

中小微企业的信贷决策

摘要

对于银行来说，向中小微企业提供信贷策略之前，应权衡该企业的偿还能力。因此需要综合考虑企业的实力、信誉、供求关系稳定性等因素来评估信贷风险，以便银行判断是否放贷和制定具体的信贷策略。本文通过建立风险决策模型，对不同企业的信贷策略进行了决策分析。

针对问题一，本文建立了基于贝叶斯决策理论的风险型多属性决策模型。首先对附件 1 中各企业的交易数据进行预处理，删除作废发票所在的交易记录和负数发票所冲销的交易记录。分析预处理后的批量数据集，提取企业规模、供求关系稳定性、企业实力、信誉等级、是否违约和盈利能力 6 个指标，通过数据可视化直观体现了各指标对不同企业的影响程度。其次，依据附件 3 中不同信誉评级下年利息和客户流失率的线性函数关系及实际情况确定了 6 种信贷决策方案，见表 3。同时，通过比对先验概率和后验概率两种情况下的风险度，证实了该模型在很大程度上降低了银行的决策风险。以企业 E1 为例，建立此风险决策模型后选择最优决策方案 q_1 。

针对问题二，本文建立了基于蒙特卡洛模拟数据的 RBF 神经网络预测模型。将特定影响因素下蒙特卡洛模拟得出的数据和问题一中真实的信贷记录数据作为 RBF 神经网络预测模型的训练集，预测得出 302 家企业的信誉评级和是否违约情况。完善决策体系后，将 302 家企业的数据量化矩阵代入问题一中建立的风险型多属性决策模型求解得出各企业的决策方案。同时，对同一决策方案下的企业根据实力大小和规模进一步划分贷款额度。以企业 E124 为例，得到最佳决策方案 q_3 。

针对问题三，通过对新冠病毒疫情下不同类别企业受影响程度的分析，基于问题一中建立的风险型多属性决策模型对原自然状态下的先验概率进行评估修正，得出非自然状态下银行对不同类别企业的先验概率。将 302 家企业的数据量化矩阵代入后得到各个企业的决策方案。同时，对同一决策方案下的企业根据实力大小和规模进一步划分信贷额度。以企业 E124 为例，受新冠病毒疫情影响后，该企业信贷策略从决策方案 q_3 调整至方案 q_4 。

关键词：贝叶斯决策理论；风险性多属性决策；蒙特卡洛模拟；期望效益值；RBF 神经网络预测模型

一、问题重述

1.1 问题背景

对于中小微企业来说，由于自身企业规模较小，抵押资产缺乏，需要向银行贷款。而对于银行来说，首先决定是否对其放贷，即根据企业的实力、信誉对其信贷风险做出评估，通常向实力强、供求关系稳定的企业提供贷款；确定放贷后，再通过企业信贷风险确定贷款额度、利率和期限等信贷策略，一般对信誉高、信贷风险小的企业提供利率优惠。其中，企业的信誉高低、实力强弱等因素可以依据附件中交易票据信息以及上下游企业的影响力得出。

附件 1~3 分别为 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。

1.2 问题提出

现某银行确定信贷策略如下：贷款额度为 10~100 万元；年利率为 4%~15%；贷款期限为 1 年。要求本团队需根据实际和附件中数据信息，通过建立数学模型研究企业的信贷策略，最终解决下列问题：

1. 通过对附件 1 中 123 家企业的数据进行量化分析，首先得出企业的信贷风险，再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，先对附件 2 中 302 家企业的信贷风险进行量化分析，再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. 一些突发因素可能会受到影响企业的生产经营和经济效益，并往往对不同行业、不同类别的企业会有不同的影响。现在需要综合考虑附件 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

银行制定中小微企业的信贷策略受多种因素影响，可以提取附表中有有效数据信息并结合实际情况确定信贷策略评估指标。通过构建的指标评价工具，衡量各企业的信贷风险等因素来确定银行是否放贷以及具体的贷款额度、利率和期限等信贷策略，最终要求给出不同条件下银行给与中小微企业的最佳信贷决策。

2.1 问题一的分析

该题目要求量化分析企业的信贷风险，并给出银行针对各企业的信贷策略。针对此类大数据问题，第一步应当对数据预处理，删除无效数据。由于解答题目的最后会根据选定的指标进行信贷决策，势必会涉及到各指标对企业的影响程度以及具体体现。所以在通读题目并查阅相关资料后，第二步应当选定并处理指标，通过图形将数据可视化，

以便更直观看出维度和多种度量的关系。最后若要进行信贷决策，其实有多种模型可以利用，但若考虑到之后问题的指标缺失或增添，力求求解模型的一体化，可以采用贝叶斯模型，自然状态、突发状态采用不同先验概率，可以大幅度的提高求解的灵活性。

2.2 问题二的分析

相对于问题一，问题二难度在于银行判断是否对企业放贷以及制定具体的信贷策略时，缺少重要指标：信贷记录，包括信誉评级和违约记录。题目提示可在上题基础上进行决策，由此可以通过化繁为简的思路，在补全缺失指标数据后，回归至上题。

可以通过构建基于蒙特卡洛模拟数据的 RBF 神经网络模型预测缺失的信贷记录，解决题目。首先可以利用附件 1 中企业的真实信贷记录和蒙特卡洛模拟的数据作为神经网络训练集，得出所求预测结果的同时，蒙特卡洛的引入也可大幅度地提高训练结果的精确性。其次可将企业的数据量化矩阵代入问题一中建立的风险型多属性决策模型，即可求出决策方案。同时，对同一决策方案下的企业根据实力大小和规模进一步划分信贷额度。最终，求解得出银行信贷总额为 1 亿元时对各个企业的具体信贷策略。

2.3 问题三的分析

和问题二相比，该题目是评估指标的增添，同样也是多属性决策模型问题。贝叶斯模型的特性在此大放异彩，由于上述问题中已经得出自然状态下的先验概率，在发生突发因素后，只需引入重新修正更新后的先验概率，依旧采用贝叶斯模型即可求解得出银行信贷总额为 1 亿元时对各个企业的具体信贷调整策略。

三个问题实则可归为一个问题，即多属性决策模型问题。或因评估指标的缺失或增添要对其进行预测或者是容纳，使问题解决变得不直观。但应找出适当的决策模型，其特征要灵活化，可控化，争取在不同条件时只是通过特定方式调控，求解方法不变。

三、模型假设

1. 假设问题一中效率函数为线性加权和形式。
2. 假设问题一、二中企业的生产经营和经济效益不受突发因素影响。
3. 假设问题二中特定因素下蒙特卡洛模拟生成的数据精度值较高。
4. 假设问题三中收集的新冠肺炎对不同类别行业的影响数据真实、准确。
5. 假设问题三中收集的数据受除新冠肺炎外其他因素的影响程度较小。
6. 假设银行对各企业的信贷策略不会因为不可控的因素而影响。

四、符号定义与说明

符号	说明
θ_t	状态变量
q_h	决策变量
f_k	决策指标
x^*	指标 f_k 的最佳稳定值
u_{ht}	方案 q_h 在自然状态下 θ_i 下的效益值
$E(U)^*$	最优方案
ρ	风险度
$\sigma(U_h)$	方案 q_h 的效益值方差
$f_m(x)$	隐含层连接函数
c_i	第 i 个径向基函数的中心值

五、问题一：分析并决策有信贷记录企业

本问题目标为：①量化分析附件 1 中各企业的信贷风险，②制定银行对中小微企业的信贷策略，最终使银行获得最大利润。本团队将该题目分为三步，即数据预处理，指标选定、处理和说明，信贷方案决策，具体步骤如下图 1 所示：

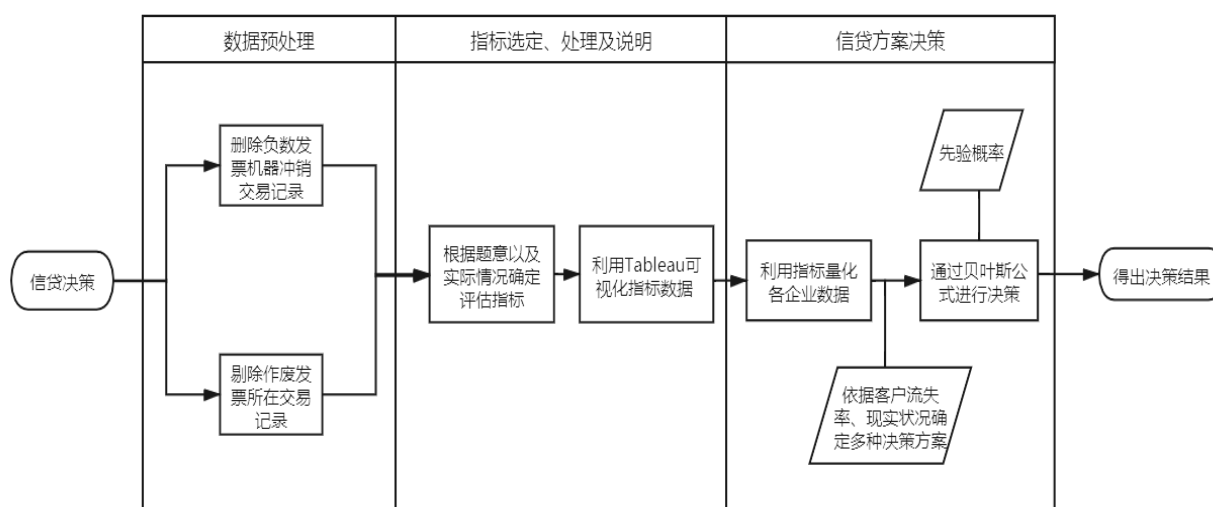


图 1 问题一的信贷策略决策步骤

5.1 数据预处理

通过观察分析附件 1 中的 123 家有信贷记录企业的交易数据，决定首先应预处理表中数据。其中包括：删除表中出现的作废发票交易记录，还应剔除由于企业入账计税后因退货退款冲销后的交易记录。以便选择和题意关联度较高的有效数据指标。

5.2 指标选定、处理及说明

5.2.1 指标的选定

根据题目中谈及到的两个步骤，即量化分析该企业贸易情况以决定是否放贷和决策各企业详细的信贷策略，本团队从该两步出发，逆向思维以选定评估指标。

首先对于量化分析交易信息、评估企业信贷风险等因素以决定是否放贷来说，由于银行更倾向于向实力强、供求关系稳定的企业提供贷款，所以决定选择企业规模（同类型交易时税率大小）、供求关系稳定性（企业发票频数或方差）、企业实力（平均单次交易额度）为评估指标。

其次对于决策信贷策略来看，最终得出的结果无非是确定的贷款额度、利率以及贷款期限。其中因为银行针对信誉高、信贷风险小的企业给予了利率优惠，并且银行的贷款年利率与各种客户的流失率有动态关联，所以在此方面选定信誉等级（直接数据量化）、是否违约（直接数据量化 1 或 0）、盈利能力（净利润）为评估指标。

最终选定指标如下：

- ①企业规模 ②供求关系稳定性 ③企业实力 ④信誉等级 ⑤是否违约 ⑥盈利能力

5.2.2 指标可视化处理和说明

①企业规模：针对企业规模量化来说，本团队利用企业同时段交易时所缴纳税率，即同时段附件 1 中销项或开项发票信息中所缴税额/该笔交易金额。由此可以得出 123 家有信贷记录的企业规模如图 2 所示：

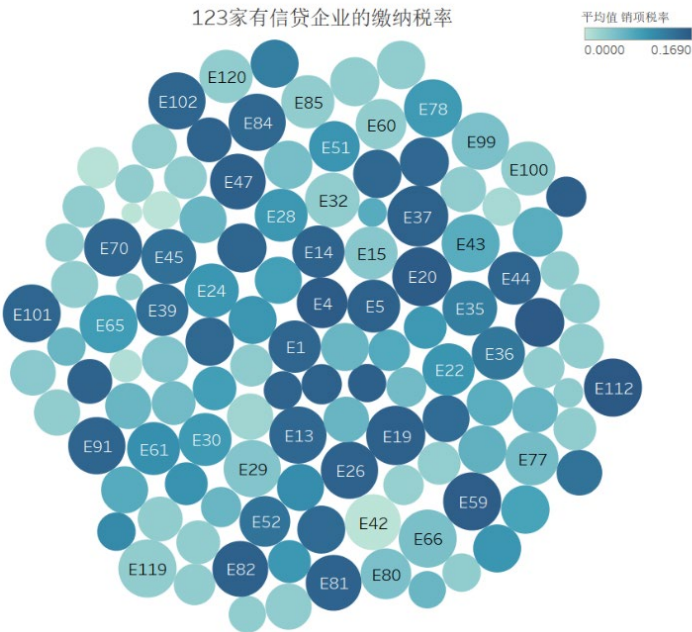


图 2 附件 1 中有信贷记录企业的缴纳税率

图 2 中图形大小表示进项税率，颜色深浅表示销项税率。通查阅税务缴纳相关资料后，规定判断标准为若该企业缴纳税率越高，则该企业规模越大。由此可直观看出，附件 1 中所涉及到的中小微企业的缴纳销项税率为(0,0.169)，其中多家集中在 0.14 左右，说明企业的规模更多集中在小型，并且中、小、微型企业组别的代号在 E1~E123 中是随机分布的。

②**供求关系稳定性：**当选定任意企业，分析一时间段的进项和开项交易频率时，如果一个企业的供求关系较为稳定，那么则要求交易频率更高，更能反映其贸易的供求的稳定性。选定企业代号、开票时间可以得出时间序列下所有企业的开票频率，如图 3：

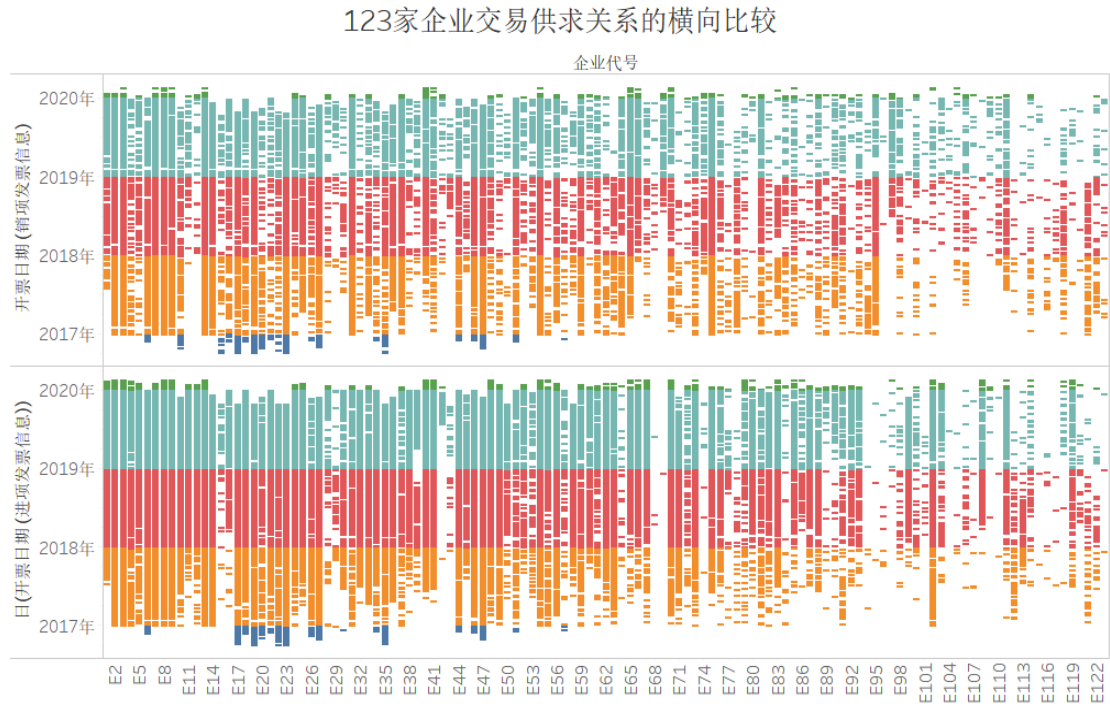


图 3 时间序列下企业供求关系比较

从图 3 中可以看出，当前所有企业的进项频率远大于销项频率，说明存在“进货的次数多但可能数目少，出货次数少但数目多”的问题，符合中小微企业的特征：该种交易类型可大程度避免货物积压，以便资金流转。另一方面，后半部分的企业 E98~E123 与前面的企业相比，其贸易频率断崖式减少，由此可以得出企业代号为 E2~E40、E32~E65 的企业供求关系更为频繁、稳定。通过各企业在时间序列季度下的有效交易次数增长率和方差关系对其进行综合量化处理。

③**企业实力：**由于企业自身性质原因，可能随时期浮动范围较大，为体现企业实力的宏观实力，可以利用企业平均单次时交易额来确定实力如何。规则如下：当企业平均单笔交易额约高则实力越强，反之越弱。

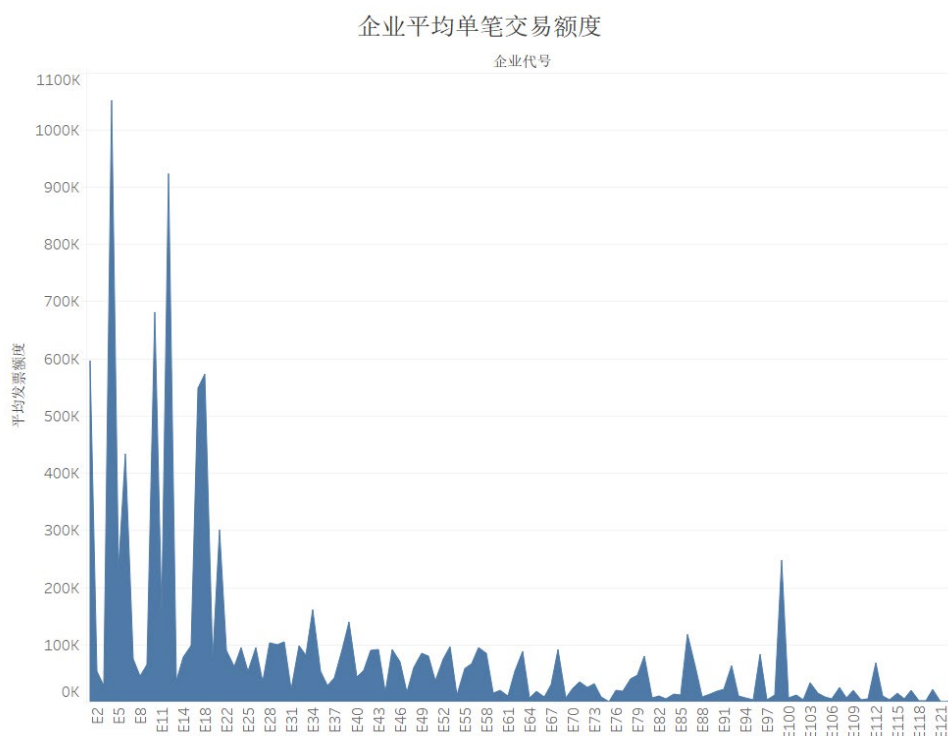


图 4 各企业平均单笔交易金额

从图 4 各企业的交易波动来看，代号为 E1~E25 的企业交易金额较高，说明其贸易能力，金额需求较大。而后半部分企业代号贸易金额较低，实力较弱。

④**信誉等级**：企业的信誉等级可以通过附件 1 直接得出。为了便于后期建模计算，可以直接进行数据量化为区间，信誉等级量化结果如表 1 所示：

表 1 信誉等级数据量化

信誉等级	A	B	C	D
量化区间	[75,100]	[50,75)	[25,50)	[0,25)

其中，企业实力和信誉等级直接决定着企业的信贷风险，也是决策信贷的因素。

⑤**是否违约**：企业是否有违约前科将会是之后是否放贷的决定性因素，企业是否违约进行数据量化，令曾经违约量化为 1，未违约量化为 0。

⑥**盈利能力**：企业盈利能力更强更能为银行带来更多的利益，符合最终的目标。银行的盈利能力=销项交易总额-进项交易总额，由此可以得出各企业的盈利能力高低，并可以通过归一化量化处理企业的盈利能力。

将附件 1 中 123 家企业对以上 6 个指标进行量化分析，得量化矩阵如下：

$$T = \begin{bmatrix} 0.131 & 0.099 & 1 & 87.5 & 0 & 0.07 \\ 0.076 & 0.159 & 0.75 & 87.5 & 0 & 0.86 \\ 0.112 & 0.043 & 0.5 & 37.5 & 0 & 0.84 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.081 & -0.579 & 0 & 12.5 & 1 & 0.34 \end{bmatrix}_{123 \times 6}$$

5.3 信贷方案决策

5.3.1 模型选择及建立

贝叶斯方法的目的是从一组竞争模型中选择一个模型，该模型最紧密地代表生成观察数据的基础过程。在贝叶斯模型比较中，选择给定数据后验概率最高的模型。

一、效益函数的确定

银行对中小微企业给出信贷策略时，目的在于要获得较高的经济收益，而信贷策略的制定往往取决于企业规模、企业供求稳定性、信誉等级、盈利能力、企业实力等诸多变量先前及未来状态。因此，信贷策略的制定本质上是一个风险性多指标决策问题。

为了较恰当的描述银行借贷企业的自然状态，我们引入“企业市场前景”，即对不同类企业市场未来发展变化的综合描述，反映了银行对企业发展未来情况的综合看法。因此，我们可将企业未来发展状态分为三种：企业市场前景较好、企业市场前景一般、企业市场前景较差^[1]。

对于此风险型多指标决策问题，记：

状态空间 $\theta = \{\theta_1, \theta_2, \dots, \theta_m\} = \{\theta_t\}$, ($t=1, 2, \dots, m$), θ_t 为状态变量。

决策空间 $Q = \{q_1, q_2, \dots, q_n\} = \{q_h\}$, ($h=1, 2, \dots, n$), q_h 为决策变量。

指标空间 $F = \{f_1, f_2, \dots, f_l\} = \{f_k\}$, ($k=1, 2, \dots, l$), f_k 为决策指标。

当状态变量 θ_t 离散的时候，设其发生概率为 $P(\theta_t)$ ，令 $\sum_{t=1}^m P(\theta_t) = 1$ 。

效益函数 U 是 l 个指标的函数， $U = u(f_1, f_2, \dots, f_l)$ ^[1]。

由于效益函数是多个指标的函数，相对复杂，其效益函数的评定需要简化。在建立本模型的过程中设定各指标函数为线性关系，效益函数可表示为线性加权形式：

$$U = \omega_1 f_1 + \omega_2 f_2 + \dots + \omega_5 f_5 + \omega_6 f_6 \quad (1)$$

其中， f_1, f_2, \dots, f_6 分别为：企业规模、供求关系稳定性、企业实力、信誉评级、是否违约、盈利能力， ω_i 是指标 f_i 的权重，($k=1, 2, \dots, 6$)。

二、指标权重 ω_k 的确定

采用最小平方和方法求权重，可以认为 a_{ij} 近似地等于指标 i 的权重与指标 j 的权重之比 ω_i / ω_j ，即 $a_{ij} \approx \omega_i / \omega_j$ ，将影响因素进行两两比较，构造判断矩阵如下形式：

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{15} \\ a_{21} & a_{22} & \cdots & a_{26} \\ \vdots & \vdots & & \vdots \\ a_{61} & a_{62} & \cdots & a_{66} \end{bmatrix} = \begin{bmatrix} 1 & 4 & 1/2 & 3 & 2 & 4 \\ 1/4 & 1 & 1/3 & 1/5 & 1/4 & 1/6 \\ 2 & 3 & 1 & 1/2 & 1/3 & 1/4 \\ 1/3 & 5 & 2 & 1 & 2 & 1/3 \\ 1/2 & 4 & 3 & 1/2 & 1 & 4 \\ 1/4 & 6 & 4 & 3 & 1/4 & 1 \end{bmatrix}$$

得出的矩阵实际上是有约束的最优化问题，可以通过构造 Lagrange 函数，要想将其转化为无约束的问题，可以利用下列式子：

$$\min L(\omega, \lambda) = \sum_{i=1}^6 \sum_{j=1}^6 (a_{ij} \omega_j - \omega_i)^2 + 2\lambda (\sum_{j=1}^6 \omega_j - 1) \quad (2)$$

进一步对 ω_k 求偏微分，得到：

$$\begin{cases} \sum_{i=1}^6 (a_{ik} \omega_k - \omega_i) a_{ik} - \sum_{j=1}^6 (a_{kj} \omega_j - \omega_k) + \lambda = 0, & (k=1, 2, \dots, 6) \\ \sum_{j=1}^6 \omega_j = 1 & \omega_j > 0, \quad (j=1, 2, \dots, 6) \end{cases} \quad (3)$$

上式可写为矩阵形式：

$$B\omega = M \quad (4)$$

采用权的最小平方和法，则构成矩阵 B

$$B = \begin{bmatrix} \sum_{i=1, i \neq 1}^6 a_{i1}^2 + 6 - 1 & -(a_{12} + a_{21}) & \cdots & -(a_{16} + a_{61}) \\ -(a_{21} + a_{12}) & \sum_{i=1, i \neq 2}^6 a_{i2}^2 + 6 - 1 & \cdots & -(a_{26} + a_{62}) \\ \vdots & \vdots & & \vdots \\ -(a_{61} + a_{16}) & -(a_{62} + a_{26}) & \cdots & \sum_{i=1, i \neq 6}^6 a_{i6}^2 + 6 - 1 \end{bmatrix} = \begin{bmatrix} 9.48 & -4.25 & -2.5 & -3.33 & -2.5 & -4.25 \\ -4.25 & 107 & -3.33 & -5.2 & -4.25 & -6.17 \\ -2.5 & -3.33 & 34.36 & -2.5 & -3.33 & -4.25 \\ -3.33 & -5.2 & -2.5 & 23.54 & -2.5 & -3.33 \\ -2.5 & -4.25 & -3.33 & -2.5 & 13.23 & -4.25 \\ -4.25 & -6.17 & -4.25 & -3.33 & -4.25 & 37.2 \end{bmatrix}$$

$$\omega^T = [0.0792 \quad 0.1317 \quad 0.1725 \quad 0.2517 \quad 0.1532 \quad 0.2117]$$

三、指标值的标准化

由于要处理不同数量级以及量纲的多种指标，各种类型指标的值代表的结果倾向也各不相同，所以必须应对指标进行标准化处理。

对于效益型指标

$$y_{ht} = \frac{x_{ht} - \min_{1 \leq h \leq n}(x_{ht})}{\max_{1 \leq h \leq n}(x_{ht}) - \min_{1 \leq h \leq n}(x_{ht})} \quad (5)$$

对于成本型指标

$$y_{ht} = \frac{\max_{1 \leq h \leq n}(x_{ht}) - x_{ht}}{\max_{1 \leq h \leq n}(x_{ht}) - \min_{1 \leq h \leq n}(x_{ht})} \quad (6)$$

对于固定性指标

$$y_{ht} = \frac{\max_{1 \leq h \leq n} |x_{ht} - x^*| - |x_{ht} - x^*|}{\max_{1 \leq h \leq n} |x_{ht} - x^*| - \min_{1 \leq h \leq n} |x_{ht} - x^*|} \quad (7)$$

四、期望效益决策

贝叶斯理论多风险决策模型流程图如图 5 所示：

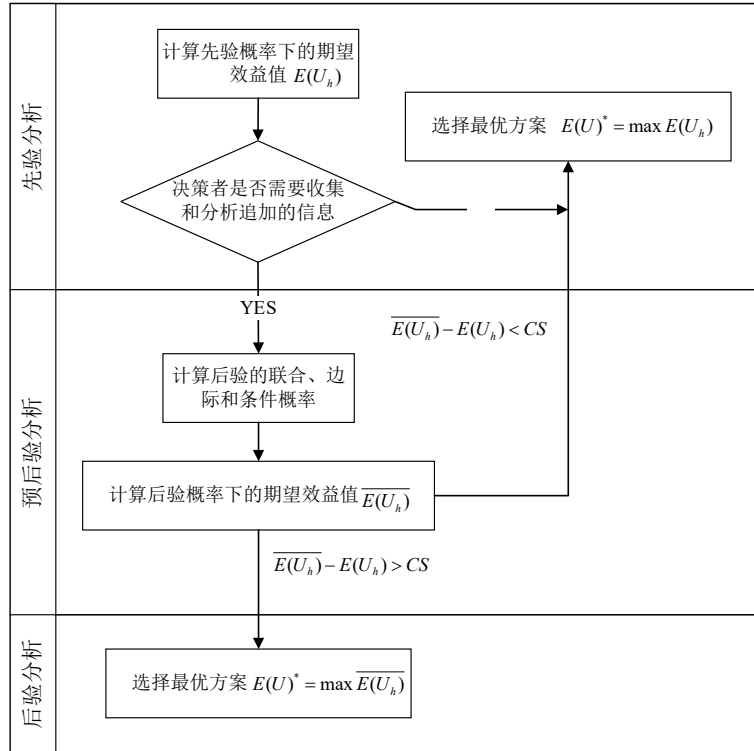


图 5 贝叶斯模型流程图

利用上述中的高风险多指标决策模型，可计算得出各方案的期望效益值：

$$E(U_h) = p(\theta_1)u_{h1} + p(\theta_2)u_{h2} + p(\theta_3)u_{h3} \quad u_{ht} (t=1,2,3) \quad (8)$$

u_{ht} 是指方案 q_h 在自然状态下 θ_i 下的效益值。

取期望效益值最大时的数据作为效益准则，便可衡量得到最优方案：

$$E(U)^* = \max_{h \in [1,n]} E(u_h) \quad (9)$$

前文选定的概率 $p(\theta_1), p(\theta_2), p(\theta_3)$ 规定为先验概率。在贝叶斯模型比较中，选择给定数据中后验概率最高的模型。为修正前段时间的先验概率分布，提高决策准确性，在决策分析中常常应用贝叶斯公式计算后验概率分布，以便得出最佳决策^[3]。

通过搜集近 10 年各银行对于不同类别企业的信贷策略，可以整理得出自然状态下

类别	θ_1	θ_2	θ_3
个体类	0.30	0.50	0.20
建筑工程类	0.25	0.6	0.15
生活类	0.28	0.55	0.17
服务类	0.30	0.45	0.25
农业园林类	0.13	0.75	0.12
医学药物类	0.15	0.70	0.15

银行对不同行业、不同类别企业的先决概率，如表 2 所示：

表 2 银行对于不同类别企业的先决概率

根据附表 3 中年利率与客户流失率的数据关系，分别拟合出信誉评级为 A、B、C 下的年利率与客户流失率线性函数关系。

信誉等级为 A 时：

$$f(x_a) = 0.1423x_a^2 - 0.02316x_a + 0.04538 \quad (10)$$

信誉等级为 B 时：

$$f(x_b) = 0.141x_b^2 - 0.01025x_b + 0.04433 \quad (11)$$

信誉等级为 C 时：

$$f(x_c) = 0.1308x_c^2 - 0.0004416x_c + 0.04389 \quad (12)$$

得到线性函数拟合图如图 6：

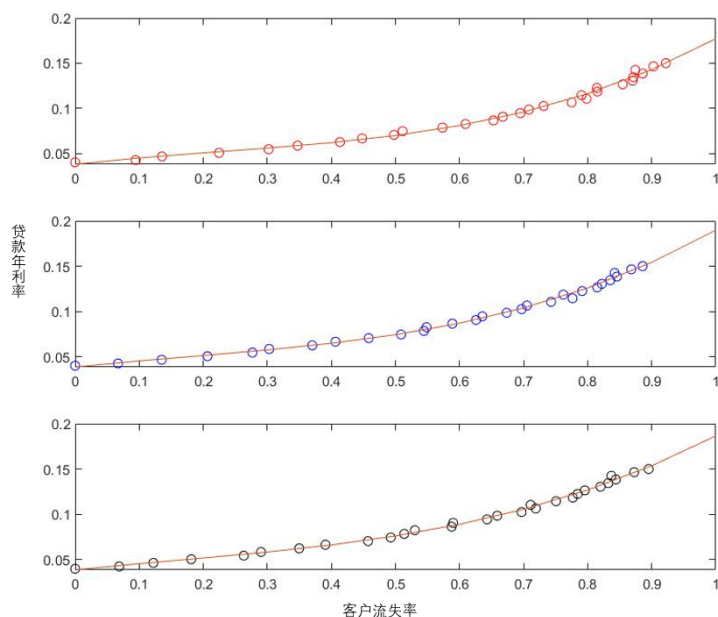


图 6 不同信誉等级下贷款年利率与客户流失率关系

其中曲线一、二和三是信誉等级为 A、B 和 C 类企业的相关关系。一方面，随着贷款年利率的提高，客户流失率也明显提高。另一方面，通过观察点状分布密度可以得出 A 级企业所能承受的贷款额度越高，客户越稳定，同贷款利率下该类企业流失率更低。信誉等级为 B，C 级的企业逐层减弱。

根据不同信誉评级下年利息和客户流失率的线性函数关系，结合银行对中小微企业所采用贷款方案的相关文献《2020 年疫情下中国新经济产业投资研究报告(艾瑞数据)》，综合得出银行信贷决策体系如表 3，其中考虑到高息高额贷款针对中小微型企业很少出现，所以暂不讨论，讨论日常中出现的低额高息，中额高息等实际问题^[2]。

表 3 银行信贷决策体系

决策标度	信贷额度	年利息	期限
q_1	[60,100]	0.0425	1 年
q_2	[10,60]	0.0505	
q_3	[20,70]	0.0665	
q_4	[10,35]	0.0825	
q_5	[45,75]	0.1305	
q_6	不予贷款		

①先验决策：根据上述分析得具体期望函数：

$$U = 0.0792f_1 + 0.1317f_2 + 0.1725f_3 + 0.2517f_4 + 0.1532f_5 + 0.2117f_6 \quad (13)$$

以企业 E1 为例，作出如下风险决策表 4：

表 4 先决决策所得期望值

方案	θ_1	θ_2	θ_3	期望值 $E(U)$
	$p(\theta_1) = 0.25$	$p(\theta_2) = 0.60$	$p(\theta_3) = 0.15$	
q_1	0.691	0.552	0.394	0.563
q_2	0.320	0.632	0.578	0.546
q_3	0.601	0.498	0.553	0.532
q_4	0.4578	0.584	0.361	0.519
q_5	0.4388	0.428	0.650	0.464
q_6	0.290	0.337	0.482	0.347

以期望效益值最大为决策准则,可确定最优方案 $E(U)^* = \max_{h \in [1,6]} E(u_h)$, 即选择方案 q_1 。

引入风险度

$$\rho = \frac{\sigma(U_h)}{E(U_h)} \quad (14)$$

其中,上式分子代表策略 q_h 的效益值方差。风险度 ρ 值的大小,表明了不同方案效益值离散程度,当方案的风险度越小,就表明着方案效益值的离散程度越小,说明未来收益的不确定性小,也就更安全。对于方案 q_2 , 其风险度:

$$\rho = \frac{\sqrt{0.25 \times (0.450 - 0.563)^2 + 0.60 \times (0.671 - 0.563)^2 + 0.25 \times (0.319 - 0.563)^2}}{0.563} = 0.281$$

②**预后决策分析:** 在先验分析时, E1 企业选择方案 q_1 时风险度达到 28.1%。此时往往暂缓作出决策,需进一步获取更详细信息来进行决策。根据过去经验的分析结果并不一定完全准确,但能估计出准确度。表 5 表示银行对不同类型的企业获得与真实自然状态相应的调查结果的条件概率^[6]。

表 5 银行对不同类型的企业的真实自然状态条件概率

调查结果 S	自然状态		
	前景较好 θ_1	前景一般 θ_2	前景较差 θ_3
前景较好 S_1	$P(S_1 \theta_1)$	$P(S_1 \theta_2)$	$P(S_1 \theta_3)$
前景一般 S_2	$P(S_2 \theta_1)$	$P(S_2 \theta_2)$	$P(S_2 \theta_3)$
前景较差 S_3	$P(S_3 \theta_1)$	$P(S_3 \theta_2)$	$P(S_3 \theta_3)$

如 $P(S_1 | \theta_1), P(S_2 | \theta_1)$ 表示实际结果较好,而调查结果一般的概率。银行根据此方案对不同类企业的先验概率进行修正,贝叶斯公式如下:

$$P(\theta_i | S_j) = \frac{P(\theta_i)P(S_j | \theta_i)}{\sum_{k=1}^3 P(\theta_k)P(S_j | \theta_k)} \quad (i=1,2,3) \quad (15)$$

③**后验决策分析:** 首先可以依据预后决策进行分析,然后针对先验概率进行修正,由此可以得出后验概率,此处以企业 E1 为例,其结果见表 6:

表 6 企业 E1 的后验概率

方案	θ_1	θ_2	θ_3	期望值
	$P(\theta_1 S_2) = 0.179$	$P(\theta_2 S_2) = 0.732$	$P(\theta_3 S_2) = 0.089$	$E(U)$
q_1	0.691	0.552	0.394	0.563
q_2	0.320	0.632	0.578	0.546
q_3	0.601	0.498	0.553	0.532
q_4	0.4578	0.584	0.361	0.519
q_5	0.4388	0.428	0.650	0.464
q_6	0.290	0.337	0.482	0.347

以期望值最大为决策方案，应选择方案 q_1 。此时该方案风险度 $\rho = 0.152$ ，利用贝叶斯风险决策分析方法很大程度上降低了银行信贷策略的风险程度。

通过贝叶斯理论对先验决策修正后。最终，我们得到 123 个企业的信贷策略如下：

表 7 所有企业的信贷策略

企业 代号	后验决策期望值						信贷 策略
	q_1	q_2	q_3	q_4	q_5	q_6	
E1	0.563	0.546	0.532	0.519	0.464	0.347	q_1
E2	0.573	0.526	0.520	0.327	0.461	0.503	q_2
E3	0.546	0.531	0.571	0.424	0.349	0.281	q_3
E4	0.521	0.526	0.583	0.506	0.339	0.315	q_3
E5	0.547	0.553	0.53	0.443	0.352	0.361	q_2
...
E123	0.274	0.201	0.442	0.439	0.37	0.528	q_6

六、问题二：预测并决策无信贷记录企业

在问题一的基础上，类比相同的方式量化分析附件 2 中各企业的相关数据，并需得出银行在信贷总额为 1 亿时针对企业的信贷策略，同样使银行获取最大利益。利用填补的思路，将特定影响因素下蒙特卡洛模拟得出的数据和问题一中真实的信贷记录数据作为 RBF 神经网络预测模型的训练集，预测得出 302 家企业的信誉评级和是否违约情况，最终通过贝叶斯理论对先验决策修正，并做出信贷决策。

6.1 RBF(径向基)神经网络预测模型

6.1.1 模型选择及建立

对于附表 2 中 302 家企业数据集，在问题一多风险决策模型得出的企业规模、供求关系稳定性、盈利能力、企业实力主要影响因素下根据蒙特卡罗动态模拟生成 600 家企业在特定条件下的预测数据，结合附表 1 中的 123 家企业的实际数据。以该 723 家企业信誉评级和是否违约结果作为训练数据。所构建函数的自变量可以选定径向基神经网络的隐节点之间采用输入模式与中心向量的距离，并规定激活函数为快速使用径向基函数。由于数据量较为较大且 BP 神经网络的学习速率固定，因此 BP 算法所需要的时间较长而且很有可能陷入局部化最优解现象。因此，选择径向基优化的 BP 神经网络模型可以在更短时间内实现权值的快速收敛且精度更高^[4]。RBF 神经网络拓扑图如下：

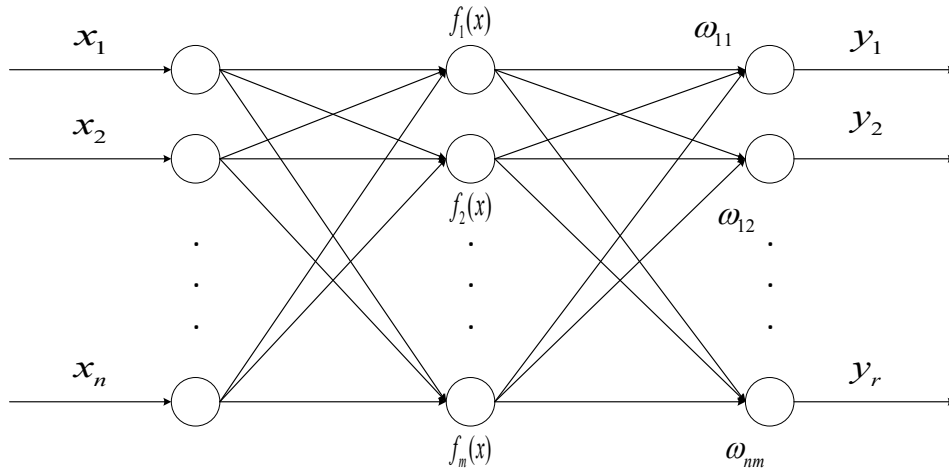


图 7 RBF 神经网络拓扑图

图 7 中， x_1, x_2, \dots, x_n 为影响企业信誉评级和是否违约的各影响因素， y_1, y_2, \dots, y_r 为企业信誉评级和是否违约的预测结果， $f_m(x)$ 为隐含层连接函数，各层神经元个数为 n_1 、 n_2 、 n_3 。

径向基神经网络激活函数

$$R(x_p - c_i) = \exp\left(-\frac{1}{2\sigma^2} \|x_p - c_i\|^2\right) \quad (16)$$

最小二乘法损失函数

$$\sigma = \frac{1}{P} \sum_j^m \|d_j - y_j c_j\|^2 \quad (17)$$

y_j 输入 RBF 神经网络得到的网络输出

$$y_j = \sum_{i=1}^h \omega_{ij} \exp\left(-\frac{1}{2/\sigma^2} \|x_n - c_i\|^2\right) \quad j=1, 2, \dots, n \quad (18)$$

其中, x 为 n 维输入向量, c_i 为第 i 个径向基函数的中心值同时与输入向量维数保持一致, σ 是径向基函数第 i 个中心点的标准化常数, $\|x_n - c_i\|$ 是向量 $x_n - c_i$ 的范数。

RBF 神经网络预测模型训练步骤如下:

- ①无监督自学习, 求解隐含层各基函数的中心及方差。
- ②监督学习, 求解隐含层传输与输出层的各权值变化规律^[5]。

6.1.2 模型求解

以蒙特卡洛模拟生成的 600 个数据集结合附件 1 中的 123 家企业的信誉记录作为初始训练集, 通过初始权值进行自学习及增量调整, 通过不断迭代得出的新的权值, 直至使输出层误差满足模型的精度要求。通过 Matlab 求解得出附件 2 中 302 家企业的信誉评级和是否违约记录。

6.1.3 结论阐述

由 Matlab 迭代训练得出 302 家企业的信誉评级及是否违约记录如表 8 所示:

表 8 企业信誉评级和违约情况预测

企业代号	信誉评级	是否违约
E124	37.1	0
E125	62.5	0
E126	47.1	0
E127	53.3	0
...
E425	26.9	0

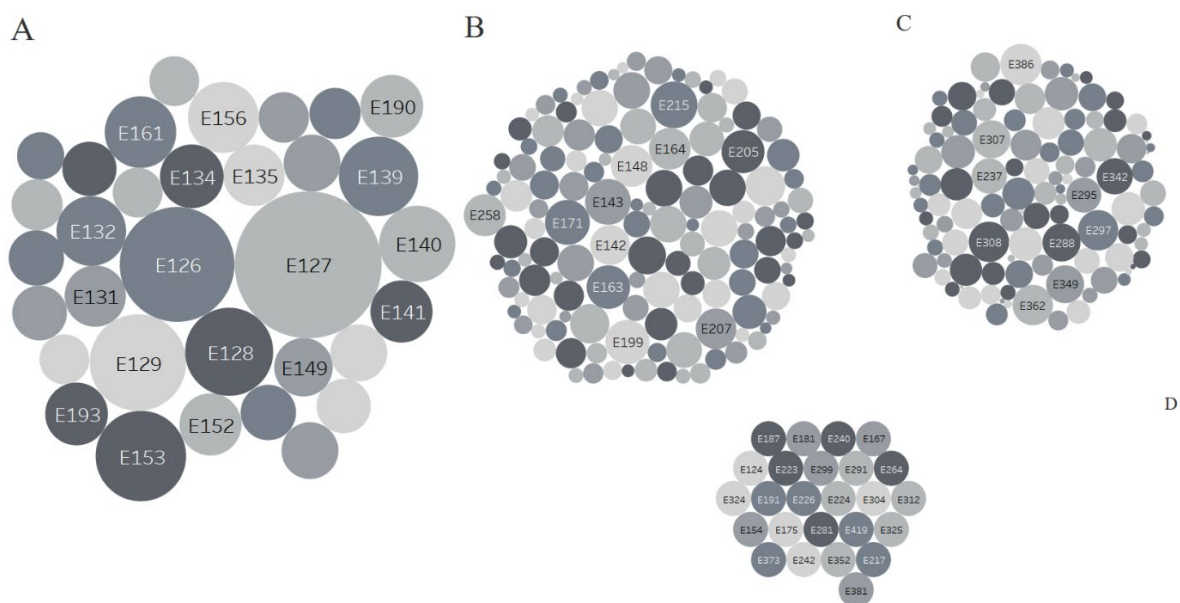


图 8 企业所属信誉评级占比分布

首先从表中可以得出各企业的信誉评级情况，利用工具进行数据可视化。在尚已分好的决策评级中，若该企业更倾向于该类评级，则形状越大。以信誉评级 A 企业为例，E127，E126 等企业形状较大，约倾向于 A 类，越值得银行信任。

6.1.4 结论检验

将预测结果与 723 个训练数据进行对比，得出各影响因素对结果的影响较为吻合。其中，对于信誉评级的 723 个训练数据有误差收敛和系数变化图 9、10：

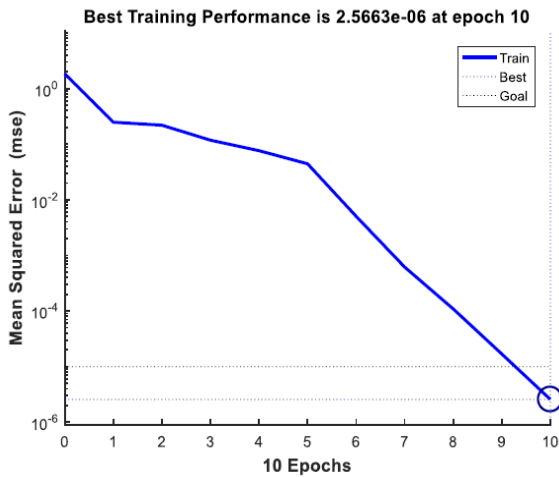


图 9 误差收敛图

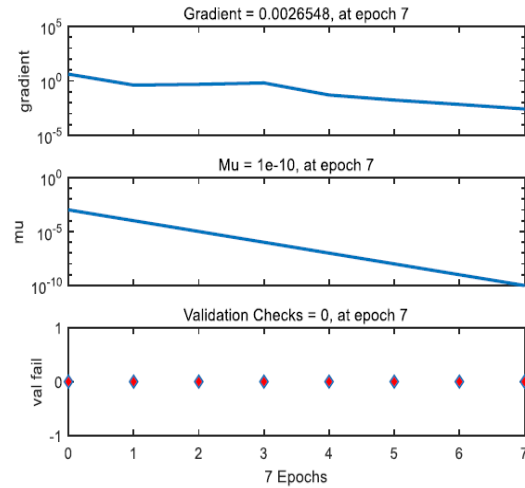


图 10 系数变化图

6.2 风险性多属性决策模型

基于 RBF 神经网络预测得出 302 家企业的信誉评级和是否违约记录后，将附件 2 中的指标进行量化分析后，得各企业的量化矩阵：

$$T' = \begin{bmatrix} 0.092 & 0.205 & 0.5 & 37.1 & 0 & 0.09 \\ 0.091 & 0.099 & 0.5 & 62.5 & 0 & 0.54 \\ 0.076 & 0.171 & 0.75 & 47.1 & 0 & 0.67 \\ 0.062 & 0.089 & 0.25 & 53.3 & 0 & 0.73 \\ 0.063 & 0.106 & 0.25 & 81.6 & 0 & 0.70 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.057 & -0.317 & 0 & 13.7 & 0 & 0.24 \end{bmatrix}_{302 \times 6}$$

将矩阵 T' 入风险型决策模型后，结合问题一中银行对不同行业的信贷先验概率，通过贝叶斯理论对先验决策修正后，得到各企业的策略选择方案^[3]。对同一决策方案下的企业根据实力大小和规模进一步划分贷款额度，此处以所有决策方案为 q_2 的企业为例进行分析，得出可以体现各企业实力、规模的条状图如下：

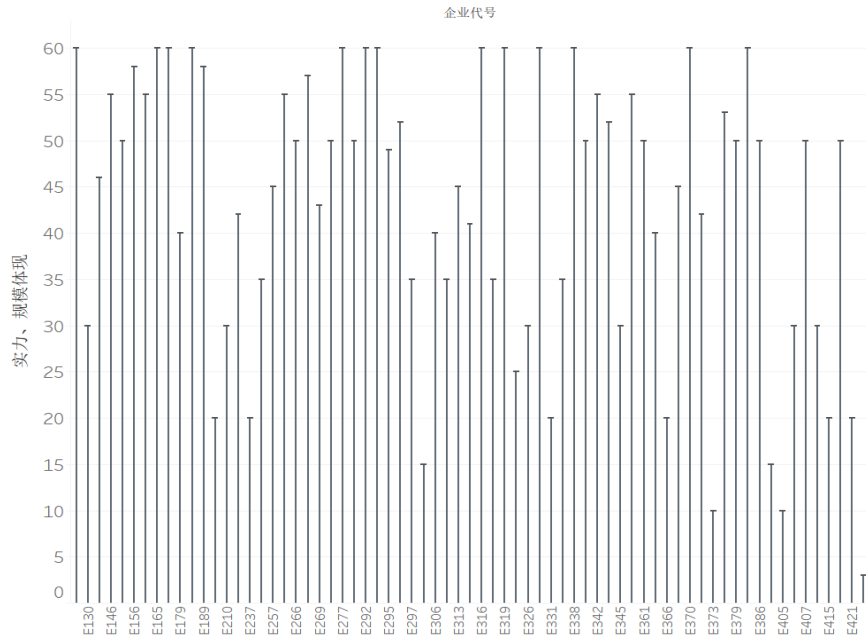


图 11 各企业实力、规模图

最终,我们得到银行在年度信贷总额为1亿时对302家企业的具体信贷策略如表9:

表 9 各企业信贷策略

企业代号	决策方案	放贷额度	年利率	期限
E124	q3	25	0.0665	1 年
E125	q1	80	0.0425	
E126	q1	75	0.0425	
E127	q2	60	0.0505	
E128	q1	72	0.0425	
E129	q1	87	0.0425	
...	
E425	q2	2	0.0505	

七、问题三：突发因素下分析并决策信贷策略

7.1.1 模型分析

在问题一、二中建立贝叶斯多指标风险决策模型时,最终是通过各个方案的期望效益值来完成具体决策的。而对于期望效益值

$$E(U_h) = p(\theta_1)u_{h1} + p(\theta_2)u_{h2} + p(\theta_3)u_{h3} \quad (19)$$

其中, $p(\theta_i)$ 是基于自然状态下的先验概率, $u_{ht}(t \approx 1, 2, 3)$ 表示在自然状态 θ_i 下方案 U_h 的效益值。而企业的生产经营及经济效益可能会受到一些突发因素影响,并且突发因

素下的不同行业 and 不同类别的企业会受到不同程度的冲击。主要参考通过查阅新冠肺炎冲击下各类行业受影响的情况及程度，可以得出三维交叉数表并通过 Tableau 进行数据可视化，图即为疫情影响下各类别企业状况：

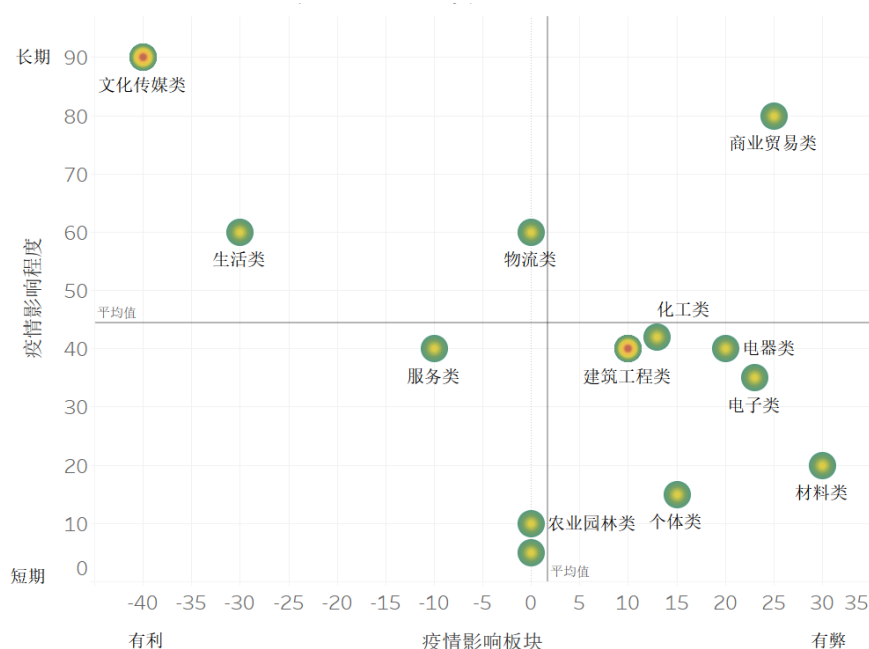


图 12 新冠疫情影响下各类别企业状况

从图 11 中散点分布可观察到，生活类、文化传媒类等搭载互联网媒介的产业受有利影响，并且受影响程度较大。而化工类、建筑工程类等传统产业则深受弊式影响，但是影响时间适中，在一段时间后可正常开工。其余行业可按同理分析。

结合可视化数据集，通过对原自然状态下的先验概率 $p(\theta_i)$ 的评估修正，得出了非自然状态下银行对不同企业的部分先验概率如下表 10：

表 10 非自然状态下银行对不同企业的先验概率

类别	θ_1	θ_2	θ_3
个体类	0.15	0.35	0.50
建筑工程类	0.20	0.5	0.30
生活类	0.08	0.5	0.42
文化传媒类	0.4	0.5	0.1
科技类	0.40	0.50	0.10
电子类	0.37	0.49	0.14
电器类	0.27	0.60	0.13
医学药物类	0.40	0.5	0.1

7.1.2 模型建立

①根据评估修正后的先验概率进行决策

基于决策体系下的效益函数

$$U = 0.0792f_1 + 0.1317f_2 + 0.1725f_3 + 0.2517f_4 + 0.1532f_5 + 0.2117f_6$$

以企业 E204 为例，得出风险决策表 11 如下：

表 11 突发因素下个方案风险决策表

方案	θ_1	θ_2	θ_3	期望值 $E(U)$
	$p(\theta_1) = 0.40$	$p(\theta_2) = 0.50$	$p(\theta_3) = 0.10$	
q_1	0.597	0.436	0.324	0.489
q_2	0.386	0.426	0.469	0.414
q_3	0.441	0.592	0.553	0.528
q_4	0.361	0.434	0.526	0.494
q_5	0.318	0.436	0.413	0.387
q_6	0.384	0.514	0.633	0.474

确定最优方案 $E(U)^* = \max_{h \in [1,3]} E(U_h)$ ，即选择方案 q_3 。此时风险度 $\rho = 0.296$ 。

②后验决策

通过贝叶斯理论对先验概率进行修正后，得到后验决策如下表 12：

表 12 修正后得到的后验决策

方案	θ_1	θ_2	θ_3	期望值 $E(U)$
	$P(\theta_1 S_2) = 0.179$	$P(\theta_2 S_2) = 0.732$	$P(\theta_3 S_2) = 0.089$	
q_1	0.597	0.436	0.324	0.455
q_2	0.386	0.426	0.469	0.423
q_3	0.441	0.592	0.553	0.562
q_4	0.361	0.534	0.326	0.512
q_5	0.318	0.436	0.413	0.431
q_6	0.384	0.514	0.633	0.501

以期望值最大为决策方案，选择 q_3 方案。此时风险度 $\rho = 0.137$ ，较先验决策时的风险度有明显的显著降低。

7.1.3 结果分析

通过分析各企业在新冠肺炎突发因素下的受影响程度^[7]，对模型原自然状态下的先验概率 $p(\theta_i)$ 的评估修正，对贷款金额的具体化体现方式同问题二。最终，求解得出银行在年度信贷总额为 1 亿元时的信贷调整策略如下表 13：

表 13 突发因素发生时针对企业的信贷策略

企业代号	决策	放贷额度	年利率	期限
E124	q4	10	0.0825	1 年
E125	q3	45	0.0665	
E126	q1	70	0.0425	
E127	q2	60	0.0505	
E128	q2	40	0.0505	
E129	q1	65	0.0425	
...	
E425	q2	2	0.0505	

八、模型的优缺点分析

8.1 模型的优点

1. 本文建立的风险决策模型中，对于效应函数的评定采用了线性加权的形式，突出了不同指标下权重的作用，更好地体现了多指标影响下银行对中小微企业的信贷策略。
2. RBF 神经网络预测模型相对于 BP 神经网络预测有更快的收敛速度和更好的精度。
3. 基于蒙特卡洛进行相关数据模拟，得出在特定指标影响下具体的数据集合供神经网络训练使用，弥补了数据过少的问题，使预测模型的构建更具有说服力。
4. 在贝叶斯风险决策模型中引入了风险度的相关概念，通过先验决策和后验决策的风险度比对，更直观地体现了模型的应用价值。

8.2 模型的缺点

1. 贝叶斯决策对先验精度的精度要求较高，具有一定的主观性。
2. 由于 RBF 神经网络训练数据量较小，通过程序模拟的数据会和真实数据的精度有所偏差，可能会导致预测结果不准确。

九、模型的推广和应用

1. 本文建立的决策模型可以广泛应用于分类问题,通过约定的决策结果实现对大批量数据的分类。

2. 基于贝叶斯决策理论的风险性多属性决策模型可以推广到公司、项目投资决策领域,对其加以改进会具有很强的现实意义。

3. 将贝叶斯推断和以模拟为基础的蒙特卡洛方法为基础合并使用,因为一些复杂的模型无法用贝叶斯分析得到解析解,而因图模式结构可以配合一些快速的模拟方法。

4. 马尔科夫·蒙特卡洛方法让贝叶斯方法的研究及应用有大幅的发展,除去了许多运算上的问题,也有利于解决非标准的复杂性问题。

十、参考文献

- [1]苗琦. 基于贝叶斯决策理论的房地产投资风险决策研究[D].武汉理工大学,2003.
- [2]庞利. 贝叶斯法则在银行信贷风险中的分析[J].赤峰学院学报(自然科学版),2015,31(07):140-141.
- [3]傅德汉,朱波,操基平.基于物联网技术的中小企业贷款方案初探[J].金融纵横,2012(02):30-33.
- [4]聂华伟,邓捷,廖晓梅.基于 RBF 神经网络的道路交通安全预测模型研究[J].科技创新,2020(27):132-133.
- [5]周俊武,孙传尧,王福利.径向基函数(RBF)网络的研究及实现[J].矿冶,2001(04):71-75.
- [6]孙彤,汪波.基于风险调整收益方法的企业信用风险管理研究[J].商业经济与管理,2009(01):45-50.
- [7]秦小虎,刘利,张颖.一种基于贝叶斯网络模型的交通事故预测方法[J].计算机仿真,2005(11):230-232.

十一、附录

一、支撑材料

- 1. 附件 1 进项发票频次.xlsx
- 2. 附件 1 平均进项发票额度.xlsx
- 3. 附件 1 平均销项发票额度.xlsx
- 4. 附件 1 销项发票频次.xlsx
- 5. 附件 1 销项进项差. xlsx
- 6. 附件 2 进项发票频次.xlsx

7. 附件 2 决策 2 类型公司放贷金额.xlsx
8. 附件 2 平均进项发票额度.xlsx
9. 附件 2 平均销项发票额度.xlsx
10. 附件 2 销项发票频次.xlsx
11. 附件 2 销项进项差.xlsx
12. 问题一企业决策信贷策略.xlsx
13. 供求关系.cpp
14. 交易规模.cpp
15. Rbf.m

二、源代码

1. 供求关系

```
#include<iostream>
#include<string.h>
#include<cstdio>
using namespace std;
string name[500];
int
ans20171[500],ans20181[500],ans20191[500],ans20201[500],sum[500],ans20172[500],ans20182[500],ans
20192[500],ans20202[500],ans20173[500],ans20183[500],ans20193[500],ans20203[500],ans20174[500],a
ns20184[500],ans20194[500],ans20204[500];
int main()
{
    int jishu = 0;
    string  n;
    string shangci = "wu";
    while(cin>>n)
    {
        if(shangci != n)
        {
            shangci = n;
            name[jishu] = n;
            sum[jishu] = 0;
            ans20171[jishu] = 0;
            ans20181[jishu] = 0;
            ans20191[jishu] = 0;
            ans20201[jishu] = 0;
            ans20172[jishu] = 0;
            ans20182[jishu] = 0;
            ans20192[jishu] = 0;
            ans20202[jishu] = 0;
            ans20173[jishu] = 0;
            ans20183[jishu] = 0;
            ans20193[jishu] = 0;
```

```

        ans20203[jishu] = 0;
        ans20174[jishu] = 0;
        ans20184[jishu] = 0;
        ans20194[jishu] = 0;
        ans20204[jishu] = 0;
        jishu++;
    }
    int riqi,x,y;
    scanf("%d/%d/%d",&riqi,&x,&y);
    if(riqi ==2017)
    {
        if(x<=3)
            ans20171[jishu-1]++;
        else if(x<=6)
            ans20172[jishu-1]++;
        else if(x<=9)
            ans20173[jishu-1]++;
        else
            ans20174[jishu-1]++;
        sum[jishu-1]++;
    }else if(riqi ==2018)
    {
        if(x<=3)
            ans20181[jishu-1]++;
        else if(x<=6)
            ans20182[jishu-1]++;
        else if(x<=9)
            ans20183[jishu-1]++;
        else
            ans20184[jishu-1]++;
        sum[jishu-1]++;
    }
    else if(riqi ==2019)
    {
        if(x<=3)
            ans20191[jishu-1]++;
        else if(x<=6)
            ans20192[jishu-1]++;
        else if(x<=9)
            ans20193[jishu-1]++;
        else
            ans20194[jishu-1]++;
        sum[jishu-1]++;
    }
}

```



```

        else if(riqi ==2020)
        {
            if(x<=3)
                ans20201[jishu-1]++;
            else if(x<=6)
                ans20202[jishu-1]++;
            else if(x<=9)
                ans20203[jishu-1]++;
            else
                ans20204[jishu-1]++;
            sum[jishu-1]++;
        }
    }
    for(int i = 0 ; i < jishu;i++)
    {
        cout<<name[i]<<" "<<sum[i]<<" "<<ans20171[i]<<" "<<ans20172[i]<<" "<<ans20173[i]<<"
"<<ans20174[i]<<" "<<ans20181[i]<<" "<<ans20182[i]<<" "<<ans20183[i]<<" "<<ans20184[i]<<" "
<<ans20191[i]<<" "<<ans20192[i]<<" "<<ans20193[i]<<" "<<ans20194[i]<<" "<<ans20201[i]<<endl;
    }
    return 0;
}

```

2. 交易规模

```

#include<iostream>
#include<string.h>
#include<cstdio>
using namespace std;
string name[500];
long long sum[500],sum2[500];
int main()
{
    int jishu = 0;
    string n;
    string shangci = "wu";
    double x;
    while(cin>>n)
    {
        if(shangci != n)
        {
            shangci = n;
            name[jishu] = n;
            sum[jishu] = 0;
            sum2[jishu] = 0;
            jishu++;
        }
    }
}

```

```
        scanf("%lf",&x);
        sum[jishu-1]+=x;
        sum2[jishu-1]++;
    }
    for(int i = 0 ; i < jishu;i++)
    {
        cout<<name[i]<<" "<<sum[i]/sum2[i]<<endl;
    }
    return 0;
}
```