

语言可信吗? 借贷市场上语言的作用

——来自 P2P 平台的证据

廖理¹, 吉霖², 张伟强¹

(1. 清华大学 五道口金融学院, 北京 100083; 2. 清华大学 经济管理学院, 北京 100084)

摘要: P2P 新兴市场为研究借款申请人所用语言提供了契机。该文从语言长度与语言内容两个维度, 以融资结果与还款行为两个层面, 试图较为全面、较为立体地解读借贷市场上语言的作用。实证结果表明: 语言长度越长, 借款成功率越高, 如约还清率越高, 语言长度特征表现出一致性; 涉及某些话题(创业、家庭、急迫、诚信), 影响借款成功率, 但与如约还清率无关, 语言内容特征表现出迷惑性。这意味着投资人在通过语言长度识别信用风险方面较为成熟, 但对语言内容的认识存在偏差。

关键词: 信用识别; 语言; Peer-to-Peer 借贷

中图分类号: O 212.1; F 832.479 文献标志码: A
文章编号: 1000-0054(2015)04-0413-09
DOI:10.16511/j.cnki.qhdxxb.2015.04.008

Role of narratives in lending market: Evidence from P2P lending platform

LIAO Li¹, JI Lin², ZHANG Weiqiang¹

(1. The PBC School of Finance, Tsinghua University,
Beijing 100083, China;

2. School of Economics and Management, Tsinghua University,
Beijing 100084, China)

Abstract: The growth of P2P industry provides new opportunities for research on the role of borrowers' narratives in lending market. This paper analyzes the narratives in dimensions of the length and content with funding and repaying considered. Empirical results show that longer content is associated with easier funding and better repaying, but the content mentioning certain topics, such as entrepreneur, family, hurry, and trustworthy, affects funding while having nothing to do with repaying. Therefore, in lending market, the role of the narrative length performs consistent but the narrative content is confusing. Lenders have better ability to identify borrowers' credit risk via the narrative length, but are biased via the narrative content.

Key words: credit identification; narrative; P2P lending

语言是日常生活中不可或缺的信息沟通方式, 也是商业行为中举足轻重的客户说服手段。在借贷市场中, 借款申请人可能运用语言, 向投资人传达信息。除借款人的客观条件之外, 借款人的文字描

述可能会对投资人的投资决策起到至关重要的作用。在传统的以银行为媒介的借贷市场上, 借款申请人提供的语言信息难以收集与量化, 为学者进行严谨的实证研究设置了障碍。然而, P2P 这一新兴借贷市场的出现, 为研究借贷市场上语言的作用提供了难得的契机。

P2P 借贷 (Peer-to-Peer lending, Person-to-Person lending) 是互联网金融的典型模式。与传统的以银行等金融机构为中心的借贷模式不同, P2P 运营商为完全陌生的借款申请人与投资人提供低成本的网络平台, 方便借贷双方实现资金的有效配置。在欧美国家, 自 2005 年以来, 以 Prosper、Lending Club 为首的 P2P 企业逐渐兴起, 迅速发展。2007 年左右, P2P 借贷引入我国。2010 年以来, 在利率市场化逐步推进、民间借贷屡遭考验的背景下, P2P 行业迎来了高速增长时期。随着大量 P2P 平台的涌现, 一些劣质企业悄然潜入。近年来, 多家平台的倒闭给我国正常的金融秩序带来了不利影响。由于 P2P 借贷主要针对违约风险相对较高的个人和小微企业, 因此, 信用风险的识别自然成为了第一要务。

借款申请人在 P2P 平台申请借款时, 除需要填写一些较为客观的项目(如借款金额、借款期限等)之外, 可选择填写文字描述一栏。该项目较为主观, 可依据借款人的主观意愿, 介绍更多的个人信息或本次借款信息, 意图体现出本借款的与众不同之处, 吸引投资人作出投资决策。因此, 借款申请人所使用的语言描述可能是投资人判断借款申请人

收稿日期: 2014-06-29

基金项目: 国家自然科学基金资助项目

(71232003, 71271214, 71441020);

北京市社会科学基金资助项目(14JGA018)

作者简介: 廖理(1966—), 男(汉), 安徽, 教授。

E-mail: liaol@pbcfs.tsinghua.edu.cn

信用风险的重要考量。而这些为实证研究借贷市场中语言的作用提供了必要的条件。

本文采集我国 P2P 行业的典型代表之一人人贷借贷平台(www.renrendai.com)的数据样本,选取 2012 年 3 月—2013 年 9 月的 47 424 个信用认证标作为研究主体。实证结果表明,在语言长度方面,语言长度越长,受到投资人的青睐越多,更容易借款成功;同时,语言长度越长,表征借款人信用越好,更可能如约还清。两相对比,说明投资人能够通过语言长度,准确识别借款申请人的信用风险。在语言内容方面,提及创业、家庭等更容易获得投资人的青睐,借款成功率更高;提及急迫更容易引起投资人的担忧,借款成功率较低。有趣的是,是否涉及这些话题对借款申请人是否会如约还清的影响均不显著。这意味着,比起语言长度,语言内容更具有迷惑性。对借款申请人而言,是否谈及某些话题更是一种出于营销策略的技巧选择,而不是基于自身风险的如实表征。投资人在通过语言内容特征识别借款申请人信用风险方面存在明显偏差。上述结论得到了各项稳健性检验的支持。

作者从语言长度和语言内容两个维度,挖掘借款申请人的语言特征;从投资人投资行为和借款申请人还款行为两个层面,探讨上述语言特征的作用,从而形成对借贷市场中语言作用的较为全面、较为立体的解读。探索中国借贷市场的自身规律,尤其是中国新兴的互联网背景下的借贷市场的特点,希望能为我国金融市场的完善提供更有针对性的借鉴,对现有研究中该领域的国际经验进行有益的补充。

下文的结构安排如下:第 1 部分总结现有相关文献;第 2 部分描述样本选择与变量情况;第 3 部分报告基准回归与稳健性检验的实证结果;第 4 部分总结。

1 文献综述

语言是日常生活中不可或缺的信息沟通方式,也是商业行为中举足轻重的客户说服手段。已有文献多在市场营销、消费行为、公司金融等领域内探讨语言的影响效果与应用技巧,如 Thompson^[1]、Levy^[2]和 Martens 等^[3]的研究。然而在借贷市场上实证检验语言作用的文章却相对较少。

在借贷市场中,借款申请人可能运用语言,向投资人传达信息。除借款人的客观条件之外,借款人的文字描述可能会对投资人的投资决策起到至关

重要的作用。在传统的以银行为媒介的借贷市场上,借款申请人提供的语言信息难以收集与量化,为学者进行严谨的实证研究设置了障碍。然而,P2P 这一新兴借贷市场的出现,为研究借贷市场上语言的作用提供了难得的契机。

在现有的 P2P 相关文献中,核心问题集中在对借款申请人信用风险识别的研究。Michels^[4]验证了借款申请人自愿披露且无法证实的信息的影响。结果显示,此类信息会降低融资成本,增加融资成功概率。一些人口特征,如性别、种族、婚姻状况、甚至相貌,可能透露出借款申请人信用风险信息(可参见 Barasinska 和 Schäfer^[5]、Herzenstein 等^[6]、Pope 和 Sydnor^[7]以及 Duarte 等^[8]的研究结果)。一些行为特征也可能揭示信用风险,例如,添加好友、加入群组或羊群行为等(参见 Berger 和 Gleisner^[9]、Lin、Prabhala 和 Viswannathan^[10]、Everett^[11]、Liu、Lu 和 Brass^[12]以及李悦雷等^[13]的研究工作)。

此外,借款申请人所使用的语言描述可能是投资人判断借款申请人信用风险的重要考量,这一问题得到了一些学者的关注。这些研究也将是本文的最主要借鉴。

Herzenstein、Sonenshein 和 Dholakia^[14]挑选了美国 Prosper 平台中 1 493 个借款申请构建样本,人工判断借款申请人的语言内容中是否涉及可靠、经济困难、勤奋、成功、品德高尚、有宗教信仰等 6 大方面。结果表明,借款申请人的信用级别越低,语言内容涉及上述内容的种类越多;控制其他变量后,语言内容涉及上述 6 类的种类越多,投资人更可能向其投资,但借款申请人如期偿付的概率反而越低。他们的研究为后续探讨语言作用提供了很多有益借鉴。但也存在一些明显不足:第一,样本规模较小,很难避免抽样过程中所引入的选择偏差;第二,人工判断的方式失于主观,很难将人工判断统一和量化;第三,语言内容中涉及的话题不易区分,某些话题间高度相关;第四,由于 6 大话题对于融资结果的影响方向可能并不相同,以涉及话题的种类总数作为关键变量可能会使所得实证结果难以解释,失去内涵。

Larrimore 等^[15]利用美国 Prosper 平台数据,通过标准的语言分析软件,解构出借款申请人的某些语言特征,探讨了这些语言特征对借款成功率的影响。结果显示,语言长度、具体描述等可能会提高融资成功率。私人问题细节和对现有财务状况的

辩解,会降低借款的融资成功率。他们的工作通过软件分析的手段,克服了人工分析中样本规模的局限,但是由于英语语言特征的影响,通过软件提取的某些语言特征略显牵强。例如,文中将冠词(a、an、the)的使用数量解读为描述是否具体,用连接词(because、hence等)使用数量表征借款申请人是否为了自身缺陷而辩护。此外,他们仅考虑了语言特征对融资成功率的影响,并未探讨语言特征对如约还款率的影响,因此无法从全局角度研究语言的真实角色。

在现有文献的启发下,本文将深入探讨借贷市场中语言的作用。相对而言,本文试图从以下几个角度完善创新:

第一,本文将基于大数据大样本进行检验,提高了结果的代表性和可信度;

第二,本文将从语言长度和语言内容两个维度,挖掘借款申请人的语言特征,从投资人投资行为和借款申请人还款行为两个侧面,探讨上述语言特征的作用,从而形成对借贷市场中语言作用的较为全面、较为立体的解读;

第三,有赖于中文词语更高的信息承载效率,在中文环境中,利用软件编程的手段,辨别语言内容中所涉及的话题,精度更高,划分更合理;

第四,探索中国借贷市场的自身规律,尤其是中国新兴的互联网背景下的借贷市场的特点,将会为我国金融市场的完善提供更有针对性的借鉴,也将对现有研究中该领域的国际经验进行有益的补充。

2 数据描述

本文的数据采集于人人贷借贷平台(www.renrendai.com)。人人贷是我国P2P行业的典型代表之一,于2010年5月成立,2010年10月正式上线,累计为30余省份的几十万客户提供网上借贷服务,近年来发展迅速。

为避免网站成立初期测试性数据的干扰,并且考虑到数据可得性与一致性,本文选取的样本时间为2012年3月至2013年9月。由于人人贷所设立的实地认证标、机构担保标、智能理财标等类别与P2P的典型形式信用认证标在担保方式、信息核实方式上有明显不同,因此本文仅将信用认证标作为研究主体。在删除明显不合理的极个别观测后,最终本文的样本中包含47424个投标记录,如表1所示。其中43934个投标未能成功融资,在成功借款

的投标中,最终291个借款未能如约还清借款($291=70+221$)。在稳健性检验中,本文将尝试使用某些子样本,进一步为基本结论提供支持证据。

表1 描述性统计:借款状态

变量	频数	百分比/%
借款失败	43 934	92.64
还付中	1 059	2.23
已还清	2 140	4.51
逾期	70	0.15
已由人人贷垫付	221	0.47

1) 被解释变量。

本文的回归部分存在两个主要的被解释变量:衡量投资人认可程度的哑变量成功借款;衡量借款人真实行为的哑变量如约还清。两个被解释变量的含义及构造方法如下。

构造的哑变量成功借款,衡量发标是否融资成功。若目前人人贷显示的借款状态为借款失败,则哑变量成功借款取值为0;若目前人人贷显示的借款状态为还付中、已还清、逾期或已由人人贷垫付(对于一些逾期时间过久的借款,人人贷平台会进行垫付),则哑变量成功借款取值为1。是否向借款人投资,是各位投资者经过分析各种信息后作出的最终决策,反映了投资人群体对于某借款人信用的识别和认可程度。样本中,如表1所示,发标借款成功率约为7.36%($7.36\% \approx 1-43934/47424$)。

构造的哑变量如约还清,衡量成功融资且已到期的借款是否如约还清。若目前人人贷显示的借款状态为已还清,则哑变量如约还清取值为1;若借款状态为逾期或已由人人贷垫付,则该哑变量取值为0。借款是否如约还清,是借款人在最终行为表现,衡量出借款人真实的信用特征。样本中,最终如约还清率约为88.03%($88.03\% \approx 2140/(2140+291)$),如表1所示。

这里需要指出的是,投资者的融资决定和借款人的最终表现,并不一定是一致的,即投资者可能存在行为和认知偏差。而这也正是本文采用两个被解释变量并进行比较分析的原因。作者希望通过两组回归的比较,来揭示借款人可能存在的迷惑行为以及投资人可能存在的认知偏差。

2) 关键变量。

本文的关键变量,即最感兴趣的变量,是与语言相关的变量。借款人在人人贷平台申请借款时,除一些较为客观的项目(如借款金额、借款期限等)

外,可填写文字描述一栏。该项目较为主观,可依据借款人的意愿,体现出本借款的不同之处。因此,文字描述是借款人传达信息的重要窗口,是投资者作出决策的珍贵依据。而本文将从两个侧面挖掘 P2P 平台上语言的价值:语言长度与语言内容。这两种语言特征可能扮演着截然不同的角色。

语言长度方面,在进行初步的有效性处理(删除一些明显无效的字符)后,99.6%的借款申请中包含文字描述(文字描述的字数大于 0)。如表 2 所示,文字描述字数的均值为 71 个中文字符,中位数为 60 个字,标准差为 42 个字,最大值为 491 个字,最小值为 0 个字。文字描述长度是下文回归中的重要考量方面,在基准回归中,本文采用了文字描述字数(百字)作为关键变量。为了检验结果的稳健性,本文去除了文字描述中英文字符、标点符号等,仅保留文字描述中的纯中文字符,计算纯中文字数后,构建指标文字描述中纯中文字数(百字)。文字描述中纯中文字数均值为 60 字,中位数为 52 字,标准差为 38 字,最大值为 454 字。

语言内容方面,根据借款申请人在借款申请时经常提及且容易区分的几类语言内容特征,作者挖掘了如下 4 种类别:创业、急迫、家庭、诚信。具体类别归属判定方法如下。若文字描述中提及“创业”、“店”、“淘宝”、“天猫”、“公司”或“买卖”,则视为该文字描述提及了创业,如表 2 所示,42%的文字描述表示本次借款是为自己开创的事业融资;若文字描述中提及“急”,则归于文字描述提及急迫的类别,该类别占比 22%;若文字描述中提及“家”、“父”、“母”、“爸”、“妈”、“妻”、“夫”、“孩”或“儿”,则认定该文字描述中提及家庭概念,该类别占比 43%;若文字描述中提及“信用”、“诚信”、“诚实”或“可靠”,则判断该文字描述中提及借款者本人重视诚信,该类别占比 19%。

表 2 描述性统计:语言特征相关变量

变量	均值	标准差
文字描述字数/100 字	0.71	0.42
文字描述中纯中文字数/100 字	0.60	0.38
文字描述是否提及创业(哑变量)	0.42	0.49
文字描述是否提及急迫(哑变量)	0.22	0.41
文字描述是否提及家庭(哑变量)	0.43	0.49
文字描述是否提及诚信(哑变量)	0.19	0.40

3) 控制变量。

影响被解释变量的因素是多种多样的。如果希

望厘清上述关键变量的作用,就应首先控制好其他变量的影响。本文的控制变量大致分为两类:借款情况变量(例如:借款金额、借款期限、借款利率等)和借款人情况变量(例如:性别、年龄、婚姻状况、收入状况、已通过人人贷认证的状况等)。此外,在稳健性检验中,还引入了省份、行业、借款用途等固定效应,作为补充证据。

表 3 就一些有关借款基本特征或借款人基本特征的变量进行了统计。从表中可以看出,在申请借款的样本中,借款金额的均值为 7.19 万元,借款期限的均值为 12.44 月,借款利率均值为年化 16.19%。借款人年龄均值为 29.70 岁,女性占比 14%,已婚人士占比 49%。

表 3 描述性统计:借款情况和借款人情况

变量	均值	标准差
借款金额/万	7.19	13.15
借款期限/月	12.44	7.70
借款利率/%	16.19	3.54
借款人年龄	29.70	7.04
性别(哑变量:女性)	0.14	0.35
婚姻状况(哑变量:已婚)	0.49	0.50

人人贷网站设置了 7 个区间供借款人描述自己的收入情况。在申请借款的借款人中,0.54%的借款人月收入不高于 1 000 元,4.46%的借款人月收入处在 1 001~2 000 元之间,41.53%的借款人月收入处于 2 001~5 000 元之间,27.14%的借款人的月收入处于 5 001~1 0 000 元之间,12.43%的借款人月收入处于 10 001~20 000 元之间,7.94%的借款人月收入处于 20 001~50 000 元之间,5.95%的借款人月收入高于 50 000 元。构建取值 1—7 的分类变量收入等级,收入越高,取值越大,用以通过分级的方式刻画借款者月收入的差异。

借款人需要向网站提供相应的信用认证材料。认证分为很多种类,在本文的样本期内,借款人可自愿选择进行更多认证。选择认证的种数,某种程度上反映了借款人信用披露以及披露意愿的情况。具体而言,认证的种类覆盖居住地证明、信用报告、购车认证、购房认证、结婚认证、手机认证、视频认证、微博认证、工作证认证、收入认证等各个方面。对于某项认证,若通过认证,则该项认证哑变量取值为 1,若未通过认证或未认证,则该项认证哑变量取值为 0。构建分类变量已通过人人贷认证种数,令其取值为各项认证哑变量加总之和。

3 实证结果

语言相关的变量是实证部分的关键变量。具体而言,作者从两个维度探讨语言的特征,一是语言长度,二是语言内容。本文衡量语言长度的变量有文字描述字数(用于基准回归)、文字描述中纯中文字数(用于稳健性检验)以及进行1%水平下缩尾处理后的文字描述字数(用于稳健性检验);衡量语言内容的变量有哑变量文字描述提及创业、哑变量文字描述提及急迫、哑变量文字描述提及家庭以及哑变量文字描述提及诚信。

本文设置了两个不同维度的被解释变量:哑变量成功借款与哑变量如约还清,分别衡量呈现某种语言特征的借款在投资人中受欢迎的程度,以及呈现某种语言特征的借款的实际信用风险。两相对比,可以看出P2P平台上的投资人是否在通过语言特征判断借款人信用风险方面是否具有行为偏差,以及语言在借贷市场中究竟发挥着怎样的作用。

在回归过程中,需要控制借款情况和借款人情况的相关变量。其中,借款情况包括借款额度、借款期限、借款利率、借款利率的平方;借款人情况包括性别、年龄、是否已婚、工资水平、已通过人人贷认证种数。在稳健性检验中,额外控制了借款人所在省份、借款人所在行业和借款用途等固定效应。

本文主要采用Probit模型,并在稳健性检验中使用OLS模型。其中Probit模型可用如下公式表示:

$$\begin{aligned} \Pr(\text{哑变量成功借款}) &= \\ \Phi(\text{语言特征, 借款情况, 借款人情况}); \\ \Pr(\text{哑变量如约还清}) &= \\ \Phi(\text{语言特征, 借款情况, 借款人情况}). \end{aligned}$$

1) 基准回归。

本文的基准回归部分使用Probit模型,采用文字描述字数长度、是否提及创业、是否提及急迫、是否提及家庭、是否提及诚信作为表征语言特征的关键指标。表4中第1列和第2列汇报了以成功借款为被解释变量的回归结果,第1列未引入控制变量,第2列加入了控制变量。第3列和第4列展示了以如约还清为被解释变量的回归结果,第3列未引入控制变量,第4列加入了控制变量。

语言长度方面,文字描述字数变量前系数在表4的4列中均显著为正。这表明,语言长度越长,受到投资人的青睐越多,更容易借款成功;语言长

度越长,表征借款申请人信用越好,更可能如约还清。两相对比,说明投资人能够通过语言长度,准确识别借款申请人的信用风险。对上述结果,可以做如下方面的理解。从信息传递的角度讲,文字描述的字数越多,传达给投资人的信息可能会更多。更多的信息会减少借贷双方的信息不对称程度,更好地打消投资人对借款申请人不确定性的担忧,从而提高借款成功率。此外,语言长度表征了借款申请人的努力程度,而努力程度可能是优质借款人发出的信号,使自身在信用风险各异的借款申请人中脱颖而出。从这个角度讲,文字描述越长,更可能是优质借款申请人,如约还清概率越高。

与语言长度方面展示的结果不同,由表4中4个语言内容相关的关键变量前的系数可知,是否涉及某些话题对投资人的决策影响较大,但与借款申请人的真实信用风险关联不大。具体而言,文字描述中提及创业、文字描述中提及家庭更容易获得投资人的青睐,借款成功率更高,而文字描述中提及急迫更容易引起投资人的担忧,借款成功率较低,文字描述中提及诚信对借款成功率的影响也是正面的,但是在控制借款情况和借款人情况后,系数并不显著。有趣的是,第4列中4个话题变量前的系数显示,是否涉及这些话题对借款申请人是否会如约还清的影响均不显著。这意味着,比起语言的长度特征,语言的内容特征更具有迷惑性,像一面引人遐想的美丽面纱,而与面纱下的真实面容并无关联。对借款申请人而言,是否谈及某些话题更是一种出于营销策略的技巧选择,而不是基于自身风险的如实表征。而表4中的结果说明,投资人对于语言内容的面纱本质认识得远非充分,在通过语言内容特征识别借款申请人信用风险方面存在明显偏差。

2) 稳健性检验。

本部分通过调整关键变量定义、转换样本选择范围、增加控制变量、改变回归模型等方式,针对基准回归部分的结果进行稳健性检验。

表5中将基准回归中的关键变量文字描述字数变为了文字描述中纯中文字数,即去掉文字描述中英文字符和标点符号等,只计数中文字符得到的语言长度统计。这样的计数方式可以一定程度上减缓借款申请人用简单字符凑字数对实证结果的干扰。

表4 基准回归:语言长度与涉及话题(Probit回归)

被解释变量	哑变量:成功借款		哑变量:如约还清	
	(1)	(2)	(3)	(4)
文字描述字数/100字	0.229*** (0.0177)	0.122*** (0.0270)	0.121** (0.0546)	0.167*** (0.0585)
哑变量:文字描述提及创业	0.196*** (0.0180)	0.129*** (0.0301)	-0.233*** (0.0691)	-0.0989 (0.0786)
哑变量:文字描述提及急迫	-0.532*** (0.0273)	-0.262*** (0.0438)	0.0183 (0.128)	0.0352 (0.136)
哑变量:文字描述提及家庭	0.0773*** (0.0179)	0.117*** (0.0285)	-0.00798 (0.0678)	0.0135 (0.0727)
哑变量:文字描述提及诚信	0.273*** (0.0205)	0.00357 (0.0323)	0.0823 (0.0751)	-0.0346 (0.0817)
借款情况				
借款金额/万元	—	-0.0443*** (0.00205)	—	0.00169 (0.00548)
借款期限/月	—	-0.0291*** (0.00263)	—	-0.0749*** (0.00938)
借款利率	—	0.206*** (0.0385)	—	-0.0839 (0.105)
借款利率平方	—	-0.00951*** (0.00122)	—	0.000494 (0.00325)
借款人情况				
性别(哑变量:女性)	—	0.0605 (0.0403)	—	0.118 (0.107)
年龄	—	0.0395*** (0.00204)	—	0.00303 (0.00534)
已婚(哑变量:已婚)	—	0.0407 (0.0336)	—	0.260*** (0.0919)
工资水平	—	0.276*** (0.0115)	—	-0.0426 (0.0300)
已通过人人贷验证种数	—	0.409*** (0.00610)	—	0.0720*** (0.0187)
常数项	是	是	是	是
观测数	47424	42562	2431	2431
伪R方	0.042	0.589	0.009	0.115

注: *、**、*** 分别代表在 10%、5%、1% 的水平下显著。括号内报告的是标准误。

从表5中的关键变量前系数可知,与基准回归的结论相同,语言长度越长,借款成功率越高,如约还清率越高;涉及某些话题,影响借款成功率,但与如约还清率无关,表现为迷惑性;投资人在通过语言长度识别信用风险方面较为成熟,在通过语

言内容识别信用风险方面存在偏差。

表6继续针对语言长度变量进行稳健性检验。这里作者对基准回归中文字描述字数变量进行1%水平下的缩尾处理(winsorizing)。

表5 稳健性检验: 文字描述中纯中文字数(Probit 回归)

被解释变量	哑变量: 成功借款		哑变量: 如约还清	
	(1)	(2)	(3)	(4)
文字描述中纯中文字数(百字)	0.274*** (0.0198)	0.152*** (0.0306)	0.127** (0.0615)	0.187*** (0.0666)
哑变量: 文字描述提及创业	0.187*** (0.0181)	0.123*** (0.0303)	-0.234*** (0.0695)	-0.101 (0.0788)
哑变量: 文字描述提及急迫	-0.533*** (0.0273)	-0.264*** (0.0438)	0.0161 (0.128)	0.0330 (0.136)
哑变量: 文字描述提及家庭	0.0696*** (0.0180)	0.113*** (0.0286)	-0.00732 (0.0679)	0.0115 (0.0729)
哑变量: 文字描述提及诚信	0.270*** (0.0205)	0.00183 (0.0323)	0.0822 (0.0751)	-0.0356 (0.0818)
观测数	47424	42562	2431	2431
伪R方	0.043	0.589	0.008	0.115

注: *、**、*** 分别代表在 10%、5%、1% 的水平下显著。与基准回归中相同, 本表中第 1、3 列中回归含常数项, 未包含控制变量; 第 2、4 列中回归包括借款情况、借款人情况等控制变量和常数项, 仅报告关键变量的结果。

表6 稳健性检验: 缩尾 1% 后文字描述字数(Probit 回归)

被解释变量	哑变量: 成功借款		哑变量: 如约还清	
	(1)	(2)	(3)	(4)
缩尾 1% 后文字描述字数/100 字	0.268*** (0.0212)	0.141*** (0.0327)	0.164** (0.0677)	0.223*** (0.0723)
哑变量: 文字描述提及创业	0.195*** (0.0180)	0.128*** (0.0302)	-0.238*** (0.0694)	-0.107 (0.0789)
哑变量: 文字描述提及急迫	-0.532*** (0.0273)	-0.263*** (0.0438)	0.0199 (0.128)	0.0366 (0.136)
哑变量: 文字描述提及家庭	0.0775*** (0.0179)	0.118*** (0.0285)	-0.0114 (0.0679)	0.00906 (0.0728)
哑变量: 文字描述提及诚信	0.272*** (0.0205)	0.00291 (0.0323)	0.0803 (0.0751)	-0.0375 (0.0818)
观测数	47424	42562	2431	2431
伪R方	0.042	0.589	0.009	0.116

注: *、**、*** 分别代表在 10%、5%、1% 的水平下显著。括号中报告的是标准误。所谓 $x\%$ 水平下的缩尾处理, 是指将上下 $x\%$ 的极端值用其近邻的内侧值代替, 从而达到减小极端值影响的目的。本表的回归中, 对文字描述字数变量采用的是 1% 水平下的缩尾处理。与基准回归中相同, 本表中第 1、3 列中回归含常数项, 未包含控制变量; 第 2、4 列中回归包括借款情况、借款人情况等控制变量和常数项, 仅报告关键变量的结果。

所谓 1% 水平下的缩尾处理, 是指将上下 1% 的极端值用其近邻的内侧值代替。缩尾处理可以达到减小极端值影响的目的, 从而进一步确认基准回归中结论的稳健性。观察表 6 中关键变量前的系数可知, 基准回归中关于语言长度和语言内容对成功借款率和如约还清率影响不同的结论得到了进一步的支持。

本文所选取的 P2P 平台曾在不同的发展阶段对文字描述提出过不同最低字数要求。由于随着平台发展最低字数要求并不一致, 且在执行中该项要求比较软性, 加之作者在原始数据的基础上已经进行了一些合理的初步处理, 因此文字描述字数下限的特征表现得并不明显。与之相应, 基准回归中作者并未刻意去掉字数低于某些字数下限的观测。而

为了检验结果的稳健性,表7选取了文字描述字数大于50字的子样本,而50字是平台历次文字描述字数下限中最严格的标准。由表7中关键变量前的

系数可知,字数下限问题并不影响基准回归中得到的基本结论。

表7 稳健性检验:文字描述字数大于50字子样本(Probit回归)

被解释变量(哑变量)	成功借款		成功借款	
	(1)	(2)	(3)	(4)
文字描述字数/100字	0.287*** (0.0183)	0.211*** (0.0279)	0.118** (0.0581)	0.172*** (0.0637)
哑变量:文字描述提及创业	0.174*** (0.0195)	0.134*** (0.0326)	-0.273*** (0.0775)	-0.130 (0.0893)
哑变量:文字描述提及急迫	-0.528*** (0.0297)	-0.288*** (0.0474)	0.144 (0.149)	0.145 (0.159)
哑变量:文字描述提及家庭	0.0856*** (0.0194)	0.119*** (0.0306)	-0.0663 (0.0752)	-0.0398 (0.0809)
哑变量:文字描述提及诚信	0.298*** (0.0220)	0.0372 (0.0343)	0.0575 (0.0815)	-0.0606 (0.0887)
观测数	40341	36295	2022	2022
伪R方	0.047	0.584	0.011	0.125

注:*、**、***分别代表在10%、5%、1%的水平下显著。括号中报告的是标准误。与基准回归中相同,本表中第1、3列中回归含常数项,未包含控制变量;第2、4列中回归包括借款情况、借款人情况等控制变量和常数项,此表仅报告关键变量的结果。

在基准回归中,控制变量为借款情况相关的基础变量和借款人情况相关的基础变量。作者加入了更多的控制变量,具体而言,加入了借款人所在省份、借款人所在行业以及借款用途的固定效应。结果表明,加入更多的控制变量不影响基准回归中的核心故事。篇幅所限,具体表格未在此处报告。

由于本文主要采用的是Probit回归模型,因此在保持被解释变量、关键变量和控制变量与基准回顾一致的情况下,运用OLS回归模型,为本文的基本结论提供侧面参考证据。结果显示,OLS回归得到结果与基准回归中的结论并无违背。

4 结束语

语言是日常生活中不可或缺的信息沟通方式,也是商业行为中举足轻重的客户说服手段。在借贷市场中,贷款申请人可能运用语言,向投资人传达信息。除借款人的客观条件之外,借款人的文字描述可能会对投资人的投资决策起到至关重要的作用。在传统的以银行为媒介的借贷市场上,贷款申请人提供的语言信息难以收集与量化,为学者进行严谨的实证研究设置了障碍。然而,P2P这一新兴借贷市场的出现,为研究借贷市场上语言的作用提

供了难得的契机。

本文采集我国P2P行业的典型代表之一人人贷借贷平台(www.renrendai.com)的数据样本,选取2012年3月至2013年9月的47424个信用认证标作为研究主体。实证结果表明,语言长度越长,借款成功率越高,如约还清率越高,语言长度特征表现出一致性;涉及某些话题(创业、家庭、急迫、诚信),影响借款成功率,但与如约还清率无关,语言内容特征表现出迷惑性。这意味着,投资人在通过语言长度识别信用风险方面较为成熟,但对语言内容的认识存在偏差。上述结论得到了各项稳健性检验的支持。

作者从语言长度和语言内容两个维度,挖掘借款申请人的语言特征;从投资人投资行为和贷款申请人还款行为两个侧面,探讨上述语言特征的作用,从而形成对借贷市场中语言作用的较为全面和较为立体的解读。探索中国借贷市场的自身规律,尤其是中国新兴的互联网背景下的借贷市场的特点,希望能为我国金融市场的完善提供更有针对性的借鉴,并对现有研究中该领域的国际经验进行有益的补充。

参考文献 (References)

- [1] Thompson C J. Caring consumers: Gendered consumption meanings and the juggling lifestyle [J]. *Journal of Consumer Research*, 1996, **22**: 388-407.
- [2] Levy S J. Interpreting consumer mythology: A structural approach to consumer behavior [J]. *The Journal of Marketing*, 1981, **45**(3): 49-61.
- [3] Martens M L, Jennings J E, Jennings P D. Do the stories they tell get them the money they need? The role of entrepreneurial narratives in resource acquisition [J]. *Academy of Management Journal*, 2007, **50**(5): 1107-1132.
- [4] Michels J. Do unverifiable disclosures matter? Evidence from Peer-to-Peer lending [J]. *The Accounting Review*, 2012, **87**(4): 1385-1413.
- [5] Barasinska N, Schäfer D. Does gender affect funding success at the Peer-to-Peer credit markets? Evidence from the largest German lending platform [R]. Working Paper, 2010.
- [6] Herzenstein M, Andrews R L, Dholakia U, et al. The democratization of personal consumer loans? Determinants of success in online Peer-to-Peer lending communities [R]. Working Paper, 2008.
- [7] Pope D G, Sydnor J R. What's in a picture? Evidence of discrimination from Prosper.Com [J]. *Journal of Human Resources*, 2011, **46**(1): 53-92.
- [8] Duarte J, Siegel S, Young L. Trust and credit: The role of appearance in Peer-to-Peer lending [J]. *Review of Financial Studies*, 2012, **25**(8): 2455-2484.
- [9] Berger S, Gleisner F. Emergence of financial intermediaries in electronic markets: The case of online P2P lending [J]. *BuR Business Research Journal*, 2009, **2**(1): 39-65.
- [10] LIN Mingfeng, Prabhala N R, Viswanathan S. Judging borrowers by the company they keep: Friendship networks and information asymmetry in online Peer-to-Peer lending [J]. *Management Science*, 2013, **59**(1): 17-35.
- [11] Everett C. Group membership, relationship banking and loan default risk: The case of online social lending [R]. Working Paper, 2008.
- [12] LIU De, LU Yong, Brass D. Friendships in online Peer-to-Peer lending: Pipes, prisms, and social herding [R]. Working Paper, 2013.
- [13] 李悦雷, 郭阳, 张维. 中国 P2P 小额贷款市场借贷成功率影响因素分析 [J]. *金融研究*, 2013(7): 126-138.
LI Yuelei, GUO Yang, ZHANG Wei. The analysis of impact factors on loan performance in Chinese P2P microfinance market [J]. *Journal of Financial Research*, 2013(7): 126-138. (in Chinese)
- [14] Herzenstein M, Sonenshein S, Dholakia U M. Tell me a good story and I may lend you money: The role of narratives in Peer-to-Peer lending decisions [J]. *Journal of Marketing Research*, 2011, **48**(SPL): S138-S149.
- [15] Larrimore L, JIANG Li, Larrimore J, et al. Peer to Peer lending: The relationship between language features, trustworthiness, and persuasion success [J]. *Journal of Applied Communication Research*, 2011, **39**(1): 19-37.