

基于机器学习的小微企业信贷供给智能决策系统研究 ——以“银税互动”政策的共享纳税数据为例

李超杰

(南京江宁经济技术开发区管理委员会, 江苏 南京 211100)

摘要:随着金融科技快速发展,如何融合现代信贷技术有效缓解小微企业的融资困境是当前金融市场亟待解决的关键问题。以“银税互动”政策作为切入点,深入挖掘小微企业纳税数据的信用特征,结合随机森林方法和深度神经网络方法对小微企业的信用违约概率和信贷风险等级进行评估,综合考虑银行的经营利润以及风险成本实现不同小微企业信贷利率和信贷额度的优化,以此构建小微企业信贷供给智能决策系统。研究表明,纳税数据能够准确识别小微企业的信贷风险,从而为银行信贷供给提供决策依据,并且纳税数据质量较高的小微企业无论是信贷额度还是信贷利率都能够获得银行层面更加有利的信贷支持。

关键词:小微企业;信贷供给;机器学习;银税互动;金融科技

中图分类号:F830.56

文献标志码:A

文章编号:1671-4970(2022)04-0066-09

一、引言

小微企业是我国国民经济和社会发展过程中不可或缺的重要力量。中国人民银行行长易纲在2018年《关于改善小微企业金融服务的几个视角》报告中指出,我国小微企业最终产品和服务价值占GDP的比例约为60%,纳税占国家税收总额的比例超过50%,提供70%以上的技术创新和城镇劳动就业贡献。总而言之,小微企业在扩大就业、促进稳定、增加收入和国家税收等方面均具有举足轻重的作用。然而,近年来在全球宏观经济下行的背景下,尤其是面对中美贸易摩擦、新冠肺炎疫情冲击、消费

增速放缓以及流动性趋于短缺等各方面因素的叠加影响,我国小微企业的利润增速严重下滑。艾瑞咨询发布的2021年《中国中小微企业融资发展报告》指出,我国2020年小微企业利润下滑同比超过30%。在这样艰难的处境中,小微企业融资难的问题更加突出。在当前严峻的市场环境下,尽管国家高度重视小微企业的生存和发展,相关部门陆续出台了多项促进小微企业融资的相关政策措施,融资环境有所改善,但小微企业获得的金融支持仍然存在不足。因此,如何进一步缓解小微企业的融资难题仍然值得深入探究。

小微企业融资难题的关键在于小微企业和银行

基金项目:江苏省社会科学基金(18GLA004)

作者简介:李超杰(1975—),男,高级经济师,博士,主要从事金融科技研究。E-mail: lichaojiephd@aliyun.com

之间的信息不对称以及缺乏合格的抵押担保等问题。近年来以人工智能和大数据技术为依托的金融科技快速发展,在颠覆传统金融服务方式的同时也为进一步缓解小微企业融资难题开辟了新的路径。金融科技使得对于小微企业细化粒度数据的充分收集成为现实,克服了银行与小微企业之间的信息不对称问题。与此同时,依据相关数据信息可以准确评估小微企业的信用状况从而实现无抵押担保的信用贷款。通过以上过程,金融科技能够全面提高小微企业的融资可得性。在实践应用层面,2015年国家税务总局提出的“银税互动”政策,试图应用金融科技缓解小微企业融资难题。“银税互动”政策允许银行获取沉淀在税务部门的企业纳税数据,达到减少小微企业与银行之间信息不对称的目的,进而帮助小微企业将纳税数据转化为融资信用,实现小微企业的无抵押担保的信用贷款。

综上所述,本文将探究如何充分利用“银税互动”政策中有关小微企业的纳税数据,从数据互联共享的角度提供通过金融科技缓解小微企业融资难题的理论框架。基于这项研究,一方面可以为消除银行和小微企业之间的信息不对称问题提供技术支撑,有效降低银行的信贷成本;另一方面能够为开展小微企业信用贷款提供可行方案,有效提高企业的信贷可得性,对于解决小微企业融资难题具有重要的理论参考价值。在此基础上,通过快速有效的小微企业信贷智能决策,进一步推动“银税互动”政策的落实。这对于提高银行的融资服务质量以及小微企业的依法纳税意识具有重要实践意义。

二、文献综述

从各国小微企业融资实践情况来看,银行信贷是小微企业最主要的外部融资渠道^[1]。然而,在传统贷款方式下银行能够从小微企业获得的硬信息相对缺乏,银行主要基于以人工调查为主获得的软信息并且通过人工审核方式进行小微企业信贷决策^[2]。受限于上述信息获取和信贷决策的高成本,银行难以实现大规模、大范围地推广以软信息为决策依据的小微企业关系型信用贷款^[3]。上述条件约束导致小微企业参与贷款市场的门槛较高,成为阻碍解决小微企业融资问题的关键。随着人工智能和大数据等技术的发展,银行纷纷通过金融科技提升获取信息的广度和深度,从而为缓解小微企业信贷供给提供了新的实

现路径^[4]。房颖从信息约束与抵押要求的角度分析了金融科技对小微企业信贷供给的赋能驱动机制^[5]。董晓林等基于疫情冲击视角进行研究,发现金融科技能够通过缓解融资约束帮助小微企业度过危机^[6]。郭丽虹等从普惠金融的角度出发,发现金融科技在支持小微企业普惠贷款方面作用明显^[7]。周光友等从P2P网贷的角度,论证了金融科技通过创造效应、信息效应和普惠效应为小微企业融资提供了重要渠道^[8]。通过以上论述,基于实际市场数据的大量研究从不同角度论证了金融科技对于缓解小微企业信贷供给的优势和促进作用^[9]。

在此基础上,李学峰等发现金融科技加剧了银行业的马太效应,即金融科技扩张了大型银行的小微信贷资产,同时对其他银行产生了一定的负面影响^[10]。金洪飞等关于银行规模异质性的研究结果也表明,金融科技降低了银行与小微企业之间的信息不对称程度,尤其克服了大型银行在获取软信息能力方面的不足,更加有利于大型银行提升小微企业信贷供给能力,并且会对其他银行产生挤出效应^[11]。同样地,盛天翔等发现金融科技能够促进信息获取、处理和传递,有利于升级大型银行的传统贷款技术从而提高小微企业的信贷可得性^[12]。关于银行对小微企业信贷供给存在异质性的原因,张一林等认为大银行在研发应用数字贷款技术上占据优势,能够服务有较多“数字足迹”的小微企业^[2]。而运用数字贷款技术以后,大型银行在为小微企业大批量地提供综合性产品和服务方面更加具有比较优势^[13]。由此可见,能否有效运用数字贷款技术降低银行与小微企业之间信息不对称程度,是影响银行运用金融科技缓解小微企业信贷问题的关键。

关于金融科技在银行信贷决策方面最为关键的数字技术应用是通过大数据和人工智能等方法实现信贷风险控制。从大数据的角度来说,邱志刚等通过理论研究发现基于大数据的信贷模式能够收集利用传统银行信贷模式中无法获取的信息,从而改善银行的信贷风险控制能力^[14]。从人工智能的角度来说,黄益平等基于信贷主体的基本特征和行为信息进行数据挖掘,通过逻辑回归、随机森林、支持向量机等机器学习算法进行信贷违约预测并以此作为信贷决策依据^[3]。目前许多银行正在运用大数据技术提升获取信息能力,进而通过人工智能实现低成本的自动化风险控制^[15]。然而,迄今为止小微企业能够获取和

利用的相关数据信息仍然相对有限,现有关于研究关于金融科技在银行信贷决策方面的应用仍然以针对个人层面的数据信息为主^[16-17]。因此,从小微企业层面来说,如何能够更加有效地运用金融科技实现信贷风险控制和相应的信贷供给决策,首先需要针对数据信息的收集利用进行更加深入的分析。

关于如何充分利用小微企业的现有数据信息进行信贷风险评估,不少学者展开了丰富的研究。张发明等提出一种基于改进动态组合评价方法的小微企业信用决策模型,并以银行的调查数据为基础对11家制造业小微企业的信贷风险评价进行了实证研究^[18]。王钊等以网页爬取的P2P平台数据信息为基础,基于混合生存分析研究了小微企业的动态信贷风险评估方法^[19]。张榕薇等以源于“央行浙江数字化平台”的首次贷款数据结合问卷方法,实证分析了信用信息共享对小微企业首次贷款可得性的影响^[20]。陈春照等以小微企业的纳税信息为样本,对银行信贷供给决策和信贷风险评估进行了研究^[21]。近年来,基于数据互联共享的“银税互动”政策也为此提供了丰富的数据信息资源。杨龙见等针对“银税互动”政策实施前后的市场数据进行回归,结果表明“以税增信”的方式能够有效降低信息不对称和交易成本,从而显著提高了小微企业信用贷款的可得性与贷款额度^[22]。陈彪等的实证研究结果同样表明,“银税互动”可以消除小微企业的融资约束^[23]。尽管如此,现有研究并没有基于数据挖掘和机器学习等技术,总结能够广泛运用小微企业纳税数据信息进行信贷供给决策的理论框架,对于如何有效利用“银税互动”共享纳税数据进行小微企业信贷供给决策目前尚且缺乏可行的实现方案。

综上所述,本文将“银税互动”政策作为切入点,构建针对小微企业的信贷供给智能决策系统,以此促进银行对纳税信用等级较高的小微企业进行快速有效的信贷支持,同时引导小微企业提高纳税意识和纳税信用。具体来说,重点对沉淀在税务部门的小微企业纳税流水大数据信息进行特征挖掘,通过机器学习算法预测小微企业的信贷风险,进而对小微企业的信贷供给决策进行组合优化,化解小微企业难以均衡的融资成本与风险控制困境。基于以上研究过程,本文的主要贡献包含以下几个方面:首先,针对银行与小微企业之间信息不对称问题,从技术层面探究了如何应用小微企业纳税数据信息进行

信贷供给决策的可行方案,研究结果对运用基于“银税互动”政策的数据来源解决小微企业融资难题的可行性进行了论证。其次,关于银行如何对小微企业无抵押担保的信用贷款进行风险控制问题,本文基于纳税数据信息融合机器学习方法对小微企业的信贷风险进行精准评估,研究结果为如何快速有效进行小微企业信贷供给智能决策提供了理论参考。

三、小微企业信贷供给智能决策系统构建

小微企业信贷供给智能决策系统主要取决于信用风险的评估,首先需要针对小微企业纳税流水的基本信用特征确定信贷风险评估指标,在此基础上通过随机森林方法对初始指标进行特征筛选和降维,进而根据深度神经网络预测小微企业的信贷风险,以此对小微企业信贷供给决策进行组合优化,通过以上过程形成小微企业信贷供给智能决策系统,并且以“银税互动”政策的共享纳税数据为例进行模型的应用研究。

1. 基于纳税流水的信贷风险评估指标

为了评估小微企业的信贷风险,首先要明确其纳税流水所包含的基本信用特征。国际上的信贷风险评估通常围绕5C要素展开,但是针对我国实际情况通过“五性”分析更加能够体现小微企业的信用特征。因此,本文围绕小微企业的纳税流水信息进行特征挖掘,根据信用状况的安全性、收益性、成长性、流动性和生产性分别构建小微企业信贷风险评估指标如下:

从小微企业安全性的角度来说,无论是从生产端还是销售端,如果关于上下游企业的交易对手因故发生退货退款等问题,均会对企业正常经营造成一定影响。因此,本文一方面考虑进项和销项中已入账的负数发票数量占比以及金额占比,另一方面考虑作废发票数量占比和金额占比,据此作为度量小微企业安全性的特征。

小微企业的收益性可以直接通过其资金流入和流出的状况进行确定,考虑到不同企业的规模异质性特征,定义销项发票和进项发票的差额为总收益,总收益与进项发票总金额之比为该企业总收益率。同时考虑不同企业的周期性经营特征,进一步对企业收益率进行时间标准化处理,通过企业收益率与企业经营期限的比值将其转化为年均收益率,以此作为度量小微企业收益性的特征。

小微企业流动性特征主要刻画其经营过程中可以用于支配的现金资产。频繁的现金交易通常意味着相对较高的资金周转率,即企业能够及时获取正常经营所需的流动性资金。因此,进项和销项发票的数量可以作为度量小微企业流动特征的指标。此外,如果流入资金大量滞后于流出资金,企业会存在严重的现金缺口从而影响正常经营,因此资金回收期限也是度量小微企业流动性的重要特征,本文以经过时间加权的企业进项、销项发票金额之差进行定义。

对于小微企业的生产性的刻画需要尽可能考虑各种形式的资产,不仅包括直接用于经营的现金资产,还需要考虑上下游的客户资源等无形资产。因此,本文根据小微企业的相关交易信息以进项、销项发票的总金额作为度量企业现金资产的生产性特征,同时以企业上下游客户的数量作为度量企业无形资产的生产性特征。

小微企业的成长性与其生产性密不可分,为了在刻画企业生产性特征的基础上进一步体现企业的成长性特征,本文对所有刻画企业生产性的特征以相邻两年的差值为基础定义企业成长性特征,即分别以企业进项、销项发票金额变动比例的均值定义现金资产成长性,同时以上下游客户数量变动比例的均值定义企业无形资产的成长性,以此作为度量企业成长性的特征。

此外,由于样本包含不同经营类型的小微企业,考虑到不同行业小微企业存在异质性,本文进一步引入了小微企业的进项发票税率和销项发票税率,以此区分不同行业的小微企业可能存在的不同差异。

2. 基于随机森林的特征筛选和降维

针对小微企业纳税流水的初始指标进行数据处理所形成的特征维度相对较高,如果以此进行机器学习评估小微企业的信贷风险,部分特征之间难免会存在线性相关的冗余变量,如此会因为过拟合现象而影响信贷风险评估质量。因此,本文首先对小微企业纳税流水的高维特征进行筛选和降维,在此基础上再通过机器学习方法进行信贷风险评估。特征筛选方法众多,大量理论和实证研究证明随机森林方法对于样本的挖掘深度要高于传统模型,并且随机森林方法对异常值和噪声具有较高的容忍度,因此得到的模型预测效果相对更好^[24]。故而本文选择随机森林算法进行特征的筛选和降维。

为了解决随机森林造成的过度拟合该问题,本

文通过 Bootstrap 抽样方法作为随机森林的训练集。同时,为了进一步体现特征的筛选过程,本文通过袋外数据误差对不同特征的重要程度进行排序并据此确定最优特征数目,具体步骤如下:每次建立决策树时,通过重复抽样得到用于训练决策树的数据集,通过没有参与决策树建立的剩余数据计算模型的预测错误率,即袋外数据均方误差。在此基础上,随机对袋外数据所有样本的某个特征加入噪声干扰,再次计算袋外数据均方误差,则该特征的重要程度可以通过前后两次袋外数据均方误差的变化幅度进行刻画。计算每个特征的重要程度并依据特征重要程度剔除相应比例的特征得到新特征集。针对新特征集重复上述过程并选择袋外误差最低的特征集,以此实现初始指标的特征筛选和降维过程。

3. 基于深度神经网络的信贷风险预测

银行对于小微企业的信贷供给决策包含信贷额度和信贷利率两个维度:首先需要通过对小微企业的信贷风险预测确定该企业是否存在违约可能,据此决定是否对其进行金融支持的信贷额度;其次需要进一步评估小微企业的信贷风险等级,以此对小微企业进行类别划分,针对具有不同风险水平的小微企业提供与之相匹配的信贷利率决策。

传统的银行信贷风险预测主要依赖于评分卡模型,与此相比机器学习模型更能刻画复杂的非线性关系以及解释变量之间的交互作用,对信贷违约预测也更加准确稳健^[3]。因此,本文通过神经网络模型对小微企业信贷风险进行预测分类,并且为了提高结果的准确性叠加了三层全连接网络作为深度神经网络的隐含层,并通过反向传播算法进行深度神经网络的学习训练。神经网络模型具体算法过程如下:以通过随机森林方法降维的小微企业信用特征作为深度神经网络的模型输入。假设模型第 l 层共有 n 个神经元,第 $l-1$ 层共有 m 个神经元,由第 l 层的线性系数 w 组成的矩阵表示为 $\mathbf{W}_{n \times m, l}$,第 l 层的偏置项 b 组成的向量表示为 $\mathbf{B}_{n \times 1, l}$,第 $l-1$ 层的输出 a 组成的向量表示为 $\mathbf{A}_{m \times 1, l-1}$,第 l 层的未激活前的线性输出 z 组成的向量表示为 $\mathbf{Z}_{n \times 1, l}$,第 l 层的输出 a 组成的向量表示为 $\mathbf{A}_{n \times 1, l}$,则第 l 层的输出表示为

$$\mathbf{A}_{n \times 1, l} = \sigma(\mathbf{Z}_{n \times 1, l}) = \sigma(\mathbf{W}_{n \times m, l} \mathbf{A}_{m \times 1, l-1} + \mathbf{B}_{n \times 1, l}) \quad (1)$$

式中 $\sigma(\mathbf{Z}_{n \times 1, l})$ 为激活函数。进一步以均方误差(MSE)来度量训练样本输出和真实训练样本输出之间的损失,假如模型第 $l+1$ 层共有 k 个神经元,则

基于梯度下降算法迭代得到梯度的递推关系为

$$\Delta_{n \times 1, l} = (\mathbf{W}_{k \times n, l+1})^T \Delta_{k \times 1, l+1} \odot \sigma'(\mathbf{Z}_{n \times 1, l}) \quad (2)$$

式中符号 \odot 为Hadamard积。根据上述递推公式,随机初始化各隐藏层与输出层的系数矩阵 \mathbf{W} 和偏置项 \mathbf{B} ,可以通过损失函数计算输出层的 Δ ,据此通过反向传播算法可以计算第 l 层的梯度 $\Delta_{n \times 1, l}$ 。此时,根据 $\Delta_{n \times 1, l}$ 更新第 l 层的 $\mathbf{W}_{n \times m, l}$ 和 $\mathbf{B}_{n \times 1, l}$,即

$$\mathbf{W}_{n \times m, l} = \mathbf{W}_{n \times m, l} - \alpha \sum_{i=1}^s \Delta_{n \times 1, i, l} (\mathbf{A}_{m \times 1, i, l-1})^T \quad (3)$$

$$\mathbf{B}_{n \times 1, l} = \mathbf{B}_{n \times 1, l} - \alpha \sum_{i=1}^s \Delta_{n \times 1, i, l} \quad (4)$$

式中: α 为迭代步长; s 为训练样本数量。当所有 \mathbf{W} 和 \mathbf{B} 的变化值都小于停止迭代的阈值时,得到各隐藏层与输出层的线性关系系数矩阵 \mathbf{W} 和偏置项 \mathbf{B} ,据此可以实现样本的预测分类。

4. 银行信贷供给决策的组合优化

由于银行的信贷额度相对有限,在进行小微企业信贷供给决策时通常需要对整体信贷收益和信贷风险进行全局规划,从而尽可能避免因为小微企业信贷违约所带来的损失,同时保证信贷业务的经营利润。为了协调小微企业信贷收益和信贷风险之间的平衡,本文根据对小微企业信贷风险的预测分类结果进一步实施具有差异化的信贷利率政策,在成本相加法的基础上考虑银行经营利润最大化对小微企业信贷利率进行定价^[25],即

$$\text{信贷利率} = \text{资金成本} + \text{目标利润} + \text{风险成本} + \text{调整项} \quad (5)$$

其中调整项主要取决于银行经营利润的最大化决策。从小微企业的角度来说,信贷利率的分级政策可能会影响其信贷需求,进而造成部分客户流失,如此反向影响银行的整体信贷供给决策,故而需要进一步刻画信贷利率差异对具有不同风险特征小微企业造成的信贷流失率。根据实际市场数据进行拟合可以得到不同风险等级小微企业信贷流失率与信贷利率之间的关系。假设上述关系以函数 $f(\cdot)$ 表示,则以银行经营利润最大化为目标可以确定不同小微企业的信贷利率,即

$$\max \{ r_k [1 - f_k(r_k)] I \} \quad (6)$$

式中: $k \in \{A, B, C, D\}$ 表示银行信贷风险等级;示性函数 I 用于刻画小微企业的信贷可得性,根据市场规则,信贷风险等级为D的小微企业通常无法获得信贷额度。

在此基础上从全局视角出发,综合考虑银行的信贷风险和信贷收益作为目标函数,以银行的信贷额度作为决策变量进行组合优化,并通过粒子群算法对银行的信贷决策进行求解。假设银行对样本 s 中小微企业 i 的信贷额度为 x_i ,对应的贷款利率为 r_i ,根据小微企业纳税流水进行预测得到其信贷违约概率为 p_i ,小微企业信贷利率通过基准利率、目标利润、风险成本和调整项共同确定,则银行对小微企业的信贷供给决策组合优化模型为

$$\max \left\{ \sum_{i=1}^s [(1 - p_i)r_i - p_i][1 - f_k(r_i)]x_i I \right\} \quad (7)$$

$$\text{s. t. } \sum_{i=1}^s x_i [1 - f_k(r_i)] \leq 1 \quad (0 \leq x_i \leq 1) \quad (8)$$

通过粒子群算法对上述组合优化模型进行求解可以得到银行对不同小微企业的信贷额度。

四、基于纳税数据的信贷供给智能决策结果分析

1. 样本描述

以“银税互动”政策的共享纳税数据为例对小微企业信贷供给智能决策系统进行应用分析,数据来源为2020年全国大学生数学建模竞赛提供的123家小微企业210947条进项和销项发票记录。通过剔除具有明显问题的数据信息,本文最终选择110家小微企业2016—2020年所有进项发票和销项发票作为纳税流水样本。上述样本共涵盖了除个体经营以外涉及农业、制造业、建筑业等各种类型的小微企业。图1给出了该样本中所有小微企业营业收入的具体统计特征。根据国家统计局发布的《统计上大中小微型企业划分办法(2017)》,对于样本所涉及行业中的大型企业营业收入划分均在10亿元以上,根据图1可以看出大部分小微企业营业收入

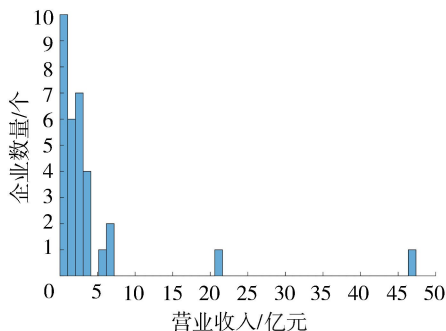


图1 小微企业营业收入分布

入均在 10 亿元以下,而且样本中营业收入超过 10 亿元的两家小微企业营业同样不符合对应行业的大型企业划分标准。因此,营业收入的统计特征表明样本均为小微企业。此外,样本所有小微企业中没有违约记录的为 87 家,另外 23 家小微企业均具有违约记录。实验过程中以 80 个样本数据作为训练集,其余 30 个样本数据作为测试集,以此验证实验结果的稳健性。经过对样本的整理和计算,得到各类变量特征的统计性描述如表 1。值得注意的是,表 1 中小微企业的资金回收期限均值为负数,说明小微企业整体存在一定程度的现金缺口。

表 1 基于纳税流水的信用特征统计性描述

信用特征	变量	均值	标准差	最大值	最小值	样本数
安全性	负数发票数量占比	0.02	0.02	0.14	0.00	110
	负数发票金额占比	0.04	0.15	1.47	0.00	110
	作废发票数量占比	0.07	0.05	0.36	0.00	110
	作废发票金额占比	0.11	0.13	0.82	0.00	110
收益性	总收益/元	3.64×10^7	2.79×10^8	1.89×10^9	-1.94×10^9	110
	总收益率	0.10	0.23	1.11	-0.01	110
	年均收益率	0.01	0.03	0.14	-0.00	110
流动性	进项发票数量/张	1780	4130	31400	2.00	110
	销项发票数量/张	1330	2950	23700	4.00	110
	资金回收期限/元	-7.33	161.23	516.98	-470.58	110
生产性	进项发票金额/元	9.42×10^7	6.33×10^8	6.64×10^9	560.00	110
	销项发票金额/元	1.31×10^8	4.99×10^8	4.70×10^9	3.67×10^4	110
	上游客户数量/个	261.71	488.28	3640	2.00	110
	下游客户数量/个	242.28	872.14	7980	1.00	110
成长性	进项发票金额变动比例	0.09	0.19	0.50	-0.49	110
	销项发票金额变动比例	0.13	0.14	0.50	-0.27	110
	上游客户数量变动比例	0.08	0.16	0.48	-0.33	110
	下游客户数量变动比例	0.05	0.20	0.50	-0.5	110
行业类别	进项发票税率	0.09	0.02	0.14	0.02	110
	销项发票税率	0.08	0.05	0.14	0.00	110

2. 信用特征筛选

通过随机森林方法对表 1 中刻画小微企业信用特征的所有变量进行筛选和降维,以均方误差作为损失函数,图 2 给出了袋外数据(OBB)均方误差与树的数量之间的关系。为了进一步区分不同特征的重要程度,图 3 根据由置换输入变量而增加的均方误差与每个变量标准差的比值对所有变量进行了排序,其中比值越大表示该变量的重要程度越高。根据图 3 的结果对所有初始变量进行筛选降维,设置变量重要程度的最低阈值为 0.5,则通过筛选得到的重要变量分别为销项发票金额、总收益以及销项发票数量,因此小微企业销项发票的数量、金额以及总收益对于预测其信用违约具有更加重要的参考价值。经过特征筛选之后,图 2 进一步给出了仅以降维变量作为样本特征的预测结果,可以看出模型的训练过程基本与以全部变量作为特征相比基本一致。

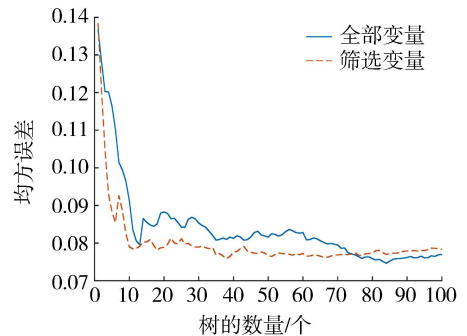


图 2 特征筛选的表现情况

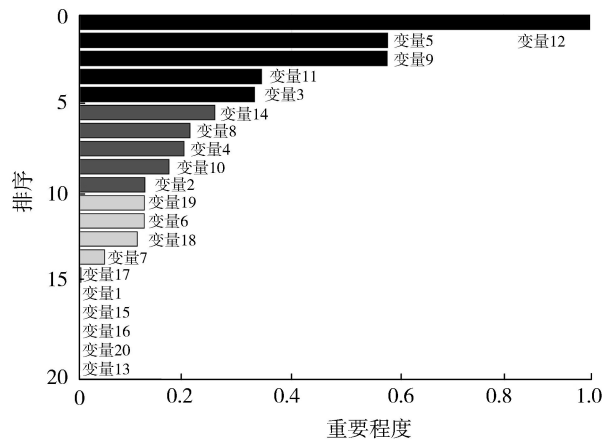


图 3 初始变量重要程度排序

为了进一步比较特征筛选的有效性,表 2 从准确率、精确率、召回率和平衡 F 分数几个方面比较了分别以全部变量和降维变量作为特征的小微企业信贷违约风险预测结果。可以看出虽然经过特征筛

表 2 信贷违约风险预测效果对比

变量	准确率	精确率	召回率	平衡 F 分数
全部变量	0.96	1.00	0.83	0.90
降维变量	0.94	0.90	0.78	0.84

选之后变量维度大幅降低,但是模型的预测效果并没有过多损失,说明特征筛选后得到的变量具有较好的代表性。因此,本文将降维后的变量作为系统输入进行小微企业信贷风险的预测。表 3 根据小微企业初始信贷违约风险等级对小微企业的信贷违约预测结果进行了统计,可以看出随着初始信贷违约风险等级增加,基于随机森林模型预测得到的小微企业信贷违约概率也逐渐增加,说明模型预测结果是有效的。由于存在严重违约风险的小微企业无法获得信贷额度,本文以随机森林方法对小微企业信贷违约概率进行预测之后,将出现违约事件的小微企业信用风险等级确定为 D,以此结合小微企业信用风险等级的预测结果确定最终的小微企业信用等级。

表 3 信贷违约预测结果的统计特征

风险等级	均值	标准差	最大值	最小值	数量
A	0.04	0.05	0.17	0.00	25
B	0.07	0.15	0.62	0.00	36
C	0.12	0.15	0.51	0.00	27
D	0.75	0.25	0.97	0.17	22
总计	0.21	0.31	0.97	0.00	110

3. 信贷风险评估

特征筛选一方面降低了样本的变量维度,另一方面排除了部分存在严重信贷违约风险的小微企业。在此基础上,不同小微企业信贷违约风险仍然存在差别。为了对具有不同信贷违约风险的小微企业提供差异化的信贷利率政策,本文以具有初始信贷评级的小微企业作为样本进行训练,通过深度神经网络进行学习并预测小微企业的风险等级。基于上述过程,对于没有任何信贷记录的小微企业来说,可以通过其纳税流水信息快速有效地实现信贷风险等级预测。如前文所述,以 80 个样本数据作为训练集,其余 30 个样本数据作为测试集,经过特征筛选的降维变量作为系统输入,小微企业的信用风险等级作为系统输出,以包含 3 个隐含层的神经网络进行学习预测。图 4 给出了样本训练的表现情况,可以看出在设定的迭代次数之内训练结果稳定且基本达到预期目标。此时得到的预测结果对于小微企业信贷风险等级的识别率为 53.33%,证明系统的预测结果具有一定的稳健性,具体如图 5 所示。在此

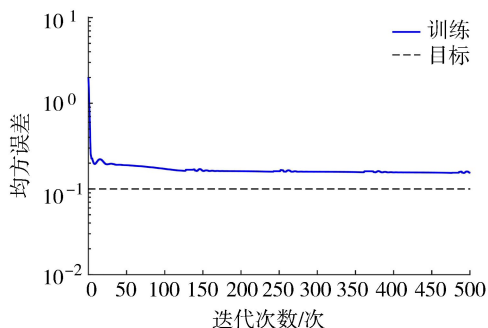


图 4 信贷风险预测的表现情况

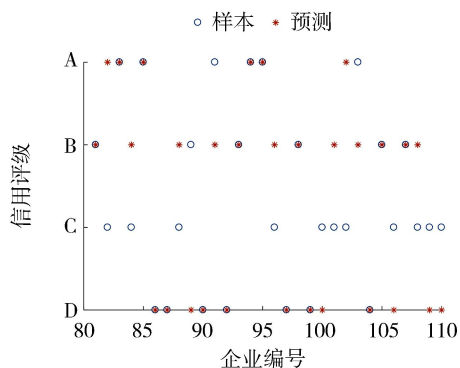


图 5 信贷风险等级的识别情况

基础上,表 4 根据小微企业初始信用风险等级,对信用风险等级的预测分类效果进行了比较,结果显示模型对于 D 类小微企业的判断更加准确,说明纳税流水能够相对有效的识别无法获取信贷额度的小微企业,这对于有效控制小微企业的信贷风险损失同时降低信贷风险成本具有重要意义。考虑到银行初始信贷风险等级并非能够准确度量小微企业的信用违约状况,而是是否发生实际违约是决定小微企业信贷可得的决定性因素,因此本文将小微企业信贷违约的预测结果与信贷风险等级的预测分类相结合,综合确定小微企业的最终信贷风险等级。

表 4 不同信贷风险等级的预测效果

风险等级	准确率	召回率	精确率	平衡 F 分数
A	0.44	0.44	0.69	0.54
B	0.83	0.83	0.50	0.63
C	0.07	0.07	1.00	0.14
D	0.82	0.82	0.56	0.67

4. 信贷供给决策

以银行经营利润最大化为目标确定不同小微企业的信贷利率,根据公式(3)定义银行的资金成本为 1 年期的贷款市场报价利率 3.85%,目标利润根据城商行的资产利润率设置为 0.62%,风险成本定义为小微企业违约概率与预期违约损失率之积,参

考《商业银行资本管理办法(试行)》假设小微企业信贷的预期违约损失率为 50%^[26]。综上所述,银行针对不同信贷风险等级小微企业的最优信贷利率如表 5 所示。银行在进行小微企业信贷利率决策时既要考虑经营利润,同时也要充分考虑信贷风险,考虑到具有相同信贷风险等级的小微企业之间存在异质性,本文通过贷款利率的调整项对小微企业超额信贷违约风险进行补偿,具体结果如表 5 中的调整信贷利率所示。可以看出对于所有类型的小微企业来说调整信贷利率普遍高于以银行最大经营利润为目标的最优信贷利率,说明针对部分小微企业的信贷利率增加了额外的风险溢价。在此基础上,定义出现信贷违约的小微企业信贷风险等级为 D,进而对通过深度神经网络预测的小微企业信贷风险等级进行修正,以此得到小微企业的实际信贷利率如表 5 所示。对比调整信贷利率和实际信贷利率可以发现,针对信用风险等级为 A 和 B 的小微企业来说信贷利率几乎没有明显区别,但是通过排除对出现违约事件小微企业的信贷支持,C 类小微企业的信贷利率明显降低,这说明综合考虑信贷违约事件和信贷风险等级的预测结果能够优化银行信贷供给决策。

表 5 不同风险等级小微企业的信贷利率 单位:%

风险等级	最优信贷利率	调整信贷利率	实际信贷利率
A	5.78	6.64	6.63
B	6.18	6.97	6.97
C	6.25	17.31	11.14

以表 5 中的实际信贷利率作为小微企业的最终信贷利率,结合公式(5)对银行的信贷供给决策进行组合优化,通过粒子群算法得到小微企业的最大信贷额度占比。表 6 给出了不同信贷风险等级小微企业最大信贷额度占比的统计结果,可以发现银行的信贷额度的决策在不同信贷风险等级的小微企业之间存在明显的偏好效应,信贷风险等级为 A、B 和 C 的小微企业平均最大信贷额度依次降低,即信贷额度整体向信用评价较高的小微企业倾斜。因此,提高纳税质量对于增加小微企业的信贷可得性具有重要的保障作用。当然,相同信贷风险等级的小微企业之间信贷利率同样存在较大差别。综合表 5 和表 6 的结果,可以根据纳税数据确定任意小微企业的实际信贷利率和最大信贷额度,在同时考虑经营利润和风险损失的基础上实现银行对小微企业的智

能信贷供给决策。

表 6 小微企业最大信贷额度占比统计结果

风险等级	均值/%	标准差	最大值/%	最小值/%
A	2.37	0.01	3.34	0.00
B	2.14	0.01	3.55	0.00
C	1.68	0.01	3.10	0.00

五、结论与启示

融资困难是众多小微企业普遍面临的问题,然而考虑到当前小微企业缺乏稳健的信用状况、可靠的财务信息及合格的抵押品等问题,如何通过精准有效的银行信贷供给解决小微企业融资困境又变得尤为困难。小微企业的传统信贷供给决策通常需要投入较多人力进行信用风险评估,昂贵的信贷决策成本使得银行针对小微企业信贷业务的经营利润受到影响,进而在一定程度上限制了小微企业的融资可得性。本文以“银税互动”政策作为切入点,通过智能化技术将小微企业沉淀的纳税数据转化为融资信用,从“以税增信”的角度为实现更低成本的小微企业智能信贷供给决策提供了可能。

以“银税互动”政策的共享纳税数据为例对小微企业信贷供给智能决策系统进行应用分析,研究结果表明依据纳税流水信息能够准确判断小微企业信贷违约事件,其中小微企业销项发票信息以及收益水平等特征具有更加重要的参考价值。纳税流水信息可以基本识别小微企业信贷风险级别,特别是明确无法获取信贷额度的小微企业,为银行的信贷供给提供决策依据。无论从信贷利率层面还是信贷额度层面,银行均会在不同程度上向纳税流水信息质量较高的小微企业提供更多信贷支持。

基于以上研究结果,得到相关启示如下:

在“银税互动”政策的背景下,各类银行可以充分利用小微企业的纳税数据对其信贷风险进行智能评估,以此降低小微企业信贷成本,同时提高小微企业信贷质量。其次,通过小微企业的纳税流水信息,可以推动银行层面小微企业主动授信业务的实现,从而大幅缩短小微企业获取信贷支持的时间,提高金融支持小微企业的效率。最后,从小微企业的角度来说,提高纳税质量对于获得银行信贷支持具有重要的保障作用,因此小微企业一定要做到依法诚信纳税。

综上所述,本文构建的小微企业信贷支持智能

决策系统,从“银税互动”的角度探究了根据纳税数据实现部分小微企业信贷决策的可行性,研究结果可以重点用于因抵押和担保不足尚未获得信贷支持,但是具有诚信良好纳税记录的小微企业,以此作为该类小微企业信贷融资渠道的重要补充。同时,针对仍然存在信贷需求但是纳税记录尚不完整或存在问题的小微企业,能够在一定程度上引导其加强依法纳税的观念和积极性。

参考文献:

[1] 盛天翔,范从来. 金融科技、最优银行业市场结构与小微企业信贷供给[J]. 金融研究,2020(6):114-132.

[2] 张一林,郁芸君,陈珠明. 人工智能、中小企业融资与银行数字化转型[J]. 中国工业经济,2021(12):69-87.

[3] 黄益平,邱晗. 大科技信贷:一个新的信用风险管理框架[J]. 管理世界,2021,37(2):12-21.

[4] DRASCH B J, SCHWEIZER A, URBACH N. Integrating the “Troublemakers”: a taxonomy for cooperation between banks and fintechs [J]. Journal of Economics and Business,2018,100: 26-42.

[5] 房颖. 金融科技赋能究竟如何影响银行小微企业信贷——基于调研数据的实证检验[J]. 金融监管研究,2021(7):69-85.

[6] 董晓林,张晔,徐虹. 金融科技发展能够帮助小微企业度过危机吗? ——基于新冠肺炎疫情的准自然实验[J]. 经济科学,2021(6):73-87.

[7] 郭丽虹,朱柯达. 金融科技、银行风险与经营业绩——基于普惠金融的视角[J]. 国际金融研究,2021(7):56-65.

[8] 周光友,罗素梅,连舒婷. 金融科技创新、网贷利率决定与小微企业融资——兼论“麦克米伦缺口”的治理[J]. 国际金融研究,2020(3):76-86.

[9] DAVID-WEST O, IHEANACHOR N, KELIKUME I. A resource-based view of digital financial services (DFS): an exploratory study of Nigerian providers [J]. Journal of Business Research,2018,88: 513-526.

[10] 李学峰,杨盼盼. 银行金融科技与流动性创造效率的关系研究[J]. 国际金融研究,2021(6):66-75.

[11] 金洪飞,李弘基,刘音露. 金融科技、银行风险与市场挤出效应[J]. 财经研究,2020,46(5):52-65.

[12] 盛天翔,朱政廷,李祎雯. 金融科技与银行小微企业信

贷供给:基于贷款技术视角[J]. 管理科学,2020,33(6):30-40.

[13] TORRE A D L, PERIA M S M, SCHMUKLER S L. Bank involvement with SMEs: beyond relationship lending [J]. Journal of Banking & Finance,2010,34(9): 2280-2293.

[14] 邱志刚,罗煜,江颖,等. 金融科技会颠覆传统金融吗? ——大数据信贷的经济解释[J]. 国际金融研究,2020(8):35-45.

[15] STULZ R M. Fintech, bigtech, and the future of banks [J]. Journal of Applied Corporate Finance,2019,31(4): 86-97.

[16] 黄志刚,刘志惠,朱建林. 多源数据信用评级普适模型栈框架的构建与应用[J]. 数量经济技术经济研究,2019,36(4):155-168.

[17] 张润驰,张谊浩,赵辉,等. 银行信用评估模型效果的量化比较[J]. 金融监管研究,2021(1):66-85.

[18] 张发明,李艾珉,韩媛媛. 基于改进动态组合评价方法的小微企业信用评价研究[J]. 管理学报,2019,16(2):286-296.

[19] 王钊,蒋翠清,丁勇. 基于混合生存分析的动态信用评分方法[J]. 系统工程理论与实践,2021,41(2):389-399.

[20] 张榕薇,朱维聪,罗琳. 信用信息共享对小微企业“首贷”可得性的影响研究——基于小微企业“首贷”数据的实证分析[J]. 征信,2021,39(7):32-40.

[21] 陈春照,谢睿,查婧怡,等. 基于大数据的中小微企业信贷风险评估和信贷策略研究[J]. 哈尔滨师范大学自然科学学报,2021,37(4):20-30.

[22] 杨龙见,吴斌珍,李世刚,等. “以税增信”是否有助于小微企业贷款? ——来自“银税互动”政策的证据[J]. 经济研究,2021,56(7):96-112.

[23] 陈彪,罗鹏飞,杨金强. 银税互动、融资约束与小微企业投融资[J]. 经济研究,2021,56(12):77-93.

[24] 王达,周映雪. 随机森林模型在宏观审慎监管中的应用——基于18个国家数据的实证研究[J]. 国际金融研究,2020(11):45-54.

[25] 匡海波,杜浩,张欣,等. 供应链金融模式下中小企业贷款定价研究[J]. 管理评论,2020,32(11):245-259.

[26] 董晓林,吕沙,张惠乾. 农村商业银行小微企业贷款定价研究[J]. 南京农业大学学报(社会科学版),2016,16(4):100-109.

(收稿日期:2022-01-07 编辑:陈玉国)