云南民族大学学报(自然科学版),2024,33(3):345 - 350　　CN 53 - 1192/N　ISSN 1672 - 8513
doi:10.3969/j.issn.1672 - 8513.2024.03.009　　http://ynmz.cbpt.cnki.net

基于自适应惯性权重 PSO—LightGBM 的信用风险评估研究

付芷宁,李慧敏,徐亚田,陶玉虎,高　伟
(云南民族大学 数学与计算机科学学院,云南 昆明 650500)

摘要: 贷款市场复杂的个人信用风险问题中,信用风险评估模型的构建是十分关键的一步.利用 Lending Club 数据集,进行信用风险评估模型的构建来预测客户的违约概率.首先进行数据处理,再通过合成少数类过采样技术(SMOTE)算法处理数据正负样本不平衡的问题,获得完备的信用贷款数据.其次采用轻量梯度提升机(LightGBM)模型进行训练,并使用自适应惯性权重的粒子群优化(PSO)算法得到 LightGBM 的最优参数.与多个主流算法进行对比,实验结果表明,构建的模型有更好的性能.

关键词: 信用风险;不平衡数据;合成少数类过采样技术;LightGBM 模型;粒子群优化算法
中图分类号:TP311.13;F832.4　　**文献标志码:**A　　**文章编号:**1672 - 8513(2024)03 - 0345 - 06

随着我国经济的不断发展和国民消费水平的不断提高,人们的消费观念也随之改变,其中一个重要表现就是对贷款的需求日益增加.随着信贷业务的飞速发展,是否向贷款人发放贷款也越来越受关注.信用风险评估模型本质上是一个二分类模型,其核心思想是根据还款情况按照不同的特征分为信用良好及信用不良的客户,从而对贷款的信用风险进行分析.一个好的信用风险评估模型可以有效降低信用机构的人工成本,提高信用机构的决策能力,从而减少损失.

信用风险评估模型当前主要有基于统计模型的方法和基于机器学习的方法.利用统计方法构建信用风险评估模型有解释性强、计算方法简单、计算成本低等优点,其中最常用的包括逻辑回归[1](logistic regression,LR)和线性判别分析[2].逻辑回归由于其稳健性和可解释性被广泛应用,但预测能力较弱;而线性判别分析则需要严格的假定条件,在实际应用中有一定局限性.随着机器学习的飞速发展,信用评分模型也越来越智能化.于晓虹[3]利用随机森林(random forest,RF)对 P2P 网络借贷数据进行实验,结果发现随机森林具有很好的实用价值和预测能力.Zhang 等[4]利用支持向量机(support vector machine,SVM)进行贷款风险识别,结果表明 SVM 在贷款风险识别中具有重要作用.Chen 等[5]提出 1 种改进的朴素贝叶斯分类器(naive bayes,NB),该分类器的预测效果优于传统的朴素贝叶斯分类器,并且比 SVM 更高效.Chern 等[6]使用决策树(decision tree,DT)分类器来处理大数据环境下的信用评估问题,实验结果显示决策树表现良好并且在预测方面接近人工神经网络(artificial neural network,ANN),同时需要较少的训练时间.Sharifi 等[7]使用人工神经网络、优化算法和决策树算法相结合,实验结果显示准确度优于传统的决策树方法.

轻量梯度提升机(light gradient boosting machine,LightGBM)模型是一种基于梯度提升树的改进模型,可以利用弱分类器迭代模型以得到最优模型,拥有决策树分类器的优势,同时有更快的训练速度、更低的内存消耗以及更高的准确率.同时 LightGBM 模型有大量超参数,这些超参数直接影响到模型的结构和模型的性能,粒子群优化算法(particle

收稿日期:2023 - 08 - 05.
基金项目:云南省研究生优质课程建设项目(云学位[2022]8 号).
作者简介:付芷宁(1999 -),女,硕士研究生.主要从事大数据分析研究.
通信作者:李慧敏(1980 -),女,博士,教授,硕士生导师.主要从事生物信息学与应用统计学研究.

swarm optimization,PSO)的结构简单且容易实现,被广泛应用于寻找模型参数. 信贷数据通常数据量较大,同时需要较高的准确率来减少损失. 鉴于此,本文提出一种基于自适应惯性权重 PSO 优化的 LightGBM 构建信用风险评估模型,采用 LightGBM 模型进行训练,并使用自适应惯性权重的 PSO 算法得到 LightGBM 的最优参数. 最后与多个主流算法进行对比,实验结果表明,构建的模型有更好的性能.

1 相关理论

1.1 PSO 理论

PSO[8]是一种进化计算技术,其概念是受一群飞向同一地点的鸟移动所启发的,它来源于鸟类的捕食行为,最早由 Kennedy & Eberhart[9]提出. PSO 是一种基于迭代的优化工具. 它首先初始化目标函数中的一组随机解,将其中的每个个体当作一个在空间中没有体积和质量的粒子,然后通过多次迭代进行搜索,得到最优解. 由于它只需要使用目标取值的信息,而不需要使用梯度信息,所以具有结构简单、参数较少、容易实现,同时收敛速度较快等优点,该算法具有快速的搜索速度和良好的初始收敛性,因此,PSO 算法[10]被广泛应用于许多领域,如路径规划、模型优化等.

PSO 算法在寻找最优解过程中是通过粒子之间的相互协助与交互,每个在粒子自身搜索空间中通过改变自身的位置和速度搜索局部最优位置(pbst)与全局最优位置(gbst)作比较,以此来找到当前全局最优解. 粒子改变自身速度的位置的公式如(1),(2)所示.

$$V_{i+1}=w_iV_i+c_1r_1(\mathrm{pbst}_i-X_i)+c_2r_2(\mathrm{gbst}_i-X_i). \tag{1}$$

$$X_{i+1}=X+V_{i+1}. \tag{2}$$

式中,X 和 V 分别表示每个粒子的当前位置和速度,X_{i+1}和 V_{i+1}表示每个粒子的新位置和速度,w_i 是惯性权重,c_1、c_2 分别是认知学习因子和社会学习因子,r_1 和 r_2 是 2 个[0,1]中的随机数.

标准的 PSO 算法容易陷入到局部最优解中. 对于 PSO 算法来说,惯性权重的选择对于算法的执行效果有很大程度的影响. 当惯性权重较大时,PSO 算法有较强的全局搜索能力,而局部搜索能力较弱;而较小的惯性权重意味着算法的局部搜索能力增强,全局搜索能力变弱. 因此我们选择的是自适应惯性权重的粒子群优化算法,惯性权重的更新公式如(3)所示.

$$w=w_{\max}-(w_{\max}-w_{\min})\times(\mathrm{iter}/\mathrm{itermax}). \tag{3}$$

式中:iter 表示迭代的次数,itermax 表示最大迭代次数,用于控制之前速度历史的影响,$w_{\max}$为惯性权重的最大值,$w_{\min}$为惯性权重的最小值.

1.2 LightGBM 模型

LightGBM[11]是一种基于梯度提升树(grandient boosting decision tree,GBDT)的改进模型,它采用了基于梯度提升(boosting)策略. Boosting 策略不仅可以将多个单一分类器得到的分类信息进行综合,从而提高分类的精度,还可将弱学习器提升为强学习器[12]. GBDT 的核心思想是通过计算负梯度计算模型,假设输入训练数据集 $D=\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}$. 其中 $x_i\in X\in R_n$,x_i 为输入的特征变量,$y_i\in Y\in R$. 定义损失函数为 $L(y_i,f(x_i))$,

首先初始化回归树 $f_0(x)$

$$f_0(x)=\arg\min_c\sum_{i=0}^{n}L(y_i,c). \tag{4}$$

式中:c 为常数,n 为样本的数量.

计算损失函数的负梯度作为残差估计,即

$$r_{ij}=-\left[\frac{\partial L(y_i,f(x_i))}{\partial f(x_i)}\right]_{f(x)=f_{m-1}(x)}. \tag{5}$$

极小化损失函数

$$c_{m,j}=\arg\min_c\sum_{x_i\in R_{m,i}}L(y_i,f_{m-1}(x_i+c)). \tag{6}$$

得到了第 m 棵回归树的预测值 $f_m(x)$

$$f_m(x)=f_{m-1}(x)+\sum_{j=1}^{J}c_{m,j}\theta. \tag{7}$$

LightGBM 是在 GBDT 的基础上进行了改进,其算法不同于其他树提升算法的特征之一是使用按叶生长(Leaf-wise)的生长策略[13],如图 1 所示. 大多数 GBDT 工具使用的是按层生长(level-wise)的生长策略,按叶生长与按层生长的生长策略有所不同. 按层生长是树从根节点开始每层的所有叶子节点都进行分裂,而按叶生长则是在每一次分裂时选择具有最大梯度的叶子节点继续分裂,这样可以更快地找到降低损失函数的分裂点. 因此,当在 LightGBM 中的相同叶子上生长时,采用带深度限制的按叶生长的生长策略具有误差更低,准确性更高、能够有效的抑制过拟合等特点. 同时,使用按叶生长的生长策略来生长树可以比按层生长减少损失,以及更高的准确性.

LightGBM 针对样本多的问题提出了基于梯度的单边采样算法(gradient-based One-side Sampling,GOSS);针对特征多的问题提出了互斥特征捆绑算法(exclusive feature bundling,EFB). GOSS 算

法[14]主要思想就是保留所有大梯度的样本,并对剩余的小梯度样本进行随机采样. 同时在计算信息增益时为小梯度数据引入了权重系数,来减少对样本点分布的影响. EFB 则是通过捆绑相似特征的方法来减少特征数量从而减少模型的计算量.

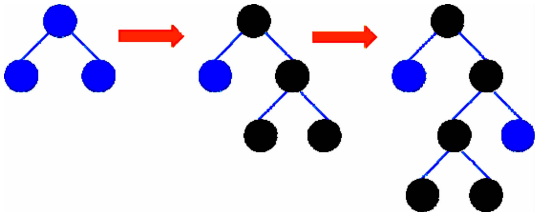

图 1　LightGBM 算法使用按叶生长(Leaf – wise)

1.3　SMOTE 算法

SMOTE 算法[15]是一种针对数据不平衡问题的处理方法,它是一种合成少数类样本的过采样技术,与随机过采样的直接复制原始少数类样本来达到正负样本平衡有所不同,它是基于少数类的 k 近邻同类样本进行线性插值的过采样方法. 它有操作简单,较好的适应性等优点. SMOTE 算法可以在一定程度上使正负样本分布均衡,来减少分类器过拟合的风险[16].

SMOTE 算法的步骤为:首先,计算少数类样本之中的每个点到其他点的欧氏距离,得到 k 近邻样本 x_p. 然后根据不平衡的比例,以此确定采样倍率,对于每个少数类样本,从其 k 近邻中随机选择 x_p. 再对每个随机选择的 x_p,分别按照以下公式生成新的少数类样本,最终合成来均衡数据集.

$$x = x_i + \text{rand}(0,1)(x_i - x_p). \tag{8}$$

其中,rand(0,1)表示区间(0,1)的随机数.

2　实验结果与分析

2.1　实验数据

数据来自于 Kaggle 网站(https://www.kaggle.com)提供的 Lending Club 借贷平台真实交易数据,Lending Club 的主要业务是为借款人提供贷款,同时为投资人提供贷款项目,是一种典型的 P2P 借贷平台,是个人或企业使用的一种借贷方式[17]. 选用的 Lending Club 真实交易数据作为样本数据集. 该数据集包含履约样本和违约样本,共计 1 379 602 条数据,其中,履约样本的数量为 1 076 751 条,违约样本的数量为 302 851 条,占比分别为 78.05% 和为 21.95%.

2.2　数据预处理

Lending Club 数据集是非结构化数据,包含多种噪声,此外,信用风险评估模型应用在实际生活中会面临一个不容忽略的问题:客户的信贷数据是不平衡的,即按期还款信用良好的申请人数量远远大于信用不良的申请人数量. 因此,在进行数据分析之前,首先对数据进行预处理. 主要方法如下:

1)特征选择. Lending Club 数据集中包含 151 个特征,特征数量较多,我们借鉴学者在研究同类问题时的特征筛选方法[18-19]进行特征选择. 首先,对于缺失值比例大于 80% 和与信用风险评估无用的字段特征直接删除. 例如,如果变量的取值只有一种,则该类变量对目标变量没有任何预测能力,删除该变量. 在消除了非效应属性后,根据皮尔逊相关系数进一步衡量两个特征变量间的相关性,最终选择了贷款金额,贷款期限,分期付款额,贷款等级等 21 个主要特征进行信用风险评估模型的构建. 使用贷款状态作为目标特征,同时对字符型数据进行标签数据化.

2)缺失值填补. 进行特征选择之后的 Lending Club 数据集缺失数据较少,但仍存在少部分缺失,选择众数填补的方法对缺失值进行填补. 即对于缺失值,采取其所属特征取值的众数进行填补.

3)不平衡数据处理. Lending Club 数据集正负样本比例为 3.5∶1,数据严重不平衡. 使用 SMOTE 算法,首先计算每个少数类样本到其他样本点的欧氏距离,得到 k 近邻样本. 然后根据 3.5∶1 的样本不平衡的比例确定采样倍率,从少数类样本的 k 近邻样本中随机选择样本,再对每个随机选择的样本生成新的少数类样本,从而数据的正负样本达到平衡. 平衡后的正负样本分布如图 2.

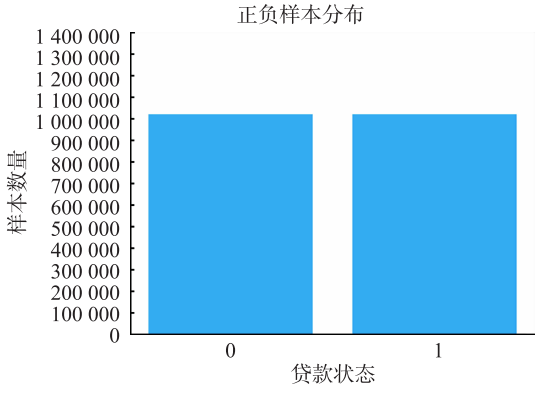

图 2　Lending Club 数据集平衡后的正负样本分布

2.3　实验步骤

使用处理后的数据构建 LightGBM 信用风险评估模型,然后使用自适应惯性权重的 PSO 算法寻找最优参数. LightGBM 模型有大量超参数,这些超参

数直接影响到模型的结构和模型的性能. 因此,对这些参数进行适当的调整就显得尤为重要. 使用 PSO 算法训练的 LightGBM 超参数含义和参数值如表 1 所示.

表 1 使用 PSO 训练的 LightGBM 模型参数的含义和参数值

参数	参数含义	参数值
num_leaves	树的最大叶子节点数	200
feature_fraction	随机抽取特征的比例	1.0
learning_rate	学习率	0.1
reg_alpha	L1 正则化项的权重系数	0.01
max_bin	最大分桶的桶个数	200
min_data_in_leaf	一个叶子上数据的最小数量	258
reg_lambda	L2 正则化项的权重系数	0.01

针对基于自适应惯性权重 PSO - LightGBM 的信用风险评估模型的方法整体流程如图 3 所示.

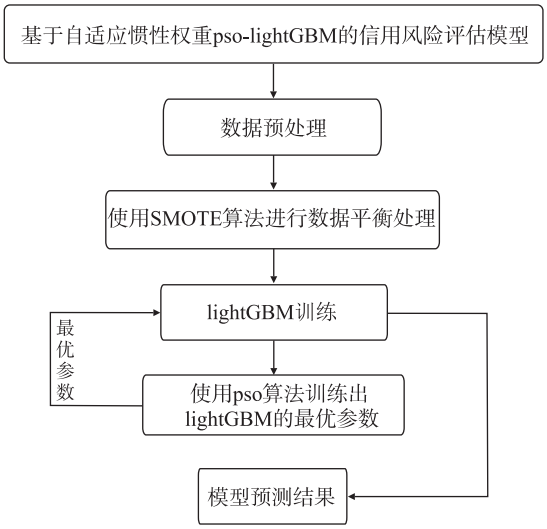

图 3 模型构建的算法流程图

为了验证 PSO - LightGBM 模型在信用风险评估中的效果,使用在信用风险评估中应用较为广泛的 4 种模型进行对比,包括逻辑回归、决策树、随机森林、朴素贝叶斯信用风险评估模型,同时也将自适应 PSO - LightGBM 模型与未经过自适应惯性权重 PSO 训练的 LightGBM 进行对比,各种模型采取 python 中 LogisticRegression 软件包、DecisionTreeClassifier 软件包、nb. BernoulliNB 软件包、RandomForestClassifier 软件包以及 LGBMClassifier 软件包执行. 模型的参数采取默认值,具体见表 2.

表 2 使用的模型参数及含义

模型	参数设置
逻辑回归	正则化强度 = 1.0,最大迭代次数 = 1 000,日志信息量控制 = 0,CPU 核数 = 1,误差容忍阈值 = 0.000 1,常数项缩放系数 = 1
决策树	划分内部节点的最小样本数 = 2,叶节点所需的最小样本数 = 1,叶节点所需的最小样本数 = 0,剪枝复杂度 = 0
朴素贝叶斯	一个叶子上数据的最小数量 = 1.0,二值化阈值 = 0.0,是否学习类先验概率 = True
随机森林	树的个数 = 100,内部节点划分最小样本数 = 2,叶子节点最小样本数 = 1,叶子节点最小样本权重 = 0.0
LightGBM	树的最大叶子节点数 = 31,随机抽取特征的比例 = 1.0,学习率 = 0.1,L1 正则化项的权重系数 = 0.0,‘最大分桶的桶个数 = 255,一个叶子上数据的最小数量 = 20,‘L2 正则化项的权重系数 = 0.0

2.4 评价指标

信用风险评估模型的构建解决的实际是一个二分类问题,因此选择准确率(Accuracy)、召回率(Recall)、查准率(Precision)作为评价指标. 同时,为了兼顾召回率和查准率,选取了综合评价指标 F1 - score. ROC 曲线可以很直观地看出分类器的性能,因此 ROC 曲线作为评价指标. 准确率、查准率、召回率和 F1 - score 见公式(9) ~ (12).

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}. \tag{9}$$

$$\text{precision} = \frac{TP}{TP + FP}. \tag{10}$$

$$\text{Recall} = \frac{TP}{TP + FN}. \tag{11}$$

$$\text{F1 - score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{12}$$

其中,TP(true positive)和 TN(true negative)分别表示正样本被模型正确分类的数量和负样本被模型正确分类的数量,而 FN(false negative)和 FP(false positive)分别表示正样本被模型错误分类的数量和负样本被模型错误分类的数量. 准确率(accuracy)、召回率(recall)、查准率(precision)值和 F1 - score 越大,说明模型的效果越好.

2.5　实验结果分析

本次实验利用 python 构建逻辑回归(LR)、决策树(DT)、随机森林(RF)、朴素贝叶斯(NB)、LightGBM 和 PSO－lightGBM 信用风险评估模型. 设定目标输出 1(不良信用的客户)和 0(信用良好的客户),整合选取的特征作为输入,从 Lending Club 数据集中按照 8:2 的比例选取训练集和测试集,进行模型训练并预测结果.

表 3　实验结果

模型	准确率/%	查准率/%	召回率/%	F1－score/%	训练时间/s
逻辑回归	78.3	98.7	73.4	84.2	14
决策树	87.2	92.2	91.5	91.9	28
朴素贝叶斯	70.2	84.1	76.5	80.1	1
随机森林	88.6	91.3	94.4	92.9	315
LightGBM	89.4	92.7	93.8	93.3	4
PSO－LightGBM	90.4	92.8	95.1	94.0	12

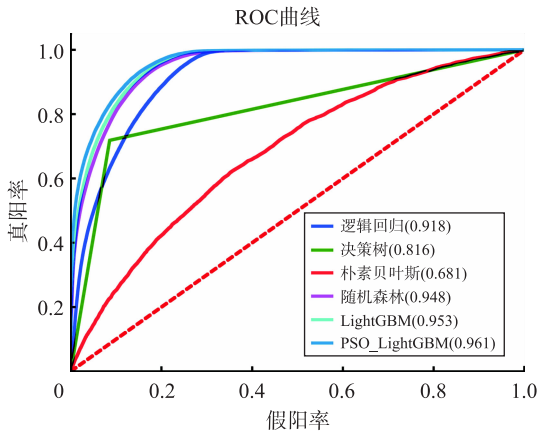

图 4　不同信用风险评估模型的 ROC 曲线

表 3 展示了不同算法得到的实验结果. 可以看到 PSO－LightGBM 的准确率和 F1－score 分别为 90.4% 和 94.0%,在所有算法中都是最高的. 与同样使用 LightGBM 模型但未经过 PSO 优化的算法相比,PSO－LightGBM 的准确率和 F1－score 分别提高了 1.0% 和 0.7%;而与其他 4 种非 LightGBM 模型相比,PSO－LightGBM 的准确率和 F1－score 提升幅度更大,分别为 1.8% ~ 20.2% 和 1.1% ~ 13.9%. 从训练时间来看,使用本文数据集进行训练时,PSO－LightGBM 用时 12 s,只比 LightGBM 少了 8 s;而相对于逻辑回归 14 s,决策树 28 s,随机森林 315 s,PSO－LightGBM 运行速度更快;虽然朴素贝叶斯算法用时只有 1 s,但其准确率和 F1－score 在所有模型中都是最低的. 图 4 可以更直观地看出 6 种模型的 ROC 曲线变化. 在 ROC 空间中,ROC 曲线越靠近 y 轴表示模型分类的效果越好. ROC 曲线与 x 轴的面积称为 AUC 值,介于[0,1]之间. AUC 的值越大,代表模型的性能越好. 从图 4 可以看出,PSO－LightGBM 模型的 AUC 值(96.1%)在所有算法中达到最高,比 LightGBM 模型提高了 0.8%,比其他模型提高了 1.3% －28%.

因此,使用自适应惯性权重 PSO(粒子群优化)算法寻找 LightGBM 模型的最优参数构建完整的信用风险评估模型,无论是从模型的准确率、F1－score 等评价指标上看,还是从训练时间上看,都取得了较好的效果,可以满足信用风险评估模型的要求.

3　结语

使用一种非常高效的集成学习模型 LightGBM 为基础来构建信用风险评估模型,首先使用 SMOTE 算法进行正负样本的平衡,以此来减小数据不平衡对模型训练造成的影响. 然后将 LightGBM 模型与其他信用风险评估模型(逻辑回归、决策树、朴素贝叶斯、随机森林)进行对比,考察了准确率、查准率和召回率等指标,实验结果表明 LightGBM 模型的的效果最好. 在 LightGBM 模型的基础上,使用自适应惯性权重 PSO(粒子群优化)算法寻找最优参数,构建本文模型 PSO－LightGBM,PSO－LightGBM 模型相比 LightGBM 模型在上述各种评价指标上都有一定程度的提升,构建的信用风险评估模型全局最优搜索能力更强,是一种有效的信用风险评估模型.

参考文献:

[1] SHON S Y, KIM D H, YOON J H. Technology credit scoring model with fuzzy logistic regression[J]. Applied Soft Computing, 2016,43:150－158.

[2] JIANG M H, JIANG L, WANG Y L. A study on personal credit scoring using linear discriminant snalysis[J]. Policy

– making Reference, 2003(01):53 – 55.
[3] 于晓虹, 楼文高. 基于随机森林的P2P网贷信用风险评价、预警与实证研究[J]. 金融理论与实践, 2016(02):53 – 58.
[4] ZHANG Z L. Identification of credit risk of personal loan in commercial bank cased on SVM[J]. Applied Mechanics & Materials, 2013,281:682 – 687.
[5] 吴陈,张明华.基于最优朴素贝叶斯分类器的个人信用预测[J]. 江苏科技大学学报(自然科学版), 2012,26(04):376 – 380.
[6] CHEM C C, LEI W U, CHEN S Y. A decision tree classifier for credit assessment problems in big data environments [J]. Information Systems and e – Business Management, 2021,19: 363 – 386.
[7] SHARIFI P, JAIN V, ARAB POSHTKOHI M, et al. Banks credit risk prediction with optimized ANN based on improved owl search algorithm[J]. Mathematical Problems in Engineering: Theory, Methods and Applications, 2021, 19:363 – 386.
[8] 郭广寒, 王志刚. 一种改进的粒子群算法[J]. 哈尔滨理工大学学报, 2010,15(02):31 – 34.
[9] KENNEDY J, EBERHRT R. Particle swarm optimization [C]//Proceedings of ICNN'95 – International Conference on Neural Networks, Perth, WA, Australia, 1995: 1942 – 1948.
[10] WEI D, WANG Z B, SI L, et al. Preaching – inspired swarm intelligence algorithm and its applications [J]. Knowledge – Based Systems, 2021, 211:106552.
[11] WANG D, ZHANG Y, ZHAO Y. LightGBM: An effective miRNA classification method in breast cancer patients[C]// Proceedings of the 2017 International Conference on Computational Biology and Bioinformatics, 2017:7 – 11.
[12] 吴照明, 胡西川. 基于LightGBM信贷风控模型的算法优化[J]. 计算机应用与软件, 2022,39(6):342 – 349.
[13] MACHADO M R, KARRAY S, DE SOUSA I T. LightGBM: an effective decision tree gradient boosting method to predict customer loyalty in the finance industry[C]//International Conference on Computer Science and Education. Toronto, ON, Canada,2019: 1111 – 1116.
[14] 谢勇, 项薇, 季孟忠, 等. 基于Xgboost和LightGBM算法预测住房月租金的应用分析[J]. 计算机应用与软件, 2019,36(09):151 – 155 + 191.
[15] CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: synthetic minority over – sampling technique [J]. Journal of Artificial Intelligence Research, 2002,16(1):321 – 357.
[16] 杨智明, 乔立岩, 彭喜元. 基于改进SMOTE的不平衡数据挖掘方法研究[J]. 电子学报, 2007,35(B12):5.
[17] EMEKTER R, TU Y, et al. Evaluating credit risk and loan performance in online Peer – to – Peer (P2P) lending[J]. Applied Economics, 2015, 47(1 – 3):54 – 70.
[18] TAN Y, ZHAO G. Multi – view representation learning with Kolmogorov – Smirnov to predict default based on imbalanced and complex dataset[J]. Information Sciences, 2022, 596:380 – 394.
[19] MA X J, SHA J L, WANG D H, et al. Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning[J]. Electronic Commerce Research and Applications, 2018, 31: 24 – 39.

Credit risk assessment based on LightGBM and adaptive inertia weight PSO

FU Zhi-ning, LI Hui-min, XU Ya-tian, TAO Yu-hu, GAO Wei
(School of Mathematics and Computer Science, Yunnan Minzu University, Kunmin, China)

Abstract: In view of the complex personal credit risk problem in the loan market, the construction of credit risk assessment model is a very important step. Using the Lending Club dataset, the credit risk assessment model is constructed to predict the default probability of customers. First, data processing is carried out, and then the problem of positive and negative sample imbalance is processed by the SMOTE (synthetic minority oversampling technique) algorithm to obtain complete credit loan data. Secondly, the LightGBM model is used for training, and the PSO (particle swarm optimization) algorithm of adaptive inertia weights is used to obtain the optimal parameters of LightGBM. After comparison with multiple mainstream algorithms, experimental results show that the constructed model has better performance.

Key words: credit risk; imbalanced datasets; SMOTE; LightGBM; PSO

(责任编辑 段 鹏)