P2P 个人风险识别研究[1]

北方民族大学　　唐姣、刘弋驰、李雪丽

摘　　要

随着我国互联网技术的快速发展，P2P 网络借贷在我国悄然兴起，但是网络借贷安全性低，信用风险较高，所以识别 P2P 网络借贷平台的风险具有重要的现实意义。本文将影响 P2P 网络借贷平台后续投标的所有变量划分为 5 个维度，包括人口特征、借款信息、信用变量、逾期信息和借款者还款情况，从这五个维度中选取了 7 个自变量构建 Logit 回归模型，基于“拍拍贷”的用户偿还贷款情况的 11314 个数据样本，利用 Logi 回归方法分析了 7 个自变量对风险识别的重要性。结论表明：信用等级和借款金额及利率对逾期且还款存在明显显著性关系，信用等级越高，逾期的概率越小。借款金额越高，逾期的概率越大。

关键字：p2p 风险识别　Logit 模型　SPSS

极大似然估计　逾期概率

[1] 注:该论文获得由中国统计教育学会举办的“2015 年（第四届）全国大学生统计建模大赛”大数据统计建模类本科生组三等奖。

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

一、问题叙述

选题 33：基于 P2P 个人借贷平台的数据，建立对个人的风险识别模型。请详细描述建模假设，模型变量，建模思路，模型方法，并采集数据验证。

二、引言

随着以互联网为代表的现代信息科技的发展，互联网金融已经成为既不同于商业银行间接融资、也不同于资本市场直接融资的新兴金融融资模式，而 P2P 网络借贷成为互联网金融模式的主要代表之一。

2.1 P2P 网络借贷介绍

P2P 网络借贷（也被译作 Person to Person，人人贷）指的是个体和个体之间通过网络实现直接借贷。具体说来，就是指有资金并且有理财投资想法的资金持有者，通过信贷平台牵线搭桥，使用信用贷款的方式将资金贷给其他有借款需求的资金需求者。这意味着，在网上通过鼠标操作，就能借钱给网友或是向别人借钱。

P2P 网络借贷起源于欧美，它的创始人是 2006 年“诺贝尔和平奖”得主——穆罕默德·尤努斯教授。网络借贷的开端起源于一次公益性质的贷款，1976 年，在一次乡村的调查中，穆罕默德·尤努斯教授发现，当地勤劳的村民因为没钱购买制作竹椅的原材料，而不得不成为商人的劳动力。为帮助当地的村民免受高利贷的剥削，将 27 美元拆分借给了 42 位需要资金的村民，用来支付他们用以制作竹凳的微薄成本。但出乎众人的意料，最后这批小额贷款实现了高达 98.7% 的偿还率，由此开启他的小额贷款之路，而这件事更标志着小额信贷在全球范围内获得了承认。

2.2 P2P 网络借贷原理

借款人和理财人在 P2P 网络平台分别注册，网贷平台对借款人的信用评级，借款人根据自身的需要，在 P2P 平台上发布借款信息，理财人对借款项目进行竞拍。P2P 平台对借款人及项目进行审核。如果有足够多的理财人竞拍了借款项目，达到借款项目的全额，且借款项目通过了 P2P 平台的审核，则该笔借款成立。理财人投标成功， P2P 网络借款并不是理财人直接打给借款人，而是第三方平台来发放。借款人按期归还本金和利息，理财人收回本息，P2P 网络借贷结束。

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

上述流程可用下图 1 来描述：

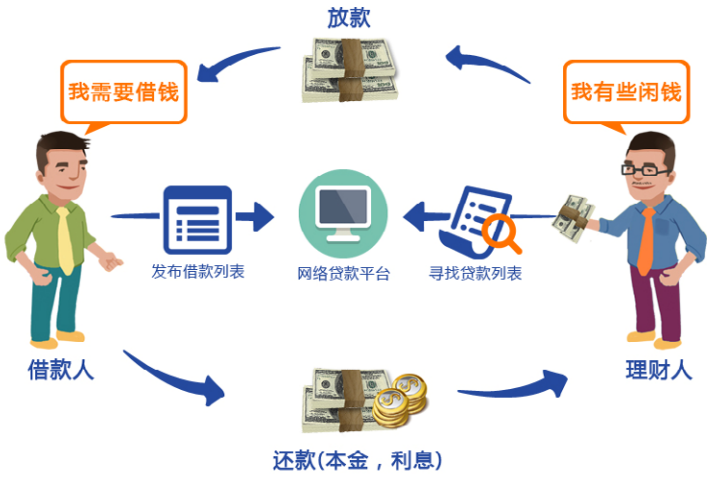

图 1 P2P 网络借贷工原理

2.3 P2P 交易流程中的风险分类

投资人、借款人和网络平台构成了 P2P 网络借贷的主体：平台软件、第三方信用评级机构、第三方支付、担保公司构成 P2P 网络借贷业务的支撑平台。不论是业务主体还是业务支撑平台，在业务发生的过程中均存在风险点，将风险影响的结果分类，P2P 网络借贷的风险归类起来分为基本风险和特定风险。

基本风险指非个人行为且对整个团队乃至整个社会产生影响的风险，是个人无法预防的风险。本文将基本风险分为法律风险、信用风险和监管风险三方面。

特定风险是指由于个人行为引起的风险，只与特定的个人或部门相关，并不影响整个团体和社会。在 P2P 网络借贷业务中，通常指企业内部因风险管理欠缺而导致的风险。主要有：信息不对称风险；投资风险；自律风险；结算风险；信息安全风险。

2.4 P2P 现状

自 2007 年国外网络借贷平台模式引入中国以来，国内 P2P 网络借贷平台蓬勃发展、百花齐放，迅速形成了一定规模。P2P 网贷最大的优越性，是使传统银行难以覆盖的借款人在虚拟世界里能充分享受贷款的高效与便捷。网贷平台数量近两年在国内迅速增长，迄今比较活跃的有 350 家左右，而总量截止到 2015 年 4 月底已有 3054 家，具体交易情况如下图表 1：

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

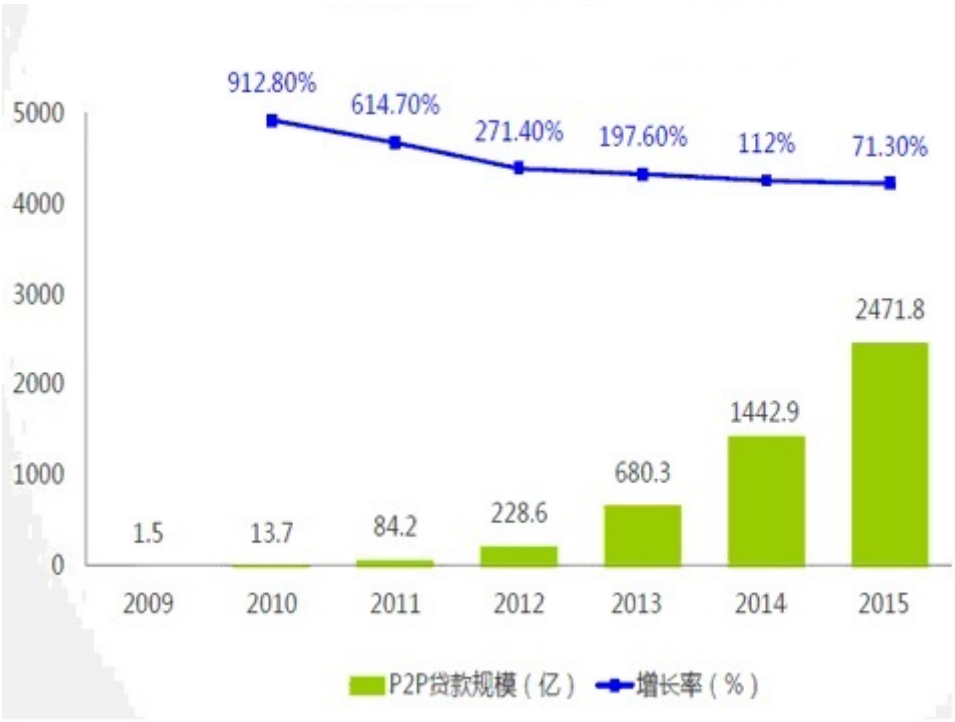

图 2　2009—2015 年中国 P2P 贷款交易量

2.5 P2P 网络借贷研究意义

P2P 网络借贷平台主要的客户是一些急需资金的个人及中小微企业，特别是一些银行业务没有覆盖到的人群。然而，由于一部分中小微企业的信用较差，还款能力相对较弱，信息不对称等问题，融资难一直成为阻碍我国小微企业发展的一大难题。P2P 网络借贷作为互联网与民间借贷结合的一种新兴互联网融资渠道，为支持中小微企业的金融服务，拓宽小微企业融资渠道，缓解小微企业的融资困难，支持和促进中小微企业的经济发展提供了新的路径。P2P 网络借贷也为以个体为单位的投资者提供了一种新的投资方式，但是，由于目前的网络借贷平台将借款人都默认为有信用的人，借款人只要按照网站的要求上传一些资料，再查询一下央行征信系统就能借钱，所以导致网络借贷安全性较低，信用风险较高。

三、数据描述

为较好地反映我国 P2P 网络借贷平台上用户行为的真实情况，保证数据分析的结果能较为客观地反映所研究的问题，我们选择了拍拍贷作为研究对象。这其中有数据获取原因，也有中国特有环境的原因。本节接下来对使用数据抓取软件获取的 2015 年拍拍贷 3 月—5 月用户偿还贷款情况的数据进行一些简单描述。

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

3.1 拍拍贷网贷平台介绍

拍拍贷于 2007 年 8 月，平台在上海正式成立，它成为我国第一家 P2P 小额无担保的网络借贷平台。平台自身的定位是作为单纯的中介平台，不参与资金的运营，主要的收入来源是借贷交易时收取的服务费。网站有自己的用户信用等级核定制度。拍拍贷主要根据用户的“线下得分”和“线上得分”两方面来核定其信用等级，线下得分主要包括用户的个人信息，还款能力等因素；线上得分包括用户在网站上的历史交易记录，各项认证等因素。网站对用户的信用评分直接影响借款人可借额度的高低和借款最后的成功率。拍拍贷网站的风险控制方式，靠强制借款人每月按时还款和提示投资者小额分散投资来降低借贷风险，同时拍拍贷的借款时限为 3-12 个月以内，且平台为了提高投资者的资金安全，已经推出了本金保障计划，这些措施都可以在一定程度上起风控作用。拍拍贷公司现有员工逾 1000 人。截至 2014 年上半年，注册用户近 360 万，是国内用户规模最大的 P2P 平台。拍拍贷从 2008 年到 2015 年成交量创 274%的涨幅，发展迅速，如下图 3：

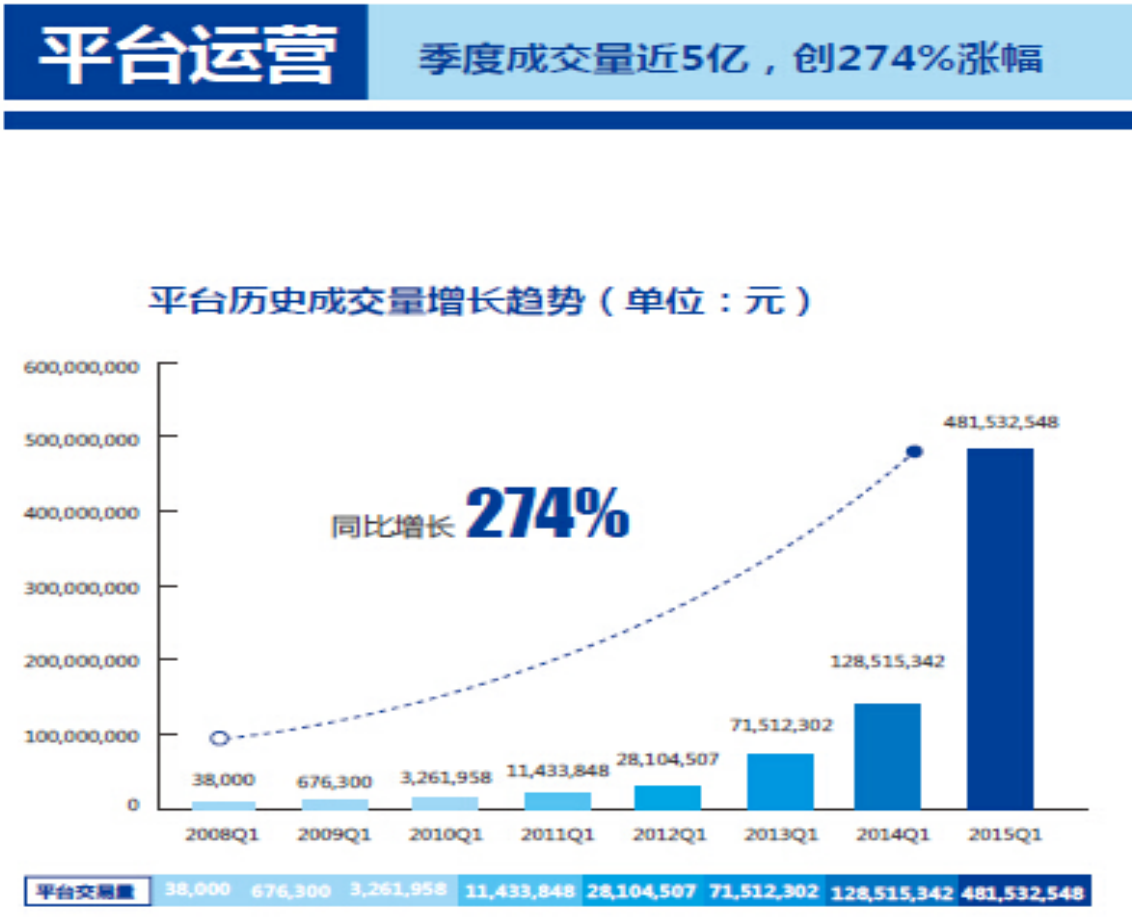

图 3 拍拍贷从 2008 年到 2015 年成交量

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

3.2 拍拍贷网络借贷机制

使用拍拍贷借款的基本流程。第一步，借款人在拍拍贷平台上注册账号，当注册成功后在平台中浏览基本的贷款种类，选择适合自己的进行贷款申请。借款人在进行贷款申请时需要填写借款人的基本信息，包括性别、学历、工作、收入等等。第二步，平台中有意愿出借的用户在拍拍贷平台上浏览借款人发布的借款信息，根据借款信息判断贷款的风险，根据本身能够接受的风险程度来投标。第三步，借款结束。当一笔借款在规定时间内成功达到满标时，拍拍贷会对本笔借款审核，审核通过后，本比借款完成。借款人的账户上会得到借款。第四步，偿还贷款。偿还贷款是借款流程的最后一步，但是也是最关键的一步。借款人应按照协议的时间进行还款。借款人应该保证账户资金充足，按时将款项还给各位出借人。还款完毕后，借贷关系结束。拍拍贷网贷借款基本的工作原理如下图 4：

图 4 拍拍贷工作原理图

3.3 数据来源

对于国内的网络借贷平台，数据基本只是通过用户借款呈现在网站上，由于 P2P 借贷在中国的发展较晚，并且没有相关检测机构进行数据检测，所以我们需要的数据没有官方的统计机构提供。为了取得网站中用户的数据结构，只能通过使用数据抓取软件对网页上呈现的数据进行抓取。在拍拍贷平台中，任意借款者的每一笔借款信息都有单独的网页对应于该笔借款信息的存储，并且对应着唯一一个以借款编号为结尾的 URL（Uniform Resource Locator 统一资源定位符），其中存储着借款人此次借款的相关信息，如借款金额、年利率、期限、还款方式、借款人历史借款成功和流标次数、借入信用以及借出信用等，此外还有借款人在网站上的用户名、目前借款状态等等。比如：在网址 http://www.ppdai.com/list/2915579 中，如图 5：

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

图 5　网址对应信息

对于数据的抓包，我们选取了 VBA 开发环境，采用了 XML Http Request 方法。将需要的借款信息的编号输入 Excel 表格的第一列，然后通过运行 ExcelVBA 函数，对相应编号的借款项目的相关内容进行数据提取 VBA 函数将网页转换为文本格式以后可以查找到文章实证所需要的数据，将其写入 Excel 表格的对应位置，就得到了我们需要的数据。

通过不断地读取对应网址页面提取数据，得到有关拍拍贷 2015 年 3 月到 5 月的 11314 个数据，依据实际情况，剔除掉一些缺失数据、未满标借款以及审核未通过数据，整理得到可用数据 11234 个。

3.4 指标说明

拍拍贷是我国第一家 P2P 网络借贷平台，目前也是国内所有 P2P 网络借贷平台中最大、用户最活跃的借贷平台。所以本文选择以拍拍贷网站上的用户作为研究对象，经过一定得分析，选择从用户的人口特征（年龄、性别、用户身份），信用变量（信用等级），借款信息（期限、利率、贷款金额），逾期信息（贷款逾期天数、逾期本息金额）和借款者还款情况（逾期且还款情况）五个方面选取出关于网络借贷风险影响因素的变量，来用作 P2P 网络借贷信用风险影响因素的实证研究。 归纳如下表 1：

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

表 1 数据指标解释

分类	名称	指标
人口特征	sex	性别
	age	年龄
	ID	用户身份
借款信息	money	贷款金额
	month	借款期限（月）
	rate	借款利息
信用变量	credit	信用等级
逾期信息	overdays	贷款逾期天数
	debt	逾期本息金额
借款者还款情况	rt	逾期且还款情况

（1）人口特征（年龄、性别、用户身份）：在拍拍贷上借款人的性别分为男、女，在实证过程中，把它们分别用 1、0 来代替；借款人年龄段分为 20-25 岁、26-31 岁、32-38 岁和大于 39 岁，在实证过程，把它们分别用 1、2、3、4 来代替；借款人用户身份分为学生、工薪族、私营业主、其他、网店卖家，实证过程中，把它们分别用 1、2、3、4、5 来代替，借款者的身份反映了借款者收入的稳定性及借款者的偿还能力。

（2）信用变量（信用等级）：拍拍贷平台运用魔镜等级来表示用户自身的信用等级，魔镜等级主要依靠实名认证，这些认证包括身份认证、视频认证、学历认证、手机认证和网上银行充值认证。网站会根据借款者相关认证完整度打出一个基础分数，在基础分数的基础上，结合借款者提供的工作证明、人民银行征信报告、房产证明、婚姻证明、银行流水等资料对借款者进行评级，其核心是一系列基于大数据的风控模型，针对每一笔借款，风险模型会给出一个风险评分，以反映对逾期率的预测。每一个评分区间会以一个字母评级的形式展示给借入者和借出者。级别分为 AAA、AA、A、B、C、D、E、F 这八个等级，风险依次上升。理论上，认证等级越高，表明个人的基本信用越高，能借到的金额越多，需要承诺的利率越低。在实证过程，把它们分别用 8、7、6、5、4、3、2、1 来代替。

（3）借款信息（期限、利率、贷款金额）：利率是拍拍贷中借款人为了获取需要的资金，愿意付出的融资成本。在投资者看来，当然是希望利率越高越好，但是借款者则恰恰相反，他会根据平台对自己评价的信用等级，提出相对适合的利率。总体来说，安全程度越高的借款人，通常可以提供较低的利率。期限是指拍拍贷中的借款者对所借金额承诺还款的时间段，按照借款者的意愿选择，有

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

1-12 个月承诺时间，借款者选择的时间越长，就越容易出现预期还款的情况，发生信用风险的可能性就相对更大。贷款金融与借款者的违约概率呈正向变动，即贷款金额越大，借款者违约的概率越高，信用风险越大。

（4）逾期信息（贷款逾期天数、逾期本息金额）：贷款逾期天数指超过应该还款最后期限的时间，一般来说时间越长，成为坏账的可能性越大；逾期本息金额，是指超过借款期限，还没有收到的还款数额与利息金额。

（5）借款者还款情况：对于在拍拍贷上贷款的用户，还款情况有能够按时还款，不能按时还款，在实证过程中，依次用 0、1 代替。

3.5 数据处理与描述性统计分析

由于上述指标中既含有定量指标，又含有定性指标，所以必须在使用样本数据前对指标进行处理，使其具有一致性。本文将定性指标的相应结论按照其与信用风险相关的程度转化为有序响应变量。如下表 2：

表 2　数据指标的量化处理

名称	指标	原始属性值以及量化处理
sex	性别	男－1；女－0
age	年龄	20－25岁－1；26－31岁－2；32－38岁－3；大于39岁－4
ID	用户身份	学生－1；工薪族－2；私营业主－3；其他－4；网店买家－5
money	贷款金额	实际值
month	借款期限（月）	实际值
rate	借款利息	实际值
credit	信用等级	AAA－1；AA－2；A－3；B－4；C－5；D－6；E－7；F－8(AAA到F风险依次上升)
overdays	贷款逾期天数	实际值
debt	逾期本息金额	实际值
rt	逾期且还款情况	已经还款－1；还有欠款－2

有关拍拍贷 2015 年 3 月到 5 月用户偿还贷款情况的 11314 个数据，依据实际情况，剔除掉一些缺失数据、未满标借款以及审核未通过数据，整理得到可用数据 11234 个数据样本，来进行一些初步的分析。其中在这 11234 个数据样本中，能够按时还款的样本数据个数为 11107 人，占总样本数的 98.88%；未按时还款的样本数据个数为 126 人，占总样本数的 1.12%。说明存在的违约率达为 1.12%。

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

具体数据如下图 6:

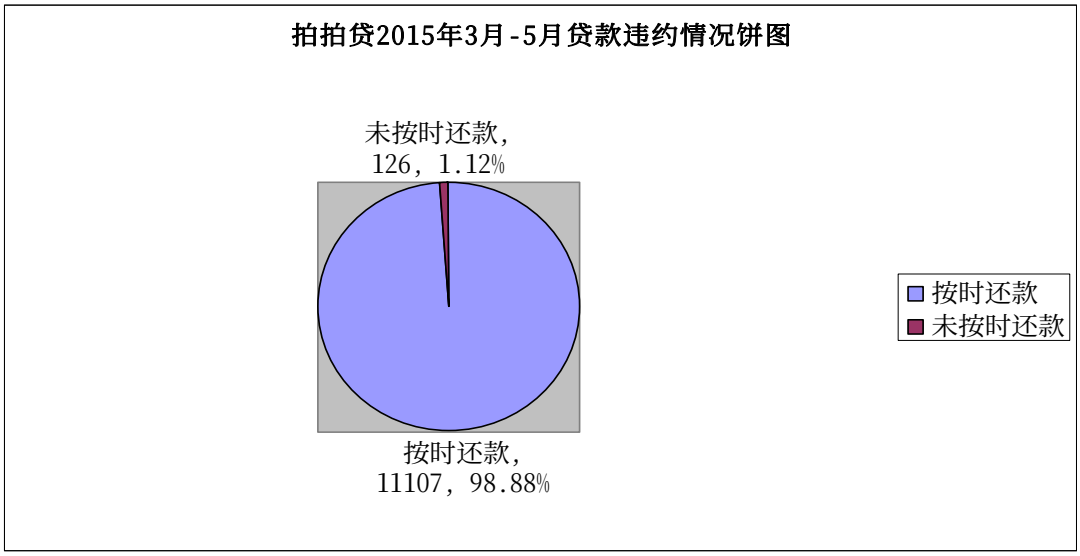

图 6 拍拍贷 2015 年 3 月-5 月贷款违约情况图

对指标的整体数据进行初步分析，其中拍拍贷的平均借款金额为 3092.25 元，借款金额比较小，拍拍贷的平均利率为 0.13，其中最大值为 0.24，最小值为 0.07，拍拍贷的平均借款期限为 8.3 个月，借款期限比较短。其中数据样本中借款人男性的频数为 9628 人，占总样本的 85.70%，女性的频数为 1606 人，占总样本的 14.30%，借款人数中男性远远多于女性。我们对数据中借款人的个人特征中的年龄与性别进行了初步分析，结果整理如下表 3：

表 3 拍拍贷款 2015 年 3 月-5 月个人特征数据

个人特征		样本数	违约	占样本比率	占违约比率
年龄	20-25岁	4596	57	1.24%	45.24%
	26-31岁	3946	42	1.06%	33.33%
	32-38岁	1909	20	1.05%	15.87%
	39岁以上	783	7	0.89%	5.56%
性别	男	11108	103	0.93%	81.75%
	女	126	23	18.25%	18.25%
总计		11234	126		

在性别方面，具有信用风险的借款人主要以男性为主，无论是占样本比率还是占违约比率，男性借款人违约概率都要远远高于女性,在借款者中，男性的比例均占 80%以上，即表示男性的违约率相对更高一些。而在年龄方面，拍拍贷借款用户的年龄在 20-25、26-31 岁两个阶段的违约人数占总人数的比例高于 80%，明显高于其他年龄段，而且随着年龄的增加，违约概率呈明显降低趋势。

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

本文对于平台里所有借款人用户在网站中的信用等级进行了划分，级别分为 AAA、AA、A、B、C、D、E、F 这八个等级，风险依次上升，在实证过程，把它们分别用 8、7、6、5、4、3、2、1 来代替。对拍拍贷 2015 年 3 月到 5 月用户偿还贷款情况的 11314 个数据的信用等级进行分析，得到结果：其中信用等级为 AA 的借款用户数最多，频数为 7993，占所以借款人数的 70.647%，等级为 B、C、D、E、F 等级的用户占总借款用户比较低，这可能是因为用户信用等级较低，给投资者带来的安全感不高，且本身的实名认证分较低，所以借到的资金有限。在拍拍贷中，用户违约风险随信用等级的提高而下降，即信用等级越高的用户，违约风险就相对越低，这也可以表示，拍拍贷网站中的信用评级，在一定程度上，对防范信用风险起到了积极作用。数据具体整理如下面图 7 ：

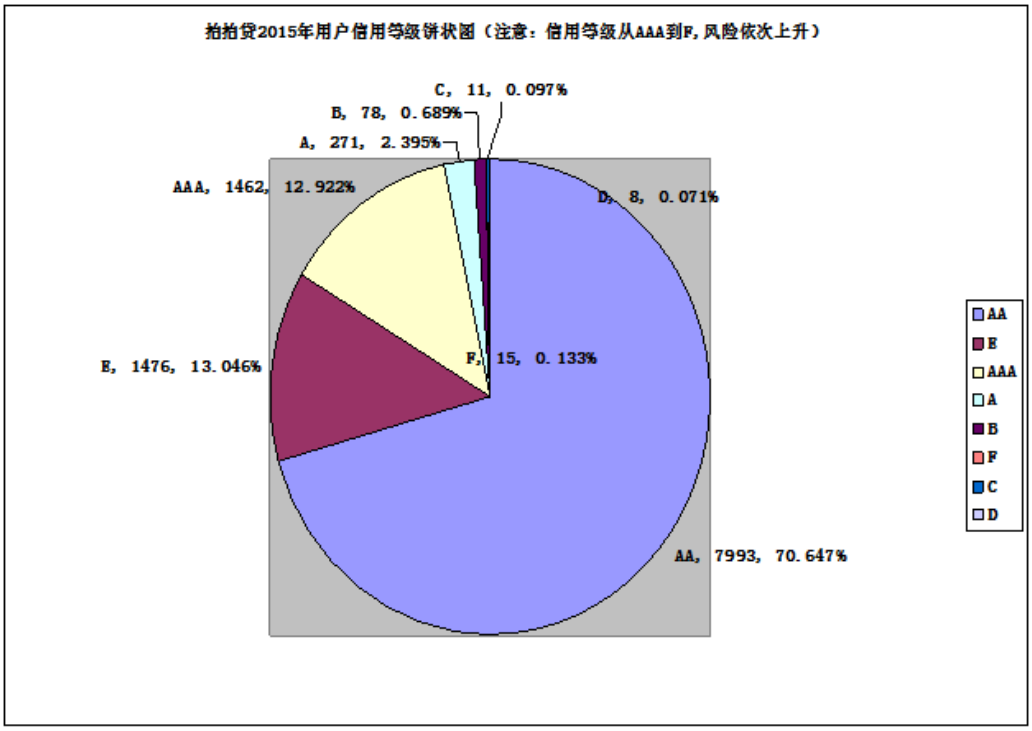

图 7 拍拍贷 2015 年 3 月-5 月用户信用等级饼状图

通过对 2015 年拍拍贷 3 月—5 月中，可用的 11234 个数据样本整理，发现其中逾期样本数据个数为 126 个，我们对这些数据进行描述性统计分析，我们发现：逾期未偿还贷款者中男性人数为 94 人，占所有逾期未偿还贷款者的 81.75%；而女性人数仅仅 22 人，占所有逾期未偿还贷款者的 18.25%。人男性违约的概率要远远高于女性。如下图 8 所示：

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

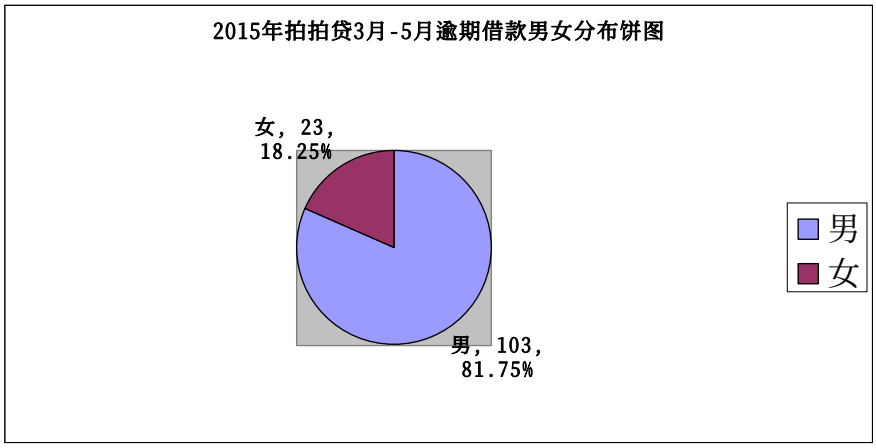

图 8 拍拍贷 2015 年 3 月-5 月逾期借款男女分布饼状图

对违约群体的年龄进行分析得到的结果有：其中 20-25 岁年龄段占逾期未偿还贷款者得 45.24%，比例最大，符合年轻群体收入不稳定的现状，随着年龄段的增长，违约人数也在不断下降，表明随着年龄的增加，人们的收入趋于稳定，所以违约的情形比例也在下降。实际数据如下面图 9：

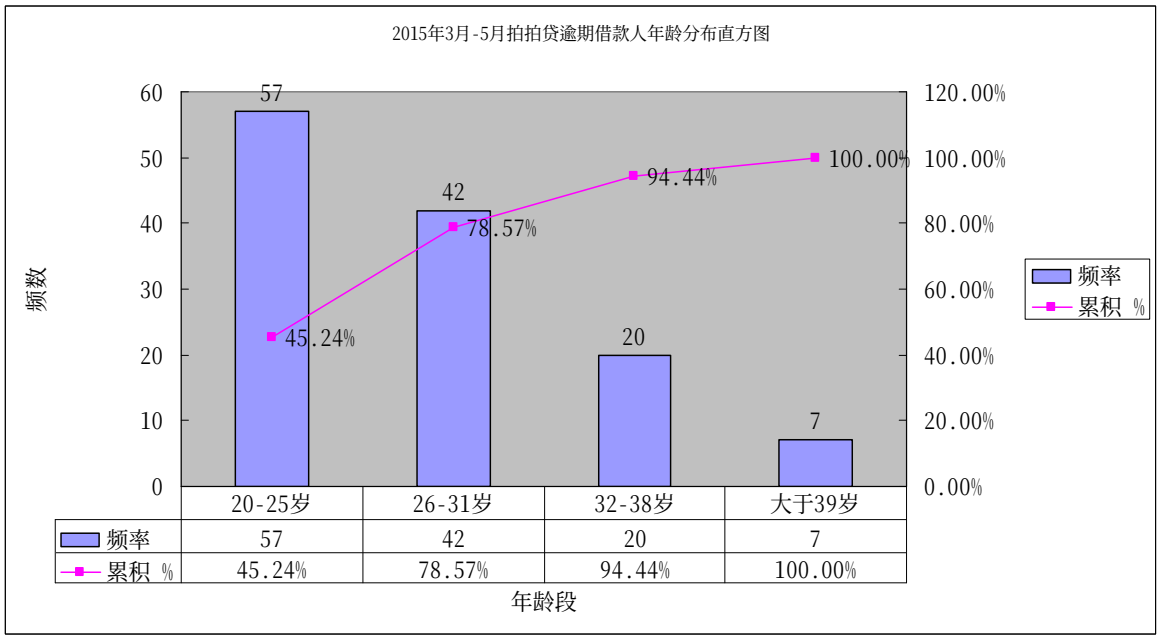

图 9 拍拍贷 2015 年 3 月-5 月逾期借款人年龄分布直方图

对违约人群的身份进行分析，得到以下结果：其中工薪族的违约者人数最多，频数为 57 人，占违约人数的 45.24%；其次是私营业主，频数是 30 人，占违约人数的 34.81%；

身份为网店卖家，其频数为 21 人，占违约人数的 16.67%；身份为学生，其频数为 13 人，占违约人数的 10.32%；其他，频数为 5 人，占违约人数的 3.96%。具

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

体数据整理如下图 10：

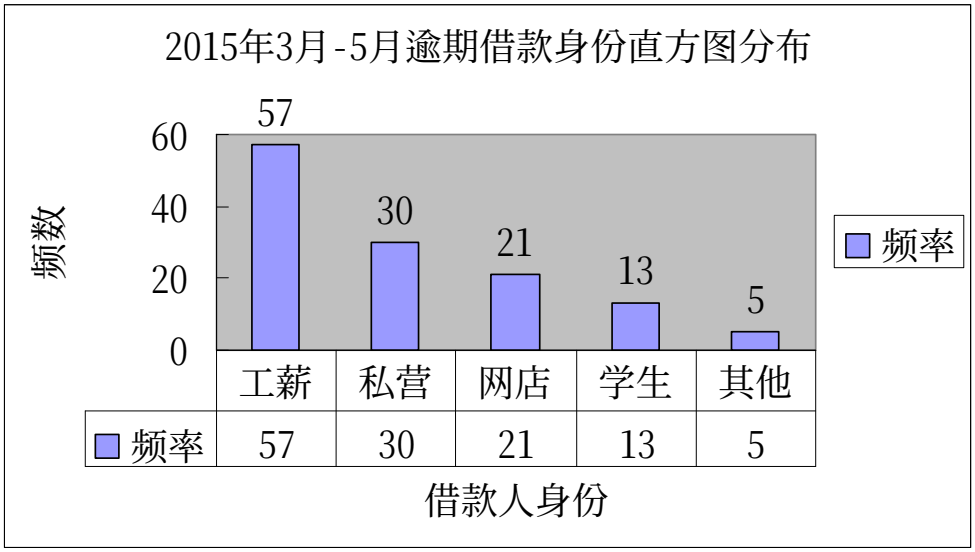

图 10 拍拍贷 2015 年 3 月-5 月逾期借款人身份分布直方图

四、P2P 网贷风险模型

4.1 Logit 模型构建

设 $X_1, X_2, \cdots, X_k$ 为一组自变量，Y 为应变量。当 Y 是不按时还款时，记为 Y=1；当 Y 是按时还款时，记为 Y=0。用 P 表示发生不按时还款的概率；用 Q 表示按时还款的概率，显然 $P+Q$=1。

Logit 回归模型为：

$$P=\frac{e^{\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k}}{1+e^{\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k}}$$

同时可以写成：

$$Q=\frac{1}{1+e^{\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k}}$$

式中 β_0 是常数项；$\beta_j(j=1,2,\cdots,\ k)$ 是与研究因素 X_j 有关的参数，称为偏回归系数。

事件发生的概率 P 与 βx 之间呈曲线关系，当 βx 在 $(-\infty,\ \infty)$ 之间变化时，P

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

或Q在（0，1）之间变化。

这样，第i个观察对象的发病概率比数（odds）为P_i/Q_i，第l个观察对象的发病概率比数为P_l/Q_l，而这两个观察对象的发病概率比数之比值便称为比数比OR（odds ratio）。对比数比取自然对数得到关系式：

$$\ln\left(\frac{P_i/Q_i}{P_l/Q_l}\right)=\beta_1(X_{i1}-X_{l1})+\beta_2(X_{i2}-X_{l2})+\cdots+\beta_k(X_{ik}-X_{lk})$$

等式左边是比数比的自然对数，等式右边的$(X_{ij}-X_{lj})(j=1,2,\cdots,\ k)$是同一因素$X_i$的不同暴露水平$X_{ij}$与$X_{lj}$之差。$\beta_j$的流行病学意义是在其它自变量固定不变的情况下，自变量X_j的暴露水平每改变一个测量单位时所引起的比数比的自然对数改变量。同多元线性回归一样，在比较暴露因素对反应变量相对贡献的大小时，由于各自变量的取值单位不同，也不能用偏回归系数的大小作比较，而须用标准化偏回归系数来做比较。

由于 Logit 回归是一种概率模型，通常用最大似然估计法（maximum likelihood estimate）求解模型中参数β_j的估计值$b_j(j=1,2,\cdots,\ k)$。

Y为在$X_1,X_2,\cdots,X_k$作用下的还款情况发生的指示变量。其赋值为：

$$Y_i=\begin{cases}1,&\text{第}i\text{个观察不按时还款}\\0,&\text{第}i\text{个观察对象按时还款}\end{cases}$$

第i个观察对象对似然函数的贡献量为：

$$l_i=P_i^{Y_i}Q_i^{1-Y_i}$$

当各事件是独立发生时，则n个观察对象所构成的似然函数L是每个观察对象的似然函数贡献量的乘积，即

$$L=\prod_{i=1}^{n}l_i=\prod_{i=1}^{n}P_i^{Y_i}Q_i^{1-Y_i}$$

式中$\prod$为i从1到n的连乘积。

依最大似然估计法的原理，使得L达到最大时的参数值即为所求的参数估计值，计算时通常是将该似然函数取自然对数（称为对数似然函数）后，用 Newton—Raphson 迭代算法求解参数估计值$b_j\ (j=1,2,\cdots,\ k)$。

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

4.2 P2P 个人风险识别模型

由于我们的因变量是一个离散的 0—1 变量，因此传统的散点图无法有效地表示同各个解释性变量的相关关系。而此时盒状图则特别有效，我们分别对各个解释变量分析如下：

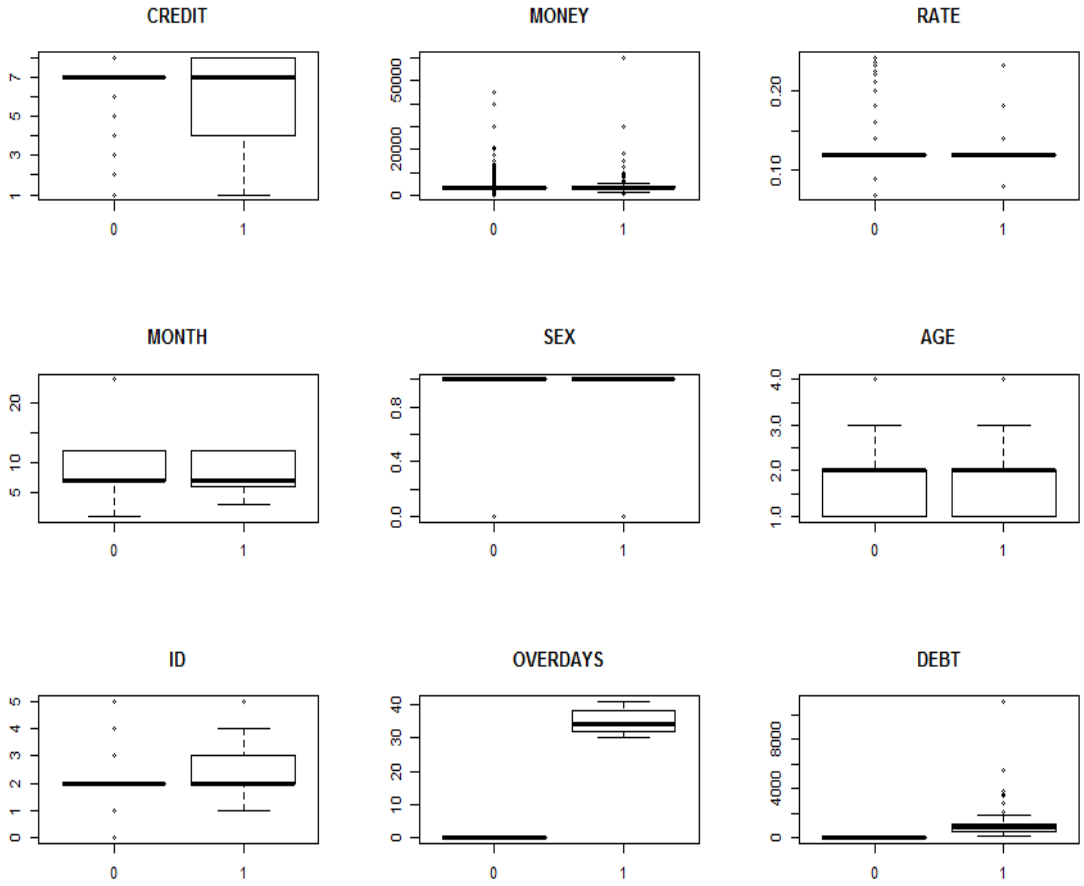

图 11 各自变量与因变量逾期且还款相关关系盒状图

从图我们可以得到以下重要结论：

①信用等级对借款人是否逾期还款有明显影响。

②借款金额对借款人是否逾期还款可能有影响。

③利率对借款人是否逾期还款可能有影响。

④借款期限对借款人是否逾期还款影响不大。

⑤性别对借款人是否逾期还款无影响。

⑥年龄对借款人是否逾期还款无影响。

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

⑦用户身份对借款人是否逾期还款有明显影响。

⑧逾期天数对借款人是否逾期还款有明显影响。

⑨逾期本息是否逾期还款有明显影响。

以上都是对数据进行初步的描述性分析。对于所得到的结论：第一，没有控制其他因素的影响；第二，没有经过严格的统计检验。而这些问题将是我们下面研究的主要内容。

特别说明：** 逾期天数（overdays）和逾期本息（debt）与逾期且还款 RT 高度相关，其 z 检验在 0.01 的显著性水平下高度显著（p=0.000），其系数为负，说明没有可能逾期，结合实际情况，拍拍贷规定，在逾期黑名单的借款者需要把逾期本息还清才有机会申请借款否则没有机会进行下次借款。，鉴于此，拍拍贷的硬性规定，虽然高度相关，AIC 准则也会通过，但不予考虑建模。

通过 spss 分析软件一步步得出结论如下：

表4各自变量与因变量rt相关关系检验图

Variables not in the Equation[a]

			Score	df	Sig.
Step 0	Variables	credit	80.625	1	.000
		money	71.879	1	.000
		rate	28.248	1	.000
		month	.208	1	.648
		sex	2.066	1	.151
		age	1.187	1	.276
		id	36.039	1	.000

a. Residual Chi-Squares are not computed because of redundancies.

将 7 个自变量全部放入模型后的得分检验结果，检验某一自变量与因变量有无关系。由该结果可见，可初步认为存在 0.05 的显著性水平下，自变量 credit，money，rate，id 与因变量 rt 有统计学意义；month，sex，age 与因变量 rt 之间的

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

联系无统计学意义。

Correlation Matrix

		Constant	credit	money	rate	month	sex	age	id
Step 1	Constant	1.000	-.118	.594	-.938	-.070	-.141	-.025	.078
	credit	-.118	1.000	.047	-.026	.070	-.064	.044	.063
	money	.594	.047	1.000	-.658	.081	-.022	.050	.152
	rate	-.938	-.026	-.658	1.000	-.137	.013	-.119	-.265
	month	-.070	.070	.081	-.137	1.000	.039	.069	.066
	sex	-.141	-.064	-.022	.013	.039	1.000	.006	.030
	age	-.025	.044	.050	-.119	.069	.006	1.000	.010
	id	.078	.063	.152	-.265	.066	.030	.010	1.000

表 5 相关系数矩阵

由相关系数矩阵表可以看出，七个变量相关系数 r 最大是-0.658，借款金额和利率的之间有一定关系，但关系不大。说明变量内部不存在线性关系。

表 6 自变量参数估计和假设检验

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	credit	-.232	.039	34.628	1	.000	.793	.734	.856
	money	.000	.000	99.452	1	.000	1.000	1.000	1.000

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

rate	-167.619	14.543	132.849	1	.000	.000	.000	.000
month	.012	.034	.120	1	.729	1.012	.946	1.082
sex	-.291	.251	1.342	1	.247	.748	.457	1.223
age	-.070	.111	.397	1	.529	.932	.750	1.159
id	.678	.101	45.180	1	.000	1.970	1.617	2.401
Constant	14.799	1.634	82.017	1	.000	2.673E6		

a. Variable(s) entered on step 1: credit, money, rate, month, sex, age, id.

这可以看成 logit 全模型，对 7 个变量进行检验，第一列 B 值是系数估计值，选取 Wald 检验，通过显著性检验的变量有 CREDIT,MONEY,R,RATE,ID 说明这四个变量对它有显著影响，其他 MONTH,SEX,AGE 对它无影响。

* 借款者信用等级和逾期且还款高度相关，其 z 检验在 0.01 的显著性水平下高度显著（p=0.014），并且可以看到，CREDIT 的系数估计值的符号为负，说明信用等级越高，逾期的风险越低。

* 借款金额和利率与逾期且还款 RT 高度相关，借款金额的系数非常小，原因在于，此次爬取的数据来自于普通应收安全标区，并没有涉及逾期就陪区，中风险收益区和高风险收益区，基本属于小额借贷，在 3000 元左右，所以导致结果是对逾期且还款不产生多大影响，这是此次统计建模的缺陷所在，由于时间关系，没有对其他标的区进行建模，但方法是相同的，给予不同区的数据进行建模，最后可以预测不同区的逾期的概率。说明借款金额越多，利率越高，越有可能不逾期，因为利率高，利息就高，那么逾期所还的利息更多，相对而言，逾期的概率就减少。

* 借款者身份与逾期且还款有一定关系，不同的身份，他的信用是不一样的。

* 没有证据证明其他指标对预测逾期且还款（RT）与否有重要作用 。

根据表 6 建立线性组合如下：

F=14.799-0.232*CREDIT+0*MONEY-167.619*RATE +0.678*ID

假设有这样一个借款者，他的各项指标分别为：CREDIT=7，MONEY=3000，RATE=0.12， ID=1，计算得到 F=-5.56956 ，然后再计算逾期的概率为：

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

$$p\{RT=1\}=\exp(-5.56956)/(1+\exp(-5.56956)=0.38\%$$

表7 模型预测精度图

Classification Table[a]

Observed			Predicted		
			rt		Percentage Correct
			0	1	
Step 1	rt	0	9873	4	100.0
		1	100	21	17.4
		Overall Percentage			99.0

a. The cut value is .500

最终模型正确预测概率达到99%, 说明建立的模型拟合效果较好。成功还款的9877个样本中有4个被错误的划为逾期未还款, 逾期未还款121个有效样本中有100个被划为成功还款。

4.3 数据验证

建立模型的一个重要目的就是预测, 我们选取 100 个预测样本预测结果如下表 9：

表 9 100 个数据样本验证表格

NAME	CREDIT	MONEY	RATE	ID	RT	F 值	p 概率（a=0.005）	预测值
2207259	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207257	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207256	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207254	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207251	8	1000	0.23	1	0	-24.93137	1.48745E-11	0

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

2207250	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207249	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207248	8	3000	0.12	2	0	-5.81528	0.002972782	0
2207247	7	3000	0.2	2	0	-18.9928	5.64328E-09	0
2207246	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207245	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207244	7	3000	0.12	3	0	-4.90528	0.007352903	1
2207243	7	3000	0.12	3	0	-4.90528	0.007352903	1
2207241	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207240	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207236	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207234	7	3000	0.12	4	0	-4.22728	0.014382162	1
2207233	7	3000	0.22	2	0	-22.34518	1.9752E-10	0
2207232	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207231	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207230	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207229	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207228	7	3000	0.22	2	0	-22.34518	1.9752E-10	0
2207227	8	3000	0.12	2	0	-5.81528	0.002972782	0
2207226	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207224	8	3000	0.12	2	0	-5.81528	0.002972782	0
2207223	8	3000	0.12	2	0	-5.81528	0.002972782	0
2207222	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207221	8	3000	0.12	2	0	-5.81528	0.002972782	0

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

2207219	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207218	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207217	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207216	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207213	7	3000	0.12	4	0	-4.22728	0.014382162	1
2207212	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207211	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207210	7	3000	0.12	4	0	-4.22728	0.014382162	1
2207209	2	4088	0.12	2	0	-4.42328	0.011852654	1
2207208	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207206	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207205	7	3000	0.12	3	0	-4.90528	0.007352903	1
2207204	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207203	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207202	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207201	8	3000	0.12	2	0	-5.81528	0.002972782	0
2207200	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207199	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207198	7	3000	0.12	4	0	-4.22728	0.014382162	1
2207197	8	3000	0.12	2	0	-5.81528	0.002972782	0
2207196	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207195	7	3000	0.12	1	0	-6.26128	0.001905164	0
2207194	8	3000	0.22	4	0	-21.22118	6.07797E-10	0
2207193	7	3000	0.12	2	0	-5.58328	0.003746126	0

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

2207191	7	3000	0.12	3	0	-4.90528	0.007352903	1
2207190	8	3000	0.12	4	0	-4.45928	0.011438342	1
2207189	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207187	8	3000	0.12	3	0	-5.13728	0.005839346	1
2207186	7	3000	0.12	4	0	-4.22728	0.014382162	1
2207185	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207182	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207181	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207178	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207177	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207175	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207173	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207172	7	3000	0.12	3	0	-4.90528	0.007352903	1
2207170	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207169	8	3000	0.12	2	0	-5.81528	0.002972782	0
2207168	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207167	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207166	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207165	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207164	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207163	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207162	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207159	7	3000	0.12	3	0	-4.90528	0.007352903	1
2207152	7	3000	0.22	4	0	-20.98918	7.66505E-10	0

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

2207149	8	3000	0.12	2	0	-5.81528	0.002972782	0
2207148	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207146	7	1000	0.22	2	0	-22.34518	1.9752E-10	0
2207143	7	3000	0.12	4	0	-4.22728	0.014382162	1
2207142	7	1000	0.23	1	0	-24.69937	1.87586E-11	0
2207141	8	3000	0.12	3	0	-5.13728	0.005839346	1
2207139	8	3000	0.12	2	0	-5.81528	0.002972782	0
2207138	5	6664	0.18	2	0	-15.17642	2.56427E-07	0
2207137	8	3000	0.22	2	0	-22.57718	1.56623E-10	0
2207136	8	3000	0.12	3	0	-5.13728	0.005839346	1
2207133	7	3000	0.12	4	0	-4.22728	0.014382162	1
2207132	7	3000	0.12	2	0	-5.58328	0.003746126	0
2207130	7	3000	0.12	2	0	-5.58328	0.003746126	0
2182891	7	3000	0.12	2	1	-5.58328	0.003746126	1
2189854	7	3000	0.12	2	1	-5.58328	0.003746126	1
2198501	7	3000	0.12	3	1	-4.90528	0.007352903	0
2167762	1	3000	0.12	2	1	-4.19128	0.014901496	0
2176293	8	3000	0.12	5	1	-3.78128	0.022285532	0
2151068	8	3000	0.12	2	1	-5.81528	0.002972782	1
1958998	8	3000	0.12	5	1	-3.78128	0.022285532	0
2185851	8	3000	0.12	2	1	-5.81528	0.002972782	1
1822196	8	3000	0.12	3	1	-5.13728	0.005839346	0
1840644	8	3000	0.12	2	1	-5.81528	0.002972782	1

具体来说, 我们利用单独的检验数据来对借款者未来的逾期且还款进行预测,

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

并用对应的的预测结果来衡量模型的预测精度。对此类离散变量的预测，我们一般用一下两个指标间接的度量预测精度：

*True Positive Rate（TPR：把(真实的借款者逾期还款)正确地预测为 RT=1 的概率。

*False Positive Rate(FPR)：把(真实的借款者还款成功)错误地预测为 RT=1 的概率

以 0.005 作为阈值，TPF 值为 5/（5+5）=50%，而 FPR 值为 18/（72+18）=20%，同样，以 0.01,0.05 作为阈值，得到不同的预测概率。

五、结论与建议

5.1 结论

1.经过对拍拍贷 2015 年 3 月到 5 月用户偿还贷款情况的 11234 个可用数据样本的描述分析，我们知道借款用户的违约率为 1.12%，风险率较小，主要原因是，拍拍贷的借款金额较少，借款期限较短，采用魔镜等级，依靠实名认证。将用户级别分为 AAA、AA、A、B、C、D、E、F 这八个等级，风险依次上升，信用等级的使用，在一定程度上，对防范信用风险起到了积极作用。借款者信用等级和逾期且还款高度相关，其 z 检验在 0.01 的显著性水平下高度显著（p=0.000），并且可以看到，CREDIT 的系数估计值的符号为负，说明信用等级越高，逾期的风险越低。

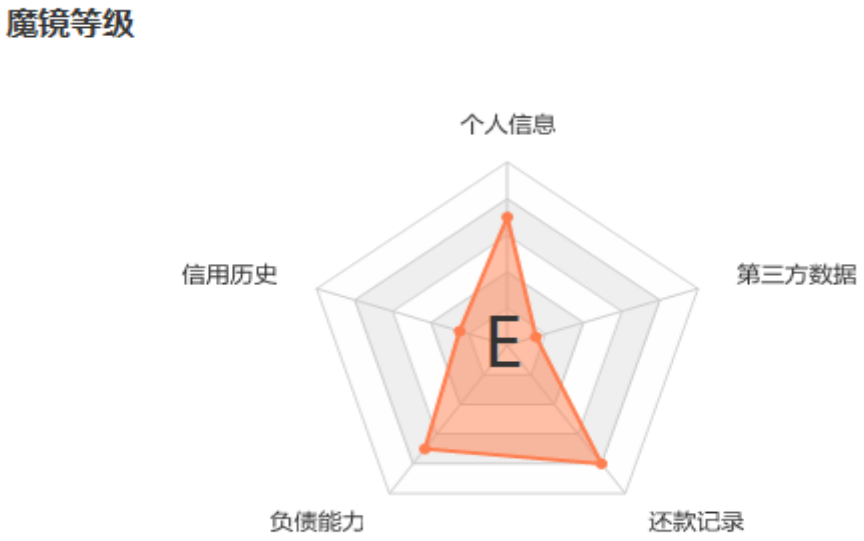

图 12 魔镜等级图

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

2.在拍拍贷 2015 年 3 月到 5 月用户偿还贷款情况的 11234 个可用数据样本中，借款人的性别中，男性远远多于女性；拍拍贷借款用户的年龄在 20-25、26-31 岁两个阶段的违约人数占总人数的比例高于 80%，明显高于其他年龄段，而且随着年龄的增加，违约概率呈明显降低趋势。

3.对拍拍贷 2015 年 3 月到 5 月用户偿还贷款情况的 126 个逾期样本数据作描述性分析，对逾期未偿还贷款者中的性别分析得到：男性违约的概率要远远高于女性。对逾期未偿还贷款者中的年龄分析得到：20-25 岁年龄段占逾期未偿还贷款者得 45.24%，比例最大，符合年轻群体收入不稳定的现状，随着年龄段的增长，违约人数也在不断下降，表明随着年龄的增加，人们的收入趋于稳定。对违约人群的身份进行分析，得到以下结果：其中工薪族的违约者人数最多。

4．借款金额和利率与逾期且还款 RT 高度相关，借款金额的系数非常小，原因在于，此次爬取的数据来自于普通应收安全标区，并没有涉及逾期就陪区，中风险收益区和高风险收益区，基本属于小额借贷，在 3000 元左右，所以导致结果是对逾期且还款不产生多大影响，这是此次统计建模的缺陷所在，由于时间关系，没有对其他标的区进行建模，但方法是相同的，给予不同区的数据进行建模，最后可以预测不同区的逾期的概率。说明借款金额越多，利率越高，越有可能不逾期，因为利率高，利息就高，那么逾期所还的利息更多，相对而言，逾期的概率就减少。

5.2 建议

1.健全行业信息共享与披露机制

目前阻碍我国 P2P 网络贷款行业进一步发展瓶颈在于信息披露不完全，尽管各网络贷款平台在各自网站发布网络贷款逾期贷款者“黑名单”，但各网站所公布的信息缺乏共享性，导致借贷双方信息高度不对称，某一个借款者可以利用相同的资料在不同贷款平台获取贷款。这增加了 P2P 网络贷款行业信用风险。网络贷款平台仅采用公布黑名单的方式并不能有效的敦促借款者归还贷款，各网络贷款平台虽然有相应措施进行贷款追讨，但追讨难度大、成本高、效率低。总之，较低的贷款违约成本和借贷信息相对封闭导致 P2P 网络贷款平台所面临的信用风险较高。

因此，建立完善的行业信息共享披露机制是我国 P2P 网络贷款行业获得长足发展的重要前提。各网络贷款平台应积极推进贷款违约信息共享，例如可以建立行业信息服务数据库，提高借款者的违约成本，增强对借款者的约束力度，并可以集中力量提升对借款违约者的惩戒力度，使 P2P 网络贷款平台在阳光下运

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

行，增强整个行业的公信度，进而降低 P2P 网络贷款信用风险，提高整个行业的运行效率。

2.明确监管主体

我国 P2P 网络贷款行业也缺乏明确的监管主体，存在监管主体缺位和在某些领域监管职能重叠的问题，这使得 P2P 网络贷款在某些方面受到了众多主体的多重监管，而又存在“监管真空”。当然这一问题的解决关键在于 P2P 网络贷款业务的法律定位。因此尽快明确 P2P 网络贷款的法律地位，进而明确监管主体，才能使 P2P 网络贷款在外部监督下健康成长，从而降低信用风险。

3.推动信用评级行业发展

我国目前信用评级行业整体发展滞后，市场规模小，经营分散，缺乏行业自律，推动信用评级行业有序健康发展对于完善我国信用体制起到了重要促进作用。对于 P2P 网络贷款而言，应积极推进对银行外信用信息的利用，推动各部门信息依法公开，进一步提升企业和个人征信系统的数据挖掘、整理、分析和服务功能。

六 参考文献

[1]何晓玲,王玫.P2P 网络借贷现状及风险防范[J].中国商贸,2013,(20).

[2]吴晓光,曹一.论加强 P2P 网络借贷平台的监管[J].南方金融，2011 年 4 期.

[3]张初兵,高康,杨贵军.判别分析与 Logistic 回归的模拟比较[J],统计与信息论坛,2010,(01).

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net