

基于卷积神经网络的客户信用评估模型研究^{*}

刘伟江^{1,2} 魏海² 运天鹤²

¹(吉林大学数量经济研究中心 长春 130012)

²(吉林大学商学院 长春 130012)

摘要:【目的】分析客户贷款信息基础上,抽取贷款客户特征并成像,利用卷积神经网络构建客户信用模型,提高客户网贷违约预测准确率。【方法】基于Lending Club客户信用数据,将反映客户信息4个方面的特征变量相互连接综合成灰度图,建立基于卷积神经网络的客户信用评估模型。【结果】实验结果表明,基于卷积神经网络的新模型在信用评估实验中特异度为99.4%,灵敏度为68.7%,G-mean值为82.7%,F1值为81.4%,AUC值为99.5%,与传统以特征处理为基础的信用评估模型相比均有显著提升。【局限】仅对比分析有限的信用评估模型,未对不平衡数据的影响做进一步研究。【结论】基于卷积神经网络的网贷客户信用评估模型,在客户违约特征信息提取和违约可能性的预测上具有良好性能。

关键词: 卷积神经网络 指标成像 信用评估 信息值 主成分分析法

分类号: TP393 G250

DOI: 10.11925/infotech.2096-3467.2019.1285

引用本文: 刘伟江, 魏海, 运天鹤. 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 Weijiang, Wei Hai, Yun Tianhe. Evaluation Model for Customer Credits Based on Convolutional Neural Network[J]. Data Analysis and Knowledge Discovery, 2020, 4(6): 80-90.)

1 引言

信用评估是指评估机构利用专家判断或数学模型,结合借款者所提供的财务状况、经营状况、历史还款情况等各类相关信息,对借款者如期足额偿还债务本息的能力和意愿进行评价,并按照其违约概率的大小以等级或分数的形式给出评估结论的行为^[1]。近年来,市场环境的优化以及消费者收入水平的提高不断刺激消费需求,同时也带动各类消费信贷需求的增加。随着各类信贷需求的不断增加,信贷违约风险也在不断提升,因此建立有效准确的信贷违约风险评估模型或者方法,无疑会对规范整个信贷行业、优化信贷市场环境以及降低投资风险

产生至关重要的作用。

国内外很多学者对于信用评估进行了大量深入研究,常用的信用评估模型包括人工神经网络(Artificial Neural Network, ANN)、随机森林、Logistic回归、支持向量机(Support Vector Machine, SVM)等^[2],其中Altman结合统计学方法建立Z-score模型评估上市公司的信贷风险最具有里程碑意义^[3]。基于Altman线性回归思想的启发,Wiginton将Logit回归模型引入到信用评估的研究中,并且对比分析逻辑回归法与线性判别分析法(Linear Discriminate Analysis, LDA),结果显示逻辑回归模型在评估中的表现要优于判别分析法^[4]。但是随着研究不断深入,影响信用的因素不断被发现,在数据维度不断增

通讯作者: 魏海, ORCID: 0000-0002-9045-7113, E-mail: weihai94@163.com。

^{*}本文系教育部人文社会科学重点研究基地项目“新常态下促进经济稳定增长的要素配置与产业升级政策研究”(项目编号: 16JJD790015)、国家自然科学基金项目“中国经济周期波动的转折点识别、阶段转换及预警研究”(项目编号: 71573105)和2020年度吉林大学东北振兴发展专项研究课题“大数据背景下吉林省外来投资情况动态监测及新冠后的对策建议”(项目编号: 20ZXZ01)的研究成果之一。

加的同时数据量也在爆炸式增长,传统基于简单统计分析方法的信用评估方法评估效果变得越来越不尽人意。随着计算机技术和人工智能的不断发展,机器学习、人工智能等新兴基于大数据分析的方法逐渐在信用评估研究中得到应用,如吴冲等利用支持向量机结合模糊积分方法构建客户信用评估模型,并对比分析不同集成模型的性能^[5];Blanco 等利用人工神经网络构建信用评估模型,并对比分析神经网络、判别模型以及逻辑回归模型等,实验结果显示利用神经网络建立的信用评估模型预测性能要远优于其他传统统计类模型^[6]。

随着数据维度的进一步增加,基于传统机器学习方法的评估模型性能也会受限,因此 Chen 等在构建信用评估模型过程中,采用不同特征筛选的方法对数据进行降维处理,以降低维度灾难,从而减少其对预测准确率的影响,还对比分析了不同模型的评估性能^[7]。然而上述模型在使用过程中存在一定缺陷:一方面是由于模型本身结构的限制,如 BP 神经网络就存在局部最优、泛化性较差的缺点,主要原因是以传统机器学习方法(如支持向量机等)为代表的这类模型本质上是一种浅层学习结构(即网络结构层次较少,如普通神经网络一般只包含一个隐层),当遇到高维、数据量庞大、结构复杂的数据问题时,这类浅层结构算法在刻画数据特征、表达复杂函数方面的能力非常有限^[8];另一方面是由于模型构建过程中对数据进行降维处理所带来的问题。降维处理在某种程度上可以减少噪音,降低数据冗余,但是同时也会损失很多有用的信息,难免会存在“误伤”问题,从而降低了模型的泛化能力和预测准确率。对同属于分类的上市公司财务风险评估问题,吴星泽在分析模型构建过程中存在的问题时,提出如果自变量对因变量具有预测能力,则自变量与因变量之间必有相关性;预测能力越强,相关性越强,反之则未必^[9]。因此,通过对数据进行简单分析处理后构建的评估模型,已经无法满足相关主体对客户信用评估准确率越来越高的需求,因此需要新的模型或方法优化建模过程,实现更高的预测准确率。

随着深度学习研究的深入,深度神经网络结构具有良好的特征信息挖掘和优化特性得到越来越多的学者青睐,在图像识别^[10]、语音识别^[11]、音频处

理^[12]和自然语音处理^[13]等领域,深度学习技术已经得到广泛应用并取得了很好的实验效果。对于客户信用评估模型存在浅层学习的缺点,深度学习的思想在一定程度上可以克服,其主要观点为:

(1)采用多隐层的深层网络结构相比浅层结构更能学习刻画复杂数据的本质特征,对可视化和分类等任务而言有很大的帮助;

(2)通过无监督的逐层初始化策略有效克服深度神经网络在训练中的困难。

卷积神经网络(Convolutional Neural Networks, CNN)作为深度学习技术应用最成熟的模型之一,一方面由于继承了深度学习自动提取特征的优点,实验中由模型自动对原始数据进行综合处理运算以提取有效特征信息后进行训练和预测,一定程度上能够以最大效度提取和利用数据特征信息,从而有效减少人为因素的干预,实现特征处理和模型训练过程的统一,因此可以很好地解决传统方法“两步”的建模过程带来的数据维度和模型性能无法有效平衡的问题。另一方面卷积神经网络模型借助局部感受野的理论进行卷积运算,可通过共享权值减少训练次数,从而大大提升模型运行效率,已经被运用到很多相关研究中,如李慧等建立基于卷积神经网络的情感分析方法,实验取得了较高的准确率同时保持了良好的运行效率^[14]。然而关于卷积神经网络的研究多是进行非数值型数据建模分析,将卷积神经网络应用于数值型数据的研究还比较少,Hosaka 尝试将卷积神经网络和上市公司财务预警相结合,对上市公司财务报表指标数据进行财务比率成像,然后利用卷积神经网络构建模型进行上市公司破产风险评估研究,对比分析 Z-score、SVM、MLP 等方法。实证结果显示新方法对比传统方法预测准确率上有较大提升,同时将卷积神经网络拓展应用于数值型数据问题分析中,进一步拓展了同类型问题的研究方法和思路^[15]。

综上所述,由于实际客户数据情况的复杂多样性,导致客户信用风险往往并不是某个或者某几个单一因素,而是由多个因素综合作用的结果。传统的客户信用评估模型构建过程中,由于人为因素的干预,并不能实现真正意义上的“统一”建模。本文借鉴国外学者将数值型数据图像化处理,利用卷积

神经网络建立风险评估模型的方法,将反映客户信息的特征变量按照客户主体特征的4个方面(贷款信息、偿付能力、历史信贷、历史申请)的序列进行相互连接综合形成灰度图,结合信用数据的实际情况,重构建新的基于卷积神经网络的客户信用评估

模型。

2 模型构建理论基础

本文提出客户信用模型研究的基本流程如图1所示。

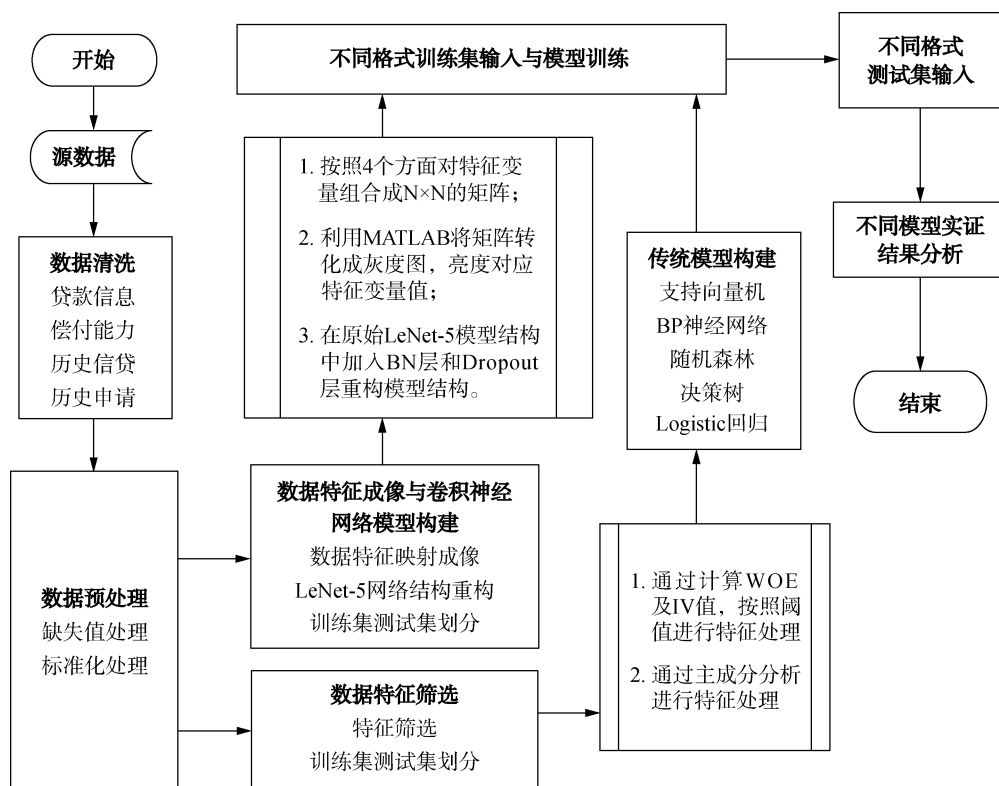


图1 结合卷积神经网络的客户信用评估研究流程

Fig.1 Research Process of Customer Credit Evaluation Combined with Convolutional Neural Network

首先是对源数据进行清洗和预处理,基于传统方法的评估模型是通过对特征变量进行共线性处理以去除具有强共线性的特征变量,由此得到初步筛选变量。而后,一方面按照传统评估模型构建方法建立模型,同时对比分析了两种常用的不同特征处理方法对特征变量进行筛选:一种方法是计算特征变量信息度的值即IV值,根据其大小设定一个阈值以对特征变量做进一步筛选(IV值的大小决定了特征变量对目标变量的影响程度);另一种方法是通过主成分分析法提取指标特征的主成分,完成指标特征的处理,最后实现测试集和训练集的划分。另一方面对于基于卷积神经网络的模型的输入数据,将经过初步预处理的变量数据利用MATLAB按照变

量所反映的特征信息的4个不同方面进行重构,得到特征矩阵,同时将特征变量的值转化为对应灰度图的亮度值,形成关于每个实例的灰度图数据集。最后对不同格式的数据集分别对应不同的模型进行训练和最后的测试以及模型评估。

2.1 卷积神经网络模型构建

卷积神经网络作为一种特殊的多层神经网络,在进行神经网络训练时和其他神经网络一样采用反向传播算法,不同之处在于网络结构。卷积神经网络的网络连接具有局部连接、参数共享的特点。局部连接是相对于普通神经网络的全连接而言,指这一层的某个节点只与上一层的部分节点相连。参数共享指一层中多个节点的连接共享相同的一组参

数。卷积神经网络的核心是由卷积层(Convolutional Layer)、池化层(Pooling Layer, 也称次抽样层)和全连接层(Full Connection Layer)三种基本结构组合而成的多层网络结构。其中卷积层和池化层一般会取若干个,通过这两种结构层在网络结构中的交替设置实现神经网络对输入数据的特征深度提取和优化,然后链接到全连接层并输出最终结果。

LeNet-5模型作为一种典型的卷积神经网络模型,最初应用于手写数字的识别并取得巨大的成功^[16]。LeNet-5模型是由输入层、卷积层、池化层、卷积层、池化层、全连接层和输出层等7层构成。针对客户信用评估问题的特点,本文对传统的LeNet-5模型进行部分改进。

(1)由于正负样本数量存在差异,为防止模型出现过拟合现象,在全连接层F6层中加入Dropout层,并将阈值设为0.5。

(2)客户信用评估的目标是区分违约用户和正常用户,属于典型的二分类问题。因此需要将传统LeNet-5模型的输出层由10个神经元修改为2个神经元。

本文设计的用于客户信用评估的卷积神经网络模型结构如图2所示。输入为经过指标变量成像处理的 32×32 的灰度图,其余各层分别为:C1卷积层、S2次抽样层、C3卷积层、S4次抽样层和C5全连接层。输出层利用Soft-max分类器,输出对样本的训练预测结果,信用正常以编码0表示,信用异常则以编码1表示。

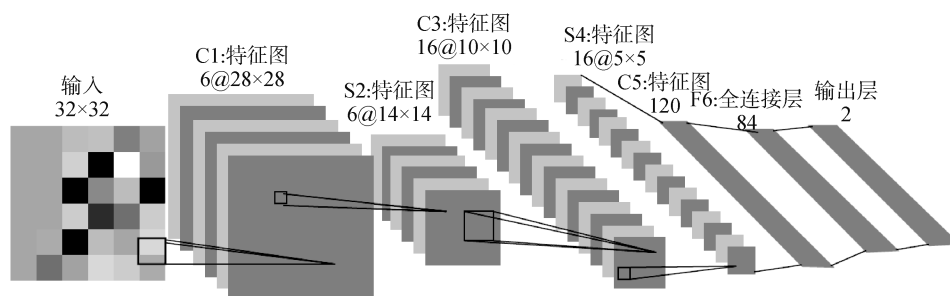


图2 本文设计的CNN网络结构

Fig.2 CNN Structure Designed in This Paper

2.2 模型输入数据处理方法

基于卷积神经网络的客户信用评估模型有效实现了指标特征选择与模型训练的统一,所以在建模时对模型输入变量要求不高,因此本文对于新模型的输入数据仅仅做了简单的数据归类、缺失值插补、清洗等预处理。传统方法的评估模型,由于模型本身原因和数据来源环境的复杂性,无法处理高维复杂数据,需要对数据做进一步筛选处理。常用的消除冗余和噪音进行指标变量筛选的方法包括基于正则化损失函数的线性模型、基于机器学习模型输出的特征重要性和基于特征信息度(Information Value, IV)等。根据数据情况,选取具有代表性的基于特征信息度和基于主成分分析的特征处理方法进行特征筛选处理。

(1) 基于特征信息度的指标特征处理

参考陶士行^[17]、刘丹等^[18]的研究,并结合数据的

分布情况,利用证据权重(Weight of Evidence, WOE)拟合度 R^2 、IV值和皮尔逊相关系数相结合的方法对数据指标变量进行筛选,具体方法如下。

① WOE值

对于某一个具体指标变量的IV值,需要先计算指标变量的证据权重值WOE。WOE作为一种对原始变量进行重新编码形式,通过对 n 个变量 $X = (x_1, x_2, \dots, x_n)$ 进行分组,划分为 M 个区间,分别计算每组WOE。对于第 i 组,WOE计算如公式(1)所示。

$$WOE_i = \ln \left(\frac{py_i}{pn_i} \right) = \ln \left(\frac{\frac{\#y_i}{\#n_i}}{\frac{\#y_i}{\#n_i}} \right) \quad (1)$$

其中, py_i 是该组中响应客户(贷款状态为违约的客户样本)占有所有样本中所有响应客户的比例, pn_i 是该组中未响应客户(贷款状态为正常的客户样本)占样本中所有未响应客户的比例, $\#y_i$ 是该组中

响应客户的数量, $\#n_i$ 是该组中未响应客户的数量, $\#y_i$ 是样本中所有响应客户的数量, $\#n_i$ 是样本中所有未响应客户的数量。

由公式(1)可知, WOE_i 表示“当前分组中响应客户占有所有响应客户的比例”和“当前分组中没有响应的客户占有所有没有响应的客户的比例”的差异。这个差异是用响应和未响应比值的数值再取比值, 然后取对数来表示。 WOE_i 越大, 这种差异越大, 该分组里的样本响应的可能性就越大; WOE_i 越小, 差异越小, 该分组里的样本响应的可能性就越小。

②IV值

为避免直接使用 WOE 值评价指标特征带来的评估值为负以及绝对值求和容易得到较高不合理评估值的情况, 需要进一步计算 IV 值即信息价值, 来评价某个具体指标特征的预测能力。对于第 i 组, 会产生一个对应的 IV, 计算方法如公式(2)所示。

$$IV_i = (py_i - pn_i) \cdot WOE_i$$

$$= \left(\frac{\#y_i}{\#y_i + \#n_i} - \frac{\#n_i}{\#n_i + \#n_i} \right) \cdot \ln \left(\frac{\frac{\#y_i}{\#y_i + \#n_i}}{\frac{\#n_i}{\#n_i + \#n_i}} \right) \quad (2)$$

对于整个变量 X_i 对应的 IV 可由各分组的 IV 进行求和取得, 如公式(3)所示。

$$IV_i = \sum_{i=1}^n IV_i \quad (3)$$

其中, n 为变量分组数量。计算所得变量的 IV, 数值的大小反映该变量对目标变量的贡献度大小。

③指标选取原则

对于变量分组数 M 的确定, 并没有一个固定值。参考文献[17-18], $M \in [5, 14]$, 结合 WOE 线性拟合度 R^2 及 IV 值确定最优分组数 M 。

对于 WOE_i 对应一个信用评估, 即一个风险因子评价, 因此需要考虑其线性拟合度情况。拟合度 R^2 越接近 1, 拟合程度越好, 评估越合理, 如公式(4)所示。

$$WOE_i = \alpha + \beta i + \varepsilon$$

$$i = 1, 2, \dots, M \quad (4)$$

对于 IV 值的计算, 参考阚士行^[17]、刘丹等^[18]的研究按照表 1 所示划分对应预测能力如表 1 所示。

(2) 基于主成分分析的指标特征处理

在构建基于传统方法的信用评估模型时, 进行

表 1 IV 值对应预测能力区间

Table 1 IV Value Corresponding to Prediction Ability Interval

IV 值	预测能力
[0, 0.02)	无预测能力
[0.02, 0.10)	预测能力低
[0.10, 0.30)	预测能力中
[0.30, +∞)	预测能力高

指标特征处理的另一种常用方法是利用以主成分提取为基础的因子分析对指标特征降维处理。在进行因子分析之前, 还需要对备选变量进行 KMO 检验和 Bartlett 球形检验。KMO 检验用于比较变量之间的相关系数。数值范围在 0~1 之间, 一般以 0.5 作为临界值。当 KMO 值大于等于 0.5 时, 说明该变量更适合因子分析; 当 KMO 值小于 0.5 时, 说明该变量不适合因子分析。Bartlett 球形检验用来检验相关矩阵变量之间的相关性(包括单位阵), 即测试每个变量是否独立。若检验的统计量较大且其 sig 值小于显著性水平, 则拒绝 H_0 , 接受 H_1 , 认为检测的变量之间存在一定相关性, 适宜做因子分析; 反之, 则接受 H_0 , 拒绝 H_1 , 不适宜做因子分析。

主成分分析(Principal Components Analysis, PCA)是将各变量之间互相关联的复杂关系进行简化分析的方法, 试图在数据信息丢失最少的原则下, 对这种多变量的截面数据表进行最佳综合简化, 即对高维变量空间进行降维处理。其主要思想是通过一定的标准提取能够代替多个变量的少数几个特征变量作为主成分进行计算, 对于主成分的提取标准有基于特征根大于 1 或特征值的累积贡献率超过一定的比例等。实际应用中, 如果通过一种标准提取的主成分不能显著表示原有变量时, 还可以将两种标准进行混合。主成分提取过程是以正交旋转变换的方式, 将其分量相关的原随机向量转化为其分量不相关的新随机向量, 即将原随机的协方差阵变换为对角形, 使之指向样本点散布开的 P 个正交方向, 再根据上面两个标准选择新变量。新变量综合地反映了原有变量所包含的信息, 因此可以在进行模型指标处理时简化数据结构和减小共线性影响, 从而提高模型泛化能力。对比特征信息度法, 主成分分析法选取的主成分能充分反映原变量的主要信息, 因此具有较强的客观性。

3 实证

3.1 样本选取和数据预处理

为更加全面地评估模型性能,本文样本数据来自 LendingClub 官网^①的公开数据集。选取 LendingClub 最新公开的 2018 年第一季度和第二季度的数据,由于本文是对客户信贷违约可能性进行评估,因此对提取的数据中贷款状态(loan_status)显示为 Issued 的样本进行剔除。对被选取数据中的文本数据进行编码分析,剔除部分缺失率大于 0.5 的指标变量。对于剩余变量中仍存在少量的样本数据缺失的问题,结合数据分布特点,采用众数对缺失数据进行填充。

根据反映贷款客户特征信息的贷款信息、偿付能力、历史信贷、历史申请等 4 个方面,预处理后初步筛选得到的数据集包含 83 个指标变量。经过处理后得到 10 613 条样本数据,其中正常履约的样本 7 692 条,异常样本 2 921 条。为避免数据的极差异常对实验结果产生不利影响,利用 Z-score 方法对数据进行标准化处理。由于卷积神经网络 LeNet-5 只能处理图像数据,因此对平衡和标准化后的数据还需要进行成像处理。按照每个指标变量对应为一个像素点的原则,将 83 个指标变量根据反映客户特征信息的不同方面重组聚合形成 10×10 的指标特征矩阵,不足的像素区域用 0 替代。指标特征的取值对应乘以 255 得到对应像素点的亮度,然后通过 MATLAB2018 形成 32×32 像素的灰度图。图 3(a)为经过图像化处理的正常履约客户所对应的特征信息,图 3(b)为违约客户所对应的特征信息产生的灰度图,可以看出正负样本在某些区域存在明显差异,因此可以利用卷积神经网络进一步挖掘分析数据深层特征信息,提升信用评估模型的预测性能。此灰度图经过缩放处理。

(1) 基于信息值的特征选择处理

对于传统评估模型,由于仍带有部分共线性冗余特征变量,会对实验建模结果产生不可预测的影响,因此需要对初始变量进行筛选。

计算指标变量的 IV 值,按照 IV 值的大小进行特征变量的第一次筛选。当指标 IV 值大于等于 0.3 且

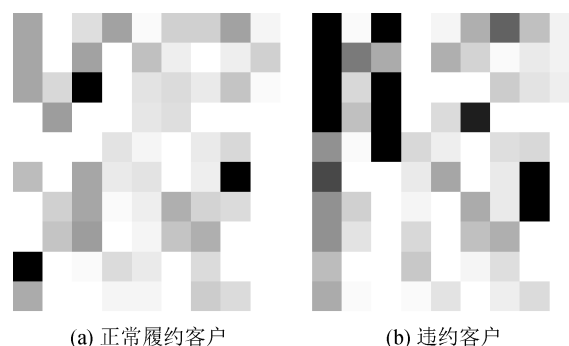


图 3 样本指标数据图像化

Fig.3 Graphical Sample Data

拟合度 R^2 大于 0.95 时,指标区分客户违约情况能力强,因此剔除 IV 值小于 0.3 的指标变量。

在上述步骤的基础上,对于筛选后的指标做进一步共线性检测,剔除变量间皮尔逊相关系数大于 0.7 的两个变量中的任意一个。

经过筛选后,共得到 26 个建模指标变量,变量及其含义如表 2 所示。

(2) 基于主成分分析的特征选择处理

对经过初步预处理的 83 个指标特征做相关性检验,剔除强相关性的指标特征中的任意一个,经过处理后得到 52 个变量。对剩余 52 个变量进行 KMO 检验和 Bartlett 球形检验,发现统计量为 0.709 (大于 0.5),且显著性小于 0.01,表明所采集的样本指标特征变量适合进行因子分析。

通常提取的主成分特征值应该大于 1,但是由于方差解释度太低,信息丢失较为严重,因此本文综合考虑以上因素提取 18 个主成分,累计贡献达到 64.9% (因子得分系数表限于篇幅保存在支撑数据中)。部分提取的主成分的表达式如式(5)所示。

$$\begin{aligned} F_1 &= 0.311V_1 + 0.169V_2 + 0.064V_3 + \dots + 0.530V_{52} \\ F_2 &= 0.369V_1 + 0.165V_2 + 0.051V_3 + \dots + 0.313V_{52} \\ &\dots \end{aligned} \quad (5)$$

$$F_{18} = -0.058V_1 + 0.075V_2 - 0.023V_3 + \dots + 0.042V_{52}$$

其中,变量 $V_i (i=1, 2, 3, \dots, 52)$ 为标准化后的指标特征变量。

对形成的样本数据进行训练集和测试集的划分时,为保证模型有足够的训练样本使得训练的卷积

^①<https://www.lendingclub.com/info/download-data.action>.

表2 指标变量体系
Table 2 Index Variable System

指标变量	变量名称	IV 值	含义
贷款信息	loan_amnt	0.561	借款人申请的贷款金额
	int_rate	0.724	贷款利率
偿付能力	annual_inc	0.560	借款人在注册期间自行报告的年收入
	dti	0.333	使用借款人的总债务偿还总额(不包括抵押贷款和要求的LC贷款)除以借款人自我报告的月收入计算的比值
	tot_cur_bal	0.555	所有账户的当前总余额
	il_util	0.685	所有固定账户的总流量余额/信用额度的比率
	max_bal_bc	0.710	所有周转账户的最大当前余额
	acc_open_past_24mths	0.488	过去24个月的消费额
	bc_open_to_buy	0.355	银行卡上可用于购买的资金
	mort_acc	0.314	抵押账户数量
	num_actv_rev_tl	0.560	当前活跃的循环交易数量
	num_bc_tl	0.333	银行卡账户数量
历史信贷	installment_feat	0.306	客户每月还款支出占月收入的比值
	open_acc	0.489	借款人信用档案中的未结信用额度
	all_util	0.534	所有交易均衡信贷限额
	total_bal_il	0.394	所有分期付款账户的当前总余额
	revol_bal	0.780	总信贷周转余额
	revol_util	0.564	循环利用率,或借款人相对于所有可用循环信贷使用的信贷额度
历史申请	pct_tl_nvr_dlq	0.489	从未拖欠交易百分比
	mo_sin_old_il_acct	0.577	自最早开立银行分期账户以来的月数
	mo_sin_old_rev_tl_op	0.441	自最早的循环账户开始以来的月数
	mo_sin_rcnt_rev_tl_op	0.349	自最近一次的循环账户开通以来的月数
	mo_sin_rcnt_tl	0.461	自最近一次开户以来的月数
	mths_since_recent_bc	0.561	自最近一次开立银行卡账户以来的月数
	mths_since_recent_inq	0.724	自最近的调查以来的月数
	mths_since_rcnt_il	0.604	自最近的分期付款账户开通以来的月数

(注:原始数据出于对贷款人个人信息的保护,未能获取借款人的个人基本情况数据,如性别、年龄等。)

神经网络模型性能达到最优,同时具有良好的稳定性,按照7:3的比例进行测试集和训练集的划分,即随机抽取30%的样本用于测试,70%用于模型训练。对基于卷积神经网络的模型,在70%的训练集中抽取15%作为验证集,剩余55%则作为最终模型的训练集,取5次实验结果平均值作为最终实验结果。

3.2 对比模型与评价指标

为更加全面地分析新模型的综合性能情况,选取在信用评估领域常用的BP神经网络、决策树、支持向量机、随机森林、Logistic回归等5种机器学习算法进行对比分析。对于基于传统机器学习方法的信用评估模型,需要对形成的模型进行参数调节,使

得模型更加稳定并保证实验结果的可靠性。基于此,结合常用的参数调优方法,利用基于Python的网格搜索法对部分关键参数进行寻优。在实验中,由于基于简单的参数寻优方法得到的参数会使得模型存在不同程度的过拟合现象,因此需要进行部分的参数修正以减小模型的过拟合问题。基于传统机器学习方法的模型部分关键参数的设置如表3所示。

由于模型应用场景和样本数据的不同,模型评价方法和指标也会存在差异。常用的模型评价指标有综合性的AUC值、正确率(Accuracy)、KS值和G-mean值等,评价对某一类样本敏感度的有灵敏度(Sensitivity)、特异度(Specificity)和查准率

表 3 机器学习模型参数设置(部分)

Table 3 Machine Learning Models' Parameter Settings

模型	参数或结构设置
LeNet-5	参考图 1 结构
BP 神经网络	采用传统三层 BP 神经网络, 参数设定参考文献[19]
决策树	max_depth: 7
支持向量机	kernel: rbf, c: 100, gamma: 0.01
随机森林	max_depth: 8, min_samples_leaf: 4
Logistic 回归	c: 0.1, penalty: l ₁

(Precision)等。为从多层次全面地评估模型性能情况,本文选取灵敏度、特异度和查准率作为第一层次细致性评估指标评估模型对不同样本的敏感程度,同时选取模型实验结果的 AUC 值、G-mean 值和 F1 值作为第二层次模型整体性能评估指标。

模型评估的混淆矩阵如表 4 所示,真正例指将正例预测为正例,假反例指将正例预测为反例,假正例指将反例预测为正例,真反例指将反例预测为反例。灵敏度、特异度、查准率、G-mean 值和 F1 值计算公式如公式(6)~公式(10)所示。评估指标 AUC 值为 ROC 曲线所覆盖曲线面积,AUC 值越大,分类器性能越好。

表 4 混淆矩阵

Table 4 Confusion Matrix

真实情况	预测结果	
	正例	反例
正例	真正例	假反例
(Positive)	(True Positive, TP)	(False Negative, FN)
反例	假正例	真反例
(Negative)	(False Positive, FP)	(True Negative, TN)

$$Sensitivity = Recall = \frac{TP}{TP + FN} \quad (6)$$

$$Specificity = \frac{TN}{TN + FP} \quad (7)$$

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$G - mean = \sqrt{Recall \times Specificity} \quad (9)$$

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (10)$$

3.3 模型实证结果及分析

将指标特征按照反映客户信息的 4 个不同方面

情况映射为灰度图,结合深度学习相关技术建立基于 LeNet-5 的信用评估模型,同时对比传统的信用评估模型构建方法。为更加全面评估对比不同模型的性能,从两个方面分别对不同模型构建评估体系。一方面,对模型的性能进行评估时,需要根据模型的实际应用需求进行不同角度的细致评价,对于本文研究的客户信用评估问题,由于将违约客户误判为信用良好的客户产生的成本要远高于将信用良好的客户误判为违约客户产生的成本,准确识别违约客户,及时采取相应的措施,尽可能减少违约损失具有更大的现实指导意义,因此选取有效区分正负样本的评估指标对模型进行第一层次的评估;另一方面,利用整体性能评估指标综合评估不同模型的性能,使得对模型的评估更加客观公正。

(1) 模型对正负样本区分度性能评估

6 种不同模型对正负样本的识别评估情况如表 5 所示。

表 5 模型正负样本识别性能

Table 5 Models' Recognition Performance of Positive and Negative Samples

特征处理方法	模型方法	灵敏度	特异度	查准率
基于信息价值特征处理	LeNet-5 ^①	0.687⁽¹⁾	0.994⁽¹⁾	0.998⁽¹⁾
	BP 神经网络	0.616	0.723	0.695
	决策树	0.625	0.706	0.679
	支持向量机	0.657⁽³⁾	0.751	0.724
	随机森林	0.663⁽²⁾	0.717	0.744
基于 PCA 特征处理	Logistic 回归	0.617	0.649	0.635
	BP 神经网络	0.648	0.889⁽³⁾	0.946⁽³⁾
	决策树	0.636	0.839	0.911
	支持向量机	0.657⁽³⁾	0.911⁽²⁾	0.956⁽²⁾
	随机森林	0.641	0.872	0.930
	Logistic 回归	0.648	0.889⁽³⁾	0.656

(注:得分后括号内数字为准确率排名前三的模型序号。)

在特异度表现上,即对负样本(信用差,出现违约情况的客户实例)识别情况,本文模型达到 0.994,而基于传统模型构建方法中表现最优的是以 PCA 对指标变量进行处理为基础,利用支持向量机构建的评估模型,特异度值为 0.911。其次是以 PCA 对指标变量进行处理为基础,结合传统 BP 神经网络的评

①基于卷积神经网络的评估模型的特征提取由模型自动完成。

估模型和结合 Logistic 回归构建的评估模型特异度值为 0.889。

在灵敏度表现上,即对正样本(信用良好,未出现违约情况的客户实例)的识别情况,本文模型表现最优,灵敏度值为 0.687,其次是以信息价值对特征处理,利用随机森林构建的信用评估模型为 0.663。对于模型识别出所有正样本的准确率,即查准率,本文模型预测值达到 0.998,对比传统评估模型,本文模型可以实现更高的预测查准率,更好的样本区分度,具有更高的实际应用价值。

(2) 模型综合性能评估

模型综合性能评估对比如表 6 所示。

表 6 模型综合性能评估对比
Table 6 Models' Performance Evaluation

特征处理方法	模型方法	G-mean	F1	AUC
基于信息价值 特征处理	LeNet-5	0.827⁽¹⁾	0.814⁽¹⁾	0.995⁽¹⁾
	BP 神经网络	0.668	0.653	0.715
	决策树	0.665	0.651	0.724
	支持向量机	0.703	0.689	0.668
	随机森林	0.705	0.701	0.778
	Logistic 回归	0.633	0.626	0.747
基于 PCA 特征处理	BP 神经网络	0.759⁽³⁾	0.769⁽³⁾	0.816
	决策树	0.730	0.749	0.837
	支持向量机	0.773⁽²⁾	0.778⁽²⁾	0.904⁽²⁾
	随机森林	0.748	0.759	0.865
	Logistic 回归	0.759⁽³⁾	0.652	0.880⁽³⁾

(注:得分后括号内数字为准确率排名前三的模型序号。)

由表 6 可知,在用于评估模型的综合性能的指标 G-mean、F1、AUC 值上,各种不同模型实现的精度都在 0.6 以上。其中基于改进卷积神经网络的客户信用评估模型在不同评估标准下,分别实现 F1 值 0.814、G-mean 值 0.827 和 AUC 值 0.995,明显优于其他 5 种常用的传统模型。可以发现,信用评估模型中,指标特征之间存在更多的是非线性相关关系,因此基于非线性的模型,在实验中往往能取得较高的准确率。进一步可以看出,由于基于信息价值的特征处理,简单地根据特征对目标变量的贡献度大小进行选择,虽降低了数据的冗余,一定程度上提高了模型运行效率,但是势必会丢失一些有价值的信息,使得模型准确率大打折扣;而基于主成分分析的特征处理方法则是将重复的变量(关系紧密的变量)删

除,建立较少的新变量,可以更有效地提取和保留有用的特征信息构建评估模型,因此基于 PCA 对特征变量进行处理后,建立的信用评估模型在实验中取得了较高的准确率。然而和基于信息价值的方法一样,基于 PCA 的特征处理方法也需要按照一定的标准(基于特征数或特征值)进行筛选和提取,因此在某种程度上也未能有效缓解特征信息丢失的问题。

利用卷积神经网络构建的信用评估模型,是通过将所有特征变量转换为灰度图,利用卷积神经网络对特征变量根据实际问题调整更新前馈神经网络的权重,自动提取更多有效信息来构建模型,从而实现信用评估建模过程的特征处理和模型训练的“统一”,不仅可以有效减少人为因素对实验的干扰,提高模型实证结果的可靠性和准确性,还可以提取更多有效的特征信息从而实现更高的准确率。对比传统的 5 种评估模型,由于支持向量机良好的数据适应性和对非线性特征的分析能力,使得模型无论是在 F1 值、G-mean 还是 AUC 值方面的表现都要优于其他 4 种模型。受限于数据源真实情况,正负样本比例处于不平衡状态,因此对于负样本的识别率要低于正样本的识别率,所以数据平衡度对实验结果的影响还需进一步研究。

4 结 语

依据客户信用数据的特征,将所对应的特征转换成相应的灰度图,利用卷积神经网络能够更好地提取图片特征而优化模型评估性能,将经典的 LeNet-5 网络做部分修改应用于客户信用评估,并依据历史样本数据形成的测试集对几种常见基于机器学习算法的客户信用评估模型性能进行对比分析。实证结果表明,本文模型与传统的特征处理方式相比能充分利用指标特征实现最大化有效信息来构建模型,实现了建模过程的“统一”,减少了不必要的人为干预,使得模型无论在整体性能表现上,还是对某一类样本的敏感度上都具有良好的模拟效果。同时对比分析了两种不同特征处理方法对实验结果的影响,发现通过 PCA 对特征进行处理,可以最大限度利用和保留特征信息,使得构建的模型比以简单的信息价值方法对特征进行处理为基础的模型具有更好的预测性能。此外本文提出的建模思想,即对指

标变量进行成像处理然后建立基于卷积神经网络的评估模型,进一步拓展了模型建立与样本数据利用的思路和方法。

此外,一方面由于国内相关信贷公司的数据不予公开、不易获得,同时不同平台的数据结构及数据正负样本比例也存在一定的差异性;另一方面信贷也会受其他因素变量(如宏观经济水平等)的影响,因此在未来研究中需要对数据结构情况和评估模型做进一步拓展分析。

参考文献:

- [1] 蒋辉,马超群,许旭庆,等.仿EM的多变量缺失数据填补算法及其在信用评估中的应用[J].中国管理科学,2019,27(3):11-19. (Jiang Hui, Ma Chaoqun, Xu Xuqing, et al. An EM-similar Imputation Algorithm for Multivariable Data Missing and Its Application in Credit Scoring[J]. Chinese Journal of Management Science, 2019, 27(3): 11-19.)
- [2] 肖进,刘敦虎,顾新,等.银行客户信用评估动态分类器集成选择模型[J].管理科学学报,2015,18(3):114-126.(Xiao Jin, Liu Dunhu, Gu Xin, et al. Dynamic Classifier Ensemble Selection Model for Bank Customer's Credit Scoring[J]. Journal of Management Sciences in China, 2015, 18(3): 114-126.)
- [3] Altman E I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy[J]. The Journal of Finance, 1968, 23(4): 589-609.
- [4] 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): 757-770.
- [5] 吴冲,夏晗.基于支持向量机集成的电子商务环境下客户信用评估模型研究[J].中国管理科学,2008,16(S1):362-367.(Wu Chong, Xia Han. Study of Customer Credit Evaluation Under E-commerce Based on Support Vector Machine Ensemble[J]. Chinese Journal of Management Science, 2008, 16(S1): 362-367.)
- [6] Blanco A, Pino-Mejías R, Lara J, et al. Credit Scoring Models for the Microfinance Industry Using Neural Networks: Evidence from Peru[J]. Expert Systems with Applications, 2013, 40(1): 356-364.
- [7] Chen F L, Li F C. Combination of Feature Selection Approaches with SVM in Credit Scoring[J]. Expert Systems with Applications, 2010, 37(7): 4902-4909.
- [8] 熊志斌,吴维焯.基于深度信念网络的信用评估研究[J].科研信息化技术与应用,2019,10(3):28-36.(Xiong Zhibing, Wu Weiye. Credit Evaluation Research Based on Deep Belief Networks[J]. E-science Technology & Application, 2019, 10(3): 28-36.)
- [9] 吴星泽.财务危机预警研究:存在问题与框架重构[J].会计研究,2011(2):59-65,97. (Wu Xingze. Problems on Research of Predicting Financial Distress and Framework Reconstructure[J]. Accounting Research, 2011(2): 59-65, 97.)
- [10] He K M, Zhang X Y, Ren S Q, et al. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification[C]// Proceedings of the 2015 IEEE International Conference on Computer Vision. 2015: 1026-1034.
- [11] Sun S N, Zhang B B, Xie L, et al. An Unsupervised Deep Domain Adaptation Approach for Robust Speech Recognition[J]. Neurocomputing, 2017, 257: 79-87.
- [12] Williamson D S, Wang D L. Time-frequency Masking in the Complex Domain for Speech Dereverberation and Denoising[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2017, 25(7): 1492-1501.
- [13] Zhang Y, Marshall I, Wallace B C. Rationale-augmented Convolutional Neural Networks for Text Classification[C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016: 795-804.
- [14] 李慧,柴亚青.基于卷积神经网络的细粒度情感分析方法[J].数据分析与知识发现,2019,3(1):95-103.(Li Hui, Chai Yaqing. Fine-Grained Sentiment Analysis Based on Convolutional Neural Network[J]. Data Analysis and Knowledge Discovery, 2019, 3(1): 95-103.)
- [15] Hosaka T. Bankruptcy Prediction Using Imaged Financial Ratios and Convolutional Neural Networks[J]. Expert Systems with Applications, 2019, 117:287-299.
- [16] Lencun Y, Bottou L, Bengio Y, et al. Gradient-Based Learning Applied to Document Recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [17] 阚士行.商业银行信用评级筛选财务指标方法效果对比与校验[D].济南:山东大学,2010.(Kan Shixing. Performance Comparison of Several Methods for Selecting Indices of Commercial Bank Credit Ranking[D]. Ji'nan: Shandong University, 2010.)
- [18] 刘丹,李战江,郑喜喜.基于WOE-Probit逐步回归的信用指标组合筛选模型及应用[J].数学的实践与认识,2018,48(2):76-87.(Liu Dan, Li Zhanjiang, Zheng Xixi. Selection Model of Credit Index Combination Based on WOE-Probit Stepwise Regression and Its Application[J]. Mathematics in Practice and Theory, 2018, 48(2): 76-87.)
- [19] 杨淑娥,黄礼.基于BP神经网络的上市公司财务预警模型[J].系统工程理论与实践,2005,25(1):12-18,26.(Yang Shu'e, Huang Li. Financial Crisis Warning Model Based on BP Neural Network[J]. Systems Engineering-Theory & Practice, 2005, 25(1): 12-18, 26.)

作者贡献声明:

刘伟江,魏海:提出研究思路,设计研究方案;
魏海,运天鹤:数据采集、清洗,程序设计,程序测试,论文起草;
刘伟江,魏海:论文修改和最终版本修订。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据由作者自存储,E-mail:weihai94@163.com。

- [1] 魏海.Lending Club.xlsx. 初步清洗后数据集.
- [2] 运天鹤.data_IV.xlsx. 经过筛选后的变量及IV值.
- [3] 运天鹤.factor.xlsx. 因子得分系数表.
- [4] 运天鹤.loan3.csv. 筛选后的数据集.
- [5] 魏海.X_test.csv. 传统方法测试集.
- [6] 魏海.X_train.csv. 传统方法训练集.
- [7] 魏海.new_train.csv. 新方法的训练集灰度图索引.
- [8] 魏海.new_test.csv. 新方法的测试集灰度图索引.

收稿日期:2019-11-29

收修改稿日期:2020-02-13

Evaluation Model for Customer Credits Based on Convolutional Neural Network

Liu Weijiang^{1,2} Wei Hai² Yun Tianhe²

¹(Center for Quantitative Economics, Jilin University, Changchun 130012, China)

²(Businesses School, Jilin University, Changchun 130012, China)

Abstract: [Objective] This paper analyzes customer loan information, and extracts their characteristics, aiming to more effectively predict customer defaults of online loans. [Methods] First, we collected customer credit data from Lending Club. Then, we integrated the characteristic variables from four aspects of customer information and created a grayscale map. Finally, we established a customer credit evaluation model based on convolutional neural networks. [Results] The proposed model had specificity of 99.4%, sensitivity of 68.7%, G-mean value of 82.7%, F1 value of 81.4% and AUC value of 99.5%. The performance of our new model was much better than those credit models based on feature processing. [Limitations] We only investigated the performance of a few models. More research is needed to study the impacts of unbalanced data. [Conclusions] The proposed model effectively predicts probability of customer defaults.

Keywords: Convolutional Neural Networks Indicator Imaging Credit Evaluation Information Value PCA