

基于 TOPSIS 和遗传算法的中小微企业信贷风险与信贷决策模型

摘要

针对问题一中信贷风险的量化分析,我们从企业实力、信誉水平、供求关系稳定性、上下游企业影响力四个方面选取六个衡量信贷风险的指标,并对其合理性、相关性进行了检验。根据指标权重计算得到 123 家企业的分类数据,与已有的可用来代表银行对企业信贷风险评级的信誉等级数据进行对比,以分类准确性为目标构建优化模型,借助遗传算法求出最优权重。再对已有的指标体系使用 TOPSIS 模型得到衡量企业信贷可靠程度的综合评价指标,剔除不予放贷的信誉等级为 D 级的企业数据后通过 k-means 聚类分为 I、II、III 三类。为简化模型,在假定每类中企业信贷额度、利率相同的前提下建立信贷策略的优化模型,拟合得到客户流失率随利率的变化关系后,以银行利润最大化为目标,在年度信贷总额确定的条件下借助遗传算法给出三类企业的额度、利率值的最优解。

基于上述模型,针对不含信誉等级、违约信息的附件 2 数据,采用相同处理流程,计算得到 302 个企业的指标数据及综合评价指标数值, k-means 聚类后分为四类,剔除不予贷款的 IV 类后建立信贷策略优化模型,以 1 亿元的信贷总额为约束条件构建优化模型,使用遗传算法求得三类企业的额度、利率值的最优解。

问题三中要求考虑突发因素影响对银行信贷策略进行调整。以新冠肺炎为例,通过查找新冠肺炎对 GDP 相关数据的影响发现其对于第一、二、三产业影响程度不同,首先对附件二中企业按照一、二、三产业进行分类,对特殊行业例如医疗、线上产业的企业单独分类。利用 2010~2019 年各产业 GDP 信息对不考虑疫情影响下的 2020 年 GDP 进行预测,与实际值之差即为疫情对各类产业的影响程度,利用影响系数预测出各类产业中企业指标,利用第一、二问中的评价体系与决策模型计算出最终的决策指标。对于普遍的突发因素,需要先进行分类与分级,建立突发性风险评价体系,预测突发因素对各类企业评价指标的影响,再利用决策模型计算出最终的决策指标。

模型检验中,我们首先进行评价体系对六项指标的灵敏度分析,六项指标数值从 5% 到 10% 变动时对评价结果影响不大,验证了评价模型的准确、稳定性。分析六项指标权重变化对评价结果的影响,发现 η_{gross} 、 var_{gross} 、 η_{del} 权重对评价结果影响不大,而 η_{neg} 、 ω_{total} 、 η_{big} 权重影响较大。

关键词: 信贷风险指标体系, 信贷策略优化, 遗传算法, TOPSIS, 聚类

目录

基于 TOPSIS 和遗传算法的中小微企业信贷风险与信贷决策模型	1
摘要.....	1
一、问题重述.....	4
二、问题分析.....	5
2.1 问题一：信贷风险综合评价与信贷策略选择.....	5
2.2 问题二：基于具体年度信贷总额数据的信贷策略选择.....	5
2.3 问题三：考虑突发因素条件下对信贷策略的调整.....	6
三、模型假设与参数定义.....	7
3.1 模型假设.....	7
3.2 参数定义.....	7
四、模型建立与求解.....	8
4.1 信贷风险综合评价体系与信贷策略优化模型.....	8
4.2 年度信贷总额给定时对缺少信誉评级企业的信贷策略的分析.....	24
4.3 考虑突发因素影响下信贷策略的调整模型.....	26
五、灵敏度分析.....	31
5.1 评价体系对六项评价指标的灵敏度分析.....	31
5.2 权重因子对评价结果的影响.....	32
六、模型评价与改进.....	33
模型优点：	33
模型缺点及改进：	33
参考文献.....	35
附录.....	36

一、问题重述

中小微企业规模相对较小，缺少抵押资产，营业收入有限，因此银行在向它们提供信贷服务时需要科学合理地分析它们各自的信贷风险，主要的参考依据包括信誉水平、供求关系、企业实力、上下游企业的影响力。银行需要以信贷风险评估的量化结果为依据，综合考虑多方面因素确定具体的放贷策略，包括是否放贷、贷款额度、利率和期限等。

（1）问题一要求我们利用附件 1 中数据量化 123 家企业的信贷风险，在年度信贷总额固定的条件下给出最优的信贷策略。

（2）问题二要求我们以问题一中建立的数学模型为基础，使用附件 2 中数据对 302 家企业的信贷风险进行量化，并给出年度信贷总额为 1 亿元的条件下相应的信贷策略。

（3）问题三要求我们考虑影响企业生产经营的突发因素对不同行业、类别的企业信贷风险的影响，并分析约束条件为 1 亿元时应对信贷策略做出怎样的调整。

二、问题分析

2.1 问题一：信贷风险综合评价与信贷策略选择

问题一要求我们完成两项任务，一项是根据附件 1 提供的 123 家企业的信誉、进销项流水信息对它们的信贷风险进行量化分析，另一项是限定年度信贷总额固定的条件下给出对这些企业的信贷策略。为了定量分析各企业的信贷风险，我们建立了与企业经营、企业信誉相关的可用来衡量信贷风险的指标体系。针对信誉水平、企业实力、供求关系稳定性、上下游企业影响力这四方面，我们选取了六个具体指标——毛利率和月均营业总额（月均销项金额）用来衡量企业影响力；年毛利率方差用来衡量供求关系稳定性；作废发票、负数发票比例（金额比例、单数比例的均值）用来衡量企业信誉水平；进销项中大单（范围限定为>1 万元）金额比例的平均值。利用 Python 和 Matlab 求出了 123 家企业的所有指标信息，并比较有无违约记录两类企业中各项指标的均值，发现二者有明显差异，验证了指标的合理、有效性。此外，我们使用 Spearman 相关系数对指标进行了相关性分析，结果显示指标间没有明显的相关性，可以共同构成综合指标评价体系。

考虑到附件中银行给各企业的信誉评分数据可以从很大程度上衡量企业的信贷风险且精度较好，建立求解各指标权重的优化模型，其中决策变量为指标各自的权重数值，利用该权重计算得到各企业的综合指标数值后，将企业分成与信誉等级中同样数量的几类，每类的数量也和信誉等级中的保持一致，以分类准确率为优化目标，采用遗传算法进行求解。

得到各指标权重后，使用 TOPSIS 法根据各指标的数值构建可反映企业信贷可靠程度的综合指标，并求出全部企业的数值，剔除掉信誉等级为 D 级的企业后进行 k-means 聚类，分割出的 I、II、III 三类信贷可靠程度依次降低。最后在优化企业信贷策略时，考虑到含有约束条件的高维优化模型难以求出最优解，为简化问题，假设每类中所有企业分配相同的金额以及利率，将信贷策略选择问题转化为以银行的利润最大为优化目标，以三类企业各自的信贷额、利率为决策变量的优化问题，仍然使用遗传算法进行求解，最终得到年度信贷总额固定时企业针对三类企业的信贷策略。

2.2 问题二：基于具体年度信贷总额数据的信贷策略选择

问题二仍然要求对企业的信贷风险进行量化分析，与问题一的不同之处在于问题一中的附件 1 给出了企业的信誉评级以及是否违约的信息，可以根据它们确定各指标权

重。附件 2 中没有上述两种信息，但问题一中求出的权重指标具有普遍意义，可以推广至对附件 2 企业数据的信贷风险分析过程。使用问题一中构建的综合评价指标体系对附件 2 中企业的信贷风险进行量化的预测分析，同样使用信贷可靠程度作为综合指标，利用 k-means 聚类将 302 家企业分为 I、II、III、IV 四类。在分析贷款策略时首先剔除其中指标数值最低的 IV 类企业对剩下的企业采用问题一中的方法转化为多维的优化问题，假定各类内部的信贷策略相同，以银行利润最大为目标函数，在年度信贷总额为一亿元、每个企业的放贷额度为 10~100 万元、年利率为 4%~15%的约束条件下，使用遗传算法求解得到 I、II、III 类企业各自的信贷额与利率先对 302 家企业进行分类，在类别内部同质性的前提下分类别确定各自的信贷策略。

2.3 问题三：考虑突发因素条件下对信贷策略的调整

根据问题一、二中建立的银行信贷策略优化模型，我们在第三问中需要考虑突发因素对信贷策略的调整，因此本质上是一个预测模型。突发因素会对企业的生产经营和经济效益造成影响，在第一、二问中的模型中可以直接表现为突发因素对评价指标中各指标数值的影响。我们需要通过已知数据预测出突发因素对评价指标中各指标数值，再带入问题一、二中建立的评价指标体系模型以及银行信贷策略优化模型，可以求得此种情况下银行的最优信贷策略。

由于突发因素种类复杂，且对于不同行业、不同类别的企业会有不同的影响，因此需要对附件二中企业进行分类；对于突发因素也需要进行分类与分级，建立突发性风险评价体系，然后利用 2010 年~2019 年各产业 GDP 相关信息对不考虑新冠疫情影响下的 2020 年 GDP 进行预测，与实际值的差值即为新冠疫情对各类产业企业的影响程度，预测得到各类企业未来一段时间各指标的数值，再带入问题一、二中建立的评价指标体系模型，仍然利用问题一、二中的智能优化算法即可求得此种情况下银行的最优信贷策略。

三、模型假设与参数定义

3.1 模型假设

1. 假设附件所给数据均真实且合理。
2. 假设附件中所给进销项数据是完整的，即进销项数据是附件中所给时间段内该企业所有的进销项数据。
3. 假设每单发票涉及的金额在 10000 元以上为大单，10000 元以下为小单。
4. 假设在指标体系中作废、负数发票的金额比例和数量比例有着相同的权重，从而可以取二者均值合并为一个指标。
5. 假设银行在制定针对本题中中小微企业的信贷政策时事先根据信贷风险水平将企业分为若干类，每类中的企业提供相同的信贷额度、利率。

3.2 参数定义

符号	定义
$\omega_{total}, \omega_{buy}$	营业总额（销项金额）、进项金额
$\omega_{gross}, \eta_{gross}$	毛利润、毛利率
η_{tax}	税负率
var_{gross}	年平均毛利率方差
$\eta_{big_buy}, \eta_{big_sell}$	进、销项大单金额比例
$\eta_{del_n}, \eta_{del_m}$	作废发票数量、金额比例
$\eta_{neg_n}, \eta_{neg_m}$	负数发票数量、金额比例
M_{total}	银行年信贷总利润
α_i	银行贷款年利率
ϵ_i	可信度系数
μ_i	客户流失率

四、模型建立与求解

4.1 信贷风险综合评价体系与信贷策略优化模型

根据题目要求，银行首先决定向实力强、供求关系稳定的企业提供贷款，然后制定贷款额度、利率和期限等信贷策略。因此问题一中我们首先需要根据附件 1 中提供的信息建立相关评价指标体系，对企业的信贷风险进行量化，从而确定银行是否向其贷款。之后建立银行贷款策略模型，根据企业的信贷风险量化指标计算出该企业的贷款额度、利率等信息。

4.1.1 信贷风险衡量指标选取

（一）影响因素分析与指标选取：

从题目中得知，银行在决定是否向中小微企业提供贷款、贷款金额及利率是多少时，主要参考企业的整体实力、供求关系是否稳定、企业的信誉情况三方面内容，进而对企业的信贷风险做出综合、全面的评判，确定自身的信贷策略。此处我们对题目提供的附件中已有的数据进行适当的分类与筛选，舍弃其中与信贷风险评估关系不大的数据，对剩余的数据进行预处理（如剔除含有作废表单的交易票据信息、含有异常值的交易票据信息等）与组合（加权综合），形成了三类评价企业信贷风险的指标。

1.企业实力衡量指标：营业总额 ω_{total} 、毛利率 η_{gross}

在从附件中挖掘信息选取指标并构造出有分析价值的新指标时，需要进行横向（与时间无关）和纵向（与时间有关）挖掘。横向分析附件一中各单据金额指标的关系可得：

①销项金额 ω_{total} 反映了企业的销售额，即营业收入；②销项金额 ω_{total} 与进项金额 ω_{buy} 的差值即为企业的毛利润 ω_{gross} ；③销项税额与进项税额的差值可以得出增值税额。

对初步得到的这三个常见的金融指标分析如下：

（1）利润率指标无法获取

利润率应为利润与成本的百分比，然而附件中并没有直接给出成本数据，用进项金额代替成本计算利润率后，结果中出现不少异常数据——许多公司存在进项金额极少而销项金额极大的情况，如下表 1、表 2，原因可能是该公司原料库存较多、储备充足，近年内不需要进货，另一个原因是进项金额仅代表进货的花销，其他成本例如员工工资、店铺租金等均未考虑在内（这一部分对许多企业来说是主要开销）。这说明进项金额不能代替企业总成本，利用附件中的已有数据得到企业真实利润率是不可取的。

表 1 利润异常企业

企业代号	利润率
E16	75450.11%
E29	28030.28%
E42	251543.82%
E64	10118.15%
E68	840370.59%
E69	115235.21%
E95	215138.48%
E104	99332.33%
E105	12335.52%
E109	13492.48%
E110	20340.31%
E115	10411.83%
E123	25167.95%

表 2 进项金额极少而销项金额极大企业

企业代号	进项金额	销项金额	利润率
E42	10425.63	26235453.61	251543.82%
E68	698.11	5867409.22	840370.59%
E69	3155.34	3639218.08	115235.21%
E95	1073.58	2310757.26	215138.48%

(2) 税负率 η_{tax} 指标建立困难

从附件所给信息中可得到 $\eta_{tax} = \frac{\text{销项税额}-\text{进项税额}}{\text{销项税额}}$ ，文献[1]指出当前各行业内通常存在增值税的平均税负率数值，税务机构习惯把各个行业的税负率做为考核企业经营是否有异常的一个指标。我们尝试将附件一中的企业按行业分类，计算出各个企业的税负率后与行业正常税负率的范围进行比较。实际处理数据过程中，附件一中企业涉及的行业数量过多，不少企业的业务是跨行业的，且同行业企业数量通常较少（4~5家），使得数据过于离散化，难以分析其中的相关性和规律性，因此该指标最终被舍弃。

(3) 建立毛利率 η_{gross} 与营业收入 ω_{total} 指标

附件一中某一企业所有销项发票的销项金额之和即为一段时间内该企业的营业收入额，此项显然可以用来衡量企业的实力，一般来讲营业收入额高的企业规模更大、资金更充足、管理更合理，因此信贷风险相对较低。这里由于每家企业的销项流水数据的

时间跨度不同，不能简单地用全部销项发票的金额总和作为营收指标数值，因此我们计算了每个企业第一笔、最后一笔订单间相隔的月份数，将月均销项发票金额收作为营收指标。

考虑到进项信息无法代替成本使得各企业实际的利润率无法获取，而销项信息可以表示总收入，我们转而选取毛利率作为企业实力的衡量指标——销售毛利是企业利润的直接来源，也能很好地体现企业实力，计算方法如下：

$$\text{毛利润}\omega_{gross} = \text{销项金额/营业收入}\omega_{total} - \text{进项金额}\omega_{buy};$$

$$\text{毛利率}\eta_{gross} = \frac{\text{毛利润}\omega_{gross}}{\text{营业收入}\omega_{total}};$$

数据处理中我们发现附件一中大部分企业的毛利率数值处于合理区间，利于后一步的分析。对于其中少数异常的毛利率数值，在后面使用指标数值确定。

2.供求关系稳定性衡量指标：年平均毛利率方差 var_{gross}

附件中给出的各公司的发票流水时间跨度差异较大，起始时间落在 16、17 两年内，终止时间落在 19、20 两年内。为了衡量企业供求关系的稳定性，我们需要对企业的供求数额随时间变化的关系进行大致分析，而利用企业进项销项总额计算的总毛利率无法反映时间上的“稳定性”，仅能粗略反映这几年企业总的供求情况。为了衡量供求关系的稳定性时，需要对附件中的信息作纵向分析，可以找出企业供求关系的变化趋势，从而判断是否稳定。

附件中给出的发票流水时间落在 2016-2020 年区间内，其中绝大多数公司的发票流水时间区间包含 17-19 这三年，可以选取这三年公司毛利率的变化情况，量化三年间的增减变动趋势。我们计算出每家企业 2017-2019 年的毛利率，通过计算方差量化年毛利率的波动情况，从而反映企业供求关系的稳定性，方差越小稳定性越高。

3.上下游企业影响力衡量指标：进销项大单金额比例 η_{big_buy} 、 η_{big_sell}

附录中提供的数据种类有限，并没有发票中出现的上下游企业的企业经营相关数据，因此在衡量它们的影响力时只能通过它们与银行所评估企业间的交易额来反映，单次交易额较多表明供需方企业的规模都较大，资金实力较强。

我们利用进项和销项交易中大单金额占该企业进项和销项总额的比例进行分析。进项发票中，大额订单金额比例可用来大致衡量该公司上游公司中规模较大的公司的比例，可用来反映上游公司的行业影响力，大额订单金额比例越高说明上游企业规模普遍较大；

销项发票中，大额订单金额比例可以用来衡量公司大客户的大致比例，反映该公司下游公司的影响力。

在确定“大单”的量化标准时，我们人为设置了一个大单数额，超过该数额则认为该订单为大单，在这里，我们设定 10000 元以上为大单。

4.信誉度衡量指标：作废发票(η_{del_n} 、 η_{del_m})和负数发票数量比例、金额比例(η_{neg_n} 、 η_{neg_m})

附件 1 中已经给出了银行对各家企业的信誉等级的评分数据，此外我们认为进项和销项信息中作废发票和负数发票信息也能一定程度上体现企业的信贷风险情况：作废发票和负数发票都是因故取消交易或退货退款的情况，且作废发票比例属于发票风控监控指标，因此作废发票和负数发票比例过多会影响企业的信誉度，增大信贷风险。在进项中作废发票和负数发票体现的是该企业的信誉情况；而销项中作废发票和负数发票体现的是该企业客户的信誉度，间接体现企业的信贷风险。

在量化信贷风险水平时，我们需要给作废发票和负数发票数量比例、金额比例赋予一定的权重。然而，造成作废发票或负数发票出现的理由多样，不难想象其中一部分是突发情况造成的，属于偶然因素，不能用来衡量企业的信誉度。一般来讲，当作废发票、负数发票比例很小时，大多数情况下是偶然因素导致的，不能作为企业信誉度的参考，且大多数拥有作废、负数发票记录的企业并不是主观上违约，因此我们需要根据二者在企业中的分布情况设定作废发票和负数发票数量比例、金额比例的阈值，当企业的该项指标低于相应的阈值则不会对信誉度产生影响，将该项数值直接置零；若高于相应阈值则需要在信贷风险评价指标中体现。使用 excel 绘制两种发票的比例分布散点图如下：

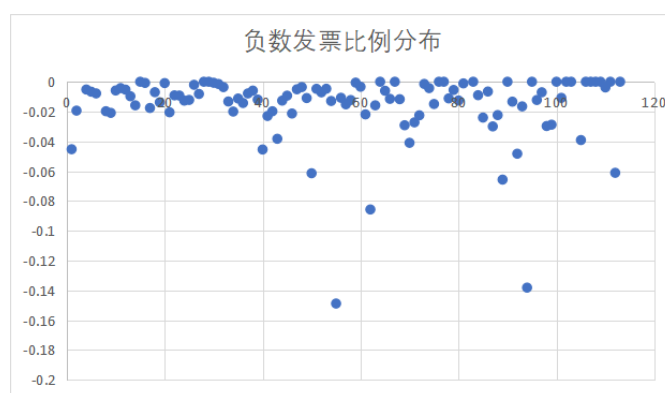


图 1 负数发票比例分布

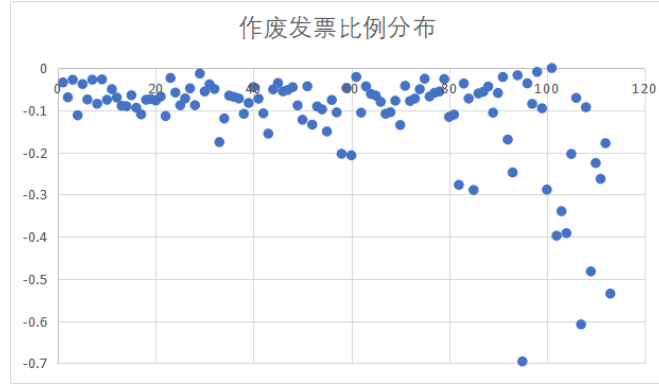


图 2 作废发票比例数据

散点图显示,大部分企业的负数发票比例都在 4%以下,作废发票比例在 15%以下,由此可以将负数发票、作废发票的比例阈值分别设置为 0.04 和 0.15,在阈值以下的企业占大多数,它们最终没有完成交易的原因主要是突发因素造成的,不应将其计入企业信誉度的评估中。在阈值以上的企业在评估信誉水平时应当考虑负数、作废发票,因为这些企业这两个指标的数值要高于平均水平。

将以上讨论的指标体系结构整理如图:

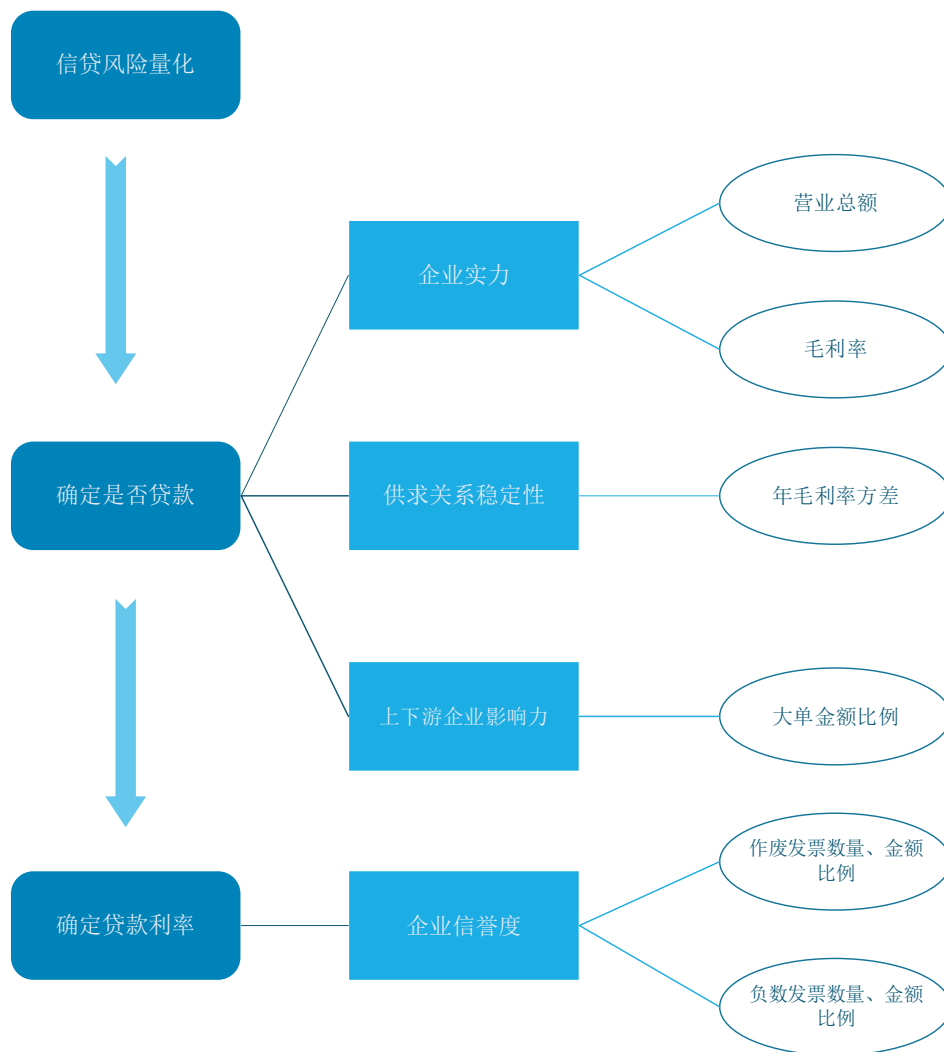


图 3 指标体系结构

（二）指标数据预处理

根据上面对六个指标的定义我们分别求出所有企业的对应指标数值，具体的表格数据见支撑材料中的文件“附件 1 指标数据整理.xlsx”，其中毛利率、年平均毛利率方差的数据中异常值较多，一些企业的毛利率数值达到负几十甚至几百（销项金额远小于进项金额）、年平均毛利率方差达到几百甚至几千（供求关系极不稳定，且年均毛利率的绝对值本身较大）。这些企业的指标数据在后面计算各指标权重时作为异常值会对最终结果造成较大影响，因此我们针对这两个指标设置阈值对数据进行清洗，舍去含有异常值指标的企业的企业数据，减小后续权重计算的误差。

经过清洗，在 123 家企业中最终选择了 113 家企业的指标数据，被舍弃数据的企业一共有十家，其指标数据如下表：

表 3 被舍弃数据的企业

企业代号	毛利率	年平均毛利率方差	作废比例	负数比例	平均销项金额	大单金额比例
E66	-6.31%	8.238	-4.85%	-0.39%	176342.77	72.931%
E69	99.91%	Invalid	-12.01%	-8.52%	191537.79	0.000%
E80	31.84%	6.573	-1.47%	0.00%	103258.34	99.932%
E83	-932.12%	11011.936	-8.43%	-1.22%	91344.03	0.000%
E89	-38.88%	7.232	-4.30%	0.00%	50358.10	34.266%
E96	-61.24%	590.251	-4.69%	0.00%	71910.19	100.000%
E104	99.90%	Invalid	-8.69%	-13.62%	17510.03	74.927%
E114	-157.95%	13.707	-11.34%	-0.66%	5935.91	0.000%
E119	-516.00%	12.339	-0.71%	-0.16%	1371.99	0.000%
E121	-958.02%	5.957	-5.53%	-0.33%	4227.58	0.000%

（三）指标的合理性与相关性检验

1.指标的合理性检验

判断衡量信贷风险的指标是否合理、有效，可以借助于附件中的历史数据，判断指标数值能否将有违约记录的企业和无违约记录的企业很好地分离开，即能否很好地刻画两类企业间的差异。根据附件 1 中的违约信息记录，将 123 家公司分成两类，分别求出两类企业各自六个指标的平均值，记录在下面的表格中：

表 4 指标合理性检验

是否违约	毛利率	年平均毛利率方差	作废比例	负数比例	销项平均金额	大单金额比例
是	1.41%	0.637897	-21.39%	-5.32%	214080.48	53.24%
否	56.96%	0.072266	-8.53%	-2.37%	4363227.25	0.758292

观察上面的表格可以发现：有违约记录的企业相比无违约记录的企业，毛利率明显较小，年平均毛利方差明显较大，作废单比例和负数单比例明显较大，销项金额和大单金额比例明显较小，即各项指标的平均值之间又显著差异，因此可以判断指标数值能有效地将有违约记录的企业和无违约记录的企业很好地分离开，从而证明了指标的合理性。

2.指标的相关性检验

由于后面在建立综合指标的量化模型时需要确定各指标的权重数据，如果指标体系中有某几个指标间相关性较高，会给权重的计算带来较大的偏差，这类似于进行典型相关分析时指标间的“多重共线性”，因此建立衡量信贷风险的指标时需要对指标间的相关性进行分析，判断是否存在相关性较强的指标，并进行剔除。考虑到指标的分布并不满足近似正态分布的情况，且 pearson 相关系数一般用来衡量指标间的线性相关性，因此我们使用 spearman 相关系数分析上述六个指标间的相关性。下面是使用 SPSS 求解

113 组数据的六个指标的 spearman 相关系数的结果：

表 5 相关性分析

	毛利率	年平均毛利率方差	作废比例	负数比例	销项平均金额	大单金额比例
毛利率	1.000	-0.523	-0.101	0.122	-0.241	-0.540
年平均毛利率方差	-0.523	1.000	-0.152	-0.173	-0.183	0.163
作废比例	-0.101	-0.152	1.000	-0.042	.281	0.072
负数比例	0.122	-0.173	-0.042	1.000	-0.074	-0.045
销项平均金额	-0.241	-0.183	.281	-0.074	1.000	0.721
大单金额比例	-0.540	0.163	0.072	-0.045	0.721	1.000

由上表可知，各个指标之间的 spearman 相关系数均不大，因此可以认为现有的六个指标是合理的，无需剔除相关的指标。

4.1.2 基于遗传算法的权重确定模型

在量化信贷风险时我们需要给上述指标赋予一定的权重，确定权重的常规方法有层次分析法、专家评估法等，但这些方法都具有强烈的主观性，在数学建模中缺乏说服力和准确性。

由于附件 1 中给出了企业信用评级，企业信用评级是银行在考虑了过去一段时间内受评企业的经营管理素质、财务结构、偿债能力、经营能力、经营效益、发展前景等方面后做出的综合评定，能够综合反映企业的整体状况，通常具有较高的准确性，而我们建立的信贷风险量化模型也是在考量了企业实力、供求关系稳定性、信誉水平等各种因素后构建的反映企业信贷风险情况、违约可能性的综合指标。因此可以利用企业信誉评级的已有数据代表银行对信贷风险的量化评估，进而用企业信誉评级的分类数据求解各指标的适宜权重。这样就将权重确定过程转化为一个优化模型，其中决策变量为指标各自的权重数值，利用该权重计算得到各企业的综合指标数值后，将企业分成与信誉等级中同样数量的几类，每类的数量也和信誉等级中的保持一致，此时我们的优化目标是使得对所有企业的分类准确率最高。考虑到此模型决策变量数量较多，且优化目标的函数表达式难以用公式表示、直接计算，我们采用智能优化算法中的遗传算法对这一问题进行求解。遗传算法具有全局搜索能力，具有较高的结果精度，适合求解优化目标较复杂、决策变量维度较高的优化问题。

（1）决策变量选取（信贷风险衡量指标选取）

我们选取了如下六个衡量企业信贷风险的指标：（为了将指标正向化，我们将作废发票、负数发票比例分别取相反数作为输入数据）

① 毛利率 η_{gross} ；②年平均毛利率方差 var_{gross} ；

③ 作废发票比例相反数 $\eta_{del} = -\frac{\eta_{del,n} + \eta_{del,m}}{2}$ ；

由于作废发票数量、金额比例间显然具有较强相关性，考虑到二者的重要性大致相同，定义作废发票比例为二者均值数据。

④ 负数发票比例相反数 $\eta_{neg} = -\frac{\eta_{neg,n} + \eta_{neg,m}}{2}$ ；与③中指标定义的原理相同。

⑤ 销项平均金额 ω_{total} ；

⑥ 进销项大单金额比例 $\eta_{big} = \frac{\eta_{big,buy} + \eta_{big,sell}}{2}$ ；

由于销项大单金额、销项大单金额分别是用来衡量下、上游企业的企业规模、企业影响力，考虑到二者的重要性大致相同，定义比例为二者均值数据。

(2) 主要求解步骤

1. **初始化种群**：将权重的取值范围限定在（-1，1）区间内，6个权重按按浮点数编码排列在一起成为一个个体，随机产生 500 个个体作为初始种群；
2. **编写适应度函数**：在 6 个指标对应的权值确定的条件下，计算出企业的综合评价指标数值后进行分类，求出所有企业中准确分类的个数，作为个体的适应度；
3. **评价、选择算子**：对个体按适应度大小进行排序，并计算种群个体的相对累积适应度；使用带有精英选择的轮盘赌法选择出若干个个体（=初始种群个体个数）构成新一代种群；
4. **交叉算子**：将个体两两配对，随机生成交叉位置，按照设定的交叉概率进行单点交叉操作；（从某点开始后面位置的数值交换）；
5. **变异算子**：对每个个体以一定的概率选择是否进行变异操作，进行变异操作时随机选择两个变异位点，生成解域内的两个随机数替代变异位点的原有值；
6. 若满足收敛条件则输出获得最佳个体、最佳适应度、最佳个体出现代，否则返回步骤 3 继续进行评价、选择、交叉和变异操作。

(3) 求解结果

通过实际运算过程中对参数值的调整，最终我们选择种群大小为 500、交叉概率为 0.6，变异概率为 0.05、迭代次数为 100。借助 matlab 运行遗传算法的优化代码后，得到的六个指标及对应的权重分别为：

毛利率 η_{gross} : $\omega_1 = 0.3029$; 年平均毛利率方差 var_{gross} : $\omega_2 = -0.0565$

作废发票比例相反数 η_{del} : $\omega_3 = 0.2818$; 负数发票比例相反数 η_{neg} : $\omega_4 = 0.3569$

销项平均金额 ω_{total} : $\omega_5 = 0.9648$; 进销项大单金额比例 η_{big} : $\omega_6 = 0.7365$

观察上述结果可以看出,在六个指标中,销项平均金额、进销项大单金额比例这两个衡量企业实力、上下游企业影响力的指标所占权重最大,在信贷风险的评估中影响较大,毛利率、作废和负数发票比例的权重差不多,年平均毛利率方差权重较小,原因可能在于数据覆盖的年份数据太少导致方差计算过程中存在异常值,不能很好地反映供求关系的稳定性。

(4) 遗传算法与神经网络的比较

此外,也可从黑箱的角度考虑权重的确定问题,我们可以借助神经网络的学习能力,根据已有的一百多家企业的数据,建立衡量企业风险的六项指标与信誉评级间的关系网络,利用已有数据进行训练。问题二中可依据这种内在关系利用附件 2 得到的六项指标数据直接预测出问题二中企业的信贷风险评价综合指标。然而,神经网络最大的缺点是不能得出具体的量化函数关系,不能从机理层面进行分析而仅仅保证预测结果的准确性,难以得到各指标相应权重,因此最终没有采取该方法。

4.1.3 基于 TOPSIS 法和 k-means 聚类的信贷风险综合评价模型

1. 基于 TOPSIS 法的信贷风险综合指标(信贷可信度指标)求解

选取合适指标并确定对应的权重后,我们已经可以量化各指标对于信贷风险的影响大小,进而可以建立衡量信贷风险的综合评价指标,求出每家企业对应的综合指标数值后进行排序、分类。对于这种多属性决策的排序问题,通常的做法包括简单线性加权、加权平方和法、主成分分析法、功效系数法等,此处我们利用 TOPSIS 法实现综合指标的构建与求解——通过计算每个方案靠近正理想解、远离负理想解的程度求出方案到理想方案的相对贴进度,从而对方案进行排序。

(1) 主要求解步骤

- ① 对原始数据中的负向指标(年平均毛利率方差)进行正向化处理(此前针对作废、负数发票已做过相同处理),对该指标的每个样本取相反数,从而每个指标都满足数值越大风险越小;

② 对原始数据向量进行规范化处理得到规范决策矩阵：决策矩阵为 $A = (a_{ij})_{m \times n}$ ，规范化决策矩阵 $B = (b_{ij})_{m \times n}$ ，则 $b_{ij} = a_{ij} / \sqrt{\sum_{i=1}^m a_{ij}^2}$ ；

③ 根据前面遗传算法模型中确定的各属性值的权重向量，将规范决策矩阵的每列与对应的指标权重相乘，得到加权规范阵 $C = (c_{ij})_{m \times n}, c_{ij} = w_j * b_{ij}$ ；

④ 确定正、负理想解 C^* 、 C^0 ，其中 $c_j^* = \max_i c_{ij}, c_j^0 = \min_i c_{ij}$ ；

⑤ 计算各方案与正理想解距离 $s_i^* = \sqrt{\sum_{j=1}^n (c_{ij} - c_j^*)^2}$ 、与负理想解距离 $s_i^0 = \sqrt{\sum_{j=1}^n (c_{ij} - c_j^0)^2}$ ；

⑥ 根据⑤中得到的两个距离计算出各方案的信贷风险综合评价指数 $f_i = \frac{s_i^*}{s_i^0 + s_i^*}$ ；

需要注意的是，由于建立指标时采用的指标均为正向化指标，因此综合评价指数 f_i 越大，信贷风险越低，违约的可能性越低。

(2) 求解结果

在 matlab 中采用 TOPSIS 法求解上述信贷风险量化模型，计算得出附件一中 123 个企业的信贷风险综合指标量化得分 $f_i (1 \leq i \leq 123)$ ，具体数值见支撑材料中，由于条数较多不在正文里展示。每个企业的综合指标确定后，即可根据指标数值大小对企业的信贷风险等级进行分类，划分为若干个区间。求解最终信贷策略时，我们无法为每个企业都设定特异的贷款数额、贷款利率，这样就要求解一个 $123 \times 2 = 246$ 个决策变量的优化模型，这是极难求出的。考虑到银行政策的制定通常不是连续的方案，不是对量化得分不同的企业采取个性化的不同方案，而是对处于某一区间的企业采取相同的政策，为简化问题，我们利用 k-means 聚类对附件一中企业的指标进行聚类。

2. 基于 k-means 聚类分析的企业信贷风险等级评估

根据上面 TOPSIS 法求解出的综合评价指数 f_i ，对附件一中的企业进行聚类，在聚类之前，由于信誉等级为 D 的企业不在银行贷款考虑的范围内，因此把信誉等级为 D 的企业及其数据剔除，再进行聚类分析。

(1) 聚类求解步骤

① 指定需要划分的簇的个数 K 值（类的个数），这里我们设定 K 值为 3，即类的个数为 3；

② 随机地选择 K 个数据对象作为初始的聚类中心，记这些点分别为 c_1, c_2, \dots, c_K ；

- ③ 计算其余的各个数据对象到这 K 个初始聚类中心的距离, 把数据对象划归到距离它最近的那个中心所处簇的簇类中, 在这里定义距离为绝对值距离 $|f_i - f_{ci}|$;
- ④ 调整新类并且重新计算出新类的中心;
- ⑤ 循环步骤三和四, 看中心是否收敛 (不变), 如果收敛或达到迭代次数则停止循环;
- ⑥ 结束。

(2) 聚类求解结果

在 matlab 中调用 kmeans 函数, 得到聚类结果, 具体聚类结果见支撑材料。从聚类结果可以看出, 企业 E1 由于综合评价指数过大, 被单独分为一类, 其余指标按 f_i 大小分别被分为两类, 利用 matlab 绘制出聚类轮廓图如下, 由图可知簇为 2 的为被分为单独一类的 E1, 其他企业被分为了 2 类。E1 单独构成一类主要原因是其销项金额数据远大于进项, 该指标的数值直接拉大了综合评级指标的数值。聚类轮廓图如下所示:

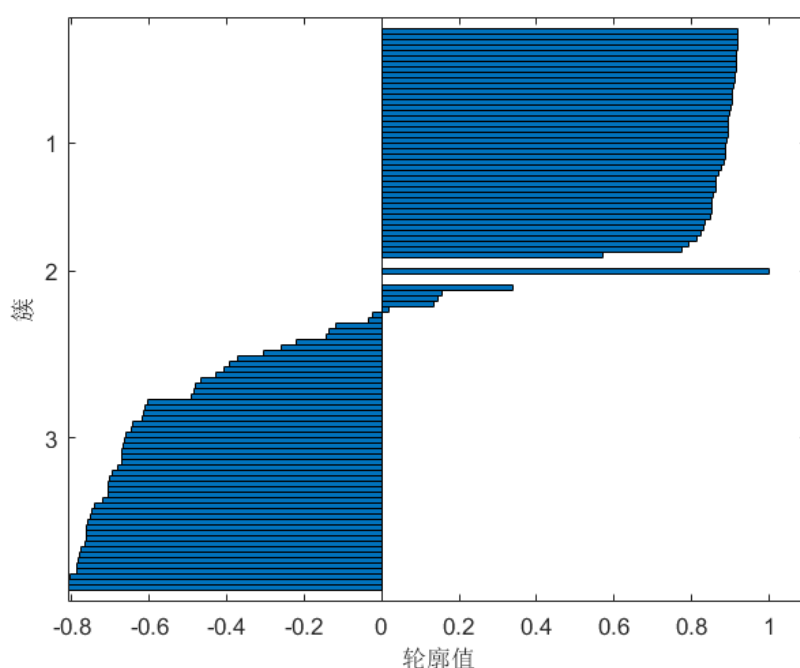


图 4 问题 1 聚类轮廓图

4.1.4 银行信贷策略优化模型

建立了信贷风险评价综合指标模型并进行分级后, 可以根据分类结果对每一类别中的企业提供不同的贷款金额并给予不同的贷款利率。此外, 根据综合评价指标的数值大小以及信誉等级划分下各类中企业的历史违约率, 可以为之前分出的每类数据添加违约

的概率值，并以此计算整个银行年利润的期望。这是一个以各类的贷款金额、贷款利率为决策变量，以年度信贷总额为约束，以年贷款利润期望最大化为目标的优化问题。

题目中规定了“银行对信誉评级为 D 的企业原则上不予放贷”的原则，因此前文在 4.1.3 模型中进行 K-means 聚类分析时剔除了所有信誉等级评定为 D 级的企业直接归为 IV 类，将剩下的企业按照信贷风险可信度指标分成三个级别—— I、II、III，对应关系如下表所示，每个风险级别的企业采取相同的贷款策略，即贷款金额 ω_i 与年利率 α_i 相同（ $1 \leq i \leq 3$ ）。

表 6 可信度水平与级别间的对应关系

级别	可信度水平
I	优秀
II	良好
III	中等
IV	差

1. 决策变量

由于题中规定贷款期限均为一年，且我们认为每个风险级别的企业采取相同的贷款策略，因此需要考虑的优化变量为银行给每类企业的贷款金额 W_i 与年利率 α_i 。（ $1 \leq i \leq 3$ ）

2. 目标函数——考虑客户流失率与信贷可信度指数的期望贷款年收入

设定的目标函数为银行一年内放贷获得总收入的期望 M_{total} （还款期限为 1 年），来源于银行从各类企业贷款金额中期望获得的年利润总和。在计算各类企业年利润时处了贷款金额、年利率外，由于企业是否能够最终还款并缴纳利息是一个未知因素，客户最终是否会选择在银行申请贷款（取决于银行贷款利率的高低）也具有不确定性，这两个影响因素可以分别用两个概率值①客户流失率 μ_i 、②可信度系数 ε_i 刻画。（ $1 \leq i \leq 3$ ）

（1）客户流失率 μ_i

附件三中给出了 A,B,C 三类企业银行贷款年利率与客户流失率关系的统计数据，根据最小二乘法对曲线进行拟合可以得出这三类企业贷款年利率与客户流失率关系的函数表达式，从而根据确定的 α_i 可计算得出相应的 μ_i 。

客户流失率的拟合计算：

根据附件 3 中所给的出的 A、B、C 三类客户流失率与银行贷款年利率之间的关系，

用 Excel 作出带平滑线和数据标记的散点图如下图，初步观察可以得出：(1)三类客户流失率与银行贷款年利率之间不呈线性关系；(2)三类客户流失率与银行贷款年利率之间的曲线关系形式基本相同。并且由于给出的数据点较多，因此对三类客户流失率与银行贷款年利率进行曲线拟合，得到客户流失率与银行贷款年利率之间的函数表达式。

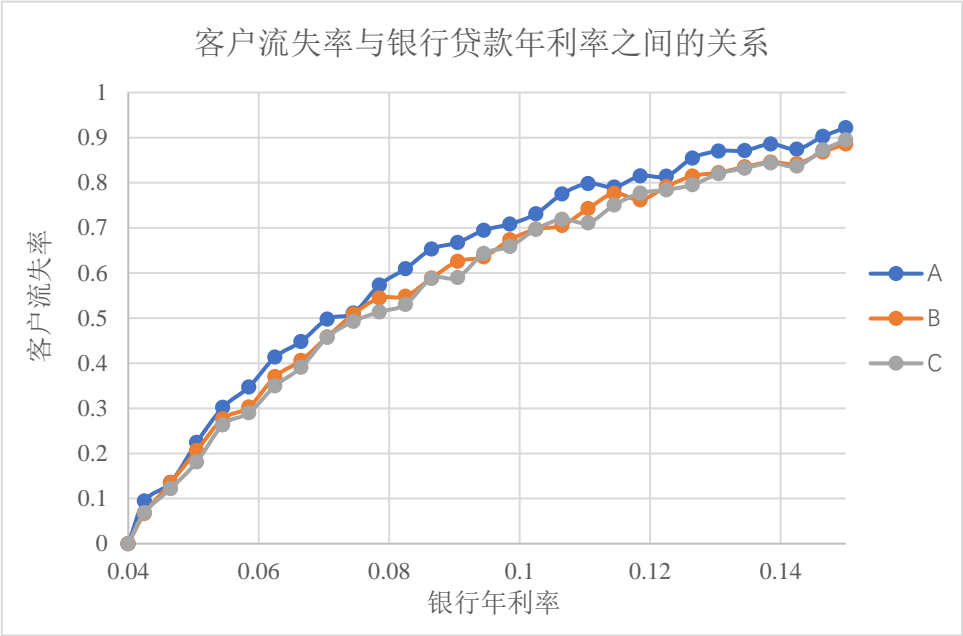


图 5 客户流失率与银行贷款年利率之间的关系

由于三种客户类型对应的曲线形状较为接近，因此其曲线拟合形式应该相同，故本小节中只对 A 进行分析就可确定拟合的形式。观察曲线可知，可能的常用拟合函数形式有对数和多项式。利用 SPSS 软件对 A 类客户流失率与银行贷款利率之间的关系，分别进行对数曲线拟合($y = a + b \ln x$)、二次曲线拟合($y = c + b_1x + b_2x^2$)、三次曲线拟合($y = c + b_1x + b_2x^2 + b_3x^3$)，得到拟合曲线和拟合参数结果如下图和下表：

表 7 三类拟合结果

	R 方	F	自由度 1	自由度 2	显著性	常量	b1	b2	b3
对数	.982	1500.776	1	27	.000	2.239	.669		
二次	.993	1847.861	2	26	.000	-.697	21.984	-76.410	
三次	.998	3690.626	3	25	.000	-1.121	37.970	-258.570	640.944

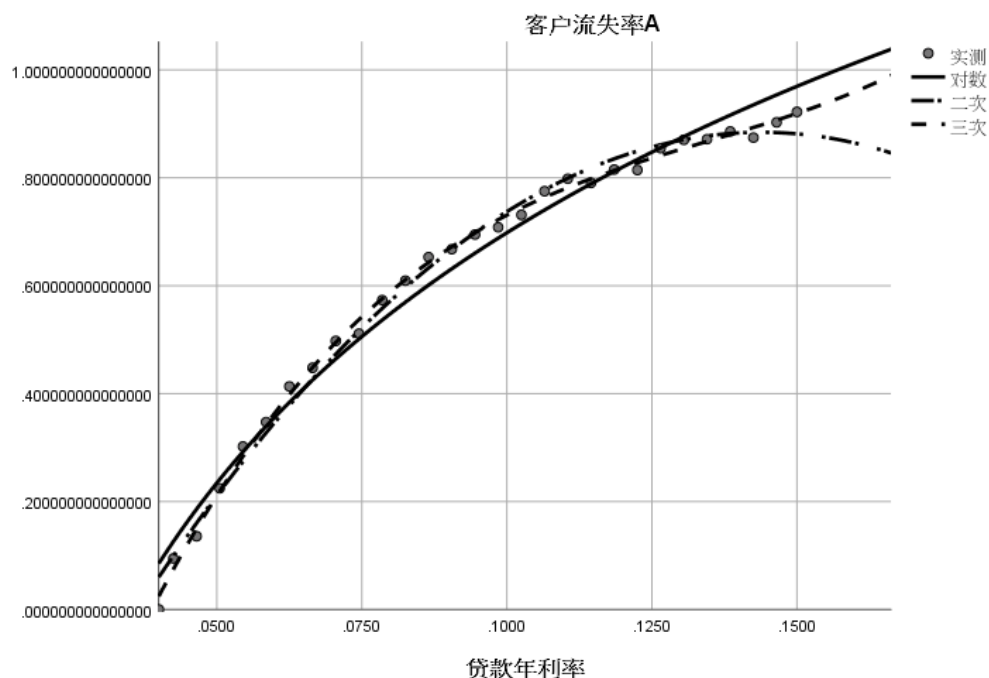


图 6 三类拟合结果

由上表可知，三次拟合时 R^2 最大， F 最小，并且由上图可知在年利率在 4%-15%时和数据点之间的贴合程度最好，因此选取三次曲线作为拟合曲线形式。

利用 SPSS 对 A、B、C 三类客户的流失率与银行贷款年利率之间的关系进行拟合，得到了三类客户流失率所对应的表达式：

$$A: y = -1.121 + 37.970x - 258.570x^2 + 640.944x^3$$

$$B: y = -1.017 + 33.995x - 225.051x^2 + 552.829x^3$$

$$C: y = -0.973 + 32.157x + 207.386x^2 + 504.717x^3$$

其中， x 为为银行贷款年利率， y 为客户流失率。

(2) 可信度系数 ε_i

这里的可信度系数表示某一类别中的企业有多大概率能够按时偿还贷款，不同类别企业的信贷可信度水平不同，所对应的信贷风险不同，前面得到的综合指标数值越大的企业可信度越高。由于我们的权重系数是根据银行对各家企业的信誉等级评分估计得到的，因此此处估计可信度数值时可以参考附件中给出的各信誉等级中企业的违约率。计算得到：

A 类企业共有 27 家，其中无企业违约；B 类企业共有 38 家，其中 1 家违约；

C 类企业共有 34 家，其中 2 家企业违约。

算出附件中 A、B、C 三类企业的实际违约率分别为 0 、 $\frac{1}{38} \approx 3\%$ 、 $\frac{1}{17} \approx 6\%$ ，考虑到实际情况下企业对银行信贷的违约概率普遍较低，因此可以将 1、97%、94% 分别作为三类的可信度系数。

综上所述，最终可得出目标函数的表达式为：

$$M_{total} = \sum_{i=1}^3 num_i W_i (1 - \mu_i)(\alpha_i \varepsilon_i - (1 - \varepsilon_i));$$

其中 $W_1 \sim W_3$ 分别表示 I、II、III 三类中每家企业可得到的贷款金额。

3. 约束条件

根据题目中所给条件，贷款额度为 10~100 万元，年利率为 4%~15%，且客户流失率 μ_i 、可信度系数 ε_i 范围均为 0~1 之间，由此得出该优化问题约束条件如下：

$$W_4 \leq W_5, W_5 \leq W_6;$$

$$100000 \leq W_i \leq 1000000;$$

$$4\% \leq \alpha_i \leq 15\%;$$

$$0 \leq \mu_i \leq 1;$$

$$0 \leq \varepsilon_i \leq 1.$$

(3) 年度信贷总额固定

问题 1 中限定年度信贷总额为一定值，但是这一定值无法确定，可能还需要分类讨论。此处为简化问题，将年度信贷总额用“1”表示， $W_i (1 \leq i \leq 3)$ 表示三类企业的总贷款额占年度信贷总额的比例，均为小于 1 的数，这样在总额未知的情况下用比例表示各类的信贷额，能够较好地满足题目要求，新的限定条件为：

$$num_1 * W_1 + num_2 * W_2 + num_3 * W_3 \leq 1$$

注：由于此时信贷额均用占总额的比例表示，这时我们不再使用 $100000 \leq W_i \leq 1000000$ 这一具体金额衡量的信贷条件。

4. 模型求解与结果

针对此模型中 6 个变量、含有约束条件且适应度函数值不能直接用一个表达式表达的优化问题，最适宜的方法仍然是遗传算法，和模型 4.1.2 中一样，利用浮点数编码的遗传算法求解此问题。在种群大小为 500、交叉概率为 0.8，变异概率为 0.2、迭代次数为 200 的参数设定下，得到的最优个体的值为：

$$W_1 = 0.0962, W_2 = 0.0133, W_3 = 0.0015, W_4 = 0.0626,$$

$$W_5 = 0.0762, W_6 = 0.0846$$

遗传算法求解结果表明，应当将年度信贷总额中的 9.62% 贷给 I 中企业、1.33% 贷给 II 中每个企业、0.1% 贷给 III 中每个企业，且给 I、II、III 三类的贷款利率分别为 6.26%、7.62%、8.46%，数值表明应当给信贷风险小的企业多提供一些贷款金额，并给予一定的利率奖励，风险越小利率应当越小。

4.2 年度信贷总额给定时对缺少信誉评级企业的信贷策略的分析

问题二仍然要求对企业的信贷风险进行量化分析，与问题一的不同之处在于问题一中的附件 1 给出了企业的信誉评级以及是否违约的信息，而附件 2 中没有上述两种信息。在问题一中，我们曾利用附件 1 中的违约信息对选定的各项指标的合理、有效性进行了检验，利用企业信誉评级对问题一中建立的信贷风险量化模型的指标权重进行了估计。可以使用问题一中构建的模型对附件 2 中企业的信贷风险进行量化的预测分析，求解贷款策略时仍采用问题一中的方法先对 302 家企业进行分类，在类别内部同质性的前提下分类别确定各自的信贷策略。

4.2.1 信贷风险衡量指标的计算

（一）指标计算与指标数据预处理：

根据问题一中对信贷风险相关指标的定义与求解方法，可以得到 302 家企业信贷风险的指标，具体数据见支撑材料。

同样地，按照问题一中绘制散点图观察的方法，对负数发票、作废发票设置了阈值为 0.05 和 0.2，小于阈值的数认为是偶然因素导致的交易终止，不纳入信誉水平的考量。此外，针对异常毛利率、年毛利率方差对企业进行清理，设置阈值为 -10 和 100，落在阈值外的数据整体舍弃，最终在 302 家企业中选择了 288 家企业的指标数据，被舍弃数据的企业一共有 14 家。

（二）基于 TOPSIS 法和 k-means 聚类的信贷风险评价综合指标模型

问题一中我们通过遗传算法计算出各个指标的权重为：

$$\omega = (0.3029, -0.0565, 0.2818, 0.3569, 0.9648, 0.7365)$$

问题二中我们继续引用此权重、利用 TOPSIS 法实现综合指标的构建与求解。与问

题一中计算综合评价指数的方法类似，得到问题二中 302 个企业的综合评价指数 f_i （具体数值见支撑材料），仍然利用 k-means 聚类法将企业进行聚类，得到的聚类结果如下面聚类轮廓图所示：

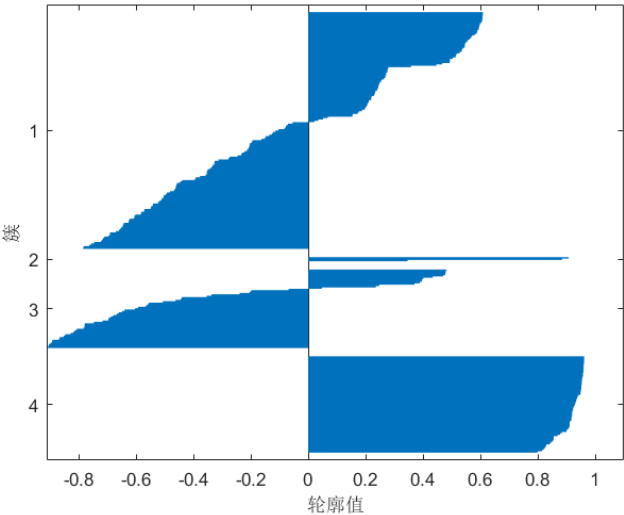


图 7 问题 2 聚类轮廓图

聚类结果显示，I、II、III、IV 四类企业的数量分别为 3、57、70、179，其中企业 E124、E125、E127 三个企业的综合评价指标明显高于其它企业，使得这三家企业可以单独构成一类，造成这一现象的原因和问题一中一样——销项金额数据远大于进项，显著拉大了综合评级指标的数值，这是进销项发票流水信息自身决定的，由于统计时间有限，不能很好地反映企业的收支情况。此外，由于 IV 类企业数量较多，我们对 IV 类企业进行了再分类，将其中一部分数据归入 III 类中，力图实现类别间的平衡。

（三）年度信贷总额数据给定条件下的银行信贷策略优化模型

通过 k-means 聚类分析将企业对应的信贷风险可信度指标划分为 4 个级别后，仿照题目中舍弃信誉等级为最差的 D 等级企业的做法，舍弃其中综合评价指标数值最小的所有 IV 类企业，利用问题一中建立的银行决策优化模型进行优化。由于本题中年度信贷总额数值给定，为一亿元，我们不需要用 W_i 代表企业贷款额占总额的比例，可以直接代表企业的信贷额，从而将约束条件 $num_1 * W_1 + num_2 * W_2 + num_3 * W_3 \leq 1$ 用

$$num_1 * W_1 + num_2 * W_2 + num_3 * W_3 \leq 10^8 \text{ 及 } 100000 \leq W_i \leq 1000000 \text{ 两个代替。}$$

用遗传算法进行优化时，选取各项参数数值为种群大小 500、交叉概率 0.8，变异概率为 0.2、迭代次数 200，连续运行 10 次后将每次得到的最优个体加和后平均，得到最优个体数值为：

$$W_1 = 6.8141 * 10^5, W_2 = 5.9876 * 10^5, W_3 = 2.7977 * 10^5, W_4 = 0.0768,$$

$$W_5 = 0.0856, W_6 = 0.0973;$$

遗传算法得到的最优策略表明，在给定年度信贷总额为一亿元的情况下，I 类企业平均每家可以收到 68.1 万元的贷款，II 类中每家企业平均可收到 59.9 万元贷款，III 类中平均可收到 28 万元贷款，且给 I、II、III 三类的贷款利率分别为 7.68%、8.56%、9.73%，数值同样表明应当给信贷风险小的企业适当增加贷款金额并给予一定的利率减免奖励，这一点是符合实际情况的。

4.3 考虑突发因素影响下信贷策略的调整模型

问题三中我们需要考虑突发因素（例如新冠病毒疫情）对各企业的影响从而调整银行信贷策略。我们从特殊到一般的思路出发，先以新冠肺炎这一突发公共卫生事件为例进行分析，再由此推广到其他一般的突发情况，最后提出银行在应对突发情况时总体的信贷策略。

4.3.1 应对新冠肺炎信贷策略

在分析新冠肺炎对各行业企业的影响程度时，我们首先通过查找 GDP 相关的变化信息拟合得出新冠肺炎对各产业企业的影响程度，根据此定量关系可预测出各类企业指标，再通过查找相关文献得出潜在信息用于分析发展趋势，最后从银行角度调整信贷策略以及分析国家层面的宏观政策调控。

1. 整体影响程度

我们通过相关网站查询到 2010 年第一季度至 2020 年第 1-2 季度的国内生产总值及同比增长率以及第一、二、三产业增加值季度累计同比增长率，如图 8 所示。从图 8 中可以发现国内生产总值以及第一、二、三产业增加值季度累计同比增长率在 2020 年第一季度出现了明显的骤降，且在第二季度时有一个较大的回升。但由于该变化属于突发

因素，且变化后的数据量太少，因此我们难以对图一中的曲线直接进行拟合等处理。

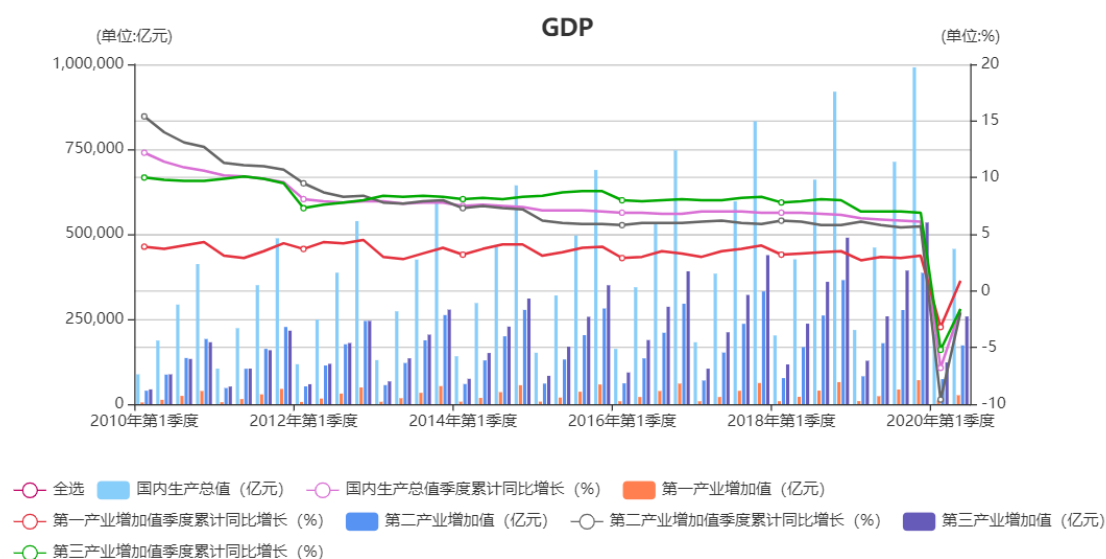


图 8 2010 年第一季度至 2020 年第 1-2 季度 GDP 信息

观察 2010~2019 年的第一、二、三产业增加值季度累计同比增长率，发现其虽然在一个范围内波动，但每年增长率均值变化较小，因此利用图一中导出的 Excel 表格可计算得出 2010~2019 年的第一、二、三产业增加值季度累计同比增长率均值，用此均值可以粗略预测在没有突发因素（新冠疫情）下 2020 年第一二季度各产业季度累计同比增长率，与实际数值作差即可得到新冠肺炎对于第一二三产业影响率分别为 175.4%，191.4%，257.4%。

2. 企业分类

通过图一我们可以发现第一、二、三产业增加值季度累计同比增长率有着较为明显的差异，并且在 2020 年第一季度下降程度也有所不同。根据题目中所述对突发因素对各企业的影响程度往往与企业所处行业有着密切相关，再通过查阅相关文献[2]，我们得出受到疫情负面影响最为快速直接且影响程度较大的行业是第三产业。例如旅游业、交通运输业、餐饮业等，受到居民对自身安全考虑，人们线下活动减少，导致服务业收到大量冲击。然而工业生产受疫情影响小幅下降，企业复工复产明显加快。同时，突发大规模疫情及成功防控往往会带动一些产业的发展。最为显著的是医疗卫生产业；由于线下活动减少使得线上活动增多，线上产业，包括新兴信息技术、数字消费、线上体验等产业都被带动。

因此我们对附件二中的企业按照第一、第二、第三产业进行分类，并且对于文献[2]中提到的被带动发展的行业例如医疗行业、线上产业等进行单独的分类，最后得到分类

结果如图 9。在之后银行制定政策时需对不同类别企业制定不同策略，例如对疫情冲击较大的第三产业需要大幅调整信贷策略，而对冲击较小的工业政策调整幅度也需相应减小。

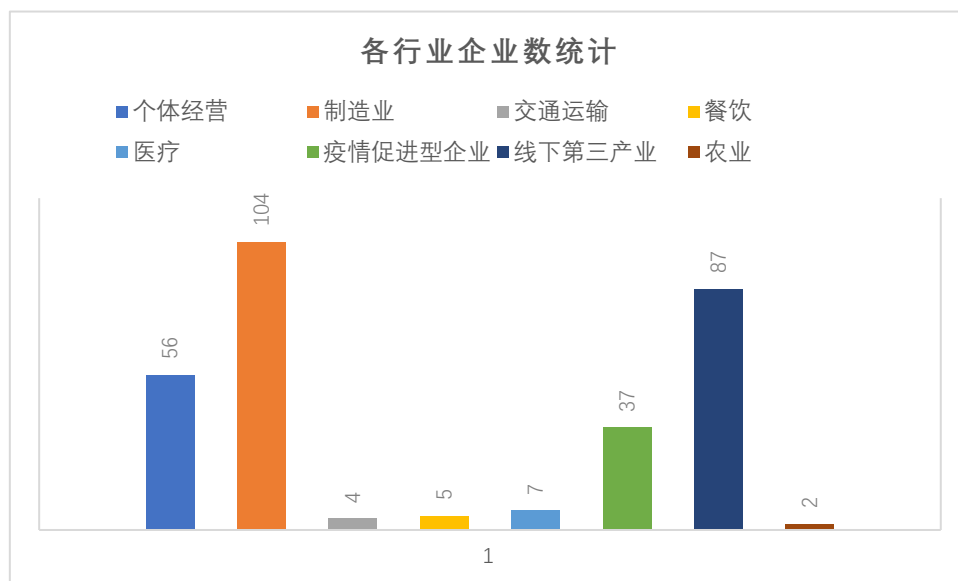


图 9 各行业企业数统计

3. 趋势分析

通过查找文献[3]，我们得出 1.疫情导致部分企业流动性趋紧，企业信用违约风险有所上升，由此在问题二优化目标函数中的可信度 ϵ_i 均会有一定程度的下降。2.疫情导致银行资产业务受限，总体信贷需求不足，因此在问题二优化目标函数中的贷款金额 W_i 与年利率 α_i 均需要进行调整。

4. 策略调整

(1) 银行调整策略

根据前文新冠疫情对各类产业企业的影响程度量化指标，预测得到各类企业未来一段时间各指标的数值，再带入问题一、二中建立的评价指标体系模型，仍然利用问题一、二中的智能优化算法即可求得此种情况下银行的最优信贷策略。

(2) 宏观调控

在此次新冠疫情中，根据图一中国内生产总值以及第一、二、三产业增加值季度累计同比增长率在 2020 年第二季度时有一个较大的回升，国家宏观政策调控在其中有着重要作用。

通过查阅相关文献[6]了解到，国家为促进疫情后的经济发展，鼓励企业复产复工，在信贷政策上对于中小微企业有相应优惠政策。例如文献[6]中提到自治区财政厅和人

民银行乌鲁木齐中心支行特制定个体工商户小额信贷政策，贷款利率按照同期银行贷款市场报价利率（LPR）执行，政府财政给予 50%的贴息。对确因不可抗力因素造成的个体工商户小额信贷损失，按照县（市、区）人民政府承担 80%，承贷金融机构承担 20%的比例分担^[3]。

国家的宏观调控政策也会在很大程度上影响银行信贷政策的制定，例如文献[3]中自治区政府财政的补贴，对于问题二优化目标函数中的贷款金额 w_i 与年利率 α_i 的调整范围会产生影响，因此需要综合考量各方因素。

4.3.2 其他突发因素

文献[4]中将银行面临的突发性风险进行了分类，将其分为了资产安全类突发性风险、营运安全类突发性风险以及信息安全类突发性风险，将每类突发性风险又分为了若干级风险，最后建立了突发性风险评价体系。

因此，针对银行在面对突发因素的总体策略步骤可以首先对突发因素进行分类，根据 2006 年 1 月国务院颁布的《国家突发公共事件总体应急预案》^[5]规定，“根据突发公共事件的发生过程、性质和机理，突发公共事件主要分为以下四类：自然灾害、事故灾难、公共卫生事件、社会安全事件。”然后建立突发因素影响评价体系量化突发因素的影响程度。根据该影响程度量化指标可预测得到各类企业未来一段时间各指标的数值，再带入问题一、二中建立的评价指标体系模型，仍然利用问题一、二中的智能优化算法即可求得此种情况下银行的最优信贷策略。流程框图如图 10 所示。

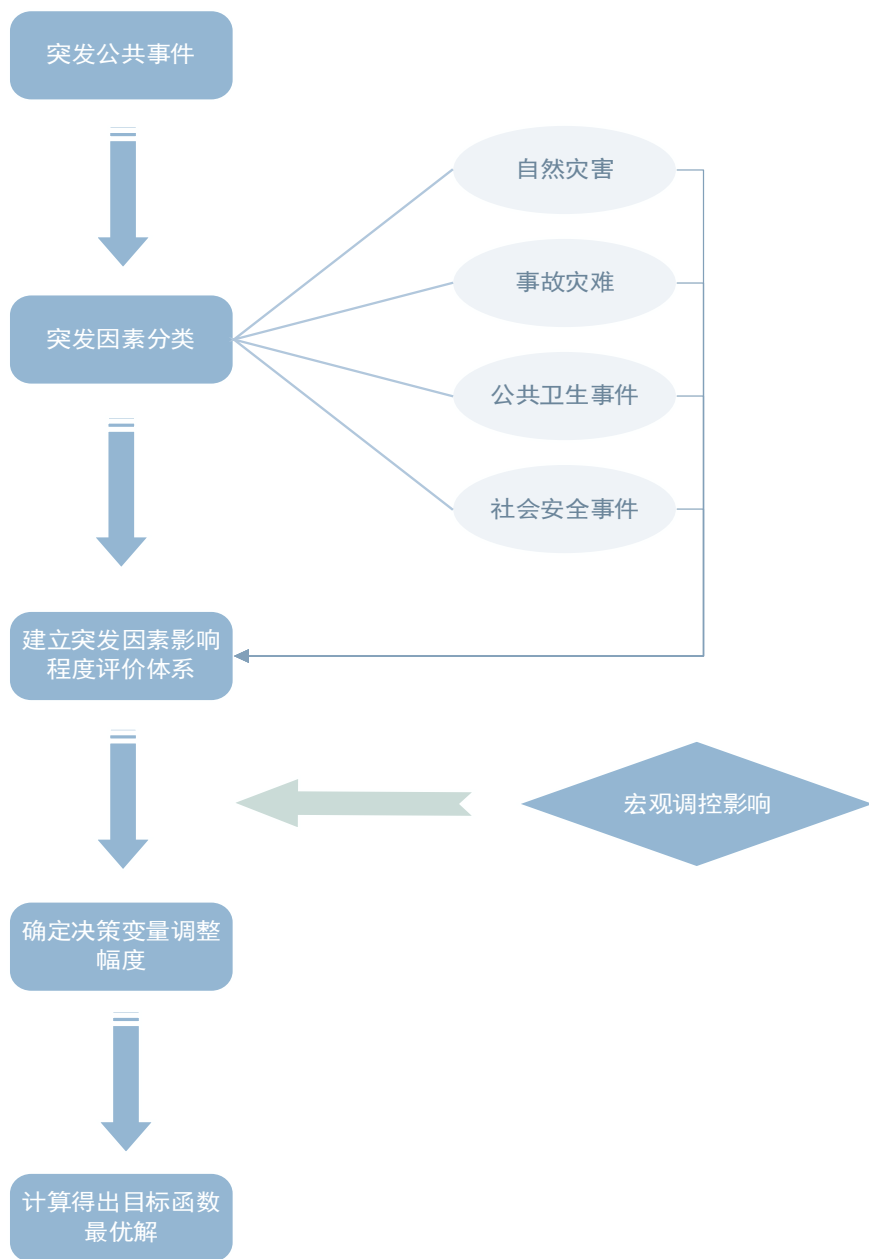


图 10 突发因素影响下信贷策略的调整模型流程图

五、灵敏度分析

5.1 评价体系对六项评价指标的灵敏度分析

在第一问与第二问建立信贷风险量化分析模型时，我们选取了六个指标——

- ① 毛利率 η_{gross} ;
- ② 年平均毛利率方差 var_{gross} ;
- ③ 作废发票比例相反数 η_{del} ;
- ④ 负数发票比例相反数 η_{neg} ;
- ⑤ 销项平均金额 ω_{total} ;
- ⑥ 进销项大单金额比例 η_{big} ;

通过加权后的 TOPSIS 法以及 K-means 聚类分析法将企业综合指标分为了四类，我们需要讨论每个指标的变化对最终 K-means 聚类得到的四类企业的分类情况的变化程度，即为每项指标对整个评价体系的灵敏度水平。

分别改变六项评价指标的数值，设置其变化幅度，得到最终评价结果的改变量，具体结果如表 8 所示。

表 8 项指标数值变化对评价结果的影响

指标	变化幅度	I 类公司数量变化	II 类公司数量变化	III类公司数量变化	IV类公司数量变化
η_{gross}	5%	0	0	-3	+3
	10%	0	0	-3	+3
var_{gross}	5%	0	0	-7	+7
	10%	0	0	-7	+7
η_{del}	5%	0	1	-1	0
	10%	1	-1	+3	-3
η_{neg}	5%	0	0	+3	-3
	10%	-1	+3	0	-2
ω_{total}	5%	2	1	-2	-1
	10%	+5	-3	+2	-4
η_{big}	5%	0	1	0	-1
	10%	1	0	-1	0

根据观察可以得出随着六项指标数值变化的幅度从 5%到 10%，各类公司数量变化较小，不超过 7 个，占 123 家公司比例为 5.69%，这表示六项指标数值变化对评价结果影响不大，验证了我们评价模型的准确性和稳定性。

5.2 权重因子对评价结果的影响

由于在确定六项指标的权重时我们采用了遗传算法计算出最优权重，我们通过改变每项指标的权重，分析最终评价结果的变化，从而可得出权重因子对银行决策方案的影响程度。

由于指标数量较多，为了简化计算，我们每次仅改变一项指标的相对权重，其余指标相对权重不变，再通过权重和为 1 计算得出每项指标的最终权重，再计算该权重下评价方案的最终结果，具体如表 9 所示。

表 9 不同权重对评价结果的影响

指标	变化幅度	I 类公司数量变化	II 类公司数量变化	III 类公司数量变化	IV 类公司数量变化
η_{gross} 权重	10%	0	0	0	0
	20%	0	3	-2	-1
var_{gross} 权重	10%	0	0	0	0
	20%	0	0	-7	7
η_{del} 权重	10%	0	0	-10	10
	20%	0	0	-11	11
η_{neg} 权重	10%	0	0	-3	3
	20%	0	71	-58	-13
ω_{total} 权重	10%	4	1	-56	51
	20%	8	0	-59	51
η_{big} 权重	10%	0	-20	-19	39
	20%	0	33	-55	22

根据观察可以得出随着前三项指标权重（ η_{gross} 权重、 var_{gross} 权重以及 η_{del} 权重）变化的幅度从 10% 到 20%，各类公司数量变化较小，不超过 11 个，占 123 家公司比例为 8.94%，表示这三项指标权重变化对评价结果影响不大。但是对于后三项指标权重（ η_{neg} 权重、 ω_{total} 权重、 η_{big} 权重），当变化幅度从 10% 到 20% 时，对于评价结果影响较大，由此可见后三项指标权重的对于评价模型的重要性，因此我们使用了遗传算法而未采取主观性较强的主成分分析法确定指标权重可以提高评价模型的准确性。

六、模型评价与改进

模型优点：

1. 在选取评价指标时分别从横向和纵向方面对附件一和附件二中信息进行了充分全面的挖掘。对于一些无法获取或建立困难的指标也进行了详细且充分的分析，思维过程清晰严密，体现了在指标选取过程中一步步思考优化的过程，指标选取考虑周到缜密。
2. 选取了评价体系的指标后，对指标进行了合理性与相关性检验，结果显示指标合理性较强且相关性较小，因此指标选取具有科学性且无需再对其进行降维等处理。
3. 问题一中在对评价体系指标赋予权重时，摒弃了常规但过于主观的层次分析法，而是找到了进销项发票信息与企业信誉评级之间的对应关系，将评价模型转化为优化模型，利用了遗传算法对权重进行优化，从而确保赋予的权重具有科学性以及说服力。
4. 在银行制定贷款政策时，利用 K-means 聚类将企业进行分级，且银行对每个等级的公司采取相同的贷款政策，即贷款金额与贷款年利率均相等，从而简少了决策模型中的变量，极大地简化了问题，并且这种分级思想与实际生活贴近，具有现实意义。
5. 在问题三中分析突发情况对银行贷款政策的调整过程中，采用了特殊到一般的思想，首先以新冠肺炎为例进行详细分析然后推广到普遍的突发情况，建立了一套详细的流程框架，具有推广价值和意义。

模型缺点及改进：

1. 在确定大单金额时，我们人为设置金额数量大于等于一万元时为大单，具有较强的主观性，并且大单金额随着行业不同而有所不同，因此在设置的大单金额数目不具有说服力。

改进措施：可根据各个行业中所有企业每笔订单数额的分布来确定大单数额的阈值。但由于在确定订单金额分布过程中订单数量极为庞大，直接处理过于繁琐、效率低下，可简化求订单金额分布的过程。可以首先将订单金额分为了若干个区间（一万以下、一万至十万、十万至一百万.....），然后分区间对订单进行分类统计，得出每家企业处于这些金额区间的订单数量，对每个区间赋予一定的数值即可建立评价指标，可计算得出每家企业的评价指标得分就可以反映上下游企业的影响力。

2. 对于异常数据的处理措施有待改进，在计算每家企业的指标时，出现了异常值的情况（毛利率的绝对值过大、年毛利率方差过大等），这些异常数据在使用遗传算法求解权重时影响过大，我们直接采取了剔除的方法。但最后使用 **k-means** 聚类分析时仍然需要对这些数据进行聚类，这样一来极大地降低了模型的准确度。

改进措施：可以对指标进行某些优化处理，使所有数据能够处在某一区间，从而减少异常数据，而不能随意剔除异常数据。

参考文献

- [1] 秦宇,李钢.新冠肺炎疫情对中国经济挑战与影响的调查综述[J].区域经济评论,2020(03):146-156.
- [2] 高国鹏.新冠肺炎疫情对经济、就业、信贷需求和物价等的影响分析——以内蒙古自治区为例[J].北方金融,2020(05):93-94.
- [3] 自治区关于应对新冠肺炎疫情影响实施小额信贷政策支持个体工商户复工复产的意见[N]. 新疆日报(汉),2020-03-18(003).
- [4] 张宇. 工商银行威海分行突发性风险管理研究[D].哈尔滨工业大学,2018.
- [5] 佚名. 国务院发布《国家突发公共事件总体应急预案》[J]. 中国卫生人才, 2006.

附录

1. 计算指标权值的代码(GA)

main.m

```
%遗传算法主函数
pop_size = 500;
chromo_size = 6;
generation_size = 100;
elitism = 1;
cross_rate = 0.6;
mutate_rate = 0.05;
for m=1:10
    low = -1;
    high = 1;
    pop = low .*ones(pop_size,chromo_size) + (high-
low) .* rand(pop_size,chromo_size); %种群初始化
    best_fitness = 0;
    best_generation = 0;
    for gen=1:generation_size
        fitness; %计算适应度
        rank; %对个体按适应度大小进行排序
        selection;%选择操作
        crossover;%交叉操作
        mutate;%变异操作
    % disp(gen);
    end
    best_ge(m,:)=best_generation;
    individual(m,:)=best_individual;
    value(m,:)=best_fitness;
end
% m = best_individual;%获得最佳个体
% n = best_fitness;%获得最佳适应度
% p = best_generation;%获得最佳个体出现代
final_individual=sum(individual,1)/10;
disp(final_individual);
```

fitness.m

```
%计算种群个体适应度，对不同的优化目标，此处需要改写
%pop_size: 种群大小
%chromo_size: 染色体长度

fitness_value = zeros(pop_size,1);

results = data*pop';
```

```

for i=1:pop_size
    result = results(:,i)';
    [after_sort,index]=sort(result,'descend');
    fit_value = cac1_fitness(index,creditrnk);
    fitness_value(i)= fit_value;
end

```

```

function value = cac1_fitness(index,rank)
    count=0;
    for i=1:113
        if i>=1&&i<=26
            if rank(index(i))~='A'
                count=count+1;
            end
        end
        if i>=27&&i<=62
            if rank(index(i))~='B'
                count=count+1;
            end
        end
        if i>=63&&i<=92
            if rank(index(i))~='C'
                count=count+1;
            end
        end
        if i>=93&&i<=113
            if rank(index(i))~='D'
                count=count+1;
            end
        end
    end
    value=113-count;
    %value = value + 2*matrix(num,5);
end

```

crossover.m

```

% 交叉算法
for i=1:2:pop_size
    if (rand < cross_rate) %cross_rate, 交叉概率
        cross_pos = round(rand * chromo_size); %交叉位置
        if (cross_pos == 0 || cross_pos == 1)
            continue;
        end
        for j = cross_pos:chromo_size %交叉

```

```

        tmp = pop(i,j);
        pop(i,j) = pop(i+1,j);
        pop(i+1,j) = tmp;
    end
end
end

```

rank.m

% 对个体按适应度大小进行排序，并且保存最佳个体

```

[fitness_value,index]=sort(fitness_value);%升序排列
pop=pop(index,:);

% 计算种群累积适应度
fitness_table = cumsum(fitness_value);
% 第G世代种群平均适应度
fitness_avg(gen) = fitness_table(pop_size)/pop_size;
% 设置最佳世代
% best_fitness 最佳适应度
% best_individual 最佳个体
if fitness_value(pop_size) > best_fitness
    best_fitness = fitness_value(pop_size);
    best_generation = gen;
    for j=1:chromo_size
        best_individual(j) = pop(pop_size,j);
    end
end
end

```

mutate.m

% 变异算法

```

for i=1:pop_size
    if rand < mutate_rate %变异概率
        mutate_pos = round(rand(1,2) * chromo_size); %变异位置
        % if mutate_pos == 0
        % continue;
        % end
        for i=1:length(mutate_pos)
            if mutate_pos(i)==0
                mutate_pos(i)=1;
            end
        end
        pop(i,mutate_pos) = low + (high - low) * rand(1,2);
    end
end
end

```

selection.m

%轮盘赌选择法

```
for i=1:pop_size
    r = rand * fitness_table(pop_size); %随机生成一个随机数，在0和总适应度之间
    first = 1;
    last = pop_size;
    mid = round((last + first) / 2);
    idx = -1;
    while (first <= last && idx == -1)%二分查找
        if r>fitness_table(mid)
            first = mid;
        elseif r < fitness_table(mid)
            last = mid;
        else
            idx = mid;
            break;
        end
        mid = round((last + first) / 2);
        if (last - first) == 1
            idx = last;
            break;
        end
    end
    for j=1:chromo_size
        pop_new(i,j) = pop(idx,j);
    end
end
% elitism : 是否精英选择(保留最优值)
if elitism
    p = pop_size-1;
else
    p = pop_size;
end
for i=1:p
    for j=1:chromo_size
        pop(i,j) = pop_new(i,j);
    end
end
end
```

2. TOPSIS 法代码 topsis.m

```
[m,n]=size(input);
input(:,2)=-input(:,2);
for i=1:n
    b(:,i)=input(:,i)/norm(input(:,i));
```

```

end
c=b.*repmat(w,m,1);
Cpos=max(c);
Cneg=min(c);
for i=1:m
    Dis_pos(i)=norm(c(i,:)-Cpos);
    Dis_neg(i)=norm(c(i,:)-Cneg);
end
f=Dis_neg./(Dis_neg+Dis_pos);
[result,index]=sort(f,'descend');

```

3. k-means 聚类分析代码 k_means.m

```

opts = statset('Display','final');
idx = kmeans(zscore(f),4,"distance",'cityblock','replicates',100,'options',opts);

```

4. python 数据处理代码

appendix_money.py

#提取附件 2/1 进项和销项发票中反映的各公司 17-19 年总金额

```

import csv
import pandas as pd
import numpy as np
import xlwt
import xlrd
from xlutils.copy import copy

file=open('附件 2 进项发票.csv','r')
lines=file.readlines()
file.close()
rows = []

for line in lines: #先将 csv 文件中数据去除 opponent 后按行导入 rows 列表
    tmp = line.strip('\n') #去掉 readlines 中包含的'/n'
    tmp = tmp.split(',') #split() 通过指定分隔符对字符串进行切片，如果参数 num 有指定值，则分隔 num+1 个子字符串
    rows.append(tmp)
del rows[0]

id=124
tmplist=[]
companies=[]
for row in rows: #将每个公司的数据编成列表储存在所有公司数据的列表 companies 中
    if row[0]=="E"+str(id) :
        tmplist.append(row)

```

```

else :
    companies.append(tmplist)
    tmplist=[]
    tmplist.append(row)
    id+=1
companies.append(tmplist)

pro17=[]
pro18=[]
pro19=[]
for company in companies:
    tmp17,tmp18,tmp19=0.0,0.0,0.0
    for row in company:
        if "2019/" in row[2]:
            tmp19+=float(row[4])
        if "2018/" in row[2]:
            tmp18+=float(row[4])
        if "2017/" in row[2]:
            tmp17+=float(row[4])
    pro17.append(tmp17)
    pro18.append(tmp18)
    pro19.append(tmp19)

wb = xlwt.Workbook()
ws = wb.add_sheet("附件 2 各年利润")
for i in range(len(pro17)):
    ws.write(i,0,pro17[i])
for i in range(len(pro18)):
    ws.write(i,1,pro18[i])
for i in range(len(pro19)):
    ws.write(i,2,pro19[i])
wb.save("附件 2 各年利润率.xls")

file=open('附件 2 销项发票.csv','r')
lines=file.readlines()
file.close()
rows = []

for line in lines:  #先将 csv 文件中数据去除 opponent 后按行导入 rows 列表
    tmp = line.strip('\n')  #去掉 readlines 中包含的'\n'
    tmp = tmp.split(',') #split() 通过指定分隔符对字符串进行切片, 如果参数 num 有
    #指定值, 则分隔 num+1 个子字符串
    rows.append(tmp)

```



```

del rows[0]

id=124
tmplist=[]
companies=[]
for row in rows: #将每个公司的数据编成列表储存在所有公司数据的列表 companies 中
    if row[0]=="E"+str(id) :
        tmplist.append(row)
    else :
        companies.append(tmplist)
        tmplist=[]
        tmplist.append(row)
        id+=1
companies.append(tmplist)

pro17=[]
pro18=[]
pro19=[]
for company in companies:
    tmp17,tmp18,tmp19=0.0,0.0,0.0
    for row in company:
        if "2019/" in row[2]:
            tmp19+=float(row[4])
        if "2018/" in row[2]:
            tmp18+=float(row[4])
        if "2017/" in row[2]:
            tmp17+=float(row[4])
    pro17.append(tmp17)
    pro18.append(tmp18)
    pro19.append(tmp19)

#list2 = []
#for (i,j) in zip(buy,tax):
#    #list2.append(i)
#    #list2.append(j)
#a = np.array(list2).reshape(len(buy),2)
#np.savetxt('new.csv',a, delimiter = ',')

old_workbook = xlrd.open_workbook("附件 2 各年利润率.xls")
workbook = copy(old_workbook)
worksheet = workbook.get_sheet(0)
for i in range(len(pro17)):
    worksheet.write(i,3,pro17[i])
for i in range(len(pro18)):

```

```

        worksheet.write(i,4,pro18[i])
for i in range(len(pro19)):
    worksheet.write(i,5,pro19[i])
workbook.save("附件 2 各年利润率信息统计.xls")

```

appendix_rebel.py

```

# 统计附件1 中违约、未违约的公司信息
import csv
import xlwt
import xlrd

file=open('附件 1 企业信息.csv','r')
lines=file.readlines()
file.close()
rows = []
for line in lines:  #先将 csv 文件中数据去除 opponent 后按行导入 rows 列表
    tmp = line.strip('\n')  #去掉 readlines 中包含的'/n'
    tmp = tmp.split(',') #split() 通过指定分隔符对字符串进行切片，如果参数 num 有
    #指定值，则分隔 num+1 个子字符串
    rows.append(tmp)
del rows[0]

norebel=[]
rebel=[]
for row in rows:
    if row[3]=="否":
        norebel.append(row[0])
    else:
        rebel.append(row[0])

wb = xlwt.Workbook()
ws = wb.add_sheet("违约")
for i in range(len(norebel)):
    ws.write(i,0,norebel[i])
for i in range(len(rebel)):
    ws.write(i,1,rebel[i])
wb.save("违约.xls")

```