
LI J X.Analysis of Personal Credit Evaluation Based on Classification Algorithm under Stepwise Logistic Regression[J].Journal of Hunan University of Arts and Sciences(Natural Science Edition),2021,33(1):5-9.
[18] 曹再辉,余东先,施进发,等.两层分类器模型应用于个人信用评估[J].控制工程,2019,26(12):2231-2234.
CAO Z H,YU D X,SHI J F,et al.Application of two-layer classifier model in personal credit evaluation[J].Control Engineering,2019,26(12):2231-2234.
[19] ZENKO B,TODOROVSKI L,DZEROSKI S.A Comparison of Stacking with Meta Decision Trees to Bagging,Boosting,and Stacking with other Methods[C]//IEEE International Conference on Data Mining.San Jose:IEEE,2001:669-670.
[20] 曹莹,苗启广,刘家辰,等.AdaBoost 算法研究进展与展望[J].自动化学报,2013,39(6):745-758.
CAO Y,MIAO Q G,LIU J C,et al.Research Progress and Prospect of AdaBoost Algorithm[J].Journal of Automation,2013,39(6):745-758.
[21] 王小伟.基于逻辑回归与 GBDT 模型的银行信贷风险评估[D].汕头:汕头大学,2021.
[22] 王思杭.威布尔分布下基于优化 GBDT-LR 模型的软件老化状态识别方法研究[D].呼和浩特:内蒙古大学,2021.
[23] YANG Z M,HE J Y,SHAO Y H.Feature Selection Based On Linear Twin Support Vector Machines[J].Procedia Computer Science,2013,17:1039-1046.
[24] 王萧萧,王亭雯,马玉玲,等.基于深度森林的 P2P 网贷借款人信用风险评估方法[J].计算机科学,2021,48(S2):430-431.
WANG X X,WANG T W,MA Y L,et al.Credit Risk Assessment Method of P2P Online Loan Borrowers Based on Deep Forest[J].Computer Science,2021,48(S2):430-431.
[25] 陈湘州,陶李红.基于 MLP 神经网络的中小企业供应链金融信用风险评估[J].湖南科技大学学报(自然科学版),2021,36(4):93-94.
CHEN X Z,TAO L Z.Financial Credit Risk Assessment of SME Supply Chain Based on MLP Neural Network[J].Journal of Hunan University of Science and Technology(Natural Science Edition),2021,36(4):93-94.
[26] 陈卫中,倪宗瓒,潘晓平,等.用 ROC 曲线确定最佳临界点和可疑值范围[J].现代预防医学,2005,32(7):729-731.
CHEN W Z,NI Z Z,PAN X P,et al.Determining the Optimal Critical Point and the Range of Suspicious Values with ROC Curves[J].Modern Preventive Medicine,2005,32(7):729-731.
[27] HAND D J.Measuring classifier performance:a coherent alternative to the area under the ROC curve[J].Machine Learning,2009,77(1):103-123.
[28] FAWCETT T.An introduction to ROC analysis[J].Pattern Recognition Letters,2005,27(2006):861-874.
[29] DUA D,GRAFF C.UCI machine learning repository[EB/OL].[2022-03-07].http://archive.ics.uci.edu/ml.
[30] XIA Y,LIU C,LI Y,Liu N.A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring[J].Expert Systems with Applications,2017,78:225-241.
[31] SEANNY.Credit Card Approval Prediction[EB/OL].[2022-03-07].https://www.kaggle.com/Datasets/caesarmario/application-data.

Personal credit evaluation based on Stacking feature enhancing multi-grained cascade logistic

Hou Tianbao, Wang Aiyin

(School of Statistics and Data Science, Xinjiang University of Finance & Economics, Urumqi 830012, China)

Abstract: Mainly aimed at the widely concerned P2P online loan credit evaluation problem, the machine learning method was used to improve the accuracy of the applicant's online loan default prediction, and the enhanced multi-granularity cascade logistic method based on the Stacking feature and its application were studied. The proposed classifier is a hybrid model, which combines the ideas of Stacking ensemble learning and cascade logistic learning. First, XGBoost, Catboost, LightGBM, AdaBoost and Gradient Boosting models are established through grid search technology, and the appropriate base evaluator as the primary learner of Stacking ensemble and the logistic model as the secondary learner are selected to build a Multi-grained Scanner based on Stacking, generate prediction results as meta-features, and to stitch into new feature data. Secondly, the new feature data and the feature enhancement of meta-features on each level of Logistic are used to establish the cascade Logistic regression model, and compare constructed model with the existing single integrated learner and each base evaluator on three different P2P network credit evaluation data sets. The experimental results show that compared with each base evaluator and other single integrated classifiers, the multi-grained cascade logistic model based on Stacking has higher accuracy and better prediction effect when evaluated by AUC, accuracy and other indicators.

Keywords: personal credit; feature enhancement; Stacking ensemble; multi-grained scanning; cascade Logistic model

[责任编校 陈留院 赵晓华]