基于中小微企业的信贷决策问题的研究

摘 要

本文主要根据中小微企业的实力、信誉等信息，通过特征工程提取特征数据，综合利用神经网络模型、多组合优化模型、曲线拟合模型、主成分分析模型，对企业的信贷风险进行量化分析及突发因素对企业生产经营和经济效益的影响。解决银行的资金供给结构中与中小微型企业的资金需求达到精准匹配的问题。

针对问题 1，利用了神经网络的思想。对问题 1 所提供的数据，首先进行预处理，提取特征数据，其次用神经网络模型，将 123 家企业的信誉评级作为标签，将处理得到的特征数据作为输入参数，搭建 BP 神经网络并进行 Softmax 映射，根据输出的放贷概率在规定内求解放贷额度，以解决对 123 家企业的具体放款额度。

针对问题 2，运用多组合优化决策的思想，利用 Matlab 软件对表中的数据进行拟合处理，寻找贷款年利率与客户流失率的关系。多次曲线拟合的结果得出客户流失率 y 与年利率 x 服从 Fourier 函数。由于银行的收益主要受到年利率和客户流失率的影响，所以根据 Fourier 函数，分别求解年利率尽可能大的情况下银行的最大收益、客户流失率尽可能小的情况下银行的最大收益，然后再用均值-方差模型解决组合优化决策问题的方法求解出最合适的年利率，以解决银行信贷的组合优化决策问题。

针对问题 3，搜集了大量关于突发因素对企业影响的信息，针对突发因素对企业的综合影响进行了定量评估，然后以新冠疫情为切入点，重点分析了新冠疫情对企业生产经营和经济效益的影响。对问题 1 和 2 中所给出的 425 家企业进行分类，可以大致分成三十个行业。调用 Python 的 jieba 库对 425 家企业的“企业名称”项进行字符串分割提取关键字、剔除通配符和无效字符后，采用 Word2Vec 算法，得到所有企业与 30 个行业进行最大同义化匹配的分类结果。采用主成分分析法对搜集到的新冠疫情对不同行业影响的数据进行统计分析，根据不同层次的贡献度对企业影响的五个程度进行加权求和得到疫情影响系数。

关键字：BP 神经网络 、主成分分析，Softmax、多组合优化决策

一、 问题重述

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。

(2) 在问题1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。

二、 问题分析

中小微企业是我国经济发展的主力军，占我国企业的95%以上，为我国实体经济的发展做出了巨大的贡献。由于我国资金供给主要是传统的商业银行，资金需求的主体为数量庞大的中小微型企业，在以银行为主体的信贷产品供给结构中，一直很难与中小微型企业的资金需求难形成准确匹配，其中很容易存在着资金资源的浪费和错误匹配现象，造成资金资源分配不合理的结果。

问题1是决策类问题，首先对问题1中123家企业的信贷风险进行量化分析，将问题1所给信息中的一些不具体、模糊的因素用具体的数据来表示，从而达到分析比较的目的。考虑其决策因素时，既要考虑银行客户流失率尽可能小的情况下银行利率最大化，又要考

虑中小微型企业在年利率尽可能小的情况下，能够准确的匹配到自己想要申请到的资金。经上述分析，首先要对各项因素进行量化分析，对附件 1 企业信息中的是否违约进行数值化将“是”记为数字“1”，“否”记为数字“0”。对进项发票信息和销项发票信息中的数据进行处理分析，然后再对其他各项因素进行赋权值，即可得到综合评价指标。由于问题涉及因素较多，且情况较为复杂，为科学合理的在银行的供给结构中为中小微型企业的需求形成精准匹配，本问题采用神经网络模型进行求解，以确立合理的决策体系。并可通过该决策体系确定银行在年度信贷总额固定时对这些中小微型企业的信贷策略。

问题2是问题1在特定条件下的问题，问题2中只给出了302家企业的资金流动数据，未对这 302 家企业进行信誉评级，在解决问题 2 时，需要先对企业进行评级，可以参考问题 1 中训练神经网络时采用的标准，由经过 123 家企业完整数据训练完成的神经网络进行评级，然后考虑实际效应对不合理的评级进行人工干预。另外，问题 2 在问题 1 的基础上增加了对贷款总额度的限制。面对企业数增加和放贷限制的情况下，需要考虑多组合优化决策问题，在组合的时候主要从两个角度考虑，以银行利益为基础的利益最大化以及客户流失率最小化。

问题 3 是在问题 2 的基础上，考虑了突发因素（如：新冠肺炎）对各行各业的冲击，银行在对企业进行放贷时要考虑“高收益，低风险”的策略，因此，在突发因素的影响下，应及时对信贷策略进行调整。例如在新冠肺炎的影响，对公开数据进行分析得到新冠肺炎对各行业的冲击，然后对问题中的 425 家企业进行分析得出其归属行业，然后对问题 2 中的信贷策略进行调整。

三、 模型假设

1. 假设仅考虑客户流失率和年利率对银行收益有影响。

2. 假设银行仅从企业的交易额度、客户数量、税额总计、成交率、信誉评级几个方面来衡量贷款项目的收益和风险。

3. 假设银行针对企业申请的贷款，做出的决策只有贷与不贷两种情况。

4. 假设在对行业进行分类时，只考虑此企业匹配度最高的行业。

5. 假设一个企业只归属于一个行业。

四、 符号说明

序号	符号	说明
1	X_1	A 等级客户信贷利率
2	X_2	B 等级客户信贷利率
3	X_3	C 等级客户信贷利率
4	y_1	A 等级客户流失率
5	y_2	B 等级客户流失率
6	y_3	C 等级客户流失率
7	$W_{总}$	银行信贷总额
8	$A_{总}$	对 A 等级客户的放贷总额
9	$B_{总}$	对 B 等级客户的放贷总额
10	$C_{总}$	对 C 等级客户的放贷总额
11	$I_{总}$	企业进行总额
12	$O_{总}$	企业销售总额
13	$P_{总}$	企业盈利总额
14	$V_{总}$	企业有效发票总数
15	r	企业成交率
16	R	取消交易单数
17	M	企业月流量
18	Pr	信贷利率
19	L	客户流失率

注：未列出符号以符号出现处为准

五、模型建立与求解

问题 1——基于企业信贷策略问题的研究

对附件 1 企业信息中的是否违约进行数值化，将违约记录映射为 0 和 1，作为神经网络中的一层参数。销项发票信息中的价税合计总和减去进项发票信息中的价税总和的结果作为神经网络中的二层参数，发票状态作为神经网络中的三层参数，利用神经网络模型对这三层参数进行拟合。将企业信息中的信誉评级赋予

权重进行归一化。将分类结果通过 Softmax 映射在[0,1]区间内，得到各中小微型企业销售总额概率，将此概率与银行放贷最高额度 100 万相乘得出银行对放贷金额，再将不符合放贷要求的企业剔除，得出最终银行对各企业的信贷策略。

1. 数据处理模型

在附件的各项数据中，我们需要对数据进行预处理，从中提取我们出对模型的建立有用的信息。

(1) 数值化

由于企业是否违约是判断企业信誉等级的一个关键性指标。因此我们对企业信息中的是否违约进行数值化处理，简化了算法的复杂度，将“是”记为数字“1”，“否”记为数字“0”。企业进项与销项中的发票状态是衡量该项交易的重要参考标准，也对其进行 0-1 化处理，将“有效发票”记为“1”，将“无效发票”记为“0”。

(2) 盈利额度

从企业的进项发票信息中提取出发票状态为有效发票状态情况下，该单发票的税额总计， 最终求出该公司的进项总额，记为$I_{总}$；同理从企业的销项发票信息中提取出发票状态为有效发票状态情况下，该单发票的税额总计， 最终求出该公司的销项总额，记为$O_{总}$。进项总额与销项总额之差，为该企业的盈利额度，记为$P_{总}$。此额度放映了该企业的量级，可作为银行对企业放贷策略的重要参考标准。

$$P_{总} = O_{总} - I_{总}$$

(3) 客户数量

提取进项发票信息中“销方单位代号”，统计出该企业的原材料商的数量。提取销项发票信息中“销方单位代号”，统计出该企业的客户数量，将该企业的两项数据进行比对，可判断出该企业的供求关系是否稳定。可作为银行考虑对企业放贷策略时的重要参考标准。

(4) 税额总计

提取发票信息中“税额”信息，统计出该企业每年的总税额。总税额反映了企业对国家财政的贡献。可作为银行考虑对企业放贷策略时的重要参考标准。

(5) 成交率

在销项发票信息表中，存在负数发票。负数发票表示客户不满意产品，对产

品进行退货处理，统计出取消交易单数，记为 R，在与有效发票总数($V_{总}$)作比，求出该企业的成交率，记为 r。退货率反映了客户对该企业产品的满意度。可作为银行考虑对企业放贷策略时的重要参考标准。

$$r = R/V_{总}$$

(6)月流量

在销项发票和进项发票中，统计出该企业在某月份的进、销货单数，为月流量，记为 M。月流量反映了企业的“健康”状态，健康的企业应当有一个较为稳定的月流量。可作为银行考虑对企业放贷策略时的重要参考标准。

2．神经网络的搭建

(1)BP 神经网络

在附件一中，对 123 家企业具有详细的信誉等级，将其余数据按照数据处理模型，进行处理。处理后得到几个有价值的参数，作为输入项，搭建 BP 神经网络。图 5-1 为 BP 神经网络的工作流程图。

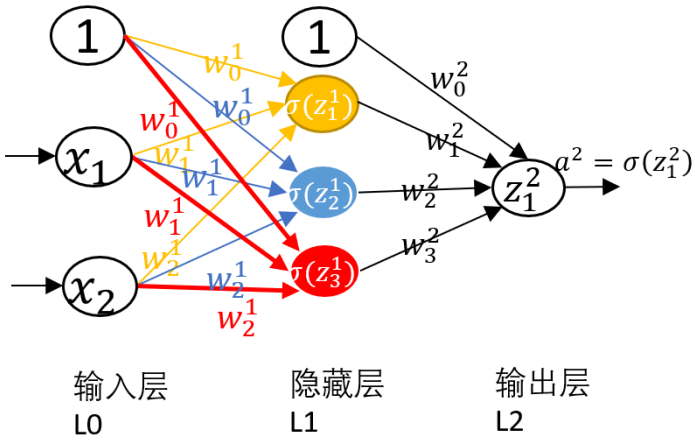

图 5-1　BP 神经网络流程图

根据以上特征工程提取的数据特征，输入样本进行神经网络训练，将已知的信誉等级作为标签，通过实际输出样本与样本输出的误差，来修订网络的连接权值以达到拟合误差的目的。最终生成一个可根据输出项判别企业信誉等级的神经网络模型。

（２）Softmax 映射

将神经网络的输出结果进行 Softmax 函数处理，将结果映射在[0, 1]区间，得到各企业在银行信贷的权重。将此权重于银行信贷总额的最大值 100 万相乘，得到各企业的信贷额度。由于银行信贷额度为 10 万-100 万。故得出的信贷额度低

于 10 万的企业，不予放贷。图 5-2 展示了部分企业额度，具体请看附录。

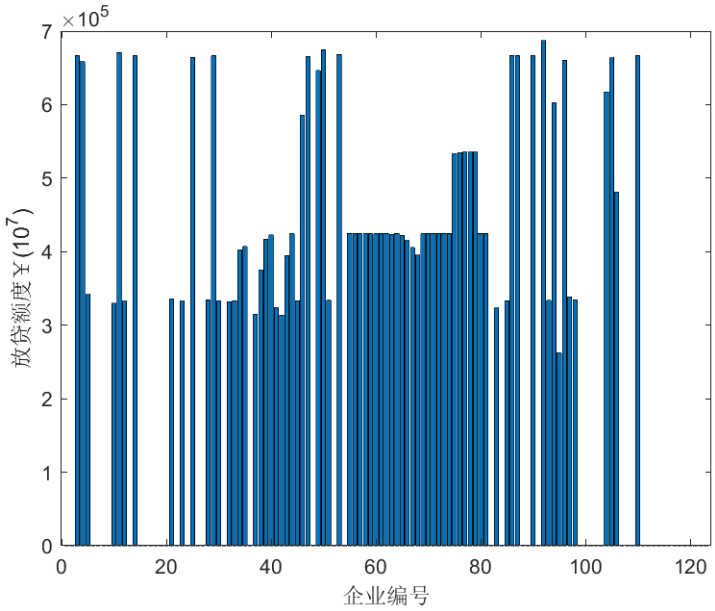

图 5-2 企业信贷额度

3. 利率-客户流失率模型

(1) 模型建立

根据材料的各项数据，并不能直接看出各项变量之间是否有线性关系。为了得到变量之间的关系，用曲线拟合对数据进行处理，寻找贷款年利率与客户流失率的关联。

设自变量 x 为贷款年利率，因变量 y 为客户流失率，经过多次数据拟合处理，得到自变量 x 是影响因变量 y 的主要因素，多次曲线拟合的结果得出客户流失率 y 与年利率 x 服从 Fourier 函数，拟合效果图如图 5-3 所示。

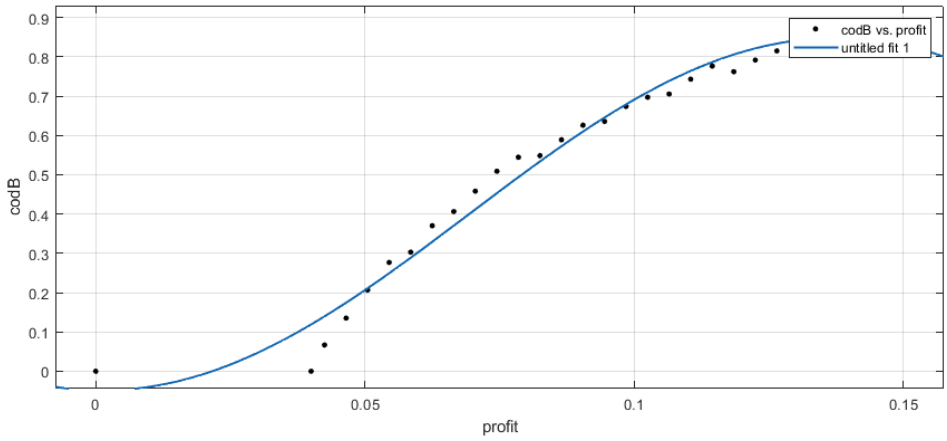

图 5-3 Fourier 函数

(2) 利率-客户流失率 Fourier 函数

客户流失率与银行贷款年利率的高低有着直接的联系，对材料中所给的贷款年利率与公司的信誉评级 A、B、C 分别进行拟合，得到三个企业信誉等级下的 Fourier 函数分别如下：

1)A 信誉评级用户

随着银行贷款利率X_1的增长，银行对 A 等级用户的流失率y_1服从下列函数表达式：

$$y_1 = 0.4167 - 0.469 * \cos（23.31 * X_1） - 0.003792 * \sin（23.31 * X_1）$$

2)B 信誉评级用户

随着银行贷款利率X_1的增长，银行对 B 等级用户的流失率y_2服从下列函数表达式：

$$y_2 = 0.3994 - 0.4481 * \cos(23.06 * X_2) - 0.01227 * \sin（23.06 * X2）$$

3)C 信誉评级用户

随着银行贷款利率X_3的增长，银行对 A 等级用户的流失率y_3服从下列函数表达式：

$$y_3 = 0.4167 - 0.469 * \cos(23.31 * X_3) - 0.003792 * \sin（23.31 * X_3)$$

(3) 模型求解

银行在规定对各企业的信贷利率时，应当参照客户信誉评级与利率函数，不同信誉评级的企业，应当拥有不同的信贷利率。

问题 2——多组合优化决策

问题 2 相对于问题 1 增加了对银行信贷总额度的限制，将 302 家的详细信息按照问题 1 中的处理方法进行处理后，放入神经网络，由神经网络对 302 家企业进行运算后，得出这 302 家的信贷额度。结合实际，对这 302 家企业中信贷额度不合适的，进行人工干预。经神经模型加人工干预后，302 家企业的信贷额度已经确定。此时需要在信贷总额确定的前提下，从银行利益最大化和银行客户流失率最低化，对这 302 家企业放贷组合及贷款额度组合进行优化。

1. 信贷额度的求解

(1) 数据处理

将问题 2 中的 302 家企业的数据，参照问题 1 中的数据处理模型进行数据处理。在对企业是否违约的处理中，由于此 302 家企业没有违约记录，故将所有企业的违约记录都记为“0”。

(2) 信誉评级

将数据处理模型处理后的数据作为输入项，经神经网络处理后，得到 302 家企业的信誉等级。

(3) 信贷总额求解

同问题 1 将神经网络的输出结果，经 Softmax 函数处理后映射在[0, 1]区间，与银行信贷额度的最大额做乘积，得出对每家企业的信贷额度。由于银行信贷额度为 10 万-100 万。故得出的信贷额度低于 10 万的企业，不予放贷，高于 100 万的按 100 万处理。

2. 基于银行利益最大化的多组合优化

银行的信贷业务从本质上来说是一种投资业务，产生高收益的背后往往需要承担高的风险，银行在选择信贷组合的过程就是在风险与收益中寻找一种平衡。

银行的收益除了与年利率有关外，也与客户流失率有着紧密联系。由于目前贷款业务的多样化，客户流失率是银行在制定信贷策略的过程中不得不考虑的问题，银行若想长久的盈利，就要尽可能的减少客户流失率，而客户流失率又与银行贷款的年利率有着直接的关系。综上分析，实现银行利益最大化最根本的问题就是针对不同中小微型企业的信誉评级分别制定合理恰当的年利率。可以采用线性规划求解出针对不同等级用户宏范围的最优年利率。

(1) 银行利益最大化

对 302 家企业的数据进行神经网络处理之后，可以得到 302 家企业的信誉评级以及相应的贷款额度，将所有 A 信誉评级用户的信贷额度求和可以得到 A 类客户信贷的总额度，记为$A_{总}$；将所有 B 信誉评级用户的信贷额度求和可以得到 B 类客户信贷的总额度，记为$B_{总}$；将所有 C 信誉评级用户的信贷额度求和可以得到 C 类客户信贷的总额度，记为$C_{总}$。

由此可知银行在信贷业务中的总收益（$W_{总}$）为：

$$W_{总} = A_{总} * X_1 + B_{总} * X_2 + C_{总} * X_3$$

则求银行利率最大化的问题可转化为以下所示的规划类问题：

$Max = W_{总}$

约束条件为：

$$\begin{cases} 0.04 \le X_1 \le 0.15 \\ 0.04 \le X_2 \le 0.15 \\ 0.04 \le X_3 \le 0.15 \\ (1-y_1) * A_{总} + (1-y_2) * A_{总} + (1-y_3) * A_{总} \le 10\text{^}8 \end{cases}$$

⑵客户流失率最小化

在保证银行收益尽可能大的情况下，需要保证客户流失率的最小化。上文中通过对材料中数据的拟合，分别得到了 A、B、C 三个信誉评级客户流失率与年利率的对应关系。X_1、X_2、X_3分别对应 A、B、C 三个信誉评级的利率，y_1、y_2、y_3分别为 A、B、C 三个信誉评级对应下的客户流失率，由此可以得到客户流失率对最小化的问题也可转化为以下所示的规划类问题：

$Min = y_1 + y_2 + y_3$

约束条件为：

$$\begin{cases} 0.04 \le X_1 \le 0.15 \\ 0.04 \le X_2 \le 0.15 \\ 0.04 \le X_3 \le 0.15 \\ (1-y_1) * A_{总} + (1-y_2) * A_{总} + (1-y_3) * A_{总} \le 10\text{^}8 \end{cases}$$

⑶均值-方差模型解决组合优化决策问题

由上述两类规划求解出在两种考虑中，在客户流失率最低的情况下考虑最大收益。用方差表示客户流失率，用均值表示银行收益。建立银行信贷业务中的均值-方差模型，解决银行信贷的组合优化决策问题。

3. 基于利率-客户流失率模型的人工干预模型

银行针对不同的信誉等级应当具有不同的利率优惠，但是为了追求利益的最大化，不能将某一等级的利率定为统一的值，即在相同的等级内也应当存在不同的区分段。以达到银行利益最大化的 s 目标。

⑴利率分段

银行的贷款利率为 4%-15%，设置 6%、8%、10%、12%、14%五个节点，将利率分为六段。

⑵客户流失率计算

将上述利率节点带入利率-客户流失率模型中，得到相应的结果，如表 1 所示：

表⑴客户流失率

L ＼ Pr	0.04	0.06	0.08	0.10	0.12	0.14	0.15
y_1	0.1341	0.3326	0.5489	0.7371	0.8568	0.8827	0.8577
y_2	0.1190	0.3040	0.4025	0.6909	0.7758	0.8469	0.8289
y_3	0.1099	0.2910	0.2971	0.6783	0.7270	0.8470	0.8933

⑶模型求解

从可以得出 A、B、C 在哪个区间内时，客户流失率较为平缓。针对不同等

级的客户，在各自相应的区间内区别对待各自等级内的用户。再由均值-方差模型求得相应区间的局部最优解，将局部最优解定位相应区间客户的信贷利率。

问题 3——突发因素对多组合决策优化的影响

由于现实环境的多样化，在现实中突发因素的出现是不受控制的，且某些突发因素对各行各业的冲击效果也不相同，因此在银行对企业的信贷策略中，应考虑该行业在某些突发因素的影响下，对该企业的冲击。

本文搜集了大量关于突发因素对企业影响的信息，针对突发因素对企业的综合影响进行了定量评估，然后以新冠疫情为切入点，重点分析了新冠疫情对企业生产经营和经济效益的影响。按照中国经济报告中的 30 类行业类型，对问题 1、2 中的 425 家企业进行分类，依据中国经济报告数据求出新冠疫情对不同行业的影响系数，供银行放贷决策使用。

1．突发因素可视化模型

(1) 数据收集

以中国经济报告在新冠肺炎中对不同企业的影响的抽样调查表中，样本分为五个等级：“非常不利”、“不利”、“没有影响”、“有利”、“非常有利”。将各种数据整理后得到附录，附件显示了在新冠肺炎的冲击下，样本反映出了不同企业受影响的程度。

(2) 主成分分析

主成分分析（Principal components analysis，PCA）的原理就是将原始多个变量数据映射到一个新的环境中，相当于矩阵分析中将一个矩阵中的数据投影到行和列中，类似于坐标系，但在这个投影后的坐标系中，不需要原来样本中的众多变量，只需要最大的一个线性无关组对应的坐标即可，即完成了降维。

将整理的数据进行主成分分析后，得到各项数据的贡献率，剔除贡献率低的维度，将贡献率作为各项样本数据的权重进行加权求和得到疫情影响指数，系如图所显示的结果。图 5-4 所示，可得出得到第 5 项属于噪声项，前 4 项即可代表全部数据空间，即在新冠肺炎的影响基本为“非常不利”、“不利”、“没有影响”、“有利”。

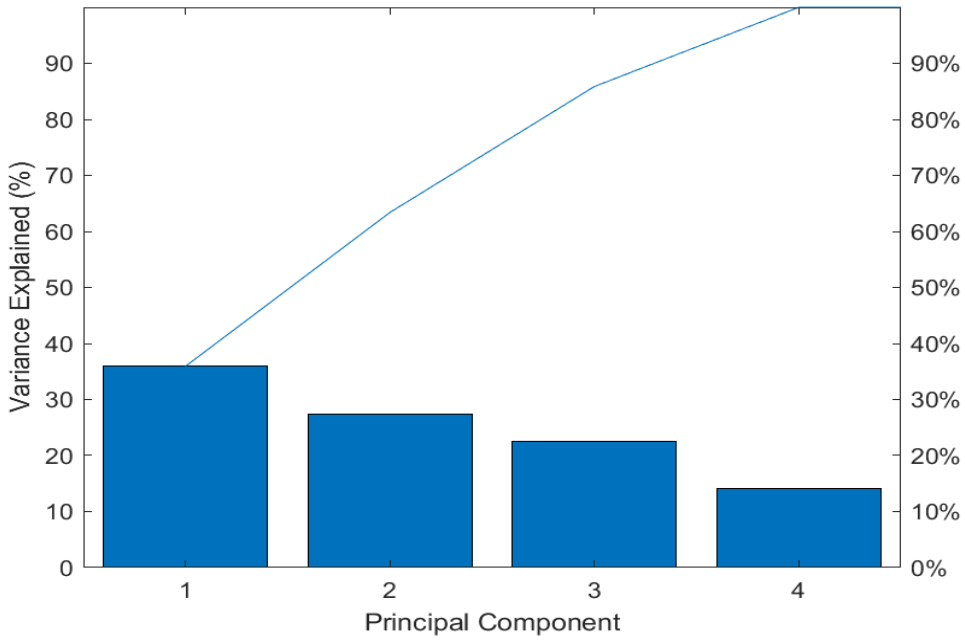

图 5-4 贡献率

(3) 分类

利用 python 提取 425 企业中“企业名称”项，并将提取出的名称进行字符串分割后，剔除通配符和无效信息后，与 30 个行业进行最大同义化匹配，对 425 家企业标注出其所属行业。统计 425 家企业中各行业的总数，其分布图如图 5-5 和图 5-6 所示：

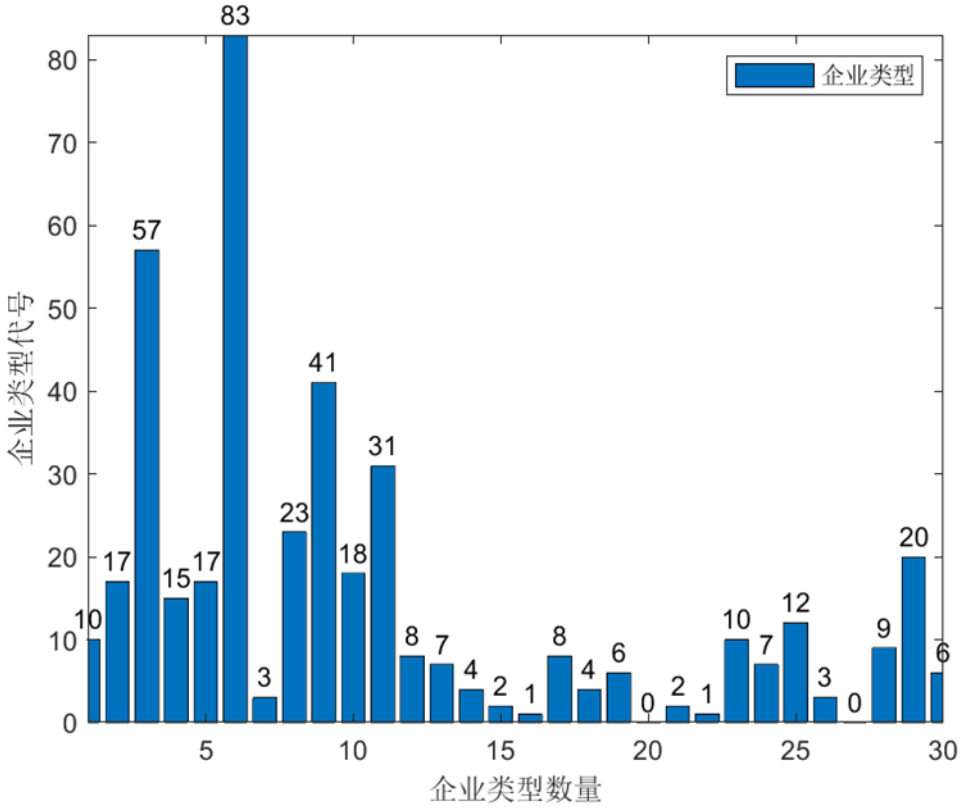

图 5-5 分布柱状图

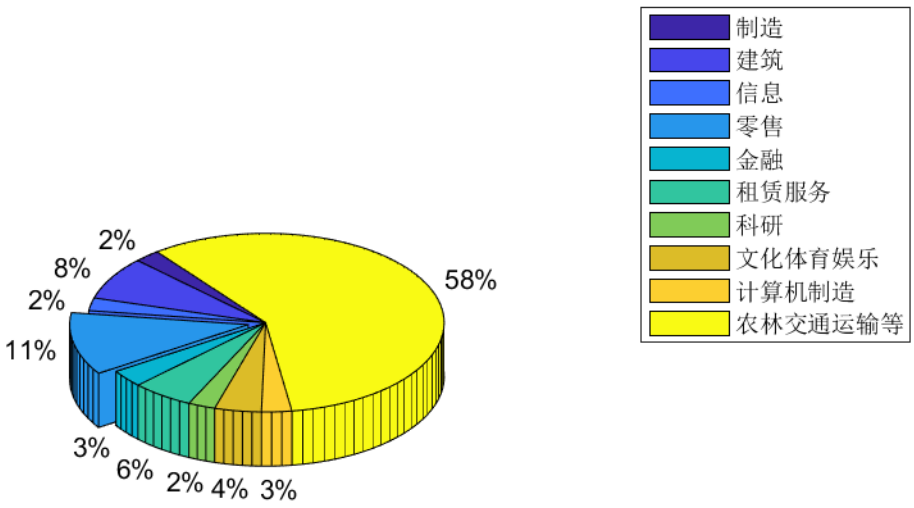

图 5-6 分布饼状图

2. 突发因素影响下的多组合优化的人工干预模型

由于突发因素属于不可控参数，但对于任何突发因素都可通过上述模型，对突发因素进行可视化处理。由于突发因素的不确定性，问题 3 的解决是在问题 2 模型的基础上进行人工干预决策模型。人工干预决策时有以下几个参考点：

(1)降低信贷等级

中小型企业的规模较低，某些突发性因素可能会对该行业内的企业造成毁灭性打击，因此对此类企业进行降低信贷等级的处理。

(2)降低信贷额度

对于突发情况有影响的但影响较小的企业，适当降低对其的放款额度。

(3)增加信贷额度

对于在突发因素中没有大的影响的企业，为了追求银行利益最大化，应当增加该企业信贷额度。

（4）注意金融变化的时效性。

六、模型的改进的推广

1.模型的改进

在搭建神经网络时，对于输入项参数的选择时，应当与实际情况相结合，考虑银行在评价某一企业的信誉额度时，对各项指标的不同侧重，在搭建神经网络时应当考虑各参数权重。

2.论文推广

此模型根据某些参考指标，具有判定级别的功能。在所有需要对客户进行评定的实际问题中，都可使用此模型。在输入的参考项足够全面的条件下，可以在一定程度上认为此模型的判定结果具有客观公正性。

七、模型优缺点

表 2 模型的优缺点

优点	缺点
（1）模型中针对不同类型数据，具体问题具体分析，为每种情况设计了不同的算法。	（1）模型中仅考虑了部分特征，模型效果仍有待提升。
（2）使用主成分分析法巧妙的得出各项数据的贡献率作为权重求出疫情冲击的影响系数。	（2）模型中仅考虑了一个企业只归属于一个行业。企业分类算法的匹配准确率对于相关参数比较敏感，在对行业进行分类时，只考虑此企业匹配度最高的行业。
（3）模型中以初始的人工评级数据作为训练集，问题数据作为测试集。得出预测结果。	（3）模型中缺乏对时效性影响的考虑，人工干预量还有待进一步减少。

八、参考文献

[1] 李航. 统计学习方法[M]. 清华大学出版社, 2012.

[2] 林乐芬, 李永鑫, LIN,等. 商业银行中小微企业信贷产品供求匹配问题研究[J]. 南京大学学报(哲学·人文科学·社会科学), 2016.

[3] 李蓓蕾. 多次自适应最小二乘曲线拟合方法及其应用[D]. 长江大学, 2014.

[4] 卓金武. MATLAB 在数学建模中的应用. 第 2 版[M]. 北京航空航天大学出版社, 2014.

[5] 周志华. 机器学习 : Machine learning[M]. 清华大学出版社, 2016.

[6] 宣士斌. 带权稀疏 PCA 算法及其应用[J]. 重庆大学学报, 2014(04):49-54.

[7] 哈林顿李锐. 机器学习实战 : Machine learning in action[M]. 人民邮电出版社, 2013.

[8] 邵峰晶. 数据挖掘原理与算法[M]. 水利水电出版社, 2003.

九、附录

附录一：主要程序源码

Problem1

```
%%loaddata.m

clc;
clear;
source1=importdata("1(1).xlsx");
src1=source1.data;
n1=size(src1,1);
score=src1(:,1);
reg=src1(:,2);
data1=[score,reg];
src2=importdata("1(2).xlsx");
str2=src2.textdata;
src2=src2.data;
n2=size(src2,1);
src3=importdata("1(3).xlsx");
str3=src3.textdata;
src3=src3.data;
n3=size(src3,1);

%计算税额
summary=0.0;
add1=[];
add2=[];
for i=1:n1
    for j=1:n2
        if src2(j,1)==i&&src2(j,3)==1
            summary=summary+src2(j,2);
        end
    end
   add1=[add1 summary];
end
summary=0;
for i=1:n1
    for j=1:n3
        if src3(j,1)==i&&src3(j,3)==1
            summary=summary+src3(j,2);
        end
    end
   add2=[add2 summary];
end
fail1=[];
fail2=[];
summary=0;
for i=1:n1
    for j=1:n2
        if src2(j,1)==i&&src2(j,3)==0
```

```
            summary=summary+1;
        end
    end
    fail1=[fail1 summary];
end
summary=0;
for i=1:n1
    for j=1:n3
        if src3(j,1)==i&&src3(j,3)==0
            summary=summary+1;
        end
    end
   fail2=[fail2 summary];
end
summary=0;
all1=[];
all2=[];
for i=1:n1
    for j=1:n2
        if src2(j,1)==i
            summary=summary+1;
        end
    end
   all1=[all1 summary];
end
summary=0;
for i=1:n1
    for j=1:n3
        if src3(j,1)==i
            summary=summary+1;
        end
    end
   all2=[all2 summary];
end
fail=fail1+fail2;
all=all1+all2;
tic_cod=(all-fail)./all;
profit=add2-add1;
profit=profit';
data=[src1';profit';tic_cod];



str2 = deblank(str2);
str3 = deblank(str3);
S2 = regexp(str2, '/', 'split');
S3 = regexp(str3, '/', 'split');
month2=[];
reg=0;
for i=1:n2
    reg=S2{i,1}(2);
```

```
    month2=[month2,str2num(reg{1})];
end
month3=[];
for i=1:n3
    reg=S3{i,1}(2);
    month3=[month3,str2num(reg{1})];
end
month2_idx=src2(:,1);


%%nn.m
T=data(1,:);
P=data(2:end,:);
N=n1;
[pn,minp,maxp,tn,mint,maxt]=premnmx(P,T);
dx=[-1,1;-1,1;-1,1];
net=newff(dx,[3,15,1]);
net.trainParam.goal = 0;
net.trainParam.epochs = 50000;
net.trainParam.lr = 0.01;
net.trainParam.showWindow = 1;
net = train(net,pn,tn);
an = sim(net,pn);
a=postmnmx(an,mint,maxt);
disp(['mse: ' num2str(mse(T-an))]);
list=[];
for i=1:n1
    if data(1,i)~=4
        list=[list,i];
    end
end
MappedData = mapminmax(a, 0, 1);
Map=[];
len=size(list,2)
for i=1:len
    Map=[Map,MappedData(list(i))];
end
figure
plot(1:len,Map,'ro')
save('out.mat','MappedData')


%%out.m
num=size(MappedData,2);
dis=MappedData*1000000;
getout=[];
get=[];
for i=1:num
    if dis(i)<100000
        getout=[getout,i];
    else
```

```
        get=[get,i];
end
end
account=[];
for i=1:size(get,2)
    account=[account,dis(get(i))];
end


%% distributionMake.m
class=round(a);
classA=find(class==1);
classB=find(class==2);
classC=find(class==3);
classD=find(class==4);
A=[];
for i=1:length(classA)
    A=[A,dis(classA(i))];
end
B=[];
for i=1:length(classB)
    B=[B,dis(classB(i))];
end
C=[];
for i=1:length(classC)
    C=[C,dis(classC(i))];
end
for i=1:length(classD)
    dis(classD(i))=0;
end
for i=1:length(dis)
    if dis(i)>1000000||dis(i)<100000
        dis(i)=0;
    end
end
summary_dis=sum(dis);
bar(dis)
hold on
xlabel('企业编号')
ylabel('放贷额度￥(10^7)')
print(gcf,'-dpng','pro1_dis.png');
sumA=sum(A);
sumB=sum(B);
sumC=sum(C);



problem2
%%loaddata.m
clc;
```

```
clear;
source1=importdata("1(1).xlsx");
src1=source1.data;
n1=size(src1,1);
score=src1(:,1);
reg=src1(:,2);
data1=[score,reg];
src2=importdata("1(2).xlsx");
n2=size(src2,1);
src3=importdata("1(3).xlsx");
n3=size(src3,1);
%计算税额
summary=0.0;
add1=[];
add2=[];
for i=1:n1
    for j=1:n2
        if src2(j,1)==i&&src2(j,3)==1
            summary=summary+src2(j,2);
        end
    end
   add1=[add1 summary];
end
summary=0;
for i=1:n1
    for j=1:n3
        if src3(j,1)==i&&src3(j,3)==1
            summary=summary+src3(j,2);
        end
    end
   add2=[add2 summary];
end
fail1=[];
fail2=[];
summary=0;
for i=1:n1
    for j=1:n2
        if src2(j,1)==i&&src2(j,3)==0
            summary=summary+1;
        end
    end
    fail1=[fail1 summary];
end
summary=0;
for i=1:n1
    for j=1:n3
        if src3(j,1)==i&&src3(j,3)==0
            summary=summary+1;
        end
    end
   fail2=[fail2 summary];
```

```
end
summary=0;
all1=[];
all2=[];
for i=1:n1
    for j=1:n2
        if src2(j,1)==i
            summary=summary+1;
        end
    end
   all1=[all1 summary];
end
summary=0;
for i=1:n1
    for j=1:n3
        if src3(j,1)==i
            summary=summary+1;
        end
    end
   all2=[all2 summary];
end
fail=fail1+fail2;
all=all1+all2;
tic_cod=(all-fail)./all;
profit=add2-add1;
profit=profit';
data=[src1';profit';tic_cod];
```

```
%%%数据 2 表%%%
table1=importdata("2(1).xlsx");
m1=size(table1,1);
table2=importdata("2(2).xlsx");
m2=size(table2,1);
table3=importdata("2(3).xlsx");
m3=size(table3,1);
%计算税额
tag=0.0;
plus1=[];
plus2=[];
for i=1+123:m1+123
    for j=1:m2
        if table2(j,1)==i&&table2(j,3)==1
            tag=tag+table2(j,2);
        end
    end
   plus1=[plus1 tag];
end
```

```
for m=1+123:m1+123
    for n=1:m3
        if table3(n,1)==m&&table3(n,3)==1
            tag=tag+table3(n,2);
        end
    end
   plus2=[plus2 tag];
end
getmoney=plus2-plus1;
getmoney=getmoney';
invalid1=[];
invalid2=[];
tag=0;
for i=1+123:m1+123
    for j=1:m2
        if table2(j,1)==i&&table2(j,3)==0
            tag=tag+1;
        end
    end
    invalid1=[invalid1 tag];
end
tag=0;
for i=1+123:m1+123
    for j=1:m3
        if table3(j,1)==i&&table3(j,3)==0
            tag=tag+1;
        end
    end
   invalid2=[invalid2 tag];
end
tag=0;
valid1=[];
valid2=[];
for i=1+123:m1+123
    for j=1:m2
        if table2(j,1)==i
            tag=tag+1;
        end
    end
   valid1=[valid1 tag];
end
tag=0;
for i=1+123:m1+123
    for j=1:m3
        if table3(j,1)==i
            tag=tag+1;
        end
    end
   valid2=[valid2 tag];
end
invalid=invalid1+invalid2;
```

```
valid=valid1+valid2;
valid_acc=(valid-invalid)./valid;
data2=[getmoney';valid_acc];

%%nn.m
T=data(1,:);
P=data(3:end,:);
P_=data2;
N=n1;
[pn,minp,maxp,tn,mint,maxt]=premnmx(P,T);
[pn_,minp_,map_]=premnmx(P_);
dx=[-1,1;-1,1];
net=newff(dx,[2,15,1]);
net.trainParam.goal = 0;
net.trainParam.epochs = 500000;
net.trainParam.lr = 0.01;
net.trainParam.showWindow = 1;
net = train(net,pn,tn);
an = sim(net,pn_);
a=postmnmx(an,mint,maxt);
%disp(['mse: ' num2str(mse(T-an))]);
Map=mapminmax(a,0,1);
plot(1:size(Map,2),Map,'ro')
print(gcf,'-dpng','nn2_dis.png');

%% distributionMake.m
class=round(a);
classA=find(class==1);
classB=find(class==2);
classC=find(class==3);
classD=find(class==4);
A=[];
for i=1:length(classA)
    A=[A,dis(classA(i))];
end
B=[];
for i=1:length(classB)
    B=[B,dis(classB(i))];
end
C=[];
for i=1:length(classC)
    C=[C,dis(classC(i))];
end
for i=1:length(classD)
    dis(classD(i))=0;
end
for i=1:length(dis)
    if dis(i)>1000000||dis(i)<100000
        dis(i)=0;
    end
```

```
end
summary_dis=sum(dis);
bar(dis)
hold on
xlabel('企业编号')
ylabel('放贷额度￥(10^7)')
print(gcf,'-dpng','pro1_dis.png');
sumA=sum(A);
sumB=sum(B);
sumC=sum(C);
```

```
problem3
%%pro_bar.m
clc;
clear;
data=importdata('classify.xlsx');
num=data.data.Sheet3';
n=size(num,2);
name=data.textdata.Sheet3';
numCopy=num;
sum_num=sum(num);
%%计算低于 4%的类型索引
low_idx=[];
for i=1:n
    if (num(i)/sum_num)<0.04
        low_idx=[low_idx,i];
        numCopy(i)=0;
    end
end
%%计算低于 4%的类型数量总和
summary=[];
for i=1:size(low_idx,2)
    summary=[summary,num(low_idx(i))];
end
%%输出高于 4%部分数量集
zero_idx=find(numCopy==0);
numCopy(zero_idx)=[];
high_num=numCopy;
%%计算高于 4%的类型索引
high_idx=setdiff(1:30,low_idx);
%%计算高于 4%的类型名称
high_name=[];
for i=1:size(high_idx,2)
    high_name=[high_name,name(high_idx(i))];
end
%%计算高于 4%的类型数量集
high_num=[];
```

```
for i=1:size(high_idx,2)
    high_num=[high_num,num(high_idx(i))];
end
all_num=[high_num,sum_num];
all_name=[high_name,"农林交通运输等"];
explode = [0,0,0,1,0,0,0,0,0,0];
pie3(all_num,explode)
legend(all_name)
print(gcf,'-dpng','pieout.png');
figure(2)
bar(num)
xlabel('企业类型数量');
ylabel('企业类型代号');
y=abs(rand(1,10)*100);
axis([1,30,min(num),max(num)]);
for i=1:30

text(i,num(i)+0.5,num2str(num(i)),'VerticalAlignment','bo
ttom','HorizontalAlignment','center');
end
legend("企业类型")
print(gcf,'-dpng','barout.png');

%%pcaAnalysis.m
clc,clear;
data=importdata('CER.xlsx');
X=data.data(:,:);
X(1,:)=[];
[x,coeff,sum_ex,latent]=pcaff(X,5);
weight=latent/100';
weight(2)=weight(2)*(1);
weight(3)=weight(3)*(-1);
X(:,5)=[];
for i=1:4
    X(:,i)=X(:,i)*weight(i);
end
Wdata=sum(X,2);
Wdata=-Wdata;
rise_num=[];
rise_idx=[];
for i=1:length(Wdata)
    if Wdata(i)>0
        rise_num=[rise_num,Wdata(i)];
        rise_idx=[rise_idx,i];
    end
end
figure(2)
area(1:26,Wdata(1:26),'FaceColor',[0 100/256
0],'EdgeColor','b')
hold on
```

```
area(26:28,[0,Wdata(27),0],'FaceColor',[192/250 42/256
42/256],'EdgeColor','b')
hold on
area(28:30,Wdata(28:30),'FaceColor',[0 100/256
0],'EdgeColor','b')
hold on
legend("不利系数","有利系数")
hold on
xlabel("企业分类编号");
hold on
ylabel("新冠疫情影响指数")
hold on
print(gcf,'-dpng','areaout.png');
%%pcaff.m
function [data_PCA, COEFF,
sum_explained,latent1]=pcaff(data,k)
% k:前 k 个主成分
data=zscore(data);  %归一化数据
[COEFF,SCORE,latent,tsquared,explained,mu]=pca(data);
latent1=100*latent/sum(latent);%将 latent 特征值总和统一为 100，
便于观察贡献率
data= bsxfun(@minus,data,mean(data,1));
data_PCA=data*COEFF(:,1:k);
figure(1)
pareto(latent1);%调用 matlab 画图 pareto 仅绘制累积分布的前 95%，
因此 y 中的部分元素并未显示
hold on
xlabel('Principal Component');
hold on
ylabel('Variance Explained (%)');
% 图中的线表示的累积变量解释程度
hold on
print(gcf,'-dpng','PCAout.png');
sum_explained=sum(explained(1:k));
```

附录二 123 家企业信贷额度表

企业代码	具体额度
'E1'	0
'E2'	0
'E3'	666667.3
'E4'	658533.6
'E5'	341646.5
'E6'	0
'E7'	0
'E8'	0
'E9'	0
'E10'	329728.1

'E11' 670718
'E12' 332731.7
'E13' 0
'E14' 666666.2
'E15' 0
'E16' 0
'E17' 0
'E18' 0
'E19' 0
'E20' 0
'E21' 335459.8
'E22' 0
'E23' 332650.6
'E24' 0
'E25' 664134.4
'E26' 0
'E27' 0
'E28' 334377.2
'E29' 666584.3
'E30' 333362.6
'E31' 0
'E32' 332161.4
'E33' 333314.3
'E34' 402175.8
'E35' 407043.9
'E36' 0
'E37' 314975.8
'E38' 375353.7
'E39' 416984
'E40' 422767.7
'E41' 323737.5
'E42' 313878.1
'E43' 394576.4
'E44' 424376
'E45' 333333.7
'E46' 586045.7
'E47' 665729.8
'E48' 0
'E49' 646435
'E50' 674473.7
'E51' 333916.6
'E52' 0
'E53' 667985.6
'E54' 0

'E55'	424615.4
'E56'	424615.6
'E57'	424588.8
'E58'	424376.9
'E59'	424613.3
'E60'	424467.8
'E61'	424627.9
'E62'	424627.4
'E63'	423627.2
'E64'	424535.7
'E65'	422018.3
'E66'	415173.2
'E67'	405671.6
'E68'	395313.4
'E69'	424462.2
'E70'	424628.8
'E71'	424621
'E72'	424628.6
'E73'	424629.2
'E74'	424629.1
'E75'	533786.2
'E76'	534775.5
'E77'	535428.5
'E78'	535531.2
'E79'	535503.3
'E80'	424567.8
'E81'	424570.1
'E82'	0
'E83'	323411
'E84'	0
'E85'	333249.3
'E86'	666674.5
'E87'	666671.4
'E88'	0
'E89'	0
'E90'	666667.4
'E91'	0
'E92'	687156.6
'E93'	333780.7
'E94'	602309.5
'E95'	262206.3
'E96'	660231.4
'E97'	338620.7
'E98'	334824.2

’E99’	0
’E100’	0
’E101’	0
’E102’	0
’E103’	0
’E104’	617339.4
’E105’	663576.9
’E106’	481353.3
’E107’	0
’E108’	0
’E109’	0
’E110’	666689.5
’E111’	0
’E112’	0
’E113’	0
’E114’	0
’E115’	0
’E116’	0
’E117’	0
’E118’	0
’E119’	0
’E120’	0
’E121’	0
’E122’	0
’E123’	0

附录三 302 家企业信誉评级

企业代号	信誉评级
’E124’	D
’E125’	D
’E126’	D
’E127’	B
’E128’	B
’E129’	B
’E130’	B
’E131’	B
’E132’	B
’E133’	B
’E134’	C
’E135’	A
’E136’	A
’E137’	A
’E138’	A

'E139'	A
'E140'	A
'E141'	A
'E142'	A
'E143'	A
'E144'	A
'E145'	A
'E146'	A
'E147'	A
'E148'	A
'E149'	A
'E150'	A
'E151'	A
'E152'	A
'E153'	A
'E154'	A
'E155'	A
'E156'	A
'E157'	D
'E158'	D
'E159'	D
'E160'	D
'E161'	D
'E162'	D
'E163'	D
'E164'	D
'E165'	D
'E166'	D
'E167'	D
'E168'	D
'E169'	D
'E170'	D
'E171'	D
'E172'	D
'E173'	D
'E174'	D
'E175'	D
'E176'	D
'E177'	D
'E178'	D
'E179'	D
'E180'	D
'E181'	D
'E182'	D

'E183'	D
'E184'	D
'E185'	D
'E186'	D
'E187'	D
'E188'	D
'E189'	D
'E190'	D
'E191'	D
'E192'	D
'E193'	D
'E194'	D
'E195'	D
'E196'	D
'E197'	D
'E198'	D
'E199'	D
'E200'	D
'E201'	D
'E202'	D
'E203'	D
'E204'	D
'E205'	D
'E206'	D
'E207'	D
'E208'	D
'E209'	D
'E210'	D
'E211'	D
'E212'	C
'E213'	C
'E214'	C
'E215'	B
'E216'	B
'E217'	B
'E218'	B
'E219'	B
'E220'	B
'E221'	B
'E222'	B
'E223'	B
'E224'	B
'E225'	B
'E226'	B

'E227'	B
'E228'	B
'E229'	B
'E230'	A
'E231'	B
'E232'	A
'E233'	A
'E234'	A
'E235'	A
'E236'	A
'E237'	A
'E238'	A
'E239'	A
'E240'	A
'E241'	A
'E242'	A
'E243'	A
'E244'	A
'E245'	A
'E246'	A
'E247'	A
'E248'	A
'E249'	A
'E250'	A
'E251'	A
'E252'	A
'E253'	A
'E254'	A
'E255'	A
'E256'	A
'E257'	A
'E258'	A
'E259'	A
'E260'	A
'E261'	A
'E262'	A
'E263'	A
'E264'	A
'E265'	A
'E266'	A
'E267'	A
'E268'	A
'E269'	A
'E270'	A

'E271' A
'E272' A
'E273' A
'E274' A
'E275' A
'E276' A
'E277' A
'E278' A
'E279' A
'E280' A
'E281' A
'E282' A
'E283' A
'E284' A
'E285' A
'E286' A
'E287' A
'E288' A
'E289' A
'E290' A
'E291' A
'E292' A
'E293' A
'E294' A
'E295' A
'E296' A
'E297' A
'E298' A
'E299' A
'E300' A
'E301' A
'E302' A
'E303' A
'E304' A
'E305' A
'E306' A
'E307' A
'E308' A
'E309' A
'E310' A
'E311' A
'E312' A
'E313' A
'E314' A

'E315'	A
'E316'	A
'E317'	A
'E318'	A
'E319'	A
'E320'	A
'E321'	A
'E322'	A
'E323'	A
'E324'	A
'E325'	A
'E326'	A
'E327'	A
'E328'	A
'E329'	A
'E330'	A
'E331'	A
'E332'	A
'E333'	A
'E334'	A
'E335'	A
'E336'	A
'E337'	A
'E338'	A
'E339'	A
'E340'	A
'E341'	A
'E342'	A
'E343'	A
'E344'	A
'E345'	A
'E346'	A
'E347'	A
'E348'	A
'E349'	A
'E350'	A
'E351'	A
'E352'	A
'E353'	A
'E354'	A
'E355'	A
'E356'	A
'E357'	A
'E358'	A

'E359'	A
'E360'	A
'E361'	A
'E362'	A
'E363'	A
'E364'	A
'E365'	A
'E366'	A
'E367'	A
'E368'	A
'E369'	A
'E370'	A
'E371'	A
'E372'	A
'E373'	A
'E374'	A
'E375'	A
'E376'	A
'E377'	A
'E378'	A
'E379'	A
'E380'	A
'E381'	A
'E382'	A
'E383'	A
'E384'	A
'E385'	A
'E386'	A
'E387'	A
'E388'	A
'E389'	A
'E390'	A
'E391'	A
'E392'	A
'E393'	A
'E394'	A
'E395'	A
'E396'	A
'E397'	A
'E398'	A
'E399'	A
'E400'	A
'E401'	A
'E402'	A

'E403'	A
'E404'	A
'E405'	A
'E406'	A
'E407'	A
'E408'	A
'E409'	A
'E410'	A
'E411'	A
'E412'	A
'E413'	A
'E414'	A
'E415'	A
'E416'	A
'E417'	A
'E418'	A
'E419'	A
'E420'	A
'E421'	A
'E422'	A
'E423'	A
'E424'	A
'E425'	A