

基于优化模型的中小微企业信贷策略研究

高 婕, 李 煜, 脱 怡

杭州师范大学, 浙江 杭州
Email: 1162209997@qq.com

收稿日期: 2020年11月28日; 录用日期: 2020年12月24日; 发布日期: 2020年12月31日

摘 要

中小微企业规模较小、缺少抵押资产, 因此银行通常是依据企业的交易票据信息和影响力等因素制定贷款策略, 并对信誉高、信贷风险小的企业给予利率优惠。基于此, 建立一个以平均客户流失率的阈值为约束条件, 银行的贷款收益最优化的模型。计算企业抗风险得分、资产实力得分、企业违约情况, 从而确定对企业是否放贷及贷款额度、利率和期限等信贷策略。

关键词

企业信贷得分, 收益最优化模型, 决策树模型, 云-Critic耦合模型

Research on Credit Strategy of Micro, Small and Medium Enterprises (MSME) Based on Optimization Model

Jie Gao, Yu Li, Yi Tuo

Hangzhou Normal University, Hangzhou Zhejiang
Email: 1162209997@qq.com

Received: Nov. 28th, 2020; accepted: Dec. 24th, 2020; published: Dec. 31st, 2020

Abstract

MSME are small in scale and lack mortgage assets. Therefore, banks usually formulate loan strategies based on factors such as the information and influence of the enterprise's transaction bills, and give preferential interest rates to enterprises with high reputation and low credit risk. Based on this, a model that optimizes the bank's loan income with the threshold of the average customer

churn rate as a constraint is established. The company's anti-risk score, asset strength score, and corporate default are calculated, so as to determine whether to lend to the company and the credit strategy such as loan quota, interest rate and term.

Keywords

Enterprise Credit Score, Profit Optimization Model, Decision Tree Model, Cloud-Critic Coupling Model

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

1. 引言

近几年来, 由于中小微企业规模小和资产较少, 是政府重点扶持的对象。政府通过银行对信誉高、风险小的企业给予贷款政策优惠。为评定不同企业的贷款方式, 需建立中小微企业的信贷决策模型。

对此我们要对以下问题进行探讨研究, 以建立中小微企业的信贷决策模型。

问题一: 在真实情况下, 银行对中小微企业的贷款通常是通过判断企业的信誉情况和风险性、稳定性等因素来确定放贷与否以及相应的额度, 期限和利率。已知123家企业的过往违约与否情况、信用等级和企业进销项的各类票据信息, 量化分析123家企业的信贷风险, 得出银行在固定信贷总额时对各企业的信贷策略。

问题二: 在问题1的基础上, 依据企业进销项的各类票据信息, 对302家没有信贷记录的企业进行风险评估的量化分析, 给出在年度信贷总额为1亿元时, 企业的信贷策略。

问题三: 考虑突发因素对企业经营和经济效益的影响, 分析突发因素对企业的影响方式和程度, 给出在年度信贷总额为1亿元时, 银行的信贷调整策略。

1.1. 问题一的分析

对于问题一, 首先, 考虑银行的贷款收益最大化, 建立一个以平均客户流失率的阈值为约束条件, 以企业贷款额度得分 Q_i 、贷款利率 α_i 为决策变量的**收益最优化模型**。为了简化模型, 将企业贷款利率 r_i 划分为稳定贷款利率 α_i 与扰动利率 β , 即 $r_i = \alpha_i + \beta$, 若 α_i 、 Q_i 可以由已知数据建立的指标确定, 则可将多变量最优化模型转化为决策变量为扰动利率 β 的单变量优化模型, 实现模型的简化。所以对处理后数据建立指标体系来唯一确定各企业稳定贷款利率 α_i 和稳定贷款额度得分 Q_i , 根据稳定贷款额度得分 Q_i 确定银行总贷款确定时的各企业的贷款额度比例 γ_i 。

由于银行首先根据中小微企业的贷款实力、信誉等因素对其信贷能力做出评定, 进而确定企业的贷款额度得分和利率。则引入企业信贷得分 Z_i 变量, 其主要取决于企业抗风险得分、企业资产实力得分、企业违规情况, 建立企业信贷得分模型, 计算每家企业的 Z_i 。再将企业信贷得分 Z_i 和企业贷款额度得分 Q_i 、稳定贷款利率 r_i' 建立关系, 从而通过企业信贷得分 Z_i 和已知数据可计算出每家企业的 Q_i 、 α_i , 进而可将最优化模型转化为决策变量为扰动利率 β 的单变量优化模型。最后只需求解该模型, 得出扰动利率 β 和企业贷款利率 r_i 。特别, 对于企业信誉评级为D的企业不予贷款, 企业贷款额度得分与利率为0。

1.2. 问题二的分析

对于没有信誉评级、是否违约记录的企业，需要根据已有信贷记录的企业的关系数据，建立这些数据与信誉评级、是否违约的关系，进而给出信誉评级、是否违约预测模型。

为了评价模型的准确性，随机取有信贷记录的企业中的 80% 的企业作为训练集，即其余 20% 作为测试集。分别对训练集中企业的信誉评级、是否违约两个变量与其对应影响因素(模型一中得到)建立分类模型：Logistic 回归模型、决策树模型、Bayes 判别模型、Fisher 判别等，并将各模型代入测试集验证，选择正确率最高的模型预测无信贷记录企业的信誉评级、是否违约。最后基于问题一的模型，带入预测的信誉评级、是否违约数据，给出这些企业的贷款额度得分、利率。特别，对于企业信誉评级为 D 的企业不予贷款，企业贷款额度得分与利率为 0。由于该银行在年度信贷总额为 1 亿元，可以通过计算每个企业的贷款额度比例，进而给出银行在年度信贷总额为 1 亿元时的贷款额度得分。

1.3. 问题三的分析

考虑不同的企业在突发因素作用下所产生的影响方式和影响的程度，为了方便量化分析，需要根据产业分类将无信贷记录企业划分为几种不同类型，分别统计探讨突发因素对不同企业的影响指数。企业在受到突发因素影响的最直观体现就是资产负债、利润、现金流量的同比增长率。我们以季度为划分，在同一突发因素的影响下资产负债了增长率越大，表示所受影响越大，利润增长率和现金流量增长率相反。银行需要根据不同行业所受影响的程度来对信贷策略中企业抗风险得分进行修正。最后，基于问题二的模型，输入修正后的，计算出突发因素作用下无信贷记录企业的贷款额度得、利率。

2. 模型的准备

2.1. 指标的量化

为了后续建模的便利性，将无法直接计算的指标进行量化。

2.1.1. 企业级别的量化

将信誉评级以此标准进行量化处理，信誉评级为 A 的企业得分为 4，信誉评级为 B 的企业得分为 3，信誉评级为 C 的企业得分为 2，信誉评级为 D 的企业得分为 1。

2.1.2. 是否违约的量化

将没有违约行为的企业记为 0，有违约行为的企业记为 1。

2.1.3. 企业类别的量化

将企业可分为建筑类、商贸类、科技类等企业，不同企业都具有不同的行业指标[1]，对应得分也就不同。

2.2. 数据的挖掘与描述

2.2.1. 企业负数发票率

负数发票，即红字发票，是指在开发票后，购买方因为退货等原因，为了保证企业税务不被影响补的发票。因此，企业的负数发票越多表示买方退货退款的行为较多，企业产品实力差，风险高。为了考虑企业的实力风险评估，计算出各企业进、销项负数发票占企业总发票数。

2.2.2. 企业历年所需缴纳的增值税

历年所需缴纳的增值税，是指历年销项发票税总和减去历年进项发票税总和。一家企业的增值税在一定程度上能够体现一家企业的盈利能力，进而能够作为衡量其经营状况、资金实力等指标。为了评估企业的资金实力，计算出企业历年的增值税。

2.2.3. 各类等级违约情况

企业的违约记录对信誉等级的影响程度可以体现为银行对企业的放贷额度和利率，我们对有信贷记录的 123 家企业的“违约记录”数量以及各信誉等级中有“违约记录”的企业占比进行了统计，如图 1。

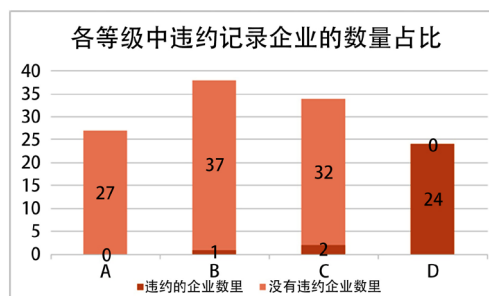


Figure 1. Percentage of companies with default records in each level
图 1. 各等级中违约记录企业的数量占比

2.2.4. 各企业进、销项发票作废率

查阅相关资料我们知道发票作废的原因可分为两大类：正常原因和非正常原因。正常原因通常有开具错误、销货退回和服务终止等，一定程度上反映企业的营业规范程度和在经营预测上的准确度；而非正常原因则有恶意隐瞒收入，销售方违规作废，虚开发票等，这种情况下，银行作为其放贷方承担很大的风险，所以在考量风险性的要求下我们对各企业进、销项发票作废率进行统计。

将统计图表中进项发票作废率和销项发票作废率两量分别倒叙排列后可看出各企业进项发票作废率均位于12.82%以下，而部分企业销项发票作废率高达68%，即票据流水有明显的结构性问题。

3. 问题一模型的建立与求解

3.1. 模型的建立

3.1.1. 优化模型建立

为了使银行的收益尽可能的大，初步建立一个决策变量为各企业贷款额度得分 Q_i 、企业贷款利率 r_i 的银行收益 H 最优化模型：

$$\begin{aligned} \text{目标函数: } \max H &= \sum_{i=1}^{123} Q_i \times r_i \\ \text{约束条件: } &\begin{cases} 10 \leq Q_i \leq 100 & \text{条件(1)} \\ 4\% \leq r_i \leq 15\% & \text{条件(2)} \\ \bar{P} \leq \sum_{i=1}^{123} P_i \leq P_0 & \text{条件(3)} \end{cases} \end{aligned}$$

其中 H 为银行一年的贷款总收益， Q_i 为第 i 家企业贷款额度得分， r_i 第 i 家企业贷款利率， P_i 为第 i 家的企业的客户流失率， \bar{P} 为所有企业客户流失率的平均值， P_0 为设定的客户流失阈值。

对于这个最优模型的决策变量有两个，为简化模型，建立各企业贷款额度得分 Q_i 、企业贷款利率 r_i 关系，或确定其中一个变量的值，将其转化为单变量优化模型。

多变量优化模型 \Rightarrow 单变量优化模型：

由于银行考虑每家企业贷款额度得分 Q_i 、企业贷款利率 r_i 的大小主要取决于中小微企业的实力、风险、违约行为等，可以将三者通过建立模型计算出企业的信贷得分，以企业的信贷得分为依据，确定银行给每家企业的贷款额度得分 Q_i 。

考虑到客户流失率的影响，将企业贷款利率 r_i 划分为稳定贷款利率 α_i 与扰动利率 β ，即 $r_i = \alpha_i + \beta$ 。经过查找相关资料公式，能够计算出最高理论授信额度，即每家企业的贷款额度得分 Q_i ，公式如下所示：

$$Q_i = \left\{ \left[\frac{R_m}{1-R_m} - \frac{R_0}{1-R_0} \right] \times Z_1 + D_0 \right\} \times RD \quad (3)$$

上述公式中所有自变量都是能够根据已挖掘处理过的数据推导计算得出，即每家企业的贷款额度得分 Q_i 是可计算的。

在利用企业信贷得分等自变量确定了企业的贷款额度得分后，使用 *RAROC* 贷款定价模型[2]来计算稳定贷款利率 α_i ，公式如下所示：

$$\alpha_i = c + e + (RAROC \times Z_1 + DC) / Q_i \quad (4)$$

同理，能够计算出每家企业所对应的稳定贷款利率 r'_i 。由于贷款额度得分 Q_i 也可以确定，则该优化模型的决策变量可转化为 β_i 一个，也就是将多变量优化模型转化为了单变量优化模型。

3.1.2. 分层指标体系的建立

根据优化模型，为了将其由多变量转化为单变量，需要确定稳定贷款利率 α_i 和贷款额度得分 Q_i 。根据贷款额度得分 Q_i 公式和稳定贷款利率 α_i 公式：

$$Q_i = \left\{ \left[\frac{R_m}{1-R_m} - \frac{R_0}{1-R_0} \right] \times Z_1 + D_0 \right\} \times RD \quad (3)$$

$$\alpha_i = c + e + (RAROC \times Z_1 + DC) / Q_i \quad (4)$$

则可将资产负债率 R_0 、银行愿意承受的客户资产负债率 R_m 、客户所属行业的付息债务比例 RD 作为贷款额度得分 Q_i 指标；将经营成本率、风险成本率、经济成本、经营成本作为稳定贷款利率 α_i 的指标。

除此之外，由于银行普遍根据中小微企业的贷款实力、信誉等指标对其信贷能力做出评定，进而确定企业的贷款额度得分和利率。因此，对贷款额度得分 Q_i 和稳定贷款利率 α_i 都引入企业信贷得分 Z_1 这个变量作为指标。其中企业信贷得分 Z_1 主要取决于企业抗风险得分、企业资产实力得分、企业违规情况。

因此为了求解模型，需建立模型逐步算出各企业抗风险得分、企业资产实力得分，再代入已给的企业违约情况数据建立模型，计算出企业的信贷得分，进而算出贷款额度得分 Q_i 和稳定贷款利率 α_i 。最后，解决决策变量为扰动利率 β 的优化模型，得到扰动利率 β ，从而算出企业贷款利率 $r_i = \alpha_i + \beta$ ，得到各企业的贷款决策。各变量关系图如图 2 所示。

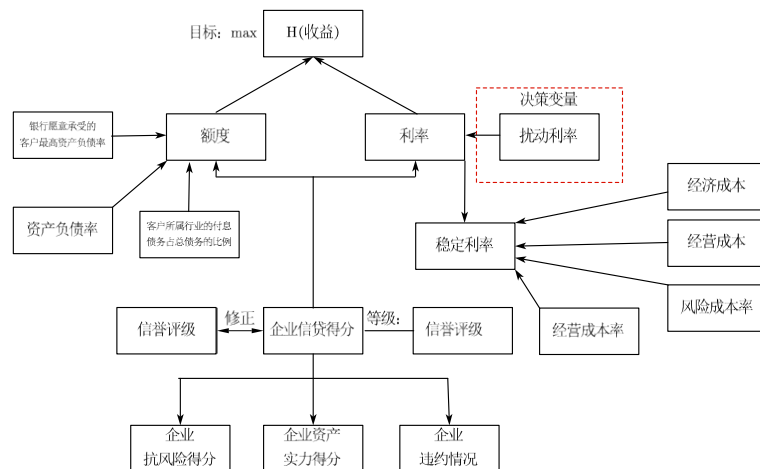


Figure 2. Relation diagram of various variables
图 2. 各变量关系图

3.2. 模型的求解

3.2.1. 变量的选取

为了计算出企业抗风险得分、企业资产实力得分，需要将已处理的变量(待测指标)进行分类和剔除，分别筛选出企业抗风险得分、企业资产实力得分的指标，进而建立模型并计算出得分。由于信誉评级是由银行内部专业员工鉴定的，与企业信贷得分有较强的关系，即信誉评级也与企业抗风险得分、企业资产实力有一定的关系，即各个待测指标与信誉评级间有一定的关联。为了更好地寻找出企业信贷得分与信誉评价内在的关联，综合题目背景以及实际数据处理情况，首先给出经过数据挖掘以及整理过的一些可能与信誉评级存在内在关联的待测变量，分别计算它们与信誉评级的相关性度量值，再根据检验 P 值与相关性度量值的正负，筛选出有内在关系的待测指标。

1) 相关性检验

信誉评级与 16 个待选指标的相关性筛选：

检验量化处理后的信誉评级是否符合正态分布。经过计算，筛选出待测指标。

2) 共线性诊断

为了提高模型的准确度，再对上述待测指标之间进行共线性分析，剔除一批共线性较高的指标。

3) 违约行为的影响因素探究

将违约行为量化为能够计算的指标，即记有违规行为的为 1，没有违规行为的为 0，将其与第四节模型的准备中计算推导筛选得出的指标进行相关性分析。

根据上述数据，得到影响因素。

3.2.2. 企业抗风险评估模型

结合实际情况与题目背景，参考相关方面的论文[3]，能够制定出下列定量因素评分，如表 1。

Table 1. Quantitative factor score sheet

表 1. 定量因素评分表

考评项目	计算公式	评价标准	得分
企业获利水平 (2019 年增值税)	增值税 = 企业销项税总和 - 企业进项税总和	增值税 ≥ 0	得分 = 企业的增值税/所有增值税大于 0 的企业的增值税平均值
		增值税 < 0	得分 = -(企业的增值税/所有增值税小于 0 的企业的增值税平均值)
企业现金流量 (销项价税和比率)	每家企业的销项价税和/所有企业的销项价税和的平均值	大于 1.0	10
		0.8~1.0	8
		0.6~0.8	6
		0.4~0.6	4
		0.2~0.4	2
营业规范程度和 经营预测准确度 (销项作废率)	企业作废发票的数量/企业总发票数量	0~0.2	0
		0.45	-5
		0.3~0.45	-2
		0.2~0.3	1
		0.1~0.2	4
		0.05~0.1	7
		0.05	9

参照定量因素评分表计算出各企业的抗风险能力得分。

3.2.3. 基于 DCF 贴现现金流量法企业资产实力的评估模型

DCF 贴现现金流量法是在企业持续经营的前提下,通过对企业合理的尤其获利能力的预测和适当的折现率的选择,计算出企业的现值,这种评估方法将企业内部的各种不相同的单项资产作为统一的不可分割的要素进行整体评估,不是各项资产的简单加总,而是企业正常经营条件下的资本化价格,整体反应企业资产的未来获利能力,揭示企业内在的价值[4]。

仿照 DCF 贴现现金流量算法[5]计算各企业价值评价,即企业实力参数。其公式为:

$$Y_1 = \sum_{i=1}^t \frac{FCF_t}{(1+G)^i} + \frac{\text{连续价值}}{(1+G)^t} \quad (5)$$

其中 G 是企业资本架构的量化评分,依照企业公司整体平均资本成本 WACC 的计算方法[6],通过其发票流水中进、销项金额和各税等之间的占比关系映射公司资本架构,设计企业资本架构量化评分的计算。将企业各种税项在企业全部税项中所占的比重乘以资本金额,然后把各数目加起来,统计其企业资本架构量化评分 G :

$$G = (x_{11}/x_{13}) \times Re + (x_{10}/x_{13}) \times Rd \times (1 - Tc) \quad (6)$$

上述式子中 Re 、 Rd 、 E/V 、 D/V 分别为进销税价和差、进项金额、增值额占销项税额的百分比、进项税额占销项税额的百分比, Tc 、 FCF 、 G 、 T 分别为企业税率(相关资料显示本题所涉及的小微企业税率为 20%)、进销税价和差、企业资本架构量化评分、年次。最后根据上述公式得出各企业资产实力得分。

3.2.4. 基于云-Critic 耦合的企业信贷得分评估模型

1) 模型的建立与求解

由于企业信贷得分影响因素众多,各种因素之间互相作用的关系、各种因素对企业未来发展的影响机理尚未完全揭示,因此企业信贷得分问题兼顾模糊性和随机性。通过引入研究模糊性与随机性关联的云理论,能够有效地解决企业信贷得分的定量评估问题。企业信贷得分评估云模型计算流程如图 3 所示,主要步骤如图 3。

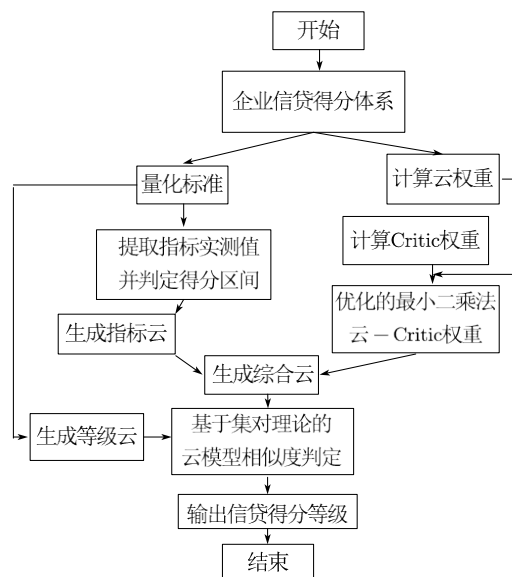


Figure 3. Calculation flowchart of cloud model for enterprise credit score evaluation
图 3. 企业信贷得分评估云模型计算流程图

为科学地分配指标云的权重，将采取主观赋权法与客观赋权法相结合。因主观权重主要依赖专家经验与主观判断，忽略了指标数据本身内在的关系；客观权重由指标实际数据决定，但易出现不符合常理的现象。本文将使用最小二乘法将主观云权重和客观 Critic 权重相结合，提出一种改进的云-Critic 耦合权重[7]。

2) 主观云权重

云权重基于云理论，将表征指标重要程度的论域设定为 $[0,1]$ ，依据专家对重要性的评价值、相关资料的查找及其对应论域区间转化为相应的权重云模型，能够计算出权重云。本节将信誉得分论域均匀划分为5个区间，分别对应5个重要性等级，并且利用(7)-(9)转化为权重云模型，如表2：

Table 2. Cloud weight classification

表 2. 云权重等级划分

专家评价值	区间	权重云
不重要	$[0,0.2]$	$(0,0.85,0.005)$
次重要	$[0.2,0.4]$	$(0.300,0.085,0.005)$
一般	$[0.4-0.6]$	$(0.500,0.085,0.005)$
较重要	$[0.6-0.8]$	$(0.700,0.085,0.005)$
极重要	$[0.8,1.0]$	$(1.000,0.085,0.005)$

经过查找资料，根据专家对指标的重要性评价，结合上表确定指标所属权重云，再由正向云发生器计算云滴的均值并归一化处理，即可得到指标云权重。

3) 客观 Critic 权重

经计算，能够得出企业抗风险能力、企业资产实力、企业违约情况三者的权重，如表3：

Table 3. Objective Critic Weight Table

表 3. 客观 Critic 权重表

	权重
企业抗风险能力	0.322
企业资产实力	0.401
企业违约情况	0.277

4) 改进的云-Critic 耦合权重研究

运用云-Critic 耦合权重计算式子，能够得出企业抗风险能力、企业资产实力、企业违约情况三者的权重，如表4。进而能够计算出企业信贷得分(归一化处理后)。

Table 4. Cloud-Critic coupling weight table

表 4. 云-Critic 耦合权重表

	权重
企业抗风险能力	0.495
企业资产实力	0.378
企业违约情况	0.127

3.2.5. 基于企业信誉评级的企业信贷得分修正

根据上述模型计算出的数据，有很多企业信贷得分高，但等级低；信贷得分低，但等级高。针对这

些异常企业数据，需要对信贷得分进行修正。

判定异常数据：企业信用评级一共有四级，其中 A、B、C、D 级企业个数占总企业数的 21.95%、30.89%、27.64%、19.51%。将企业信贷得分从大到小按照企业信用评级的占比原则，也分为 A、B、C、D 四个企业信贷等级。判断该企业信用评级与企业信贷等级是否一致，若不一致，判定为异常数据。

异常数据进行修正：给出异常企业信贷得分 Z_1 修正原则为

$$Z'_1 = Z_1 + Z_1 \times \Delta d$$

其中 Δd 为等级差百分率。例：企业信用评级 A 但信贷等级为 C，中间则等级差百分率 $\Delta d = +30.89\%$ ；企业信用评级 D 但信贷等级为 C，中间则等级差百分率 $\Delta d = -(27.64\%/2 + 19.51\%/2)$ 。将修正后的企业信贷得分替换原值。具体流程如图 4：

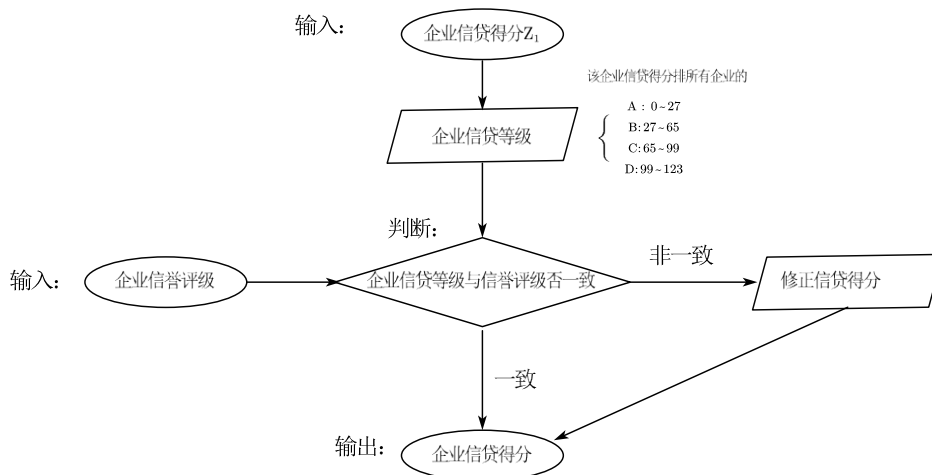


Figure 4. Flow chart of corporate credit score correction
图 4. 企业信贷得分修正流程图

如表 5 所示，修正后企业信贷得分与信誉等级 spearman 相关系数显著提升。由于信用评级是根据银行企业人员内部评定的，对企业信贷得分有较大的参考价值，同时修正后的企业信贷得分为银行选择信贷政策。同时，筛选和修正了企业实力强但信誉低下的异常企业，所以修正后的企业信贷得分更有价值，如表 5。

Table 5. Correlation coefficient before and after correction
表 5. 修正后前后相关系数

	spearman 相关系数
修正前	0.377589
修正后	0.648954

3.2.6. 贷款额度得分

计算企业贷款额度得分的公式如下所示：

$$Q_i = \left\{ \left[R_m / (1 - R_m) - R_0 / (1 - R_0) \right] \times Z_1 + D_0 \right\} \times RD \quad (3)$$

Z_1 为企业信贷得分， R_m 为商业银行愿意承受的客户最高资产负债率， R_0 为客户基期的资产负债率， D_0 为客户基期的负债总额， RD 为客户所属行业的付息债务占总债务的比例。结合《企业评价标准值》，

令 $RD = 50\%$, $R_m = 70\%$, 企业的进项税价总和来体现企业的负债总额。计算出企业的贷款额度得分 Q_i , 特别, 信用评级为 D 等的企业, 对他们不予贷款, 额度为 0。根据贷款额度得分 Q_i , 可以得到银行年度信贷总额一定的情况下, 贷款额度比例 γ_i 。

3.2.7. 稳定利率

经过查阅相关资料, 在利用企业信贷得分确定了企业的贷款额度得分后, 使用 RAROC 贷款定价模型来确定稳定贷款利率 α_i , 公式如下:

$$\alpha_i = c + e + (RAROC \times Z_1 + DC) / Q_i \quad (4)$$

上式中, α_i 为贷款利率; c 为经营成本率; e 为风险成本率; EC 为经济成本; DC 为企业信贷得分; OC 为经营成本。按照国际惯例, 商业银行的 RAROC 通常处于 20%~30%, 最低为 15%, 为方便计算比较, 令 $RAROC = 25\%$ 。经济成本通过销项金额与进项金额等信息能计算。通过查找资料, 银行的经营成本设为定值, 在此种设定下, 对于每家企业而言影响是一样的。运用上述转换理解计算, 能够得出企业的稳定贷款利率 α_i 。

3.2.8. 基于模拟退火算法的优化模型求解

1) 基于回归模型的客户流失率模型

银行的客户流失率主要取决于贷款年利率和企业的信用评级。由于贷款年利率是定量变量, 企业信用评级是定性变量, 有 A、B、C、D 四个等级, 其中 D 等不予贷款。为了建立客户流失率与贷款年利率和企业的信用评级的关系, 考虑回归模型。由于企业信用评级是定性变量, 将企业信用评级设置为虚拟变量, 即将业信用评级为 A、B、C 的虚拟变量分别记为 $k_1 \sim k_3$ 。

$$k_1 = \begin{cases} 0 & \text{信用等级为A} \\ 1 & \text{信用等级不为A} \end{cases}, \quad k_2 = \begin{cases} 0 & \text{信用等级为B} \\ 1 & \text{信用等级不为B} \end{cases}, \quad k_3 = \begin{cases} 0 & \text{信用等级为C} \\ 1 & \text{信用等级不为C} \end{cases}$$

为了避免完全多重共线性的影响, 将 A 级设置为参照类, 利用 OLS 法可得回归模型: $P = 7.447852r - 0.0361891k_2 - 0.0452414k_3 - 0.0907254$ 。

回归模型中 F 检验统计量 $F(3,83) = 337.04$, p 值为 0.0000, F 检验通过, 拟合优度 $R^2 = 0.9241$, 接近于 1, 拟合效果较好。如表 6 所示各变量回归系数的 t 检验的 p 值均小于 0.05, 各变量均显著, 即客户流失率回归模型合理, 如表 6。

Table 6. Relevant parameters of customer churn rate regression model

表 6. 客户流失率回归模型相关参数

P	系数	标准误差	T 统计量	P 值
r	7.447852	0.2349475	31.70	0.000
k_1			0 (omitted)	
k_2	-0.0361891	0.0191942	-1.89	0.043
k_3	-0.0452414	0.0191942	-2.36	0.021
常数	-0.0907254	0.0260292	-3.49	0.001

2) 基于模拟退火算法扰动利率 β 和企业的贷款利率 r_i 求解

已知求得各企业的贷款额度得分 Q_i 和各企业稳定贷款利率 α_i , 由于 $r_i = \alpha_i + \beta_i$, 则下列优化模型的决策变量只有扰动利率 β 。

$$\begin{aligned} \text{目标函数: } \max H &= \sum_{i=1}^{123} Q_i \times r_i \\ \text{约束条件: } &\begin{cases} 10 \leq Q_i \leq 100 & \text{条件(1)} \\ 4\% \leq r_i \leq 15\% & \text{条件(2)} \\ \bar{P} = \sum_{i=1}^{123} P_i \leq P_0 & \text{条件(3)} \end{cases} \end{aligned}$$

其中 P_i 为第 i 家的企业的客户流失率, \bar{P} 为所有企业客户流失率的平均值, P_0 为设定客户流失阈值。 P_i 可以根据模型的准备中基于回归模型的客户流失率模型给出, 根据《2019 年中国银行业调查报告》, 将中小微企业的 P_0 设置为 0.38。

模拟退火算法求解[8]:

考虑到银行利率一般精确到小数点后两位, 为了是模型更简化, 设 β 的取值为 $\beta_i = 0.04:0.01:0.15$, 即每个 β_i 有 12 个取值。由于 β 是一个 123 行的列向量, 使用蒙特卡洛模拟搜索一次的时间约为 16 s~18 s。这里采用启发式搜索中的模拟退火算法, 进行银行收益 H 最大值搜索。其中选择 Matlab 模拟退火工具箱的新解产生规则: 首先生成一组随机数 $(\beta_1, \beta_2, \dots, \beta_{123})$, 随机选取 y_i 服从 $N(0,1)$, 将其标准化的 (z_1, z_2, \dots, z_n) , 其中 $z_i = y_i / \sqrt{y_1^2 + y_2^2 + \dots + y_n^2}$, 对于每个 $i \in [1, n]$, 进行 $\beta_i^{new} = \beta_i + T \times z_i$ 得到新解 β^{new} , 判断新解是否满足约束条件:

$$r_i = r'_i + \beta_i \Rightarrow \text{约束条件 1: } r_i \in [0.04, 0.15]$$

$$P = 7.447852r - 0.0361891k_2 - 0.0452414k_2 - 0.0907254 \Rightarrow \text{约束条件 2: } \bar{P} \leq P_0$$

若满足约束条件, 计算对应银行收益 $H(\beta_{new})$, 选择 Metropolis 准则接受新解:

$$\text{若 } H(\beta_{new}) \geq H(\beta), \text{ 将 } \beta_{new} \text{ 赋值给 } \beta, \text{ 并重复以上步骤; 若 } H(\beta_{new}) < H(\beta), \text{ 令 } P_t = e^{-\frac{|H(\beta) - H(\beta_{new})|}{100 \times 0.95^t}},$$

随机生成 $a \in [0,1]$, 若 $a < P_t$, 将 β_{new} 赋值给 β , 并重复以上步骤; 否则, 重新生成 β_{new} 。达到迭代 10,000 次停止搜索。具体流程如图 5。

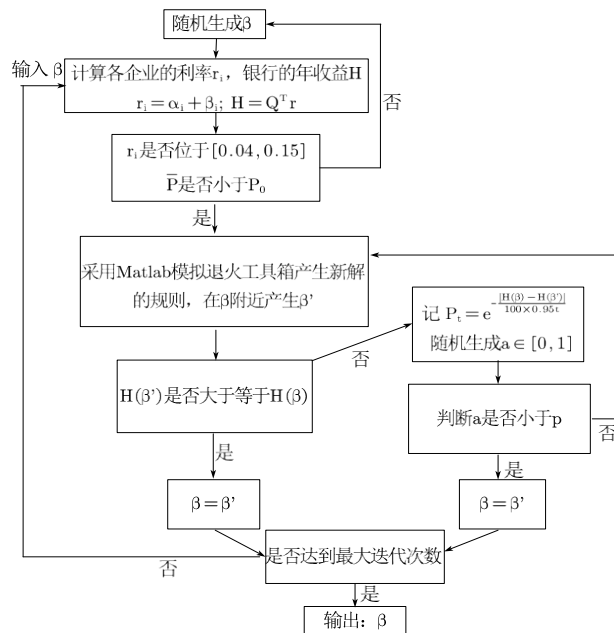


Figure 5. Flow chart of simulated annealing algorithm
图 5. 模拟退火算法流程图

通过迭代 10,000 次模拟退火算法，可以计算出最终的取值。再根据可以计算出各企业的利率，特别信贷等级为 D 级的企业不予贷款，利率为 0。

4. 模型二的建立与求解

4.1. 模型的准备

对于没有信誉评级、是否违约记录的企业，需根据有信贷记录的企业数据，建立这些数据与信誉评级、是否违约的关系，从而得到企业的信誉评级、是否违约预测模型。为了评价模型的准确性，随机取有信贷记录的企业中的 80% 的企业作为训练集，即其余 20% 作为测试集。分别对训练集中企业的信誉评级、是否违约两个变量与其对应影响因素(模型一中得到)建立分类模型：Logistic 回归模型、决策树模型、Bayes 判别模型、Fisher 判别等，并将各模型代入测试集验证，选择正确率最高的模型预测没有信誉评级、是否违约记录企业的信誉评级、是否违约。经过筛选，选择决策树模型预测是否违约，Fisher 判别模型预测信誉评级。

4.1.1. 基于决策树模型的是否违约预测

审查信用违约的方法是 C5.0 (CRT) 算法，由 J. Ross Quinlan 创建。该算法用于生成决策树。C5.0 算法在评估缺省值方面有很大的优势，例如可以在没有数学背景的情况下对结果进行解释，并且可以用于小数据集或大数据集[9]。利用 R 软件中“C50” packages，可以对数据进行拆分，决策树可以持续增长。考虑决策树的修剪，由于决策树的目的是深入了解数据，而在让决策树增长到最佳大小之前对其进行预修剪可能会错过重要的模式。因此，选择将决策树深度最大化后进行修剪。

算法流程如图 6:

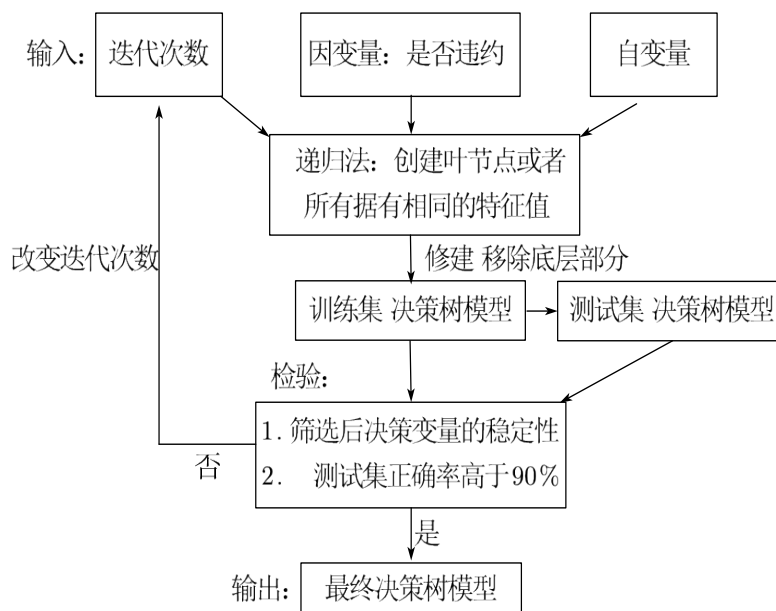


Figure 6. Decision tree model algorithm flow

图 6. 决策树模型算法流程

通过多次运算可得到是否违约的决策树模型的相关信息。如表 7 所示，通过决策树模型，是否违约主要取决于销项发票作废率、2019 销项税价总和，进项发票作废率 对其影响较小，如表 7。

Table 7. Information of the final decision tree model

表 7. 最终决策树模型的信息

变量	数据
节点数	7
终端节点数	4
深度	2
自变量	原始
	最终

进项发票作废率 x_1
 销项发票作废率 x_2
 2019 销项税价总和 x_9

销项发票作废率 x_2
 2019 销项税价总和 x_9

同时可以画出训练集决策树模型运行图，如图 7。

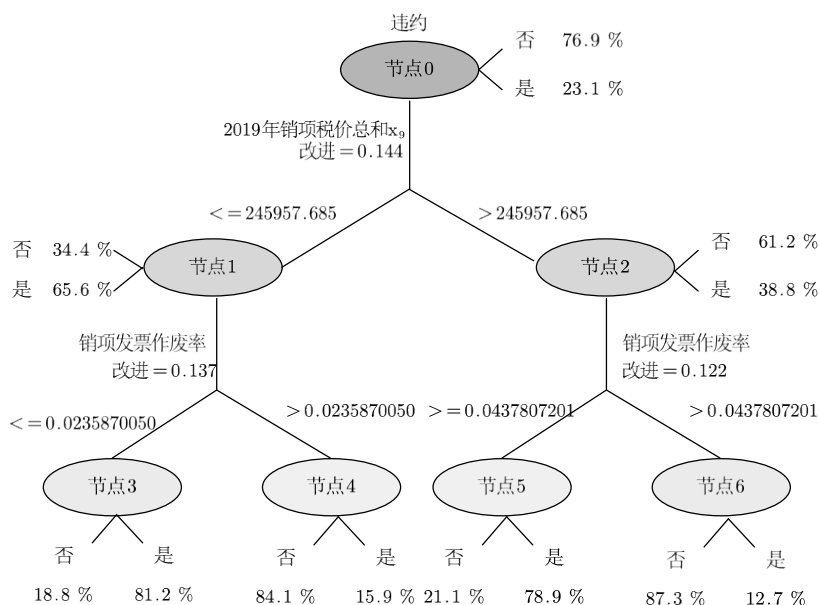


Figure 7. Training set decision tree model running process

图 7. 训练集决策树模型运行流程

决策树运行流程如下：

Step 1: 判断 2019 销项税价总和 x_9 ，若 $x_9 \leq 245957.685$ ，跳转到 step 2-节点 1；若否则，跳转到 step 2-节点 2。

Step 2: 节点 1: 判断销项发票作废率 x_2 ，若 $x_2 \leq 0.0235870050$ ，跳转到 step 3-点 3；否则，跳转到 step 3-节点 4。

节点 2: 判断销项发票作废率 x_2 ，若 $x_2 \leq 0.0437807201$ ，跳转到 step 3-点 5；否则，跳转到 step 3-节点 6。

Step 3: 节点 3——违约；节点 4——不违约；节点 5——违约；节点 6——不违约

为了保证模型的准确，还需对测试集的样本运行同样的决策树模型，运行结果如图 8。

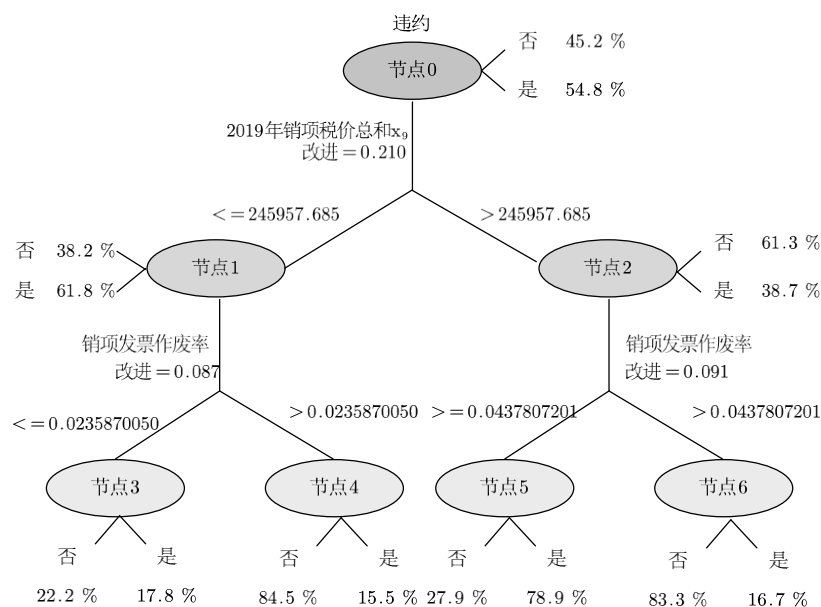


Figure 8. Test set decision tree model running process
图 8. 测试集决策树模型运行流程

分别统计训练集和测试集实测数据和决策树模型预测数据的正确率，得到训练集和测试集的数据正确率分别为 93.9%、92.6%，与 Logistic 回归模型、Bayes 判别模型、fisher 判别等模型的正确率相比更高，即选择决策树模型预测效果更好。同时，决策树模型体现了比拟合等方法更能体现出变量间的内在关系，如表 8。

表 8. Decision tree model prediction accuracy
Table 8. 决策树模型预测准确性

	实测	否(预测)	是(预测)	正确百分比
训练集	否	67	3	97.7%
	是	8	13	61.9%
	总体百分比	82.4%	17.6%	93.9%
检验集	否	25	1	96.2%
	是	2	4	66.7%
	总体百分比	84.4%	15.6%	92.6%

最后，根据决策树模型，可以给出 302 个无信贷记录的企业的违纪记录。

4.1.2. 基于 Fisher 多分类模型的信誉评级

Fisher 判别分析是一种线性判别方法，通过将训练集中的变量数据投影到 k 维，进而达到降维和分类的目的。根据模型一的假设与分析，企业的信誉评级取决于企业的实力 Y_1 、信贷风险 Y_2 、违约行为 Y_3 ，其中 123 家企业的三个变量得分在模型一中均已得出。并随机取 123 个企业样本的 80% 作为训练集建立企业的信誉评级的 Fisher 多分类模型，并对另 20% 测试集做预测并给出正确率。

算法流程：

Step 1: 对于 Fisher 多分类模型，通过引入全局散度矩阵 S_b ，可测得测试集企业的实力 Y_1 、信贷风险 Y_2 、违约行为 Y_3 的散度矩阵 S_i ：

$$S_t = S_w + S_b = \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T; \quad S_b = S_t - S_w = \sum_{i=1}^n m_i (\mu_i - \mu)(\mu_i - \mu)^T$$

Step 2: 为了使样本向量投影到确定的类别, 并使得投影下与同类中心的距离尽可能的小, 与异类中心距离尽可能地大。则计算投影矩阵 W , 并解决以下优化目标: $\max_w \frac{\text{tr}(W^T S_b W)}{\text{tr}(W^T S_w W)}$

Step 3: 加入拉格朗日乘子并求导, 可推导出: $S_w^{-1} S_b W = \lambda W$

得到投影矩阵 W , 可以达到分类的目的。

将训练集的企业实力 Y_1 、信贷风险 Y_2 、违约行为 Y_3 数据代入建立企业的信誉评级的 Fisher 多分类模型, 得到结果如下所示。由表 9 可知, 最终 Fisher 多分类模型的线性系数向量分别为 $\omega_1 = (-0.566, -0.308, -0.925)^T$; $\omega_2 = (-0.308, 0.948, 0.193)^T$; $\omega_3 = (0.925, 0.196, 0.107)$, 如表 9。

Table 9. Standardized canonical discriminant function coefficient
表 9. 标准化典则判别函数系数

	1	2	3
企业资金实力 Y_1	-0.566	-0.308	0.925
企业抗击风险得分 Y_2	-0.261	0.948	0.196
是否违约 Y_3	1.108	0.193	0.107

如表 10 所示, 计算出贝叶斯判别函数系数。根据该表, 可将企业实力 Y_1 、风险 Y_2 、违约 Y_3 数据代入贝叶斯判别函数比较对应的函数值, 按照函数值大小将样本数据的信誉评级归类, 如表 10。

Table 10. Bayesian discriminant function coefficients
表 10. 贝叶斯判别函数系数

信誉评级	D	C	B	A
企业资金实力 Y_1	-0.003	0.011	0.014	0.210
企业抗击风险得分 Y_2	-0.017	0.248	0.262	0.387
是否违约 Y_3	51.459	2.104	-0.361	-1.282
常量	-27.734	-2.606	-2.691	-4.167

代入训练集样本数据得出 Fisher 模型后, 需要对模型进行检验和评价。如表 11 所示, 将测试集数据带入模型, 可以得到预测的信誉评级与实际数据的正确率为 81.3%。与 Logistic 回归模型、决策树模型等方法相比 Fisher 模型预测的准确率有 10%~20% 的提高, 即 Fisher 模型选择合理, 如表 11。

Table 11. The accuracy result of the Fisher model prediction
表 11. Fisher 模型预测的准确率结果

预测 实测		D	C	B	A	总计
计数	D	21	0	0	0	21
	C	2	6	22	4	34
	B	0	8	19	10	37
	A	0	2	8	15	25
百分比	D	100.0	0.0	0.0	0.0	100.0
	C	5.9	17.6	64.7	11.8	100.0
	B	0.0	21.6	51.4	27.0	100.0
	A	0.0	8.0	32.0	60.0	100.0
总计		正确地对 81.3% 的测试集个案进行了分类				

最后, 根据 Fisher 模型, 可以给出 302 个无信贷记录的企业的信誉评级。

基于第一问的企业贷款额度得分、利率模型, 并带入 302 家企业样本预测出的企业信誉评级和是否违约数据, 同理可得银行应对这 302 家企业设定的贷款额度得分和利率, 即信贷策略。

5. 模型三的建立与求解

5.1. 模型的准备

突发因素对中小型企业的打击较大企业更强, 突发因素产生普遍性影响时, 银行作为放贷方势必需对企业信贷做出一定的调整。据百融云创统计数据显示, 疫情爆发后短期内给线下消费带来了巨大的消极影响, 相关业务短期内风险上升, 服装鞋帽针织纺织品类、金银珠宝、汽车、建筑及装潢材料增速下滑明显, 中西药品类、日用品增速明显, 其他分项表现较为平缓[10]。

1) 汽车产业

根据中国汽车流通协会调查结果显示, 调研中的 2895 家经销商中仅有 573 家 4s 店完成复工, 复工率不到 20%, 而疫情期间, 购车需求急速下降。汽车销售市场经营受到极大影响。

2) 医疗和生命科学产业

在疫情的大条件下, 研发各种应对病毒的检测、抑制和治疗的相关药品受到了更多重视, 疫苗类的产品使用量也有较大提高, 医疗相关行业受到的其负影响较少。

3) 科技产业

和众多行业相比科技受到的影响较小, 疫情防控措施限制了人口流动, 而由于减少了线下交流的机会, 以通信、云端、人工智能等服务为代表的科技行业得到了较快的发展。但工厂员工返工晚等因素使得电子供应链受到一定影响。

4) 建材铸造产业

受疫情防控暂停施工、建筑工人返工放慢等因素, 建材铸造业投资和收益节奏也被迫放缓, 作为杠杆率较高的行业, 快速资金回笼有着重要的意义, 疫情影响使得其所受风险有所加大。

5) 文化传媒产业

疫情防控工作限制了人口的大量流动, 对于媒体行业而言, 有一定的促进作用, 不论是电视节目还是各类社交平台, 关注度都有所上升, 仅线下院线产业遭受到猛烈冲击。

6) 其他零售实体产业

疫情直接影响零售实体店营收, 餐饮、购物中心、百货超市和专业品牌店, 院线和旅游出行关停, 导致时尚奢品、酒类等场景体验消费类目受挫, 经营利润大幅下跌, 直接挑战生存。

以疫情为例, 我们以以上行业类型为标准把企业分为六类, 分别统计疫情影响下各产业在 2020 年前两季度的资产负债、利润、现金流量同期增长率, 以探讨疫情对不同行业的冲击度。

由以上各个产业 2020 前数据可以得出, 在季度维度: 疫情对于汽车相关产业、科技产业的影响程度第二季度高于第一季度; 疫情对于其他零售产业、文化传媒产业、医药和生物科技产业、建材产业的影响程度中第一季度的影响远大于第二季度。在整体影响程度维度: 对于汽车产业和家具产业的影响较大。

5.2. 模型的建立

综合上述分析, 企业的同期资金负债增长率、利润增长率、现金流量增长率这三个指标是并列存在的, 因而这三者的影响因素是无法互相重叠的。结合实际情况与题目背景, 参考相关论文[11], 能够制定出下列定量因素评分表, 如表 12。

Table 12. Quantitative factor score sheet
表 12. 定量因素评分表

考评项目	计算公式	评价标准	得分
负债增长率	$(2020 \text{ 年同季度负债金额} - 2019 \text{ 年同季度负债金额}) / 2019 \text{ 年同季度负债金额}$	大于 1	5
		(0.5, 1]	3
		(0, 0.5]	1
		[-0.5, 0]	-2
		小于 -0.5	-4
利润增长率	$(2020 \text{ 年同季度利润} - 2019 \text{ 年同季度利润}) / 2019 \text{ 年同季度利润}$	小于 -0.3	5
		[-0.3, -0.2]	3
		[-0.2, 0]	1
		[0, 0.2]	-2
		大于 0.2	-4
现金流量增长率	$(2020 \text{ 年同季度现金流量} - 2019 \text{ 年同季度现金流量}) / 2019 \text{ 年同季度现金流量}$	小于 -0.3	5
		[-0.3, -0.2]	3
		[-0.2, 0]	1
		[0, 0.2]	-1
		大于 0.2	-3

5.3. 模型的求解

根据以上定量因素评分表对六类产业进行计算，得出各类产业疫情中的影响程度，如表 13。

Table 13. The degree of impact of various industrial epidemics
表 13. 各类产业疫情中的影响程度

产业	影响程度
汽车	2.6
其他零售	-2.1
文化传媒	-2.6
医药生物	-3.5
科技	2.7
建材铸造	-2.5

基于第一问的企业贷款额度得分、利率模型，并带入 302 家企业样本预测出的企业信誉评级和是否违约数据，同理可得银行应对这 302 家企业设定的贷款额度得分和利率，即信贷策略。

参考文献

- [1] 雷霞飞. 基于行业风险系数的科技型中小企业风险评估模型的研究与实现[D]: [硕士学位论文]. 长春: 吉林大学, 2016.
- [2] 陈振业, 杨吉峰, 赵莹. 小微企业无还本续贷风险评价与贷款定价研究——以山东省城市商业银行为例[J]. 华北金融, 2020(7): 68-79.
- [3] 房斌. P 银行小微企业信贷风险评价体系研究[D]: [硕士学位论文]. 西安: 西安石油大学, 2020.
- [4] 李秋浩. DCF 模型中稳定状态现金流量计算[J]. 商, 2014(26): 136-136.

-
- [5] 李波. DCF 估值模型探讨: 以净利润代替自由现金流[J]. 全国流通经济, 2018(32): 29-30.
- [6] 李泳慧. 基于 MIRR 和 WACC 的战略投融资决策模型[J]. 产业创新研究, 2018(9): 59-61.
- [7] 徐昕军, 杨峰. 基于动态跟踪与集对理论的地下病害风险评估云模型研究[J]. 矿业科学学报, 2018, 3(6): 575-586.
- [8] 肖思和, 鲁红英, 范安东, 等. 模拟退火算法在求解组合优化问题中的应用研究[J]. 四川理工学院学报(自然科学版), 2010, 23(1): 116-118.
- [9] 陈思. C5.0 算法在个人信用评估中的研究与应用[D]: [硕士学位论文]. 武汉: 华中科技大学, 2019.
- [10] 史善胤. 新冠肺炎疫情对我国中小企业的影响及应对[J]. 时代经贸, 2020(7): 6-7.
- [11] 陈建. 信用评分模型综述[J]. 中国信用卡, 2005(1): 46-49.