

Research on Default Prediction of Online Lending Borrowers Based on Machine Learning

Xiangting Wang, Zixuan Zhao, Shutan Wang, Ningning Liu

University of International Business and Economics, Beijing
Email: wangxtfighting@163.com

Received: Dec. 26th, 2018; accepted: Jan. 8th, 2019; published: Jan. 15th, 2019

Abstract

The rapid development of the online lending industry has made traditional risk control increasingly prominent in the timeliness, comprehensiveness and hierarchy of data. At present, the booming machine learning enables the online lending platform to build an intelligent risk control model by using multi-dimensional big data so that it can assess personal credit status more accurately and reduce default risk more effectively. Based on the borrower loan risk data provided by CCX Credit Technology, this paper uses Logistic, XGBoost and NN to construct a forecasting model and compares the results. The XGBoost algorithm has a high degree of flexibility and allows custom optimization goals and evaluation criteria, and it also has more parameters, the scope of adjustment is large. So the model built based on XGBoost algorithm has higher accuracy for default prediction of online loan borrower. At the same time, this article uses the automated tuning tool to traverse all the parameter combinations, which brings great convenience to the model tuning.

Keywords

Machine Learning, Online Lending, Default Prediction

基于机器学习的网贷借款人违约预测研究

王相婷, 赵子轩, 王殊檀, 刘宁宁

对外经济贸易大学, 北京
Email: wangxtfighting@163.com

收稿日期: 2018年12月26日; 录用日期: 2019年1月8日; 发布日期: 2019年1月15日

摘要

网络借贷行业的飞速发展使得传统风控在数据时效性、全面性和层次性上的短板日益凸显。而目前,

机器学习的蓬勃兴起使网贷平台能够利用多维大数据构建智能风控模型,更加准确的评估个人信用状况,有效地降低违约风险。本文基于中诚信征信有限公司提供的借款人贷款风险数据,利用Logistic、XGBoost和NN构建预测模型,并将结果进行对比。由于XGBoost算法具有高度的灵活性,允许自定义优化目标和评价标准,并且参数较多,可调整的范围大,因此基于XGBoost算法构建的模型对网贷借款人违约预测的准确率较高。同时,本文利用自动化调参工具遍历所有参数组合,给模型调参带来了极大的便利。

关键词

机器学习, 网络借贷, 违约预测

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

互联网技术和移动通信技术的不断创新以及与金融行业的快速融合,使得互联网金融中的重要发展模式——网络借贷得到了飞速的发展。网络借贷是指互联网公司利用互联网技术和移动通信技术搭建一个线上平台,为资金提供者和资金需求者提供直接的资金融通服务。这种基于网络平台的借贷方式相较于银行借贷,不仅手续简便、效率高,且允许信用贷款,为自有资本不足、缺乏担保的小额借款人提供了新的出路。它的出现满足了人们多元化的投融资需求,提高了社会闲散资金的利用率。

在以往,商业银行对贷款用户进行信用风险评估时,往往依靠风控人员依赖5C分类法来主观判断,从其个人品格、信用额度、偿付能力、市场经济状况这5个因素对贷款用户进行判断和权衡,以此作为是否贷款给该用户的参考,决定是否发放贷款,这种依赖主观判断的方法显然效率低下,而且评估的时候十分依赖风控人员的主观判断能力,从公司内部管控的角度来看,甚至存在风控人员内部作弊的可能性,不能适应市场经济的快速发展,满足贷款用户的需求,也不能满足网贷平台风险管理的需求。面对数以万计甚至是数以十万计的申请借款的用户时,网贷平台则需要采用各种机器学习的方法来减少监控与检测过程中的人工参与部分,利用自动化的方法提高放款审核的准确率和效率。

目前,网贷行业运用大数据技术进行风险控制管理已经取得了一定的成效,比较成熟的产品有Zest Finance公司所开发的基于数据挖掘和机器学习理论的分析模型,以及美国使用最广泛的个人信用评分系统——FICO信用评分,都是美国借贷行业贷款决策的重要参考标准。而在国内,2014年10月,宜人贷将采用了大数据技术的“极速模式”添加进“宜人贷借款”APP中,积木盒子根据建立的“读秒”标准判断用户的信用等级,拍拍贷于2015年推出的魔镜风控系统被认为是行业内首个基于大数据的风控模型,首个能准确预测借款标的风险概率的风控系统,爱钱进在其两周年发布会上,推出了基于机器学习、深度学习等技术的全新风控体系——“云图动态风控系统”。由此可见,机器学习的蓬勃兴起使网络借贷平台利用多维大数据构建智能风控模型,更加准确的评估个人信用状况,有效地降低违约风险。

2. 文献综述

近年来,学者们研究了多种机器学习方法如决策树、神经网络、支撑向量机和集成算法等在信用评

估方面的应用。Aledanjro (2016)等[1]发现在各种特征变量呈现出复杂的非线性关系的情况下,传统的 Logistic 回归的效果并不如以前的表现良好。虽然 Logistic 模型可能在预测精度方面不如机器学习模型,但其在变量解释性和稳定性方面有较强的优势,因此也有学者将 Logistic 回归改进后应用到对借款人的违约行为预测中。熊正德等人(2017) [2]将客户由“违约”与“不违约”两类细分为“提前结清”、“当前正常”、“可疑”、“损失”四类,构造出信用状态转移方程,反向推测每类客户的潜在违约概率,利用有序多分类 Logistic 模型及 ROC 检验进行网贷客户违约预测实证研究,结果显示,模型在客户违约预测方面,尤其针对“当前正常”和“可疑”两类客户,具有更高的准确率。阮素梅,周泽林(2018) [3]利用 L1 惩罚 Logit 模型,实证检验 P2P 网贷信用违约的关键影响因素;并利用混淆矩阵与 ROC 曲线等分类评价方法,检验模型的违约预测效果,研究发现 L1 惩罚 Logit 模型具有很好的变量选择功能,可以有效地识别影响信用违约的关键因素,降低管理者的监管成本;L1 惩罚 Logit 模型能够获得比普通 Logit 模型、支持向量机等更好的预测效果,既能够从总体上实现对信用违约状态的准确预测,又能够细致分析关键影响因素对违约概率造成的影响,有助于预测和控制信用风险的发生。

Sideny (2014)等[4]提出了使用决策树等有监督学习方法的预测欺诈行为,并通过变量筛选的方法,针对真实的在线商业数据,挑选出所有变量中解释性最强的 10 个变量,构造出一个比用所有变量更稳定、准确率更高的决策树模型。王茂光等(2016) [5]以决策树算法为核心建立起风险监控模型,并通过对模型的参数进行调整,提出以保证整体错误率的前提下,尽可能的降低错误率的评价标准。王程龙等(2016) [6]同时发现决策树模型在解释贷款违约原因、划分信用等级以及降低违约率等方面具有适用性强、精确度高、可解释强的优势。

部分研究将神经网络应用到信用风险评估,实验结果其对借款人信用风险有较为准确的判断。Desai 等人(1996) [7]使用神经网络技术,构建个人信用风险模型,通过实证分析的方法证明了在这种变量间关系错综复杂,且大多数为非线性关系的情况下,神经网络技术有着明显的优势。吴斌等人(2017) [8]在考虑 P2P 网贷中“软信息”较多、“硬信息”缺失的特点上,提出了基于 BP 神经网络的信用评估模型。为了提高 BP 神经网络的收敛速度和精度,将改进的果蝇优化算法作为 BP 神经网络的学习算法,对神经网络的权重进行训练,通过“人人贷”平台收集的样本数据进行实验验证,结果表明改进果蝇神经网络评估模型比传统 BP 神经网络模型有更强的学习能力和预测能力,是 P2P 网贷个人信用风险评估的有效方法。李昕等人(2018) [9]应用 BP 神经网络模型对 P2P 网贷借款人信用评估进行实证测评,结果表明该模型具备良好的特征抽取和知识发现能力,在借款人评估指标体系中存在虚拟信息指标时,仍能对借款人信用风险做出较为准确的判断,具有较强的评估预测能力。

Fernandez-Delgado (2014) [10]利用一家大型银行的实际数据,测试了 179 种分类模型在 UCI 所有的 121 个数据集上的性能,发现 SVM 在综合性能上最优。

由于较高的准确率,运用机器学习中的集成算法是近期的一种趋势,目前接触较多的集成学习主要有 3 种:基于 Boosting 的、基于 Bagging 的和基于 Stacking 的。基于 Boosting 的代表算法有 Adaboost、GBDT 和 XGBOOST,基于 Bagging 的代表算法主要是随机森林,这两种都是基于一种算法,而 Stacking 集成是基于多种算法。张宁静等(2018) [11]基于美国 Lending Club 网贷平台,通过探究 P2P 校园网络贷款真实数据,运用数据挖掘 Adaboost 算法识别借款人违约风险重要因素指标,结果显示除借款利率外,社会信用、社会关系资本和还款行为都是影响借款违约风险的重要指标。蒋翠清等学者(2017) [12]提出一种融入软信息的网络借款违约预测方法,即分析不同软信息变量对借款违约的影响关系进而对软硬信息进行组合筛选,引入随机森林算法构建融入软信息的违约预测模型,并结合 P2P 平台的真实数据进行实证分析,结果表明,在 P2P 借款的违约预测模型中融入有价值的软信息可以提高预测准确率。丁岚等(2017) [13]以 logistic 回归、决策树、支持向量机作为初级学习器,以 SVM 作为次级学习器,构建基于 Stacking

集成策略的评估模型来预测 P2P 网贷中借款人的违约风险,构建模型时采用 10×10 折嵌套交叉验证方法以克服交叉学习现象,通过爬虫技术抓取人人贷的交易数据进行的实证研究表明,相较单一的 logistic 回归、决策树或者 SVM 模型,基于 Stacking 集成策略的预测模型能显著降低一类错误和二类错误比例,提高预测正确率。具体研究方法整理如表 1。

Table 1. Method of loan default prediction

表 1. 贷款违约预测方法

分类	方法/模型	研究内容	作者
基于机器学习的 信用风险评估	Logistic	传统的在非线性下表现一般,有序多分类 Logistic 和基于 L1 惩罚的 Logit 能有效识别影响信用违约的关键因素	Aledanjro (2016); 熊正德等(2017); 阮素梅等(2018) Sideny 等(2014);
	决策树	决策树能较好地解释贷款违约原因,较为准确地预测违约率	王茂光等(2016); 王程龙等(2016) Desai 等(1996);
	神经网络	在非线性情况下,对借款人信用风险能做出较为准确的判断	吴斌等(2017); 李昕等(2018)
	支持向量机	SVM 预测准确性高,但算法复杂,解释性较差	Fernandez-Delgado (2014)
集成学习	Boosting	运用 Adaboost 算法识别借款人违约风险重要因素指标	张宁静等(2018)
	Bagging	引入随机森林算法构建融入软信息的违约预测模型,可以有效提高预测准确率	蒋翠清等(2017)
	Stacking	以 logistic 回归等作为初级学习器,以 SVM 作为次级学习器,显著降低错误比例,提高预测正确率	丁岚(2017)

基于集成学习的方法在近几年备受关注,其核心在于针对不同领域数据使用不同机器学习对数据进行处理和建模,产生一组“个体学习器”,集成模型通过将多个学习器进行组合,常可获取比单一学习器显著优越的泛化能力。因此在本研究中将采用集成学习进行贷款人违约行为的预测。

3. 数据分析

3.1. 数据描述

中诚信征信公司主办的第二届“风云杯”机器学习建模大赛提供了网络借贷行业的贷款风险数据供参赛者下载;其中,总共提供了基于两个相似金融场景的数据集,scene_A 与 scene_B 有相同的字段名称,对应的特征内容相同,包括借款人的基本信息和历史行为信息,比如借款人的性别、年龄、学历、借款次数、借款金额,还款情况等,总计约 2300 个字段。大赛主办方提供了 21,245 条训练集数据 scene_A,包括违约标的 6982 条,非违约标的 14,263 条,以及 8094 条测试集数据 scene_B。scene_A 用于比赛初期进行模型提交测试,scene_B 用于比赛最后一天模型泛化能力测试,评判指标为 AUC 值。

参赛者需要依据提供的数据,对每个标的逾期情况进行分析、建模,最终给出指定借款人逾期概率的预测值,并通过提交的预测结果、模型理论基础、算法创新型和表现稳定性等多个维度进行综合评估。

3.2. 问题分析

借款人的违约预测实质上是二分类问题,x 是用户的基本信息和历史行为数据等,预测变量 y 是每个借款人的贷款逾期情况,正常还款标记为 0,违约标记为 1。在对比赛的数据进行数据清洗和特征工程后,选择相应的模型进行学习,并进行模型融合,最终的评估指标为 AUC 值。

问题解决思路如图 1:

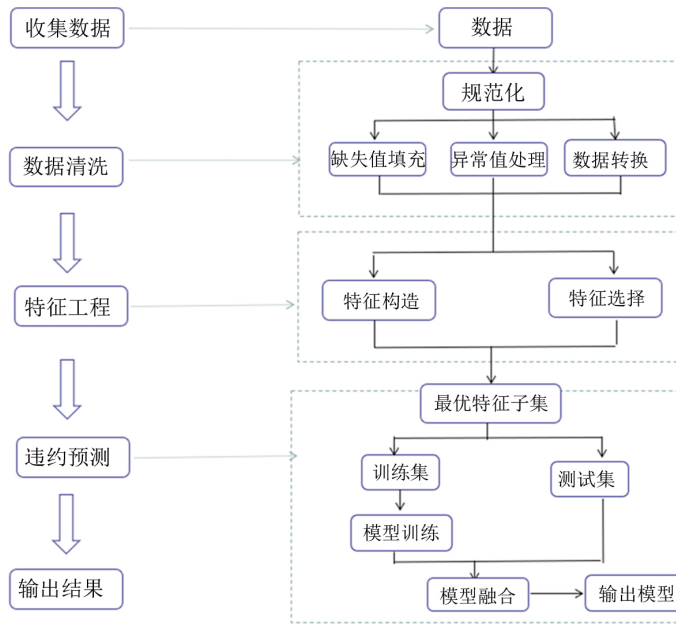


Figure 1. Solution of the problem
图 1. 问题解决思路

3.3. 数据清洗

在征信领域，用户信息的完善程度可能会影响该用户的信用评级。一个信息完善程度为 100% 的用户比起完善程度为 50% 的用户，会更加容易审核通过并得到借款，因此需要进行数据清洗。在分析过程中，通常主要的关注点在于：缺失值、异常值以及不一致数据等方面，需要通过采用适当的方法对数据进行有效的清查与转换，使之成为符合建模要求的数据源。从这一点出发，首先对缺失值进行了多维度的分析和处理。在对数据表进行缺失值统计后，结合字段属性本身的特征和信息缺失的多少，对应了不同的缺失处理方式。若字段的缺失值占总记录的 99.9%，仅有小部分具有数值，其所具有的信息非常有限，因此将其剔除；对于剩余的缺失值所占比重较低的字段，根据关键属性列的性质，对数值型数据，分别用 0 和中位数填充；对取值为有限个数的类别数据，用 0 和众数填补其缺失值。其次，针对极少数的异常值进行处理，超出正常范围的数值 231-1 以空值填充。最后，通过观察发现日期型数据中出现了不一致的形式，即部分年份数据出现了具体的日期如 20170301 等，需要去掉其日期，保留年份。

3.4. 特征工程

接下来需要对清洗后的数据进行特征构造和选择。特征构造是特征工程的第一步，也是最为重要的一步，它需要深入分析综合数据特征及所属业务场景，为后续的分析与建模提取关键信息。首先是将主从属性合为一类，如针对地理位置信息，合并省市信息，并统计出现每个 id 下城市出现的次数等。其次，对于日期型数据，统计所有日期的时间跨度作为一系列新属性；经观察发现有的 id 对应多条记录，但时间却是相同的，将其频率与时间跨度相除的结果作为一系列新属性。再次，对于数值型数据，统计每个属性最小值、最大值、均值、中位数、方差等作为新的属性；属于类别型的数据在进行值处理时为了防止过拟合，将这些列出现频率低于 0.01 的所有值视为同一类别，随后进行 One-Hot 编码。

结合特征构造来看，由同一类或同一个变量衍生出来的变量必然存在较大的相关性，而其他变量之间的相关性微弱，基于此本文选取 Pearson 相关系数 0.90 为阈值作进行特征选择。

相关系数公式如下：

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

其中 (X_i, Y_i) 是样本点, \bar{X} 和 \bar{Y} 是样本均值。部分相关分析结果如表 2 所示。经过特征删选后, 保留了 1514 个字段的的相关属性作为模型输入因子。

Table 2. Correlation analysis

表 2. 相关性分析

	var1740	var1812	var1734	var1806	var197	var1163	var1169	var1208	var170	var1214
var1740	1									
var1812	0.99	1								
var1734	0.99	0.98	1							
var1806	0.98	0.99	0.99	1						
var197	0.98	0.97	0.96	0.96	1					
var1163	0.97	0.96	0.96	0.95	0.98	1				
var1169	0.97	0.96	0.95	0.94	0.98	0.98	1			
var1208	0.96	0.95	0.95	0.94	0.97	0.96	0.98	1		
var170	0.96	0.94	0.95	0.93	0.98	0.98	0.96	0.97	1	
var1214	0.96	0.94	0.95	0.93	0.97	0.97	0.98	0.96	0.95	1

4. 模型的建立

4.1. Logistic 回归

Logistic 回归是最简洁、快速、稳健的分析方法, 可解释性强; 但是借款人违约预测应以精度为标准, 而 Logistic 回归对变量关系的线性限制, 难以达到精度最优, 但是在建模时可以充分发挥它的特性: 1) 作为基准, 对数据清洗效果和模型表现进行快速评估; 2) 与结构不同的模型加权组合预测, 补充原模型精度和稳健性。

Logistic 回归通过

$$f(x) = \frac{1}{1 + e^{-x}}$$

将因变量值域 $(-\infty, +\infty)$ 映射到 $(0, 1)$, 解决二分类问题。当因变量值大于 0.5 时, 认为因变量属于 1 类; 当因变量小于 0.5 时, 认为因变量属于 0 类。本文以 Logistic 回归对其他模型表现进行评估, 与结构不同的模型加权组合预测。

建立 Logistic 回归模型如下:

$$y = \frac{1}{1 + e^{-(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots)}}$$

其中 y 代表借款人违约概率, x_i 借款人相关信息, 包括基本信息和行为信息等。

4.2. XGBoost 算法

提升树在数据挖掘领域有着相当广泛的运用, 包括提升树(BT)和梯度提升树(GBDT)。XGBoost 是在 GBDT 基础上改进的, 改进的点有: 1) 使用二阶泰勒展开式逼近学习的目标, 与 GBDT 的一阶泰勒展开式相比, 能学习到更多的东西, 所以在泛化能力及模型的稳定相面有很大的提升; 2) 目标函数之外加上了正则化项整体求最优解, 用以权衡目标函数的下降和模型的复杂程度, 避免过拟合。考虑网络借贷违

约预测是一个非线性的分类问题，变量成分较多元，样本和变量间无固定模式关联。如果以精度为目标，综合考虑稳健性、速度、通用性等因素可以首选 XGBoost。

XGBoost 的目标函数

$$L(\theta) = \sum_i^n l(y_i, \hat{y}^{(t-1)} + f_{t(x_i)}) + \Omega(f_t)$$

目标函数二阶泰勒展开整理后得最优解

$$L(\theta) = -\frac{1}{2} \sum_{j=1}^T \frac{\left(\sum_{i \in I_j} g_i \right)}{\lambda + \sum_{i \in I_j} h_i} + \gamma T_j$$

4.3. 神经网络

XGBoost 的出发点是各变量完全独立，而神经网络的出发点是各变量充满复杂的非线性关联，从而不断去优化网络权重向真实关联趋近。两种模型结构具有较高的互补性，因此本实验也选取了神经网络模型。一个基本的神经网络由输入层、隐藏层、输出层构成，相邻两层间由权重矩阵连接。通过不断提供训练样本，神经网络会学习最优权重参数，从而减小拟合误差，利用输入数据建立模型并模拟输出。本文选取借款人的性别、年龄、学历、借款次数、借款金额，还款情况等信息指标作为神经网络模型的输入值，模型的输出值是借款人的违约概率。

4.4. 参数优化

本文使用 hyperopt 库的 Tree of Parzen Estimators 接口来做参数优化，它可以找到所有模型的最佳参数，从而得到最佳的模型并进行比较，相较于 random search 和基于全空间搜索的 grid search 方法，hyperopt 兼顾了速度和精度。给出最优参数的概率分布，hyperopt 即可自动拟合参数空间，从而寻找全局最优参数。XGBoost 的最优参数如表 3，神经网络的最优参数如表 4。

Table 3. List of best parameters
表 3. XGBoost 最佳参数列表

类别	参数名称	参数含义	取值
提升参数	gamma	空值叶子个数	7.26
提升参数	max_depth	每棵树最大深度	5
提升参数	min_child_weight	每个叶子的最小权重和	1

Table 4. List of best parameters
表 4. NN 最佳参数列表

参数名称	参数含义	取值
nhidlayer	隐藏的层	2
rdrop	每棵树最大深度	0.5
nhidnode	隐藏节点数	128
outnode	最后一层全连接层的节点数	64
maxnorm	限定单个权重的最大值	4
batch_size	每训练一次抽取的样本数	64
earlystop	在 N 个回合中没有验证集的损失下降就提前停止训练	5

5. 实验结果

本文采用 AUC 值来评判模型的最终效果, AUC 即以 False Positive Rate 为横轴, True Positive Rate 为纵轴的 ROC 曲线下方的面积的大小。AUC 值在[0,1]区间, 一般认为, 一个分类模型的 AUC 值要大于 0.5 才会有实际上的应用效果; 若分类模型的 AUC 值在 0.9 以上, 则应该对模型进行检查, 看是否有异常的变量进入模型, 导致 AUC 值过高。所有表格应尽可能采用“无环绕”环绕方式, 尽量避免采用“环绕式”。

$$AUC = \frac{1}{2} \sum_i^{m-1} (x_{i+1} - x_i)(y_i + y_{i+1})$$

利用 Logistic、XGBoost 和 NN 构建预测模型, 三种模型用相同数据集训练模型后, 对比结果如表 5。从结果中可以看出, XGBoost 模型的 AUC 值高于其他两种模型, 在网贷借款人违约预测中是一种较为有效的方法。

Table 5. Comparison of modeling results
表 5. 模型结果对比

模型	Logistic	XGBoost	NN
类型	逻辑回归	梯度强化树	神经网络
平均 AUC	0.621	0.635	0.615
最差 AUC	0.615	0.629	0.607
可解释性	好	中	差

6. 总结

本文基于中诚信征信有限公司提供的借款人贷款风险数据, 在进行数据清洗后, 依照特征工程的分析与构造流程, 按照地理信息和消费信息等深入研究并设计批处理方案, 随后利用 Logistic、XGBoost 和 NN 构建违约预测模型, 并将结果进行对比, 发现由于 XGBoost 算法具有高度的灵活性, 允许自定义优化目标和评价标准, 泛化能力稳定, 因此基于 XGBoost 算法构建的模型对网贷借款人违约预测的准确率较高。同时, 通过特征重要性分析发现借款人的消费行为数据对于违约预测存在积极意义, 因此建议网贷平台将借款人的消费行为数据作为辅助决策依据纳入风控体系中。

致 谢

感谢我的导师刘宁宁老师, 在整个论文写作过程中给予我的大力帮助。刘老师的悉心指导贯穿了论文写作的方方面面, 在他的指导下我认识到了自己很多不足, 并在这一过程中取得进步。

参考文献

- [1] Correa Bahnsen, A. (2016) Feature Engineering Strategies for Credit Card Fraud Detection. *Expert Systems with Applications*, **51**, 134-142 <https://doi.org/10.1016/j.eswa.2015.12.030>
- [2] 熊正德, 刘臻焯, 熊一鹏. 基于有序 logistic 模型的互联网金融客户违约风险研究[J]. 系统工程, 2017, 35(8): 29-38.
- [3] 阮素梅, 周泽林. 基于 L1 惩罚 Logit 模型的 P2P 网络借贷信用违约识别与预测[J]. 财贸研究, 2018, 29(2): 54-63.
- [4] Tsang, S. (2014) Detecting Online Auction Shilling Frauds Using Supervised Learning. *Expert Systems with Applications*, **41**, 3027-3040. <https://doi.org/10.1016/j.eswa.2013.10.033>
- [5] 王茂光, 葛蕾蕾, 赵江平. 基于 C5.0 算法的小额网贷平台的风险监控研究[J]. 中国管理科学, 2016, 24(S1):

345-352.

- [6] 王程龙, 陈程. 基于决策树的 P2P 网贷平台信用评级体系研究[J]. 农村金融研究, 2016(12): 45-50.
- [7] Desai, V.S., Crook, J.N. and Overstreet, G.A. (1996) A Comparison of Neural Network and Linear Scoring Models in the Credit Union Environment. *European Journal of Operational Research*, **95**, 24-37.
[https://doi.org/10.1016/0377-2217\(95\)00246-4](https://doi.org/10.1016/0377-2217(95)00246-4)
- [8] 吴斌, 叶菁菁, 董敏. P2P 网贷个人信用风险评估模型研究——基于混合果蝇神经网络的方法[J]. 会计之友, 2017(21): 32-35.
- [9] 李昕, 戴一成. 基于 BP 神经网络的 P2P 网贷借款人信用风险评估研究[J]. 武汉金融, 2018(2): 33-37.
- [10] Fernandez-Delgado, M., Cernadas, E., Barro, S., *et al.* (2014) Do We Need Hundreds of Classifiers to Solve Real World Classification Problems. *The Journal of Machine Learning Research*, **15**, 3133-3181.
- [11] 张宁静, 顾新, 杨铖. P2P 校园贷款个人违约风险因素指标探析[J]. 财会月刊, 2018(6): 82-89.
- [12] 蒋翠清, 王睿雅, 丁勇. 融入软信息的 P2P 网络借贷违约预测方法[J]. 中国管理科学, 2017, 25(11): 12-21.
- [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): 41-54.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2324-7908, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: ssem@hanspub.org