

# 银行对中小微企业的信贷决策模型

## 摘要

中小微企业信贷问题一直是国家和社会普遍关注的问题。中小微企业为社会提供了就业岗位，为社会经济发展作出了重要贡献。同时，贷款是银行最重要的盈利业务。因此，如何在向中小微企业提供足够贷款的同时达到银行利益最大化就显得尤为重要。

针对问题一，我们首先对企业的信贷风险进行分析与评估。利用 matlab 对题目给出的 123 家中小微企业的发票信息进行预处理，剔除经营状况异常的企业，对发票数据进行汇总，利用营运资金量公式预测下一年各个企业生产经营所需资金，并以此推断企业的可能贷款金额。同时，利用在不同信誉等级，不同年利率条件下的客户流失率和违约率计算银行放贷的可能收益率，并在银行贷款收益率最大化的前提下建立单目标优化模型确定放贷的最佳年利率。最后根据企业可能的贷款金额以及银行贷款收益率计算银行贷款预计产生的收益并进行贷款额度分级，并分别确定 **10 级贷款额度**。

针对问题二，我们首先利用 matlab 对 302 家企业的发票信息进行预处理，并利用**主成分分析方法**，得出影响信誉等级与违约情况的 **7 个主成分因素**。然后对信誉等级与违约情况统一量化，并利用**机器学习决策树的方法**。通过分析问题一中 123 家企业的信誉等级与违约情况对 7 个主成分因素进行分层，并通过生成的决策树得到问题二中各个企业的信誉等级与违约情况。接着按照问题一中的**单目标优化模型与贷款额度分级模型**，对企业的贷款年利率与实际贷款额度进行计算。

针对问题三，我们首先根据资料将突发事件（如：新冠肺炎疫情）下各个行业受到的负面影响进行量化并代入模型中，将负面影响指数对客户流失率以及违约率进行调整；同时，考虑到国家对中小微企业的扶持，对银行的贷款年利率作适当的调整。两个因素共同作用对银行的放贷策略产生影响。

**企业贷款额度评级**是贷款投放的重要依据条件，本文使用 matlab 对数据进行预处理后对企业的资金需求量，以及银行的预计收益进行了预测，根据预计收益值进行贷款额度的确定并将所有企业的实际贷款额度存放在支撑材料中的 result.xlsx。

**关键词：**信贷策略；数据挖掘；主成分分析；机器学习决策树；贷款额度评级

## 1 问题重述

在实际中，由于中小微企业规模相对较小，也缺少抵押资产，因此银行通常是依据信贷政策、企业的交易票据信息和上下游企业的影响力，向实力强、供求关系稳定的企业提供贷款，并可以对信誉高、信贷风险小的企业给予利率优惠。银行首先根据中小微企业的实力、信誉对其信贷风险做出评估，然后依据信贷风险等因素来确定是否放贷及贷款额度、利率和期限等信贷策略。

某银行对确定要放贷企业的贷款额度为 10-100 万元；年利率为 4% – 15%；贷款期限为 1 年。附件 1-3 分别给出了 123 家有信贷记录企业的相关数据、302 家无信贷记录企业的相关数据和贷款利率与客户流失率关系的 2019 年统计数据。该银行请你们团队根据实际和附件中的数据信息，通过建立数学模型研究对中小微企业的信贷策略，主要解决下列问题：

- (1) 对给出信贷记录的 123 家企业的信贷风险进行量化分析，给出该银行在年度信贷总额固定时对这些企业的信贷策略。
- (2) 在问题 1 的基础上，对未给出信贷记录的 302 家企业的信贷风险进行量化分析，并给出该银行在年度信贷总额为 1 亿元时对这些企业的信贷策略。
- (3) 企业的生产经营和经济效益可能会受到一些突发因素影响，而且突发因素往往对不同行业、不同类别的企业会有不同的影响。综合考虑附件 2 中各企业的信贷风险和可能的突发因素（例如：新冠病毒疫情）对各企业的影响，给出该银行在年度信贷总额为 1 亿元时的信贷调整策略。

## 2 问题分析

### 2.1 问题一的分析

问题一要求我们对附件 1 中 123 家企业的信贷风险进行量化分析，给出该银行在年度信贷总额固定时对这些企业的信贷策略。我们可以建立以银行自身收益最大化为目标的单目标优化模型，并以此为参考进行放贷。影响银行收益的因素包括企业预计贷款额度与贷款利率，在加入风险因素时应主要考虑客户流失率以及企业违约率。我们首先利用 matlab 进行数据的初步处理并剔除事实性倒闭的企业（2019-2020 期间没有发票产生的公司），预测出企业下一年可能需要的资金作为可能贷款金额。然后通过不同行业和不同信誉等级共同影响的违约率计算出具体企业违约率，将其代入我们建立的单目标优化模型中，选取银行利润率最高时的贷款利率作为最优利率。再以银行向企业贷款预计产生收益为指标做出企业贷款额度分级，并将信贷总额根据各企业的贷款额度级数等比投放，最后联系各企业的贷款利率得到银行针对每一企业的信贷策略。

## 2.2 问题二的分析

问题二要求我们在问题一的基础上，对附件 2 中 302 家企业的信贷风险进行量化分析，并给出该银行在年度信贷总额为 1 亿元时的信贷策略。比较第一题和第二题的差异，主要在于第二题的 302 家企业没有信贷评级与信用记录，并且给定了 1 亿元的信贷总额。因此，我们首先将 123 家企业的信誉等级与是否违约进行统一量化，得到具体风险等级，然后用 SPSS 进行主成分分析，得到影响风险等级的 7 个主要成分因素。利用机器学习决策树的算法，将附件 1 中 123 家已有信贷风险企业的风险等级与 7 个主要成分因素进行运算，得到求解风险等级的决策树，再将 302 家没有信贷记录企业的同样属性的 7 个主成分代入决策树中，得到各自的风险等级。接着按照第一题的方法进行信贷策略商定，并根据信贷总额 1 亿元，进行信贷策略调整。

## 2.3 问题三的分析

问题三要求在第二问的基础上，给出在突发事件下银行对信贷策略的调整，以新冠疫情为例，在疫情发生后，各行业都受到了不同程度的打击，由于行业属性的不同，受到的影响程度也不同，将疫情对各行业的负面影响转换为量化指标来反映负面影响的大小。在我们建立的模型中，负面影响主要作用于客户流失率和违约率，由此对银行的最终收益产生影响。另一方面，在疫情发生后，央行鼓励银行向中小企业提供低息贷款，因此在求得银行贷款最优利率后应根据中央的指导意见进行下调利率的操作。两个因素共同影响了银行向各中小企业贷款的额度与策略。

## 3 问题假设

- (1) 假设在企业运营过程中当年所有留存资金都用于下一年的继续投资
- (2) 假设所有企业均以持续稳定运营为目标
- (3) 假设所有评级为 D 的企业均不提供任何贷款
- (4) 假设两年内进销项发票数量极少或为 0 的企业濒临或已经破产，不提供贷款
- (5) 假设经济持续稳定发展
- (6) 假设所有票务数据均真实可靠

## 4 符号说明

符号	意义
$R$	银行贷款预期收益率
$int$	银行一年期贷款利率
$a$	客户流失率
$d_k$	企业违约率
$i$	企业所处行业
$g$	企业信誉评级
$G$	银行给企业的贷款额度级别
$d_i$	某行业平均贷款违约率
$d_g$	某信誉评级的企业平均贷款违约率
$n$	营运资金周转次数
$p$	银行预期利润
$N$	营运资金量
$S$	销售收入
$SR$	销售利润率
$SRI$	预计销售收入年增长率
$SC$	销售成本
$X$	销项发票总金额
$J$	进项发票总金额
$f$	疫情为行业带来的负面影响指数
$A$	企业实际获得贷款额

## 5 问题的建模与求解

### 5.1 模型准备

#### 5.1.1 银行向中小微企业贷款考察因素

贷款一般是银行的主要业务之一，也是银行的主要盈利业务之一。贷款在本质上是一个商业行为，在银行发放贷款时，面对大型企业与小微企业有着不同的贷款定价策略，由于小微企业在其他条件相同的时候意味着更严重的信息不对称，不透明，因此如果只有市场行为参与这一调节过程，则银行可以通过提高利率来对冲小微企业由于信息不对称带来的高风险，但小微企业在面对商业银行提高贷款利率时同样拥有选择权，业务状态良好的企业可以选择拒绝高息贷款转而寻找优质投资，而业务状态较差的小微企业为了企业的生存就有可能选择接受高息贷款，并为了偿还高息贷款进行一系列高风险投机行为，由此产生恶性循环，最终导致银行的“垃圾贷款”增多，银行的根本利益受到侵犯。因此在对小微企业进行信誉评级时，企业规模，业务状态(数量，履约情况，业务

金额等), 运营状态均是重要的参考因素 [4]。

贷款业务的核心目的是盈利, 在贷款业务发生的过程中各种可能损害银行利润的因素都是负面因素。例如, 过高的贷款利率会引起客户的流失, 而过低的贷款利率会直接导致收入降低。因此仅针对风险提高或降低利率不能追求利润的最大化, 针对不同状态的小微企业, 银行都有着不同的最优利率。

目前我国商业银行主要以行业为划分依据确定信贷政策, 在过去的很长时间内, 大量的贷款投放主要集中在基建, 制造, 通讯设备等行业, 这代表了我国商业银行的主要投资习惯, 但一成不变的投资习惯导致了银行贷款结构的不合理, 风险控制不完善等问题。[8] 近些年来, 银行的风险评估系统逐渐完备, 贷款的投放行业也逐渐丰富起来, 互联网企业等新兴产业在疫情期间的强势崛起也让资本重新正视了互联网经济的潜力。在投放贷款时, 也需要对企业所处行业的前景各上下游企业进行评估, 最终决定放贷额度与利率等策略。

### 5.1.2 数据预处理

附件一和附件二分别给出一批企业的基本信息, 信贷状况和票据状态等, 附件三提供了不同利率下各中评级企业的客户流失概率。但并不是所有给出的数据都能直接用于分析。由于大量数据中存在的缺失值, 极端值, 无效数据等, 直接分析会使结果偏差较大。因此在对各企业的风险评级进行量化分析之前, 需要对给出数据进行一定的数据预处理。

#### Step 1: 破产与倒闭企业的剔除

在 2019 年 1 月 1 日之后进项或销项发票小于等于一的企业因为没有任何缴税记录, 我们认为这样的企业已经事实性倒闭 [5], 在贷款投放策略中不予考虑(不投放贷款), 在之后的分析中这些企业的相关数据也要进行剔除。

#### Step 2: 确定负数发票以及作废发票的处理方案

负数发票是由于原定交易完成后, 客户对产品不满意退款退货产生的发票, 对于这样的票据, 除了特定的统计数据, 其余处理数据的过程中应对这样的发票进行剔除; 作废发票是在开具发票后客户因故取消了交易活动因而产生的票据, 一般不对其进行分析, 但短时间内出现大量的作废发票也应引起投资者的警觉, 这可能意味着该企业的产品质量出现重大问题。

#### Step 3: 对数据进行基本的统计性描述

综合判断企业的生产状况时需要用到很多有关企业经营基础性的统计描述, 例如某企业某年销售总收入, 有效发票张数, 作废发票比率等, 这些指标附件中未明确给出, 但可以经过一系列的基础操作求得备用。

指标	指标值	指标	指标值
企业代号	E9	销项发票作废占比	0.024720623
企业名称	*** 生活用品服务有限公司	销项发票总税额	49819902.51
信誉评级	3	销项发票总金额	319592135.4
违约量化	0	总增值税税额	48022003.97
进项发票有效开票次数	4199	2019 销项发票有效开票次数	1875
进项发票作废次数	94	2019 进项发票有效开票次数	1585
进项发票作废占比	0.02189611	2019 销项发票总金额	115300066.9
进项发票总税额	1797898.54	2018 销项发票总金额	122031761.7
进项发票总金额	24535615.55	2019 进项发票总金额	8163648.98
销项发票有效开票次数	5760	2019 进项税额	520084.06
销项发票作废次数	146	2019 销项税额	14187989.71
行业	消费品	2019 增值税额	13667905.65

上表是以代号为 E9 的企业为例列出的一些统计性描述指标，其中银行对企业的信誉评级 (A,B,C,D) 为了便于分析统一转换为数字 (A=3,B=2,C=1,D=0); 违约指标量化则使用 (0-1) 进行区分，具体对应关系为：0-无违约记录，1-有违约记录；进，销项发票有效开票次数与总金额反映了企业近几年的业务状况是否良好，现金流状况是否健康；进项发票作废占比反映了企业在商业行为上的诚信指数；销项发票作废占比反映了外界对企业产品的承认度或企业产品的质量指数 (作废占比越高质量指数越低)；2018 与 2019 的年度销项发票总金额可用于分析企业近年的发展趋势，也就是企业发展的前景预测；单年度的增值税总金额表示了企业的盈利规模。

### 5.1.3 预计未来销售成本

根据银行业监督管理委员会 2010 年发布的《流动资金贷款管理暂行办法》剔除，贷款需求量应基于借款人日常运营企业时的流动资金缺口确定，可由下面的参考公式计算：

$$N = \frac{S_{2019} \cdot (1 - SR_{2019}) \cdot (1 + SRI)}{n} \quad (1)$$

即：

营运资金量 =

上年度销售收入 × (1-上年度销售利润率) × (1+ 预计销售收入年增长率) / 营运资金周转次数 [6]

在联系本体数据状况和估算需求后，营运资金周转次数并不能依据现有数据计算得出，因此我们将这一公式简化后得到了用于预测未来一年的销售成本公式：

$$SC_{2020} = S_{2019} \cdot (1 - SR_{2019}) \cdot (1 + SRI) \quad (2)$$

即：

未来一年的销售成本 =

$$2019 \text{ 年销售收入} \times (1 - \text{上年度销售利润率}) \times (1 + \text{预计销售收入年增长率})$$

上式中，上年度的销售收入我们用企业 19 年的销项发票金额总和进行描述，上年度销售利润率 = (2019 年销项发票总金额 - 2019 进项发票总金额) / 2019 进项发票总金额，预计销售收入年增长率 = (2019 年销项发票总金额 / 2018 年销项发票总金额) - 1。

## 5.2 问题一的分析

问题一要求在一定的信贷总额下建立信贷投放模型，确定向中小企业的信贷策略。银行发放贷款的最终目的是在控制风险的前提下利益最大化，因此首先以银行预计收益率最高为目标，计算出在企业信誉等级和对应行业双重影响下的最优贷款利率，并在使用 matlab 整理附件一的票务数据后计算出各企业下一年的预计资金需求量，假设资金需求全部由贷款补足的话，计算出银行与每个企业完成贷款业务后的预计收益，将企业按照预计收益规模和信用评级做出贷款额度分级，并将信贷总额根据各企业的贷款额度级数等比投放，最后联系各企业的贷款利率得到银行针对每一企业的信贷策略。

## 5.3 问题一的模型建立

### 5.3.1 针对各企业的最优贷款利率

题目中提到，银行提供的信贷总金额是一定的，贷款的本质是商业行为，最终的目的是获利，因此在一定的信贷总额前提下，信贷策略的目标即为在考虑风险的前提下追求最大的利润率。

针对每个企业不同的企业状况，银行选择不同的贷款利率时会造成不同程度的客户流失，客户流失率  $a$  为由离散函数  $a(R,g)$  决定，其中  $R$  为贷款利率， $g$  为企业信誉评级。当银行追求最终贷款收益率的最大化时，其目标为：

$$\max R = [int \cdot (1 - a)] \cdot (1 - d_k) \quad (3)$$

上式中，银行最终收益率等于银行贷款利率乘以客户留存率与客户履约率的乘积。客户留存率决定了企业能否接受这样的利率，在一定的贷款利率条件下是否会放弃该银行贷款，在式中表示为  $(1-a)$ ；客户履约率 决定了在将贷款投放后，银行在一年内能收

回贷款利息的概率，在式中表示为  $(1 - d_k)$ ,  $d_k$  的计算公式如下：

$$d_k = \frac{d_i + d_g}{2} \quad (4)$$

$d_i$  为行业贷款违约率， $d_g$  为某评级企业贷款违约率。各行业以及各评级企业的违约率数据均来源于标普关于 2019 年度全球企业违约和评级的过渡研究。[3] 其中各行业和企业信誉评对应的违约率如下：

	Rating	A, N	A, Y	B, N	B, Y	C, N	C, Y	D
<b>Global Corporate Default Rates By Industry (%)</b>		<b>0</b>	<b>0.0002</b>	<b>0.0005</b>	<b>0.0016</b>	<b>0.0061</b>	<b>0.0333</b>	<b>0.2708</b>
<b>Aerospace/automotive/capital goods/metal</b>	<b>0.0167</b>	<b>0.00835</b>	<b>0.00845</b>	<b>0.0086</b>	<b>0.00915</b>	<b>0.0114</b>	<b>0.025</b>	<b>0.14375</b>
<b>Consumer/service sector</b>	<b>0.026</b>	<b>0.013</b>	<b>0.0131</b>	<b>0.01325</b>	<b>0.0138</b>	<b>0.01605</b>	<b>0.02965</b>	<b>0.1484</b>
<b>Energy and natural resources</b>	<b>0.0384</b>	<b>0.0192</b>	<b>0.0193</b>	<b>0.01945</b>	<b>0.02</b>	<b>0.02225</b>	<b>0.03585</b>	<b>0.1546</b>
<b>Financial institutions</b>	<b>0.006</b>	<b>0.003</b>	<b>0.0031</b>	<b>0.00325</b>	<b>0.0038</b>	<b>0.00605</b>	<b>0.01965</b>	<b>0.1384</b>
<b>Forest and building products/homebuilders</b>	<b>0.0116</b>	<b>0.0058</b>	<b>0.0059</b>	<b>0.00605</b>	<b>0.0066</b>	<b>0.00885</b>	<b>0.02245</b>	<b>0.1412</b>
<b>Health care/chemicals</b>	<b>0.0156</b>	<b>0.0078</b>	<b>0.0079</b>	<b>0.00805</b>	<b>0.0086</b>	<b>0.01085</b>	<b>0.02445</b>	<b>0.1432</b>
<b>High technology/computers/office equipment</b>	<b>0.0136</b>	<b>0.0068</b>	<b>0.0069</b>	<b>0.00705</b>	<b>0.0076</b>	<b>0.00985</b>	<b>0.02345</b>	<b>0.1422</b>
<b>Insurance</b>	<b>0.0012</b>	<b>0.0006</b>	<b>0.0007</b>	<b>0.00085</b>	<b>0.0014</b>	<b>0.00365</b>	<b>0.01725</b>	<b>0.136</b>
<b>Leisure time/media</b>	<b>0.0144</b>	<b>0.0072</b>	<b>0.0073</b>	<b>0.00745</b>	<b>0.008</b>	<b>0.01025</b>	<b>0.02385</b>	<b>0.1426</b>
<b>Real estate</b>	<b>0</b>	<b>0</b>	<b>0.0001</b>	<b>0.00025</b>	<b>0.0008</b>	<b>0.00305</b>	<b>0.01665</b>	<b>0.1354</b>
<b>Telecommunications</b>	<b>0.028</b>	<b>0.014</b>	<b>0.0141</b>	<b>0.01425</b>	<b>0.0148</b>	<b>0.01705</b>	<b>0.03065</b>	<b>0.1494</b>
<b>Transportation</b>	<b>0.0068</b>	<b>0.0034</b>	<b>0.0035</b>	<b>0.00365</b>	<b>0.0042</b>	<b>0.00645</b>	<b>0.02005</b>	<b>0.1388</b>
<b>Utility</b>	<b>0.0032</b>	<b>0.0016</b>	<b>0.0017</b>	<b>0.00185</b>	<b>0.0024</b>	<b>0.00465</b>	<b>0.01825</b>	<b>0.137</b>

图 1: 违约率数据收集

在这一图表中，rating 列表示仅受行业影响时违约率的值，即  $d_i$ , Global Corporate Default Rates By Industry 行表示仅受企业信誉状况影响时的信誉，即  $d_g$ ，例如 A,N 列表示某企业信誉评级为 A 且没有违约记录。

### 5.3.2 企业收益评级

在模型准备过程中，我们估计了每个企业的下一年的销售成本。假设企业所有的销售成本全部来自银行贷款时，可以用银行贷款乘以银行针对各企业的最大利润率得到银行预期获得的利润。

$$p = SC \cdot R_k \quad (5)$$

得到各企业给银行带来的预期收益后，根据预期收益的量级进行各企业贷款额度的制定。

### 5.3.3 模型总结

根据银行贷款业务的目标与约束条件，通过联立上述条件可以得到下面的单目标优化模型：

$$\begin{aligned} \max R &= [int \cdot (1 - a)] \cdot (1 - d_k) \\ s.t \left\{ \begin{array}{l} SC_{2020} = S_{2019} \cdot (1 - SR_{2019}) \cdot (1 + SRI) \\ p = SC \cdot R_k \\ d_k = \frac{d_i + d_g}{2} \end{array} \right. \end{aligned} \quad (6)$$

## 5.4 问题一的模型求解

使用 matlab 对附件一中给出的票据进行初步的统计并完成统计性描述。并将这些企业分类为航空航天/汽车/生产资料/金属材料消费品/服务，能源与自然资源，金融机构，农林牧渔/建筑业，医疗/化工，技术/信息/办公用品，个体经营，娱乐/媒体，房地产，通讯，交通物流，公共设施等十三个行业，，处理结果如图：

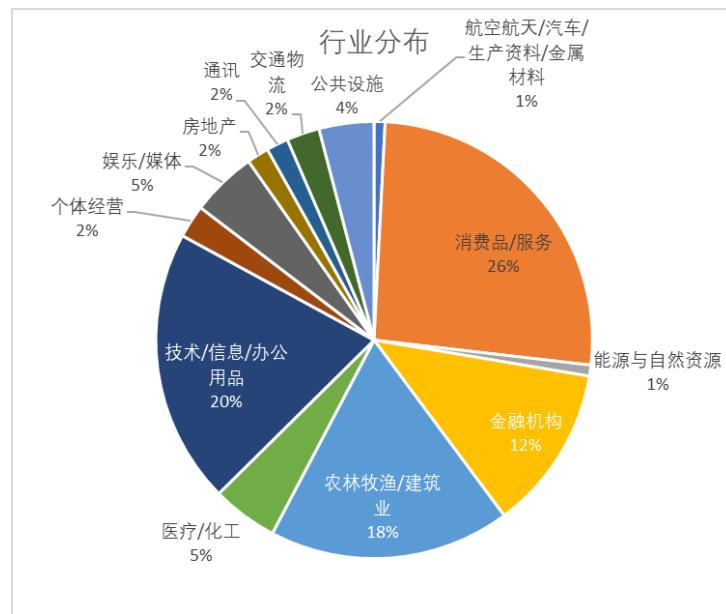


图 2: 123 家企业所属行业分布图

得到企业分类后，将其与企业评级进行联立，得到行业信誉评级状况，

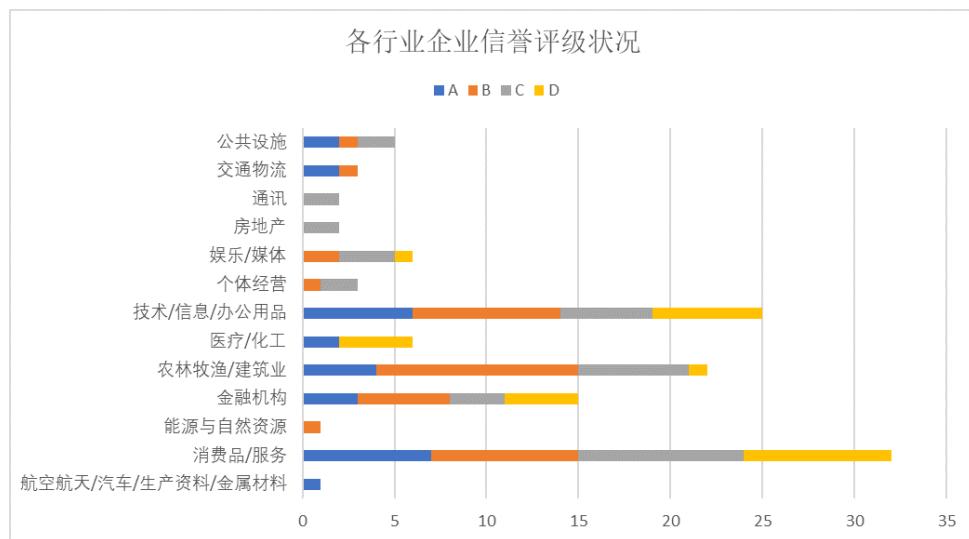


图 3: 各行业企业信誉评级分布

从图中我们可以得出，消费品/服务业的企业信誉普遍较好，而通讯，房地产等行业则极易产生信誉评级较差的状况，因此在投放贷款时，整体信誉较差的行业会得到更少的信贷额度。

使用 matlab 求出银行对各个行业各类评级企业的最优贷款利率，经过计算可得，行业在第一题的条件下对企业的最优利率影响几乎为 0，因此在后续的计算中同一评级的企业均采用同一最优贷款利率。

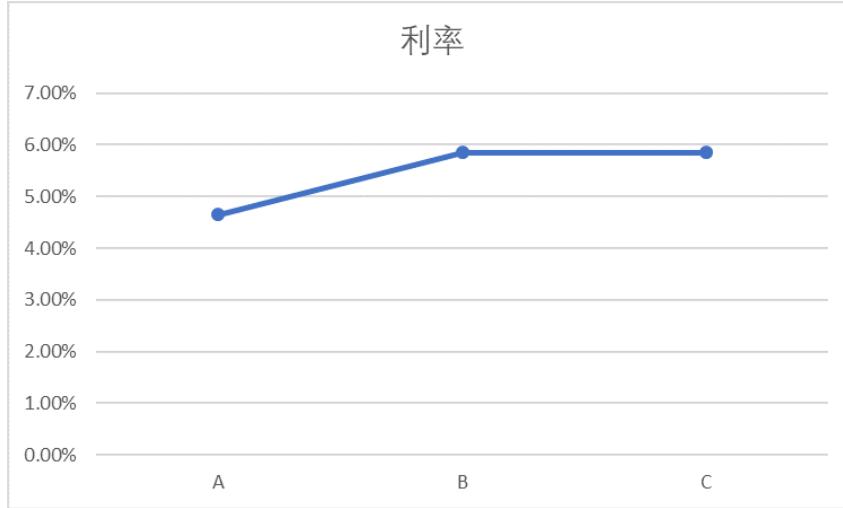


图 4: 不同评级下银行对企业贷款的最优利率

从折线图中可以得出，对于信誉评级为 A 的企业，银行最优贷款利率为 4.65%，对于评级为 B 或 C 的企业，银行最优贷款利率为 5.85%.

在具体确定银行信贷策略时，相关的数据处理流程可以由数据流程图进行表示，如下：

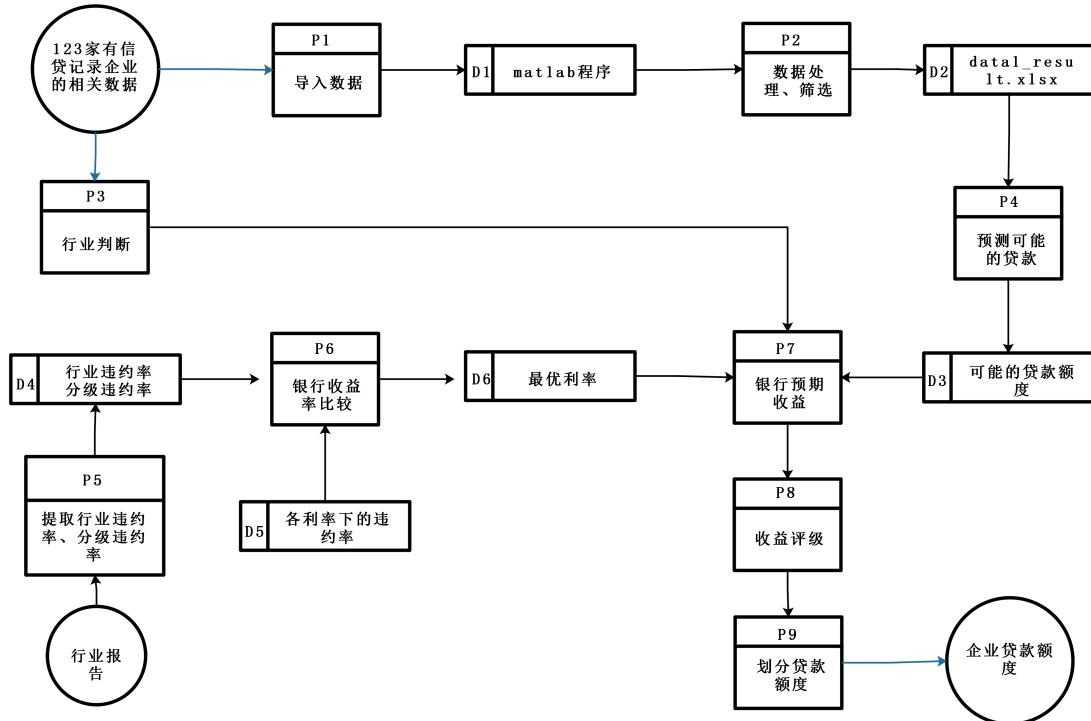


图 5: 问题一求解数据流程图

在数据预处理过程中，我们发现部分企业资金量过于庞大，对最终的评级有较大影响，因此我们将银行预期收益高于 500W 的企业提前设定为最高贷款额度等级后，剔除这部分企业的相关数据。计算出各企业能为银行带来的预期收益，并将预期收益和信誉评级作为决定银行投放贷款额度评级的判断标准建立决策树。

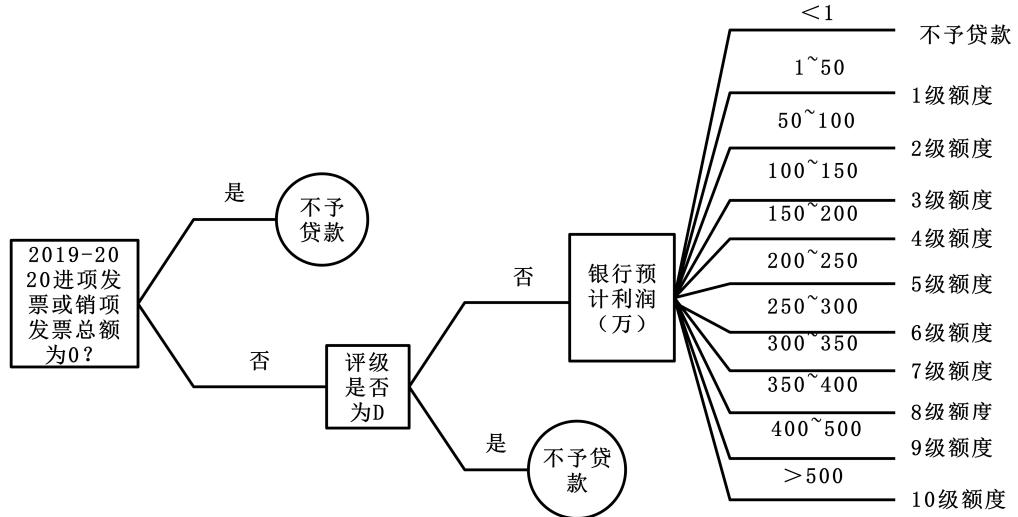


图 6: 贷款额度分级决策树

问题一说明了银行信贷总量是不变的，我们假设信贷总额以字母  $T$  表示，则在实际分配贷款投放策略时，我们假设每个企业对应的信贷额度等级为  $G_k$ ，我们将银行的信贷总额按企业的信贷额度等级等比例分配，即：

$$A_k = \frac{G_k}{\sum_{i=1}^k G_k} \cdot T \quad (7)$$

将各企业的信贷额度等级代入后即可得到所有企业实际获得贷款金额  $A_k$ ，从结果中我们可以看出，共有 87 个企业获得了贷款资格，其中 44 个企业获得一级贷款金额，即  $0.00333T$  元，11 个企业获得 10 级贷款金额，即  $0.03333T$  元。企业未获得贷款资格的原因主要有二：一是信誉评级为 D 或企业发生事实性倒闭；二是企业预计为银行带来收益小于 10000 元。获得最高级贷款的原因也主要分为两种，一是企业年流水金额过大；二是为银行带来的预计收益大于 500 万元，两者有所重复，但从不同方面反映了这样的企业规模大，前景好的特点。

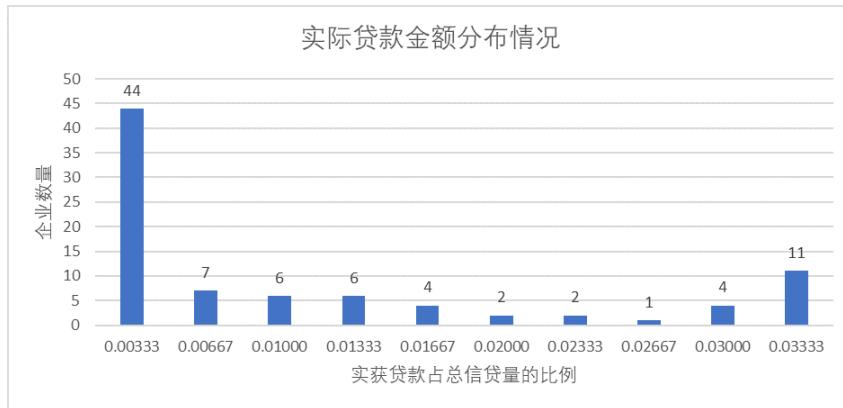


图 7: 企业实获贷款金额分布

注: 通过本题建模运算出的所有相关企业的信贷额度均存放于支撑材料下的 result.xlsx

以企业代号为 E9 的企业举例, E9 信誉评级为 A, 无违约记录, 因此银行贷款利率为 4.65%; 预计下一年的资金需求量为 42458805.05 元, 能为银行带来的预计收益为 1684180.857, 贷款额度评级为 4 级, 实际获得贷款额度为 0.013333333T。

## 5.5 问题二的分析

问题二是基于问题一的一个拓展, 首先, 附件二中的 302 家企业没有进行企业信用评级, 我们需要利用已有数据进行判断与分析。首先, 从经过统计性描述的附件一中的 22 个指标值中挑出 12 个数据重复率较低且有可能作为信誉评级参考的指标, 使用 SPSS 软件对这 12 个指标进行主成分分析后, 得到 7 个与信誉评级相关度最高的指标。使用机器学习的决策树算法, 以附件一中的数据为训练集, 对 7 个主成分进分级并判断出附件二中企业的信誉评级状况和是否有违约记录。得到各企业信誉评级后, 引入第一题的模型, 可以解得银行针对每一企业的信贷策略。

## 5.6 问题二的建模与求解

### 5.6.1 主成分分析

模型准备过程中, 我们已经对附件一, 二的数据进行了预处理与统计性描述, 但数据预处理中我们为每个企业计算得到的 22 个指标中, 由于需要满足问题一的一些特定情境下的分析要求, 许多指标有着较高的数据重复度。因此在进行主成分分析之前, 首先要剥离一部分高数据重复度的指标。经过筛选, 我们最终选择了十二个相对数据重复度较低的指标作为主成分备用指标, 将相关数据与已知的企业信誉评级导入 SPSS 软件后, 分析相关系数矩阵特征向量的方差贡献率, 得到下表:

总方差解释						
成分	初始特征值			提取载荷平方和		
	总计	方差百分比	累积 %	总计	方差百分比	累积 %
进项发票有效开票次数	6.236	51.965	51.965	6.236	51.965	51.965
进项发票作废开票占比	1.699	14.158	66.123	1.699	14.158	66.123
进项发票总金额	1.470	12.253	78.376	1.470	12.253	78.376
销项发票有效开票次数	.997	8.307	86.683			
销项发票作废开票占比	.814	6.781	93.464			
销项发票总金额	.479	3.993	97.457			
总增值税税额	.184	1.536	98.992			
2019销项有效开票次数	.064	.537	99.529			
2019进项有效开票次数	.040	.336	99.865			
2019销项总额	.013	.106	99.971			
2019进项总额	.003	.022	99.993			
2019增值税税额	.001	.007	100.000			

图 8: 主成分分析表格

在所有的指标中，为了尽量不遗漏影响信誉评级的因素，我们取累计方差贡献率达到 99% 时的 7 个主要成分，分别是：进项发票有效开票次数，进项发票作废开票占比，进项发票总金额，销项发票有效开票次数，销项发票作废开票占比，销项发票总金额，总增值税税额。我们以这 7 个指标为基础进行 302 家企业的信誉评级 [7]。

### 5.6.2 机器学习与决策树

问题 2 的附件 2 中，有着 302 家无信贷记录的企业数据，根据第一题中我们建立的模型，如果想要对它们进行量化分析并给出银行的信贷策略，那么需要对它们的信誉评级与是否违约进行评估。上面我们已经对影响企业信誉评级和是否违约的因素进行的主成分分析，得出了 7 个主要影响企业信誉评级的因素。在这里，我们使用了机器学习之决策树算法，对信誉评级未知的 302 个企业进行评估。

机器学习是基于人类学习的习惯和模式进行总结，学习和预测的，其本质为一种函数映射。在本题中我们使用了 python 语言中的 scikit-learn 数据挖掘工具包进行辅助学习。决策树是一种树形结构，其中每个内部节点表示一个属性上的判断，每个分支代表一个判断结果的输出，最后每个叶节点代表一种分类结果，本质是一颗由多个判断节点组成的树。而机器学习中的决策树算法是一种监督学习算法，即对结果进行标记的算法。[1]

在使用 python 对 7 个主成分进行层级排序并对企业信誉评级进行评估前，首先将

附件 1 中已知的 123 家企业的数据进行整理收集，将信誉等级与是否违约转化为具体的等级数字，方便我们对其进行操作，如下表：

信誉等级/是否违约	否	是
A	8	7
B	6	5
C	4	3
D	2	1

我们把等级作为输出结果，将 7 个主要因素作为输入变量，通过计算它们的信息增益进行特征选择，从而生成离散型的分类决策树。

随机选取 22 家企业作为测试集，95 家企业作为训练集，通过训练集生成的决策树，对测试集进行评估，生成的决策树的准确率大多高于 95%。经过反复测试，我们选取准确率最高的决策树，作为判定附件二中企业信誉等级的决策树。

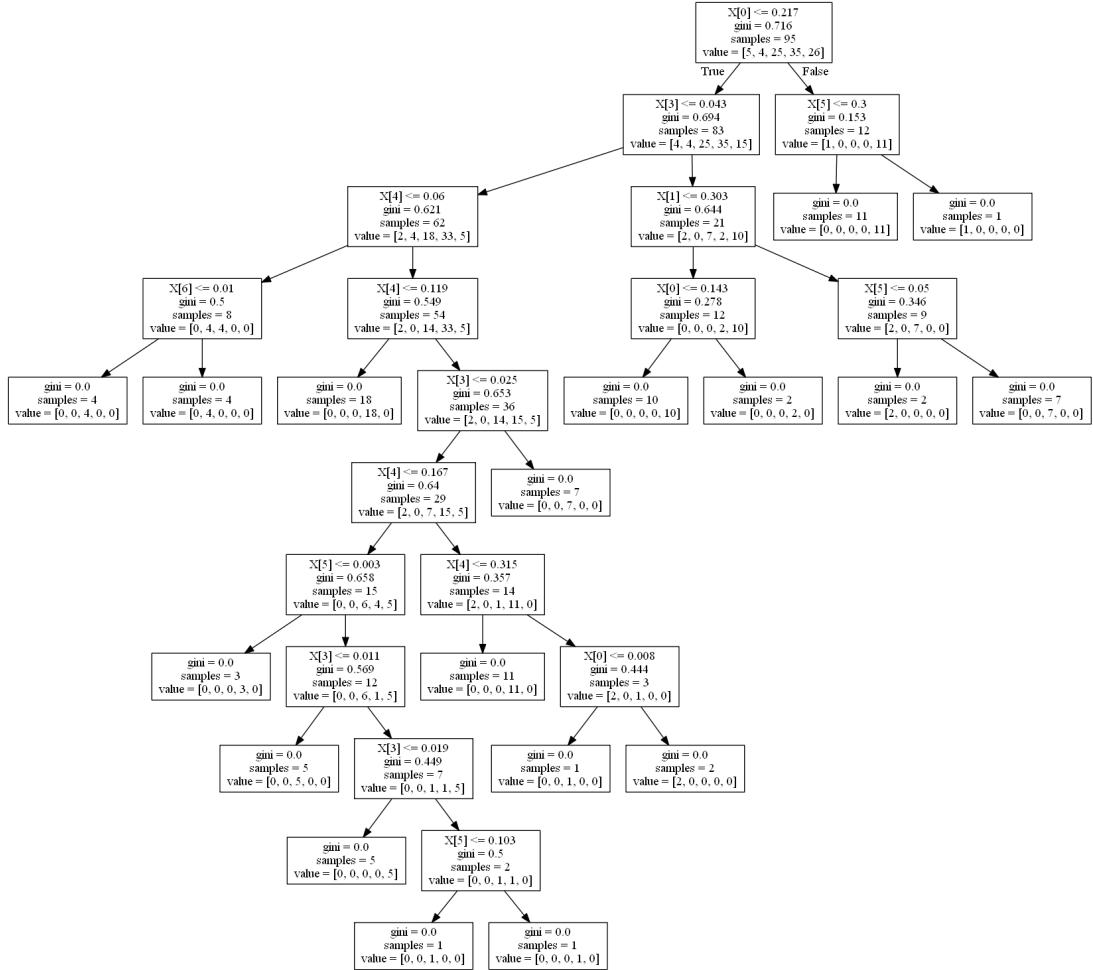


图 9: 机器学习决策树过程示意图

由图 10 可以看出，D 等级的企业全都存在违约记录，A、B 等级的企业没有任何违约记录，因此经过机器学习后仅存在五种等级数字，这五个数字分别对应信誉评级为 A 且不违约、等级为 B 且不违约、等级为 C 且不违约、等级为 C 且违约、等级为 D 且

违约四种情况。其中决策树中的  $X$  为  $1 \times 7$  的向量，代表 7 个主要成本因素。

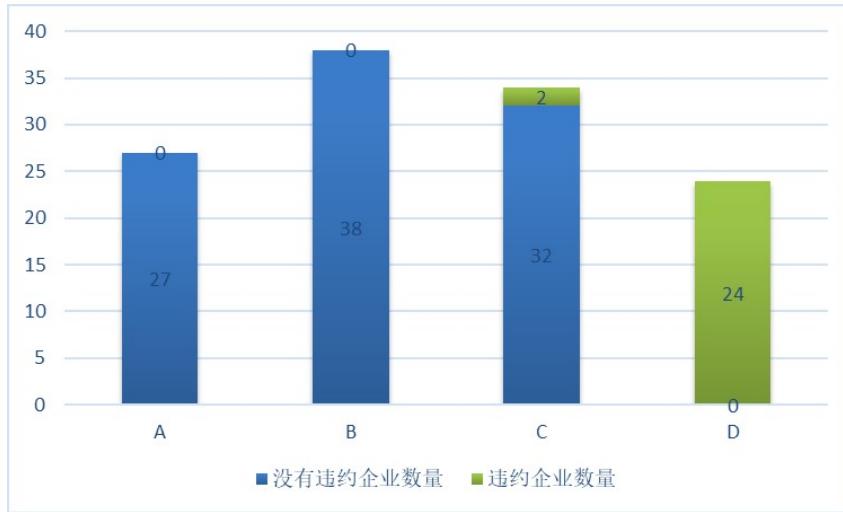


图 10: 各评级企业违约状况

将附件 2 中的 302 家企业进行数据处理，将每一家的 7 个因素代入决策树中，得到企业等级，根据我们设定的企业等级对应关系转化为信誉评级与是否有违约记录。

### 5.6.3 模型总结

在银行信贷策略的选择上，我们选择了与第一问相同的解题步骤，利用单目标优化模型得到在行业与信誉评级多重影响下的银行最优贷款利率，然后通过下一年的销售成本来预估企业下一年的贷款量与银行的预计收益。单目标优化模型如下：

$$\begin{aligned} \max R &= [int \cdot (1 - a)] \cdot (1 - d_k) \\ s.t &\left\{ \begin{array}{l} SC_{2020} = S_{2019} \cdot (1 - SR_{2019}) \cdot (1 + SRI) \\ p = SC \cdot R_k \\ d_k = \frac{d_i + d_g}{2} \end{array} \right. \end{aligned} \quad (8)$$

## 5.7 问题二的模型求解

相比起问题一的求解过程，问题二给定了信贷总金额，因此最终结果会直接以金额展示出来。问题二的求解过程可以用以下数据流程图进行表示：

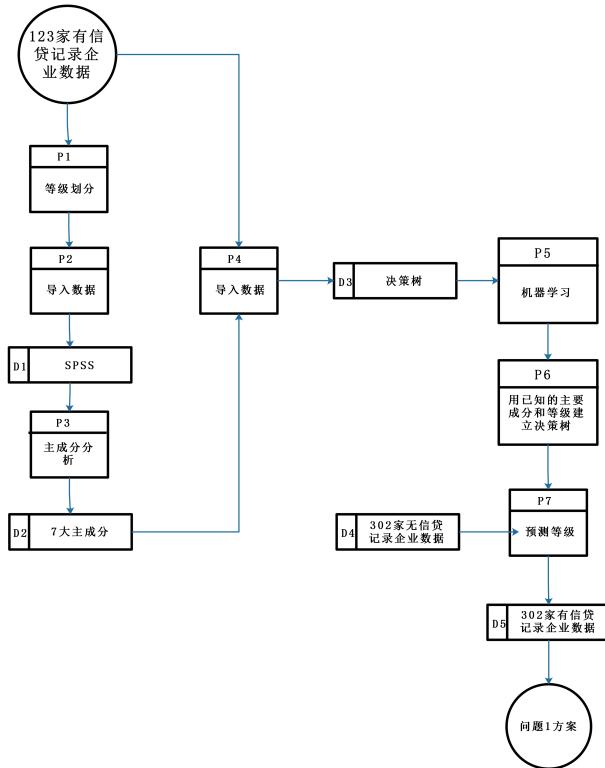


图 11: 问题二求解数据流程图

得到企业信誉状况分级后，首先以银行最终收益率最大化为目标计算得到各个行业各企业信誉评级所对应的最优贷款利率，经过比对后发现与第一问中求出的最优利率差距较小，因此采用与第一问相同的贷款利率策略。具体表示为：

当企业信誉评级为 A 时，贷款利率为 4.65%；

当企业信誉评级为 B 或 C 时，贷款利率为 5.85%。

确定最优利率后，估算每个企业的资金需求量与给银行带来的预期利润，按照预期利润对企业贷款额度进行分级，具体分级标准可参照以下决策树：

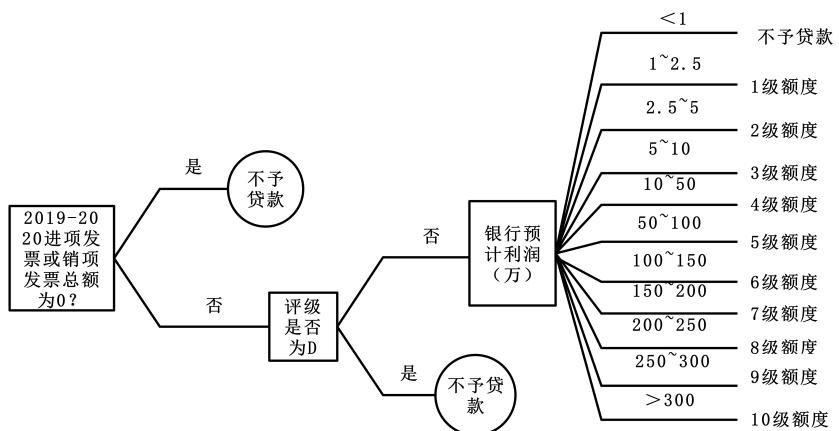


图 12: 贷款额度分级决策树

在根据贷款额度等级对信贷金额进行实际分配时，分配原理为：

$$A_k = \frac{G_k}{\sum_{i=1}^k G_k} \cdot T \quad (9)$$

$A_k$  为企业实际得到的贷款， $G_k$  为各企业的贷款额度评级， $T$  为信贷总额，在本题中即为 1 亿元。经过计算后，发现将所有资金按照贷款额度等级等比分配时，一部分企业的实际贷款金额小于 10 万元，因此我们将这些特殊企业提取出来，对实际贷款金额小于 10 万元的企业赋予 10 万元的贷款额度。

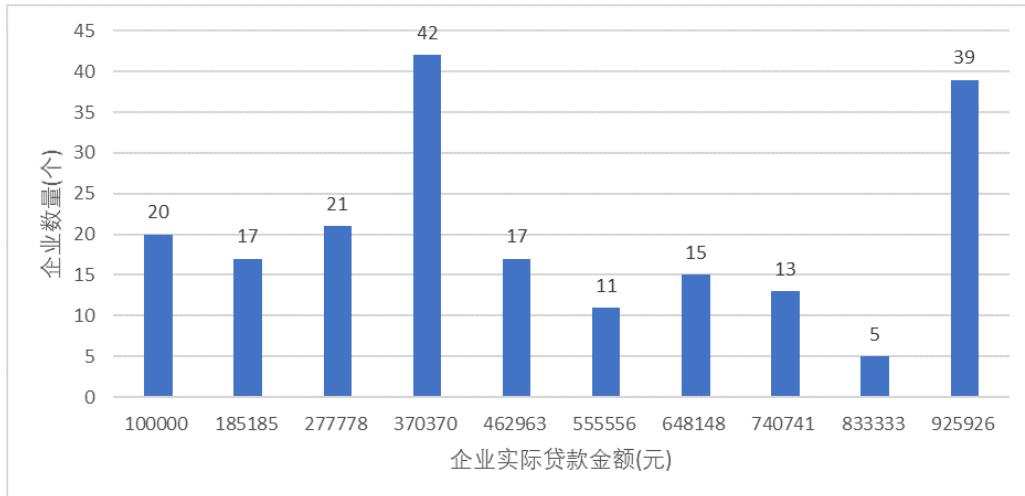


图 13: 各企业实际贷款金额分布

注：通过本题建模运算出的所有相关企业的信贷额度均存放于支撑材料下的 result.xlsx

从上图中我们可以得知，302 家企业中，仅有 200 家企业能够获得贷款资格，所有企业中得到的最大贷款额度为 925926 元，获得这一贷款额度的企业有 39 家，占有贷款资格企业数目的 19.5%。企业中获得最小的贷款额度为 100000 元，也就是银行预计贷款额度的下限，获得这一贷款额度的企业有 20 家，占有贷款资格企业数目的 10%。

以企业代号为 E125 的企业为例，经过机器学习算法的评估，E125 的信誉评级为 A，且没有违约记录。下一年 E125 的预估资金需求量为 1654872433 元，为银行带来的预估利润为 66507148.57 元，贷款额度评级为 10 级，实际获得的贷款额度为 925925.9259 元。

## 5.8 问题三的分析

问题三要求给出在突发事件下银行对信贷策略的调整，以新冠疫情为例，在疫情发生后，各行业都受到了不同程度的打击，由于行业属性的不同，受到的影响程度也不同，将疫情对各行业的负面影响转换为量化指标来反映负面影响的大小。在我们建立的模型中，负面影响主要作用于客户流失率和违约率，由此对银行的最终收益产生影响。另一

方面，在疫情发生后，央行鼓励银行向中小企业提供低息贷款，因此在求得银行贷款最优利率后应根据中央的指导意见进行下调利率的操作。两个因素共同影响了银行向各中小企业贷款的额度与策略。

## 5.9 问题三的建模与求解

由于行业属性的不同，新冠疫情对于各行业造成的影响也不尽相同，为了能够尽可能准确地描述负面影响的大小，我们选用了较为权威的标普公司发布的研究报告中的数据作为量化指标。<sup>[2]</sup> 将研报中提供的数据进行整理后，得到下图：

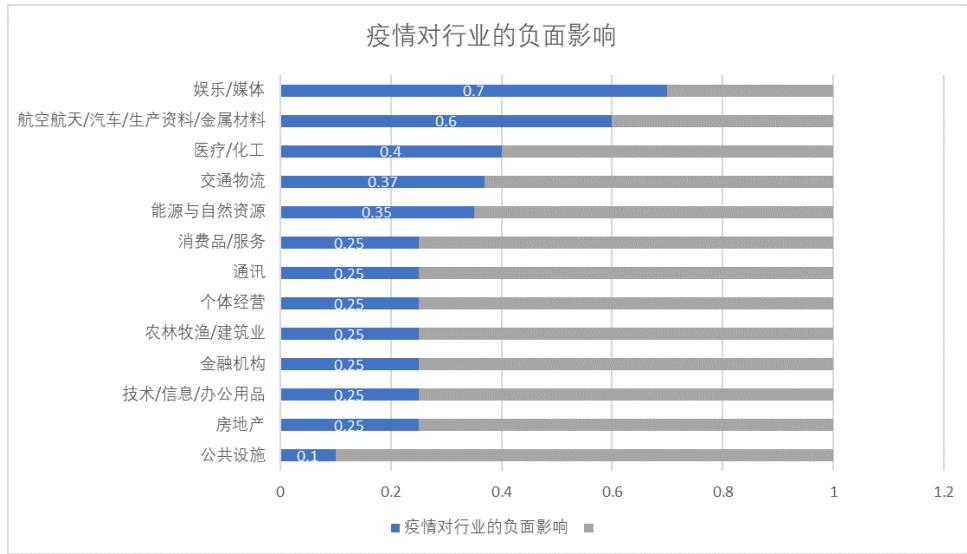


图 14: 疫情对行业造成的影响

疫情对各行业产生的影响主要作用在用户流失率与用户违约率上，疫情期间，许多业务无法开展，意味着企业将更有可能无法拥有健康的现金流，自然也就无法偿还银行的贷款本息，由此造成更高的违约率；而另一方面，疫情期间许多企业将没有营业收入，却需要维持着诸如房租之类的固定支出，由此会造成更高的企业倒闭概率，也就是客户流失率。因此，假设负面影响指数为  $f$ ，则经历疫情影响后的客户流失率则为  $a(1 + f)$ ，客户违约率为  $d_k(1 + f)$ 。

在建立以银行预期利润率最大化为目标的单目标优化模型时，将负面影响指数考虑进去之后则会得到：

$$\begin{aligned} \max R &= [int \cdot (1 - a(1 + f))] \cdot (1 - d_k(1 + f)) \\ s.t \quad &\left\{ \begin{array}{l} SC_{2020} = S_{2019} \cdot (1 - SR_{2019}) \cdot (1 + SRI) \\ p = SC \cdot R_k \\ d_k = \frac{d_i + d_g}{2} \end{array} \right. \end{aligned} \quad (10)$$

除此之外，为扶持中小企业，尽可能使企业在疫情中得以存活，我们对利率和贷款门槛都做出了一定调整。首先，依据中国人民银行发布的《2020 年第一季度货币政策执行报告》，分别对评级不同，信誉状况不同的企业进行降息，具体降息政策如下：

分级降息标准			
评级	是否违约	降息后利率	(政策优惠)
A	否	0.0425	=0.0465-0.004
	是	0.0465	
B	否	0.0505	=0.0585-0.008
	是	0.0545	=0.0585-0.004
C	否	0.0545	=0.0585-0.004
	是	0.0585	

图 15: 分级降息标准

然后，取消贷款审核时的预估盈利门槛，为尽可能多的企业提供疫情期间的低息贷款，修改后的贷款额度评级决策树如下：

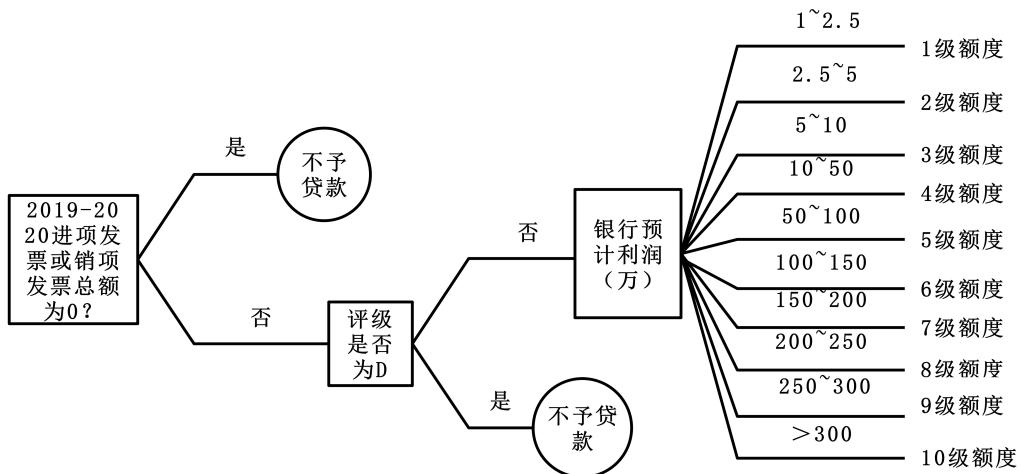


图 16: 取消准入门槛后的贷款额度评级决策树

经过针对贷款额度评级的信贷分配后，再将不足 10 万元贷款额度的贷款等级调整为 10 万元，最终企业实际得到贷款的情况分布为：

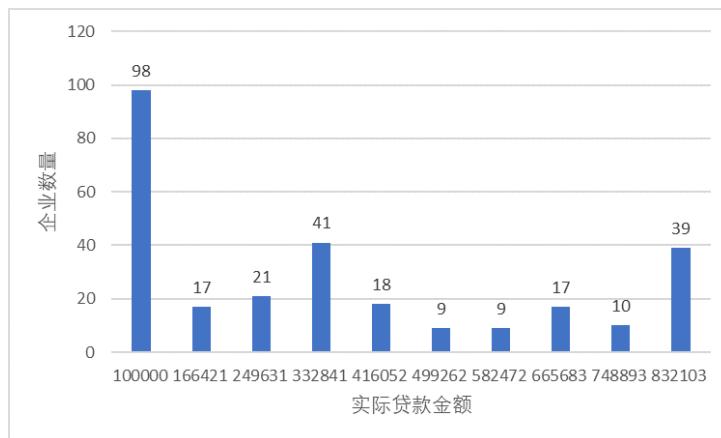


图 17: 疫情下实际贷款投放情况

注：通过本题建模运算出的所有相关企业的信贷额度均存放于支撑材料下的 result.xlsx

以企业代号为 E125 的企业为例，E125 的信誉评级为 A，且没有违约记录。下一年 E125 的预估资金需求量为 1654872433 元，为银行带来的预估利润为 66507148.57 元，政策调整后为银行带来的预期收益为 65559090.98 元，贷款额度评级为 10 级，政策调整前实际获得的贷款额度为 925925.9259 元，政策调整后获得的实际贷款额度为 832103.321 元

经过政策的放宽后，得到贷款资格的企业由第二问的 200 个增加到了 279 个，同比增长了 39.5%，其中获得最低档贷款额度（10 万元）部分的增长幅度最为显著，由调整政策前的 20 个增长为 98 个，同比增长了 390%。贷款资格获取门槛的降低，尽可能地帮助了更多的中小企业，为中国市场增添了一定活力。

## 6 模型评价与推广

### 6.1 模型评价

#### 6.1.1 模型优点

- 1 所建立的模型普适性强、考虑全面、思路严谨清晰、贴近现实。
- 2 模型采用专业的数学分析软件，例如 matlab、SPSS 和 excel 等软件进行数据的预处理与汇总，使得数据的集成性高，主成分因素可靠性强。
- 3 使用机器学习决策树的算法对 123 家有信贷记录的企业训练，建立决策树，然后对 302 家无信贷记录的企业进行信誉评估、预判违约情况，测试集的正确率大多位于 95% 以上，评估可信度高。
- 4 模型中的参数选择参考了专业评级机构的研究报告，权威性强。

#### 6.1.2 模型缺点

附件所给企业信息种类较少，因此银行预计贷款收益公式中的自变量相比于实际情况中影响因素的数量较少，考虑不够完全。

#### 6.1.3 模型的改进

- 1 对模型中贷款额度更加细化，以充分考虑各行业各企业的贷款情况。
- 2 加强对企业信息进行搜查力度，以便于更加深入的了解企业内部的具体情况，建立一个更加全面和准确的贷款策略评价模型。

## 6.2 模型的推广

本文是在固定银行的贷款总额前提下，通过对中小企业的信贷风险进行量化分析，在考虑突发因素的情况下，提出了各个企业的信贷策略。

本文建立的模型对现实中银行放贷的策略有着一定的参考价值和现实意义，并且可以将其推广到银行对大型企业的信贷策略中，同时也对国家在重大突发事件下进行宏观调控具有一定的参考意义。

## 参考文献

- [1] C. M. Bishop. *Pattern recognition and machine learning*. Springer, 2006.
- [2] X. Jean. Asia-pacific corporate and infrastructure midyear outlook 2020. Report, Standard and Poor's, 2020.
- [3] N. W. Kraemer. Default, transition, and recovery: 2019 annual global corporate default and rating transition study. Report, Standard and Poor's, 2019.
- [4] 于洋. 中国小微企业融资问题研究. 成都: 西南财经大学, 2013.
- [5] 侯琦. 企业倒闭的原因及预测. 党政干部学刊, (2):23, 1996.
- [6] 李唐瑜. 关于流动资金贷款额度测算的思考. 海南金融, (4):53–55, 2012.
- [7] 李小胜 and 陈珍珍. 如何正确应用 spss 软件做主成分分析. 统计研究, (8):105–108, 2010.
- [8] 陈华清. 我国商业银行小微企业信贷风险管理研究. 中小企业管理与科技 (中旬刊), (3), 2015.

## A 附录一 支撑材料清单及说明

- | 主目录
- | SPSS 主成本分析结果 //第二问的主成本分析
  - | 主成本分析.sav
  - | 主成本分析报告.spv
- | 查阅的数据
  - | 相关链接.docx //由于查阅资料比较大，因此我们将相关的链接放入 word 中
- | 结果
  - | 结果.xlsx //由于三道题的企业数目过多，正文放不下，因此我们将三道题的结果放入 excel 表中
- | 图表 //论文中所用到的图
  - | 插图
- | latex 代码 //论文 latex 源程序
  - | Paper.tex //latex 源程序
  - | Refe.bib //参考文献管理文件
- …//其余文件
- | 源代码
  - | dataprocess
    - | data\_process.m //数据预处理的代码，如果要运行请修改代码中文件的路径以及选择的区域
    - | machineLearning //机器学习决策树的代码文件夹
      - | Decision\_tree.py //机器学习 python 代码
      - | iris\_simple.doc //训练后决策树生成的 word 文件
      - | iris\_simple.png //训练后决策树的结构
    - | solveddata.xlsx //第二问 302 个无信贷记录的企业
    - | solveddatad.xlsx //通过决策树对 302 个无信贷记录的企业，生成的等级
    - | traindata.xlsx //第一问中训练的数据——123 个有信贷记录的企业

## B 附录二 程序代码

### B.1 数据预处理代码

```
%本程序是对附件一和附件二的企业信息进行预处理  
%统计出关于有效开票次数、作废开票次数、发票总金额等总体情况  
%适用于附件一与附件二，每次修改 excel 表的名字和读取区域范围  
即可完美运行
```

```

clc;clear;
[~,~,data11]=xlsread('data2.xlsx',1,'A2:D303');

[~,~,data12]=xlsread('data2.xlsx',2,'A2:H395176');

[~,~,data13]=xlsread('data2.xlsx',3,'A2:H330836');

startnum=1;
% endnum=211077;
endnum=length(data12(:,1));
for i=1:length(data11(:,1))
sum_moneny=0;%进项金额总数
sum_tax=0;%进项税额总数
count_effect=0;%进项有效开票次数
count_ineffect=0;%进项无效开票次数
outtax19=0;%2019进项税额
outcome19=0;%19年进项总额
count_effect19=0;%2019进项开票次数
for j=startnum:endnum
if strcmp(data11{i,1},data12{startnum,1})
if strcmp(data12{startnum,8}, '有效发票')
sum_moneny=cell2mat(data12(startnum,5))+sum_moneny;
sum_tax=cell2mat(data12(startnum,6))+sum_tax;
count_effect=count_effect+1;
end
else
count_ineffect=count_ineffect+1;
end

if cell2mat(data12(startnum,3))>=43466
&&cell2mat(data12(startnum,3))<43831
outtax19=outtax19+cell2mat(data12(startnum,6));
outcome19=outcome19+cell2mat(data12(startnum,7));
count_effect19=count_effect19+1;
end
else
count_ineffect=count_ineffect+1;
end

if cell2mat(data12(startnum,3))>43466

```

```

data11(i,11)=num2cell(1);
else
data11(i,11)=num2cell(0);
end
startnum=startnum+1;

else
break;
end
end
data11(i,5)=num2cell(count_effect);
data11(i,6)=num2cell(count_ineffect);
data11(i,7)=num2cell(sum_moneny);
data11(i,14)=num2cell(outtax19);
data11(i,16)=num2cell(outcome19);
data11(i,17)=num2cell(sum_tax);
data11(i,19)=num2cell(count_effect19);

end

startnum2=1;
endnum2=length(data13(:,1));

for i=1:length(data11(:,1))
sum_moneny=0;%销项金额总数
count_effect=0;%销项有效开票次数
count_ineffect=0;%销项无效开票次数
sum_tax=0;%销项税额总数
income19=0;%2019销项总额
income18=0;%18年销项总额
intax19=0;%2019销项税额
count_effect19=0;%2019销项开票次数
for j=startnum2:endnum2
if strcmp(data11{i,1},data13{startnum2,1})
if strcmp(data13{startnum2,8}, '有效发票')

```

```

%2019 销售总额， 税额
if cell2mat ( data13 ( startnum2 ,3))>=43466
&&cell2mat ( data13 ( startnum2 ,3))<43831
income19=income19+cell2mat ( data13 ( startnum2 ,7));
intax19=intax19+cell2mat ( data13 ( startnum2 ,6));%2019 销项税额
count_effect19=count_effect19+1;
end

%2018 销售总额
if cell2mat ( data13 ( startnum2 ,3))<43466
&& cell2mat ( data13 ( startnum2 ,3))>=43101
income18=income18+cell2mat ( data13 ( startnum2 ,7));
end

sum_moneyn=cell2mat ( data13 ( startnum2 ,5))+sum_moneyn;
sum_tax=cell2mat ( data13 ( startnum2 ,6))+sum_tax;

count_effect=count_effect+1;

else
count_ineffect=count_ineffect+1;
end

startnum2=startnum2+1;
else
if cell2mat ( data13 ( startnum2 -1,3))<43466
data11 ( i ,11)=num2cell ( 0);
end
break;
end
end

data11 ( i ,8)=num2cell ( count_effect );
data11 ( i ,9)=num2cell ( count_ineffect );
data11 ( i ,10)=num2cell ( sum_moneyn );
data11 ( i ,12)=num2cell ( income19 );
data11 ( i ,13)=num2cell ( income18 );
data11 ( i ,15)=num2cell ( intax19 );

```

```

data11(i,18)=num2cell(sum_tax);
data11(i,20)=num2cell(count_effect19);
end

result= xlswrite('data2_result.xlsx',data11,1,'A2');

```

## B.2 机器学习决策树代码

```

# 此代码是通过对123家企业进行训练，建立决策树，
并对302家无信贷记录的企业进行信誉等级和违约情况进行判定
# 本程序是在python3.6的环境下进行运行，
需要安装pydot、sklearn、pandas等一系列的库
# 可通过修改excel的名字进行运行

import numpy as np
import scipy as sp
from sklearn import tree
from sklearn.metrics import precision_recall_curve
from sklearn.metrics import classification_report
from sklearn.model_selection import train_test_split
import random
import xlrd
import pandas as pd
from sklearn.externals.six import StringIO
import pydot

data = xlrd.open_workbook('traindata.xlsx')
data=data.sheet_by_name('Sheet1')

# 待预测的数据
data2 = xlrd.open_workbook('solveddata.xlsx')
data2 =data2.sheet_by_name('sheet1')

data_solved=np.array(data2.row_values(1,5,12))
data_train=np.array(data.row_values(1,4,12))

```

```

for i in range(data2 .nrows):
    if i>=2 :
        data_solved= np .vstack (( data_solved ,
        np .array (data2 .row_values (i , 5 , 12))))


    # 待预测的数据
    for i in range(data .nrows):
        if i>=2 :
            data_train= np .vstack (( data_train ,
            np .array (data .row_values (i , 4 , 12))))


    # 训练集 95, 测试集 22
    set_num = 22
    train_num = 95


    # 归一化
    max=data_train .max( axis=0)
    min=data_train .min( axis=0)
    for i in range(len(max)):
        for j in range(data .nrows-1):
            if i>0:
                data_train [j , i]=( data_train [j , i]-min [i])/ (max [i]-min [i])


    # 待预测的数据
    for i in range(len(max)):
        for j in range(data2 .nrows-1):
            if i >0:
                data_solved [j , i-1]=( data_solved [j , i-1]-min [i])/ (max [i]-min [i])


    #打乱次序
    random .seed (10101)
    random .shuffle (data_train )
    data_train=np .array (data_train )

    set_data=data_train [0:set_num]
    train_data=data_train [-train_num :]

```

```

x_train=train_data[:,1:]
y_train=train_data[:,0]

# 待预测的数据
x_solved=np.array(data_solved)

# print(x_solved)

#建立模型
clf=tree.DecisionTreeClassifier()
clf=clf.fit(x_train,y_train)

x_set=set_data[:,1:]
y_set=set_data[:,0]

predict_test_y=clf.predict(x_set)

# y_solved=clf.predict(x_solved)

print(x_set)
print(predict_test_y)

# print(y_solved)

# 保存数据
# excelframe=pd.DataFrame(y_solved)
# excelframe.to_excel('solveddata.xlsx')

#画图
dot_data = StringIO()
tree.export_graphviz(clf, out_file=dot_data)
graph = pydot.graph_from_dot_data(dot_data.getvalue())
graph[0].write_dot('iris_simple.dot')
graph[0].write_png('iris_simple.png')

```