**中小微企业的信贷决策**

1. 问题背景

某银行对确定要放贷企业的贷款额度为万元；年利率为4%15%；贷款期限为1年。附件13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。

1. 有信贷记录企业的信贷决策模型

2.1 问题分析

2.1.1数据数据预处理

先对对123家有信贷记录企业的相关数据进行数据预处理，包括企业信誉评级、违约情况、进项发票信息及销项发票信息进行多维度分析，发现并分析无关的数据，对数据进行数据清洗和变换处理。

2.1.2决策指标的确定

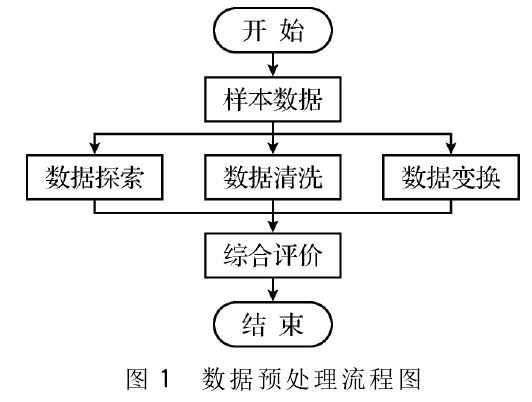
从企业的经营实力及信用水平等方面确定可以刻画出企业信贷风险的指标，得到预处理后的数据。

2.1.3基于熵权法的信贷决策模型

建立熵权法模型得到每个企业的综合得分，并将其进行归一化处理。在银行年度信贷总额固定时，根据企业归一化处理后的综合得分给予相应的贷款额度和各个企业相应的贷款年利率，得到银行的信贷策略。

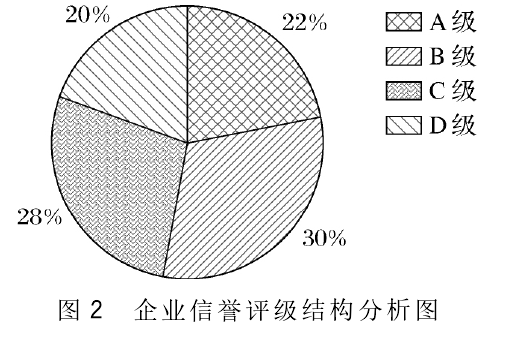
2.2问题求解

2.2.1 数据预处理

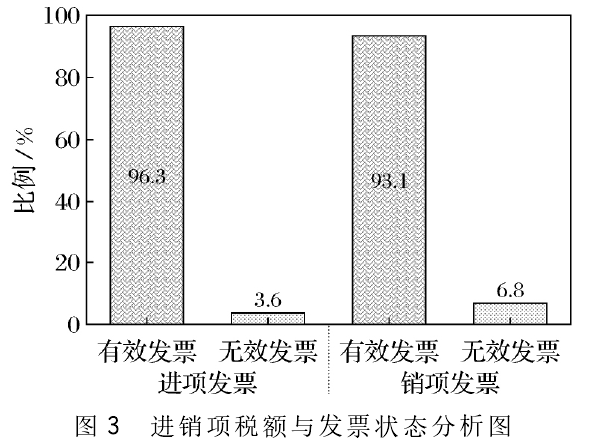


（一）数据探索

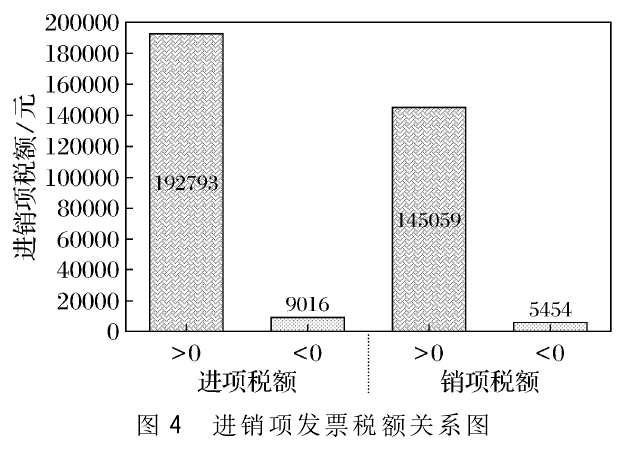
第一，企业的信誉评级。中小微企业信誉评级结构如图２，对信誉评级为D的企业，银行原则上是不予放贷的。



第二，进项发票与销项发票的发票状态。由图３可知，销项发票的作废发票比率比进项发票的高。作废发票比率越高，反映企业的失误率越高，也提高了企业做假账的嫌疑，影响企业的信誉。为了降低银行信贷风险，应该把作废发票比率高的中小微企业筛除。



第三，进项发票与销项发票的税额关系。由图４可知，税额小于０的进项税额与销项税额发票占总发票数的比例较少，显然该类发票会影响企业的信誉。



（二）数据清洗

在对数据进行探索分析的过程中，去掉一些无关数据。

第一，错误数据。利用统计软件对123家企业信息数据进行分析，由发票价税合计＝发票金额＋发票税额，剔除进销项发票信息中存在的错误数据。

第二，信誉评级为Ｄ的企业数据。对信誉评级为Ｄ的企业原则上银行是不予放贷的。因此，在后续确定放贷企业时，不考虑此类企业。

第三，发票状态为作废发票的数据。作废发票是在交易活动开具发票后，因故取消了该交易。考虑到某些企业会出现故意刷票的行为，此操作在一定程度上会影响企业的信誉程度，因此将此类数据剔除。

第四，发票税额为０的发票数据。发票税额为０的发票数据属于作废发票，交易活动开具发票后因故取消了该项交易，使发票作废。

（三）数据变换

对数据进行规范化处理，将数据转换成“适当的”形式。在这里，主要是对数据信息进行离散化。

第一，信誉评级。根据给出的中小微企业的信誉评级对 Ａ、Ｂ、Ｃ 三个等级进行赋值。将信誉等级为 Ａ 的企业赋值为４，信誉等级为 Ｂ 的企业赋值为３，信誉等级为 Ｃ的企业赋值为２。

第二，以月份为单位合并企业的流量数据。目的是为了计算销项、进项的波动及稳定情况，在下文指标中有所体现。

第三，绝对指标的对数化处理。为了避免过度异常波动的影响，将进销项发票价税合计及利润等与经济计量有关的变量取对数。经过自然对数处理后，异常点会变得比较小，会缩回期望范围内；当模型中有“价格”、“收入”之类的变量时，用自然对数的形式将变量数据进行处理。

第四，进销项作废发票率和负数发票率。在为交易活动开具发票后，企业因故取消了该项交易，使发票作废；而购方因故取消了交易，由于企业已入账记税，此时需开具负数发票。考虑到某些企业会出现故意刷票的行为，此操作在一定程度上会影响企业的信誉程度，分别定义作废发票率和负数发票率，可以间接反映出企业的信誉程度大小。

**2.2.2 指标的确定**

从企业的经营实力及信用水平两个方面，结合123家企业的相关数据，构建关于企业信贷风险的综合评价指标。

表1：企业信贷价值评估体系的指标体系

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 一级  指标 | 二级  指标 | 三级  指标 | 四级指标及符号 | 计算方法 |
| 信贷价值  评估体系 | 经营实力 | 资金  流量 | 进项总金额（X1） | 样本期内进项税价合计 |
| 销项总金额（X2） | 样本期内销项税价合计 |
| 上下游企  业实力 | 进项复合额（X3） | 所有进项企业的进项额与  其市场份额的加权平均值 |
| 销项复合额（X4） | 所有销项企业的进项额与  其市场份额的加权平均值 |
| 供求稳  定程度 | 进项波动率（X5） | 月进项额的标准差系数 |
| 销项波动率（X6） | 月销项额的标准差系数 |
| 盈利  能力 | 年总利润（X7） | 年销项额-年进项额 |
| 利润率（X8） | （年销项额-年进项额）/  年销项额 |
| 信用水平 | 信誉等级（X9） | | 附件1数据确定 |
| 是否违约（X10） | | 附件1数据确定 |
| 作废及负数发票比例（X11） | | 进项、销项作废及负数  发票数/总发票数 |

**2.2.3 基于熵权法的企业信贷额度模型**

“熵”最早作为热力学的概念提出，后来被引入信息理论中来，表明系统中的混淆程度。对于某项指标，可用熵值来判断某个指标的离散程度。熵值越小，指标的离散程度越大，该指标对综合评价的影响(即权重)越大。

熵值法的计算方法与步骤：

（一）构建原始指标数据矩阵。设有n个待评对象，m项评价指标，构成原始指标数据矩阵，表示第i个待评对象的第j项评价指标值，

（二）对数据标准化处理。在本文选取的11个指标，消除不同指标的量纲和数量级的不同会造成数据之间的差异影响，在利用熵值法进行赋权前需对所有的指标数据进行无量纲化处理，以消除量纲的影响，本文选择极值处理法对指标数据进行无量纲化处理。则归一化后为：

正向指标 

逆向指标 

其中，，，，分别为第i个待评对象的第j项评价指标原始值，第i个待评对象的第j项评价指标标准化值，第j个指标的最小值，第j个指标的最大值。

（三）确定权重，先计算第j项指标的信息熵



式中，为第j项指标的信息熵，，，。为第j项指标下第i个对象指标值发生的概率。



（四）然后计算第j项指标的权重。



其中，为第j项指标的差异系数



（五）最后根据各项指标权重与待评对象值计算各评价对象的得分。为使结果方便观看，将得分乘以100，换成百分制。



其中，，，分别表示第i个待评对象的综合得分，第j项指标权重，第i个第j项指标标准化值。的取值范围为0-100，值越大说明该对象的信贷价值越高。在信贷总额M一定的条件下，企业i能获得贷款的额度为：



**2.2.4信贷利率模型**

本文以利润最大化为原则进一步讨论银行针对中小微企业的放贷利率策略。根据信誉评级体系，构建银行对信誉评级为j(j=A,B,C)的企业i放贷的期望收益模型：



其中，为银行对信誉评级为j的企业i的贷款额度，为银行对信誉评级为j的企业i贷款利率，为企业i针对贷款利率为的流失率。

满足的约束条件：

 （或者 ）

满足的约束条件：



其中，分别为银行对信誉评级为j类的企业的最大与最小贷款额度，分别为银行对信誉评级为j类的企业的最高与最低贷款利率。即假定银行的贷款策略是：越是等级得分高的企业，其贷款额度就越高，相应的贷款利率就低；反之就越高。并假定企业的贷款利率随着贷款额度的降低而成比例升高。于是，银行的利率决策问题转化为，如何确定最高利率与最低利率，即为：



s.t  （或者 ）





**2.2.5基于聚类分析信贷期限模型**

根据等级得分，基于前文的企业运用实力及信用水平为分类变量，运用聚类分析分成三类，综合得分最高一类期限为三年，居中的一类为两年，最低的一类为一年。

1. 无信贷记录企业的信贷决策模型

3.1 问题分析

与有信贷记录的企业比较，银行对无信贷记录的企业缺少信誉评级与是否违约的记录。因此，只要将无信贷记录的企业的信誉评级与是否违约的数据估计出来，该问题就可运用问题一的方法求解。

3.2问题求解

3.2.1 基于Logistic回归的违约概率模型

Logistic回归模型的表达式为：

 （1）

可转化为：

（2）

其中，P示违约的概率，Xi表示表1中的指标。

表2：Logistic回归模型的数据形式

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 因变量 | 自变量 | | | |
| Y | X1 | X2 | … | Xm |
| 1 | x11 | x21 | … | xm1 |
| 0 | x12 | x22 | … | xm2 |
| 1 | x13 | x23 | … | xm3 |
| 1 | x14 | x24 | … | xm4 |
| 0 | x15 | x25 | … | xm5 |
| … | … | … | … | … |
| 1 | x1n | x2n | … | xmn |

方法：运用123家有信贷记录企业的相关数据，估计出Logistic回归模型（1）的未知参数，再将302家无信贷记录企业的相关数据代入Logistic回归模型（2），当其概率P>0.5时，即认为该企业将违约。

3.2.2基于判别分析的信用评级模型

（一）欧氏距离的定义

设有两组样本，每组有k项指标：A1 (x11, x12, …… , x1k), A2 (x21, x22, …… , x2k)，则定义A1与A2 之间的欧氏距离定义为：



（二）基于欧氏距离的判别分析

设 有 t 个 总 体 G1, G2,…, Gt ， 第i个总体Gi样本均值为，则新样品M (xm1, xm2, …… , xmk)到总体Gi的距离为：



当时，则M∈Gj.

1. 突发公共事件对信贷决策的影响分析

4.1 问题分析

（一）行业分类

将无信贷记录的样本企业根据行业性质进行划分，并考虑新冠疫情这一外生冲击变量对不同性质、不同行业影响来调整优化银行对该 302 家企业信贷策略。

（二）不同行业的影响系数

企业生产经营和经济效益会受突发因素影响，且往往对不同行业、不同类别企业影响不同。

（三）影响系数查找及指标调整

本文根据新冠疫情暴发以来统计局公布的各行业利润和营业额数据，调整行业的有关数据。

4.2 问题求解

表3：行业调整系数

|  |  |  |  |
| --- | --- | --- | --- |
| 行业 | 进项调整系数 | 销项调整系数 | 信用等级调整系数 |
| 食品餐饮行业 |  |  |  |
| 石油、煤炭加工业行业 |  |  |  |
| 电气机械和器材制造类 |  |  |  |
| 废弃资源利用业 |  |  |  |
| 文教、服务业 |  |  |  |
| 产品零售业 |  |  |  |
| 通信、电子科技业 |  |  |  |
| 印刷和记录媒介业 |  |  |  |
| 医药器械业 |  |  |  |