

# 基于 K-means 聚类的银行效应最大化信贷决策模型

崇雨琪<sup>1</sup> 朱文君<sup>1</sup> 苗田硕<sup>2</sup>

1 安徽财经大学金融学院 2 安徽财经大学艺术学院

DOI:10.12238/ej.v4i3.675

**[摘要]** 在贷款者众多的现状下,银行通常根据当前信贷政策、企业提供的交易票据信息和合作企业情况,优先向实力强、应用质量高、信贷风险小的企业提供贷款和利率优惠。本文选取某银行的2019年的部分有过信贷交易的公司数据,并进行脱敏处理,运用K-Means聚类、因子分析等方法,构建了信贷风险指标体系以及基于效用最大化理论的企业信贷策略模型;综合运用经济学相关理论,构建了企业信贷风险评估机制,得出适合银行选择小微企业信贷对象的策略。

**[关键词]** 信贷策略; 效用最大化; K-means聚类; 因子分析; 数据挖掘

**中图分类号:** F0-0 **文献标识码:** A

## Bank Utility Maximization Credit Model Based on K-means Clustering

Yuqi Chong<sup>1</sup>, Wenjun Zhu<sup>1</sup>, Tianshuo Miao<sup>2</sup>

1 School of Finance, Anhui University of Finance and Economics

2 School of art, Anhui University of Finance and Economics

**[Abstract]** Under the current situation of many lenders, banks usually give priority to providing loans and interest rate concessions to enterprises with strong strength, high application quality and small credit risk according to the current credit policy, transaction bill information provided by enterprises and cooperative enterprises. Selecting some credit data in 2019 with desensitization treatment and using K-Means clustering and factor analysis, this paper constructs the credit risk index system and the enterprise credit strategy model based on utility maximization theory. It constructs the mechanism of enterprise credit risk assessment using the comprehensive economic theories and gets strategy that suitable for bank to choose credit objects for small and micro enterprise.

**[Key words]** credit strategy; utility maximization; K-means clustering; factor analysis; data mining

### 引言

政府拨出大量专项资金,对小微企业使用进行引导,鼓励地方将其纳入扶持项目当中去。国家税务部门认真落实税收优惠的政策,并根据国家形势变动的需要进行有针对性的改进。

为了响应号召,央行在制定相关政策的时候,加强了扶持力度,要求商业银行增加信贷额度。在一片向好的同时,也存在一些天然的高风险问题。小微企业的规模相对较小,管理体系不够完善,缺少具有分量的抵押资产,虽然国家政策上面给予大力的支持,但由于贷款条件比较严格,融资方面仍然十分困难。此外,商业银行本就是经营风险的机构,清楚应该将款贷给什么样的企业才是合适

的,如果强行违背银行的风险控制原则,降低银行的贷款审核标准,将会大幅提高银行的坏账率,给银行带来巨大的风险。

因此,如何选择合乎标准的小微企业进行贷款是银行现行需要解决的重要问题。

### 1 有记录公司的信贷策略

#### 1.1 数据的预处理

数据来源于123家有信贷记录企业的相关信息。首先使用EXCEL的数据透视表功能和相关代码语句,对数据进行筛选,剔除企业交易数据中的异常值,并参考文献<sup>[1]</sup>,统计提取出能够反映企业实力的相关指标。

#### 1.2 研究流程

在得出信贷策略前,首先要将123家有信贷记录企业的信贷风险用具体的数据表现出来。在年度信贷总额固定时,为银行做出对这些中小企业的信贷策略。针对这一问题,我们分为五步进行研究:

STEP1:对附件一中的数据进行数据挖掘,选择合适的指标,建立信贷风险指标体系。

STEP2:使用因子分析,构建出企业运营的综合实力评价模型,计算出各企业的综合实力得分,得分越高,风险系数越低。

STEP3:基于以上结果,使用K-means聚类分析,对企业的各项经营实力指标和综合实力得分进行聚类。

STEP4: 将聚类结果与附件一所提供的信誉评级进行对比,若高度吻合,则说明以上模型分析效果良好,可以评价各企业的信贷风险大小。

STEP5: 参考各企业的信誉评级和聚类结果,建立边际效用最大化模型,在安全性原则和效益最大的目标下,构建银行对于中小微企业信贷策略的规划,在保证总额全部贷出的基础上,使贷款利率最大,收益最高。

1.3基于因子分析的企业综合实力评价模型的建立

参考<sup>[2]</sup>等文献,我们选取以上十四个风险指标,将数据代入SPSS 24软件中,进行因子分析,结果如下所示:

### 1.3.1 KMO和Bartlett检验

表1 KMO检验

KMO 取样适切性量数	0.691	
巴特利特球形度检验	近似卡方	2347.453
	自由度	91
	显著性	0

由以上KMO和巴特利特检验,可以看出,KMO值约为0.7,且显著性小于0.05,原始变量之间存在相关性,适合进行因子分析。通过总方差解释表和碎石图我们可以看出,前5个因子对应的特征值变化比较陡峭,从第六个因子开始,变化逐渐平坦,因此我们选择五个因子进行分析。

### 1.3.2公共因子载荷分配

我们对因子采取方差最大正交旋转后,得出旋转成分矩阵和因子载荷散点图,如下所示:

### 1.3.3模型结论

通过SPSS计算,得出各企业综合实力水平得分,并进行由高到低进行排序,部分结果如下所示:

表3 123家企业因子综合得分

因子	综合得分
E43	5.892083266
E7	5.110128396
E1	4.608017917
E3	4.310426519
E86	3.478972433
.....	
E107	-2.189882795
E117	-2.293019381
E123	-3.098612911
E101	-3.167837331
E120	-4.594475857

企业综合实力水平得分越高的企业,例如E7,说明企业的体量,经营,管理等实力越强,信贷风险系数越低。

### 1.4基于K-means聚类的企业信贷风险等级评估模型的建立

我们将以上所得到的企业综合实力水平的因子综合得分,并选取进方有效发票率、销方有效发票率、进方负数发票率、销方负数发票率、进货税金比率和销售税率,共七个指标。将数据导入SPSS24软件中,得出聚类结果可以看

表2 旋转后的成分矩阵 a

成分	因子				
	1	2	3	4	5
交易天数(销方)	-0.015	0.889	0.124	-0.059	0.047
销售税率	-0.037	-0.11	-0.973	-0.042	-0.033
销方平均日流水金额(元)	0.813	0.051	0.082	0.152	-0.217
销方日平均纳税(元)	0.822	0.065	0.061	0.148	-0.211
有效发票率(销方)	0.067	0.159	0.965	0.019	0.028
负数发票率(销方)	-0.025	0.116	0.053	0.127	0.845
交易天数(进项)	0.083	0.792	0.139	0.171	-0.02
进项平均流水金额	0.935	0.105	0.002	-0.013	0.13
购方单位代号	0.173	0.814	0.06	-0.056	0.314
进项日平均税额(元)	0.957	0.075	0.01	0.011	0.088
进项平均日流水金额(元)	0.956	0.086	0.013	0.02	0.088
有效发票率(进方)	-0.119	-0.184	-0.06	-0.835	0.119
负数发票率(进方)	0.054	-0.044	0	0.828	0.268
进货税金比例	0.14	0.505	0.003	0.271	-0.316
提取方法: 主成分分析法					
旋转方法: 凯撒正态化最大方差法					
a 旋转在 5 次迭代后已收敛。					

出, 1, 2, 3, 4代表聚类后各公司的信贷风险等级, 等级越小, 代表公司的信贷风险程度越低, 公司综合实力水平越高。第1、2等级的公司数量为42家, 第3、4等级公司数量总计为81家, 分别与附件中银行对各公司的信誉评级为A、B等级和C、D等级的公司数量高度吻合。说明K-means聚类模型有效, 能够准确识别各企业的信贷风险等级。

### 1.5 信贷策略的制定——实现效用最大化

当银行的全部的信贷资金全部贷出, 且分给的企业数量最少时, 银行得到的收益是最高的。

为了达成收益最大化, 银行必须将资金全部贷出。若是将资金持有在手里, 会存在持有成本。贷款给企业, 虽然会有一定风险带来损失, 但是只要其期望回报率是正值, 将资金贷出, 就会产生正收益。

题目中的年度信贷总额已经固定, 根据消费者均衡定理可知, 当最后一元钱给予A、B、C级公司获得的实际利率相等时, 银行收益是最高的。此时

$$r_A = r_{B实} = r_{C实}$$

由于B、C级公司存在一定的违约率, 因此在推算B、C级公司的利率时, 我们应该考虑风险带来的损失, 在名义利率中增加一定量的违约风险溢价。

$$1 + r_A = (1 + r_{B名}) (1 - r_{B违}) = (1 + r_{C名}) (1 - r_{C违})$$

所得利率之间的相关关系:

$$r_A = r_B (1 - 2.63\%) - 2.63\% = r_C (1 - 5.88\%) - 5.88\%$$

随着贷款年利率的不断增大, 此时将钱贷给A、B、C级公司, 所得到的收益和所产生的风险是一样的, 因此不用考

$$\begin{cases} A \cdot F(r_A) \cdot (T/[T/100]) + B \cdot F(r_B) \cdot (T/[T/100]) + C \cdot F(r_C) \cdot (T/[T/100]) = T(\text{信贷总量}) \\ A \cdot F(r_A) \cdot (T/[T/100]) \cdot r_A + B \cdot F(r_B) \cdot (T/[T/100]) \cdot r_C + C \cdot F(r_C) \cdot (T/[T/100]) = R(\text{总收益}) \end{cases}$$

$$\begin{cases} A \cdot F(r_A) \cdot 100 + B \cdot F(r_B) \cdot 100 + C \cdot F(r_C) \cdot 100 = T(\text{信贷总量}) \\ A \cdot F(r_A) \cdot 100 \cdot r_A + B \cdot F(r_B) \cdot 100 \cdot r_C + C \cdot F(r_C) \cdot 100 = R(\text{总收益}) \end{cases}$$

虑A级公司的风险相对较小的情况。

因为  $r_B$ 、 $r_C$  与  $r_A$  具有比例关系, 贷款总额和总收益可以用  $r_A$  表示出来。

$$n(A) \cdot F_A(r_A) \cdot K \cdot r_A + n(B) \cdot F_B(r_B)$$

$$\cdot K \cdot r_B \cdot (1 - p_B) + n(C)$$

$$\cdot F_C(r_C) \cdot K \cdot r_C \cdot (1 - p_C) = R$$

$$n(A) \cdot F_A(r_A) \cdot K + n(B) \cdot F_B(r_B)$$

$$\cdot K + n(C) \cdot F_C(r_C) \cdot K = T$$

由上式可以看出收益  $r_A$  与呈现正相关的关系, 且  $r_A$  越大,  $r_B$ 、 $r_C$  也相对地增大。

由于客户流失度和利率成正比关系, 此时的客户人数最少。单个客户的贷款上限是100万元, 问题一中A企业一共27家, B企业一共38家, C企业一共34家。D企业有24家, 不纳入考虑范围。

$$t = \left\lceil \frac{T}{A+B+C} \right\rceil, \text{ 向上取整, 使得顾客}$$

总数恰好是  $t$  人。

$$[27 \cdot F(r_A)] + [38 \cdot F(r_B)] + [34 \cdot F(r_C)] = t = [T/(A+B+C)]$$

此时的银行贷款策略受到信贷总量的限制。

第一种情况: 当银行的信贷总额超过  $A+B+C$  百万时, 此时可以满足所有企业的信贷需求, 每个企业的信贷量都是100万元, 此时需要满足  $R$  最大值的  $r_A$  就是银

行的最优利率。满足:

第二种情况: 当银行信贷总规模没有达到  $A+B+C$  百万时, 设此时的信贷总规模为  $T$ 。根据之前的推论, 此时利率最高时, 人数最少, 应当只能让  $T/100$  万(向上取整)的人满足贷款需求, 此时由于只考虑违约风险, 由于边际效用相等, 此时贷给A、B、C企业都是没有任何差异的, 所以贷给每一家企业的金额都是相等的, 等于  $(T/[T/100])$  万元<sup>[3]</sup>。满足:

第三种情况: 当信贷总量小于100万大于10万时, 只需要选择一家能够接受最高有效利率的企业即可。

## 2 结束语

本文依据企业流水状况, 类型等数据给银行信贷决策规划提供了依据, 合理调配资金量, 在保证银行流动性基础上, 获得尽可能多的收益。在面对新客户时, 提供了系统性的评价体系, 不仅考虑到了违约风险还考虑了不同企业对于风险的抵御程度不同带来的突发风险。

### [参考文献]

[1] 张左敏, 李文婷. 大数据在商业银行中的应用——基于风险控制的视角[J]. 科技与经济, 2020, 33(04): 61-65.

[2] 赵楠. 商业银行个人信用风险评估模型研究[D]. 江西理工大学, 2019.

[3] 戴国强. 《商业银行经营学》[M]. 高等教育出版社, 2007.

### 作者简介:

崇雨琪(2000—), 女, 汉族, 安徽蚌埠人, 大四学生, 研究方向: 金融学。