

基于银行流水数据的个人客户信用风险评估

赵胜利¹, 张力芝¹, 钟妤玥²

¹重庆理工大学理学院, 重庆

²重庆市渝高中学, 重庆

Email: zhaoshengli@cqut.edu.cn

收稿日期: 2020年11月25日; 录用日期: 2021年1月6日; 发布日期: 2021年1月13日

摘要

对客户进行信用风险评级, 是当前几乎所有银行降低信贷风险的方法。现行的客户风险评级方法大多是根据客户个人信息、历史信用信息以及财务状况等数据展开的, 而这些数据结构不统一, 形式复杂, 无法进行批处理, 增加了银行的审核时间, 容易导致客户的流失。但客户的流水数据结构统一, 形式相对简单, 可以用批处理的方式提取包含在其中的信息, 并在此基础上进行初步的个人信用风险评级, 为进一步精确评级做准备。根据重庆市某商业银行个人客户的流水数据, 构建了银行个人客户信用风险评价指标体系, 通过因子分析法对指标进行降维, 建立了新的个人风险评价指标体系, 然后运用熵值法对评价指标进行赋权, 对各个评价指标进行加权求和得到综合风险值, 最后根据综合风险值对银行客户进行了分类。依据综合风险的大小, 可以将客户分为优质客户、普通客户和不良客户三类, 而三类客户对银行产生的利润比率符合帕累托80/20法则。

关键词

个人信用风险, 因子分析法, 熵值法, 帕累托80/20法则, 风险评估模型

Personal Customer Credit Risk Assessment Based on Bank Flow Data

Shengli Zhao¹, Lizhi Zhang¹, Yuyue Zhong²

¹School of Science, Chongqing University of Technology, Chongqing

²Chongqing Yugao Middle School, Chongqing

Email: zhaoshengli@cqut.edu.cn

Received: Nov. 25th, 2020; accepted: Jan. 6th, 2021; published: Jan. 13th, 2021

文章引用: 赵胜利, 张力芝, 钟妤玥. 基于银行流水数据的个人客户信用风险评估[J]. 金融, 2021, 11(1): 1-9.

DOI: 10.12677/fin.2021.111001

Abstract

Credit risk rating of customers is the current method of almost all banks to reduce credit risk. Most of the current customer risk rating methods are based on the personal information, historical credit information and financial status of bank customers. However, these data are not only disunion in structure, but also complex in form. Therefore, they cannot be processed in batches, which increases the audit time of banks and easily leads to the loss of customers. The flow data of customers are not only unified in structure, but also relatively simple in form, so the information contained in it can be extracted by batch processing, and the preliminary personal credit risk rating can be carried out on this basis to prepare for further accurate rating. According to the flow data of individual customers of a commercial bank in Chongqing, an evaluation index system of credit risk of bank individual customers is constructed. The factor analysis is used to optimize and reduce the dimensions of the index system, and then to establish a new evaluation index system of personal credit risk. Subsequently, using the entropy method to obtain the weight of each evaluation index, and the comprehensive evaluation value is obtained by weighted sum. Finally, the bank customers are classified according to the comprehensive evaluation value. According to the level of comprehensive risk, the bank customers can be divided into good customers, ordinary customers and bad customers, and the profit ratios generated by the three categories of customers to the bank conforms to the Pareto 80/20 rule.

Keywords

Personal Credit Risk, Factor Analysis, Entropy Method, Pareto 80/20 Rule, Risk Assessment Model

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来,随着中国经济的飞速发展,住房按揭、商业贷款以及信用卡消费等各种业务逐渐增多,个人信贷业务的规模迅速扩大,已经成为商业银行获取利益的重要途径之一。对客户进行风险评级,将评级结果作为发放贷款的标准之一,是当前几乎所有银行降低信贷风险的方法。银行大多是根据客户个人信息、交易信息以及信用历史情况来对客户进行评级。这种评级的方法依赖于客户的个人信息和信用历史,信息数据形式复杂,效率较低,容易导致客户的流失。因此,利用银行客户的流水信息,对客户个人信用风险作出快速的评级,减少审核时间,对提高银行的工作效率和收益都有重要的意义。本文根据重庆市某商业银行的客户个人流水数据,构建基于定量数据的评价指标体系,利用因子分析法和熵值法建立了个人信用风险评估模型,并对个人客户风险进行了评级。所给出的评级结果,能对个人客户信用风险情况进行迅速、有效的筛选,减少进一步详细审核的工作量,提高银行的工作效率。

2. 研究现状

国外对个人信用的评价与管理已有一百多年的历史,形成了较为完善的评估模式和运作机制。我国的个人信用评价较国外起步晚,许多学者针对该问题进行了深入研究。一些研究者通过建立个人信用风险评价指标体系,采用传统的统计方法来对个人信用风险进行评估,如模糊层次分析法、Logistic 回归[1]

[2] [3]等,虽然这些方法能够较快的进行信用评估,但主要适用于较小的数据量,其准确性和有效性还有待检验。随着技术的持续进步,银行客户的流量越来越大,使用传统的统计方法可能无法满足大量客户的信用评估,因此越来越多的学者开始采用处理大数据的机器学习方法对银行个人客户信用风险进行研究。为了提高评估的准确率,一些学者采用集成算法[4] [5] [6] [7]来建立个人信用风险模型,但仍存在一定的缺陷,即虽然比单个模型预测的结果好,但需要大量的维护工作,使得评估过程过于繁琐。还有一些学者在针对信用评估的传统方法的方向上,提出新的评价方法来提升评估效果,如 Group-LASSO 方法[8]、多源数据融合[9]、LGB-BAG 方法[10]和群决策[11]等,这些方法虽然在信用评估过程中提高了准确率,但大多是基于提取客户各种基本信息作为评价指标进行研究,而基于银行客户纯数据的交易信息的评价方法还鲜有研究。

上述关于银行个人客户信用风险的研究都是利用客户的基本个人信息和历史信用信息来进行的。客户的基本个人信息变量多,形式复杂,格式不统一,难以进行批量的计算和分析。本文旨在在银行客户的个人流水数据的基础上,建立个人客户信用风险评价模型,从大量客户中快速筛选出重要客户,为进行更精确的风险评估做准备。本文的结构安排如下:一、针对客户个人流水数据构建银行个人客户的信用风险评价指标体系;二、利用因子分析进行优化降维,并构建新的个人信用风险评价指标体系;三、基于熵值法计算评价指标体系中的各指标权重,并加权求和得到综合信用风险得分;四、根据综合得分用帕累托 80/20 法则[12]对客户的个人信用进行等级划分。

3. 银行个人客户信用风险评价模型

3.1. 基于客户个人信息的信用风险评价指标体系

在国内外已有的相关文献中,关于个人信用风险评估指标一般分为 3 类:侧重贷款者还款能力的经济分析、侧重贷款者还款意愿的信用分析以及贷款者的综合分析,其中大多数采用的是综合分析方法。在对客户个人风险进行评估的过程中,合适的评价指标的选取对于整个评价模型的建立尤为重要。国内外的相关研究所采用的评价指标主要包括银行客户的个人信息、财务信息、贷款信息以及信用历史,这些都是基于客户的基本信息所选取的指标。例如,冯琼[7]根据我国商业银行个人信用风险评估指标选取的原则,确定了 5 个二级指标以及 24 个相应的三级指标,如表 1 所示,这里展示了基于客户个人基础信息、经济状况、职业信息、信用状况以及贷款信息的各个风险评价指标,不仅涉及到了银行流水数据,还有个人的基础信息数据,从各个方面来考虑了影响银行客户的个人信用风险的因素。

Table 1. Individual credit risk assessment index system

表 1. 个人信用风险评估指标体系

一级指标	二级指标	三级指标
综合评价指标	基本信息指标	年龄
		性别
		婚姻状况
		教育程度
		健康状况
	职业信息指标	住房状况
		单位类型
		行业情况
		职位或职称
		在本职位工作时间

Continued

经济状况指标	借款人个人月收入
	家庭月收入
	家庭平均月收入
	金融资产
	其他资产
信用状况指标	保险能力
	是否有不良信用记录
	在本银行的存款账户
	与银行业务关系
	是否为本银行职员
贷款产品基本信息指标	是否存在其他贷款
	贷款期限
	贷款金额
	还款方式

3.2. 基于银行流水的信用风险评价指标体系

重庆市某商业银行个人客户的流水数据包含账户号、交易日期、交易时间、交易类型、交易金额以及交易对手等数据信息，但是没有个人客户的基本信息和历史信用信息。本文将在表 1 所给的指标体系的基础上，从银行客户的流水数据提取出相应的个人信用风险评价指标，最终确定基于银行流水的个人信用风险评价指标体系。根据对商业银行客户信用情况的考虑，遵循一定的指标选取原则[13]，依据已有的评价指标体系，从经济状况方面对客户的银行流水数据进行指标的选取，建立科学、规范、合理的个人信用风险评价指标体系，如表 2 所示。其中展示了交易信息指标、收入状况指标、支出状况指标和资金变化指标这 4 个二级指标以及交易总金额、交易总次数和月均收入等 11 个三级指标。

Table 2. Evaluation index system of credit risk of bank individual customer

表 2. 银行个人客户信用风险评价指标体系

一级指标	二级指标	三级指标	指标解释
信用风险	交易信息指标(Y ₁)	交易总金额(X ₁)	所有交易的总金额，单位：元
		交易总次数(X ₂)	所有交易的总次数，单位：元
		平均现金交易额(X ₃)	平均每次存取金额，单位：元
	收入状况指标(Y ₂)	月均收入(X ₄)	平均每月的收入，单位：元
		平均固定收入(X ₅)	有固定收入的平均收入金额，单位：元
		存款金额(X ₆)	当前账户的存款金额，单位：元
	支出状况指标(Y ₃)	倾向性消费的平均金额(X ₇)	有倾向性消费支出的平均交易金额，单位：元
		平均固定支出(X ₈)	有固定支出的平均支出金额，单位：元
		其他消费金额(X ₉)	其他消费支出的总额，单位：元
	资金变化指标(Y ₄)	收入占比(X ₁₀)	总收入与总交易额之比
		平均理财金额(X ₁₁)	平均每次理财的金额，单位：元

表 2 中的评价指标与个人信用风险都呈正相关或负相关, 其中与个人信用风险呈正相关的评价指标有倾向性消费的平均金额、平均固定支出和其他消费金额, 呈负相关的评价指标有月均收入、交易总金额、平均理财金额、平均固定收入、交易总次数、平均现金交易额、收入占比和存款金额, 它们都在一定程度上影响着个人信用风险的评估。

3.3. 评价指标的降维优化

先对上述指标进行因子分析找出公因子, 并选出累计贡献率在 75% 及以上的各个因子。这些因子都是由原指标体系中的评价指标组合而成, 并且他们可以代替所有的指标来分析客户个人信用风险的问题, 因此, 最终可以由这些因子作为评价指标建立新的个人信用风险评价指标体系。

本文利用 SPSS 统计软件, 对商业银行客户的个人信用风险评价指标进行因子分析, 数据来源于重庆某商业银行 2018 年的客户个人流水数据, 由于原始数据之间量纲的不同, 需要对其进行标准化处理。

① 指标数据的适用性检验。我们在做因子分析之前需要检验数据的适用性, 本文采用巴特利特球度和 KMO 值两种检验方法, 来验证评价指标数据是否适合做因子分析。根据结果显示, 该指标体系的巴特利特球度检验统计量的观测值为 602326.671, 相应的 p 值为 0.000, 小于 0.5, 则拒绝原假设, 并且 KMO 值为 0.623, 大于 0.5, 总的说明该指标体系可以做因子分析。

② 提取公因子。根据输出结果中的总方差解释显示, 前六个因子的累计贡献率已达到 79.190%, 说明前六个因子共解释了原有变量总方差的 79.190%。总体来说, 前六个因子能够代表最初的 11 个评价指标来分析客户的个人信用风险情况, 如表 3 所示。

Table 3. General variance interpretation

表 3. 总方差解释

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Cumulative %
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	
	2.927	26.610	26.610	2.927	26.610	26.610	26.610
	1.659	15.081	41.691	1.659	15.081	41.691	41.691
	1.152	10.476	52.167	1.152	10.476	52.167	52.167
	1.081	9.829	61.996	1.081	9.829	61.996	61.996
	0.985	8.954	70.950	0.985	8.954	70.950	70.950
	0.906	8.240	79.190	0.906	8.240	79.190	79.190
	0.877	7.971	87.161				
	0.561	5.102	92.262				
	0.370	3.366	95.628				
	0.330	3.005	98.632				
	0.150	1.368	100.000				

③ 将因子载荷矩阵进行旋转。旋转后的因子载荷矩阵对主成分的解释更加清楚明了, 表 4 为旋转后的因子载荷矩阵。根据表 4 可将原始的 11 个三级评价指标综合为 6 个因子评价指标, 分别为其命名, 并列如表 5 中。其中消费因子和支出因子与个人信用风险呈正相关, 即值越大, 信用风险越高, 而其它因子则与个人信用风险呈负相关, 即值越大, 信用风险越低。

Table 4. Factor load matrix after rotation
表 4. 旋转后的因子载荷矩阵

Rotated Component Matrix	Component					
	1	2	3	4	5	6
Zscore (存款金额)	0.784	0.371	-0.015	-0.023	-0.003	0.026
Zscore (平均固定收入)	0.715	0.146	0.001	-0.106	0.001	0.001
Zscore (平均现金交易额)	0.630	-0.297	0.048	0.247	0.034	-0.008
Zscore (交易总次数)	-0.026	0.822	0.112	-0.030	-0.062	0.003
Zscore (月均收入)	0.506	0.723	0.014	0.187	0.105	0.090
Zscore (交易总金额)	0.392	0.599	0.096	0.397	0.381	0.006
Zscore (倾向性消费的平均金额)	0.016	0.008	0.912	0.010	0.010	-0.013
Zscore (其他消费金额)	0.010	0.132	0.902	-0.010	-0.004	-0.013
Zscore (平均固定支出)	-0.006	0.085	-0.014	0.939	-0.020	-0.023
Zscore (平均理财金额)	0.001	0.020	-0.004	-0.018	0.979	-0.030
Zscore (收入占比)	0.017	0.039	-0.024	-0.023	-0.030	0.997

Table 5. Evaluation index system of personal credit risk based on factor analysis
表 5. 基于因子分析的个人信用风险评价指标体系

变量	评价指标	原始指标
F_1	收入因子	存款金额、平均固定收入、平均现金交易额
F_2	交易因子	交易总次数、月均收入、交易总金额
F_3	消费因子	倾向性消费的平均金额、其他消费金额
F_4	支出因子	平均固定支出
F_5	理财因子	平均理财金额
F_6	收入占比因子	收入占比

3.4. 基于熵值法的信用风险评价模型

本文基于熵值法来对商业银行客户的个人信用建立信用风险评价模型，科学地计算所选取的个人信用风险评价指标权重，并根据各指标权重对银行客户的个人信用风险进行评价。

根据熵值法确定指标权重，基本步骤如下：

① 对评价指标进行标准化处理，消除量纲。

② 计算第 j 项指标下第 i 个客户对应值占该指标的比重： $P_{ij} = F'_{ij} / \sum_{i=1}^n F'_{ij}$ ($j=1, 2, \dots, m$)。

其中： F'_{ij} 为标准化后的指标数值。

③ 计算第 j 项指标的熵值： $e_j = -k \times \sum_{i=1}^k P_{ij} \ln(P_{ij})$ 。其中 $k = 1/\ln(n)$, $0 \leq e_j \leq 1$ 。

④ 计算第 j 项指标的差异系数。针对第 j 项指标，指标值的差异越大，对整个评价体系的左右就越大，相应的熵值就越小。差异系数： $g_j = 1 - e_j$ 。

⑤ 求第 j 项指标的权重： $w_j = g_j / \sum_{j=1}^m g_j$

根据上述的熵值法确定权重的步骤，可利用前面因子分析所得到的各个因子得分来计算相应的评价指标的熵值与权重，则计算得到的每项评价指标的熵值与权重，如表 6 所示。

Table 6. Entropy value and weight of evaluation index
表 6. 评价指标的熵值与权重

评价指标	熵值	权重
收入因子(F_1)	0.999995	0.001376
交易因子(F_2)	0.999997	0.000851
消费因子(F_3)	0.999790	0.071265
支出因子(F_4)	0.999958	0.014455
理财因子(F_5)	0.999997	0.000682
收入占比因子(F_6)	0.997254	0.911371

利用熵值法计算得到的指标权重的大小能够反映评价指标对整个评价体系影响的大小，从表 6 中可以看出，指标收入占比因子 F_6 的熵值最小且为 0.997254，它所对应的评价指标权重是相对最大的，为 0.911371；而指标理财因子 F_5 的熵值最大且为 0.999997，它所对应的评价指标权重是相对最小的，为 0.000682。

根据所得的权重，将标准化后的因子指标 F'_{ij} ($i = 1, 2, \dots, n; j = 1, 2, \dots, m$) 进行加权求和，得到个人信用风险的综合风险值，即

$$F_i = \sum_{j=1}^m F'_{ij} * w_j \quad (i = 1, 2, \dots, n) \quad (1)$$

3.5. 结果展示与分析

本文根据加权求和得到的银行客户的个人信用综合评价值被展示于表 7 中，这里由于客户人数较多，客户账户号较为复杂，为了便于查看，可对所有客户的账户号进行编号，展示了综合评价值前十名客户和后十名客户的相关数值。

Table 7. Comprehensive evaluation value of personal credit risk of bank customers
表 7. 银行客户的个人信用风险综合评价值

前十名客户		后十名客户	
客户编号	综合评价值 F	客户编号	综合评价值 F
162278	0.917648	101175	0.129903797
182338	0.749550	92124	0.124324661
73568	0.604905	120862	0.12340393
197451	0.576922	87387	0.121567462
172424	0.564564	184872	0.120566952
59220	0.551617	29124	0.10221091
88538	0.535600	91777	0.084995814
50682	0.531816	84315	0.08188248
150112	0.488859	128826	0.04055243
147240	0.477693	43466	0.020058014

计算得到的个人信用风险综合评价值的区间范围大概为[0.91,0.02]，最大值为 0.917648，最小值为 0.020058014，并且大部分银行客户的综合评价值都在[0.21,0.37]的区间范围以内。商业银行可以由算得的客户个人信用风险综合评价得分来对客户进行简单的排序，可快速地根据个人流水数据掌握客户的资金状况，来判断客户的个人信用是否良好，从表 7 中可以看出，个人信用风险较高的排行前十的银行客户，他们的综合评价得分相差较大，说明这些客户可能存在个人代替公司交易，如给员工发放工资或商业采购等，这样就造成该客户的总支出较大，从而使得综合风险值较大。

4. 银行客户信用等级分类

根据帕累托 80/20 法则(又称 ABC 客户分类法),企业可以依据客户给企业带来的利润大小进行分类: A 类客户占比 20%左右,企业 80%的利润来源于该类客户,因此 A 类客户也被称为优质客户或者 VIP 客户; B 类客户占比 70%左右,只为企业带来了不足 20%的利润, B 类客户也被称为普通客户; C 类客户占比 10%左右,给企业带的利润很低,也称为不良客户。本文首先将每个客户的风险值从小到大排序,在利用帕累托 80/20 法则中的 20%, 70%, 10%的比率进行分类,可以将客户分为优质客户、普通客户和不良客户三类,如表 8 所示。

Table 8. Classification and classification of customer personal credit ratings

表 8. 客户个人信用等级分类及划分范围

指标	划分范围	客户类型	客户数量	客户占比	银行利润贡献率
综合风险值 F	$F \leq 0.215633$	优质客户	39883	19.9812%	92.5103%
	$0.215633 < F \leq 0.368491$	普通客户	140135	70.2068%	33.0760%
	$F > 0.368491$	不良客户	19585	9.8120%	-25.5863%

注: 因为利用年余额作为银行利润指标, 所以利润贡献率中有负值, 可以理解为不良客户可能会给银行造成坏账带来损失。

为了验证上述的分类结果是否符合帕累托 80/20 法则, 对客户对银行带来的利润进行了计算。本文将银行客户的年余额(总收入减去总支出)作为评估银行利润的指标, 计算出各类客户给银行带来的利润比例, 如表 8 所示。从中可以看出, 根据每类银行客户的年余额占总的年余额的比例来替换他们给银行所带来的利润比, 其中优质客户能给银行带来 92.5%左右的利润, 普通客户能给银行带来 33%左右的利润, 而不良客户非但没有给银行带来任何利润, 还降低了 25.6%左右的利润。由此表明, 上述对银行客户的分类结果基本符合帕累托 80/20 法则, 从而说明本文的分类结果是有意义的。上述分类结果也说明: 优质客户值得银行重点关注, 但也不能忽视普通客户和不良客户对银行的影响。

5. 结论

面对商业银行对客户个人信用风险评估难、数据多、操作繁杂以及流程慢等问题, 本文通过银行客户的个人流水数据, 利用因子分析法, 提出了一种基于熵值法的个人信用风险评估模型, 旨在能够使商业银行可以快速准确地对银行客户进行分类。本文从银行的个人流水数据中提取出关键的评价指标, 构建指标体系并建立了银行个人客户风险综合评价模型。依据风险值大小排序将客户分成了三类优质客户、普通客户和不良客户, 通过计算这三类客户对银行的利润贡献率, 说明了本文的分类结果基本上符合帕累托 80/20 法则, 从而也说明了本文模型和分类结果的正确性。

基金项目

重庆市社会科学规划项目(2018BS56)。

参考文献

- [1] 易传和, 彭江. 基于 FAHP 的个人信用评分模型[J]. 统计与决策, 2009(15): 67-70.
- [2] 张国政, 陈维煌, 刘呈辉. 基于 Logistic 模型的商业银行个人消费信贷风险评估研究[J]. 金融理论与实践, 2015(3): 53-57.
- [3] 冯振涛, 冯梦媛. 基于 AHP 的信用卡评分模型研究[J]. 金融理论与实践, 2016(1): 74-77.
- [4] 陈云, 石松, 潘彦, 等. 基于 SVM 混合集成的信用风险评估模型[J]. 计算机工程与应用, 2016, 52(4): 115-120.
- [5] 陈舒期, 梁雪春. 改进的 SSVM 集成算法在信用风险评估中的应用[J]. 计算机工程与设计, 2019, 40(10): 2822-2826.
- [6] 莫赞, 张灿凤, 魏伟, 等. 基于 Bagging 集成的个人信用风险评估方法研究[J]. 系统工程, 2019, 37(1): 143-151.
- [7] 冯琼. 基于组合预测模型的银行个人信用风险评估体系应用研究[D]: [硕士学位论文]. 武汉: 中南民族大学, 2011.
- [8] 张娟, 张贝贝. 基于 Group-LASSO 方法的广义半参数可加信用评分模型应用研究[J]. 数理统计与管理, 2016, 35(3): 517-524.
- [9] 方匡南, 赵梦峦. 基于多源数据融合的个人信用评分研究[J]. 统计研究, 2018, 35(12): 92-101.
- [10] 李淑锦, 嵇晓佳. LGB-BAG 在 P2P 网贷借款者信用风险评估中的应用[J]. 技术经济, 2019, 38(11): 117-124.
- [11] 姜雪莹, 秦进. 基于群决策的 P2P 借贷信用风险评估[J]. 计算机系统应用, 2019, 28(5): 226-231.
- [12] 理查德·科克. 帕累托 80/20 效率法则[M]. 李汉昭, 译. 北京: 海潮出版社, 2001.
- [13] 胡爱平, 邱世芳, 伍度志. 重庆市城镇化综合评价指标体系应用研究[J]. 重庆理工大学学报(自然科学), 2016, 30(3): 140-148.