

# 借款陈述会影响借款成本和借款成功率吗？

## ——基于网络借贷陈述的文本分析

彭红枫 赵海燕 周 洋

(武汉大学经济与管理学院, 湖北武汉 430072)

**摘 要:** 本文以 Prosper 网络借贷平台上 2007 年 12 月至 2012 年 2 月的交易数据为样本, 从文本分析的视角对比研究利率形成机制变动前后借款陈述对借款成功率和实际借款利率的影响。实证表明: 在利率竞拍模式下信用等级越低的借款人越倾向于提供借款陈述, 借款陈述的迷雾指数和文本长度都与借款成功率呈现“倒 U 型”关系, 并与实际借款利率存在线性正相关关系; 其次, 在固定利率模式下, 借款利率与借款陈述长度存在显著的“U 型”关系, 与迷雾指数只表现出线性正相关关系; 最后, 在两种利率模式下借款人提供借款陈述都能降低借款成本, 但是不一定能增加借款成功率, 且信用等级较低的借款人提供借款陈述对借贷行为的影响更大。

**关键词:** 借款陈述; P2P 借贷; 文本分析

**JEL 分类号:** G11, G12, G14 **文献标识码:** A **文章编号:** 1002-7246(2016)04-0158-16

## 一、引 言

网络借贷(Peer-to-Peer Lending, P2P)是指不以银行等金融机构为媒介、借贷双方直接通过互联网进行交易的无担保借贷, 其本质是向中低收入群体或小企业等被传统信贷排斥的个体提供金融服务。这种新型的借贷模式虽然有利于解决中小规模经济活动的资金短缺问题, 但在结构、制度和网络信息的完整度等方面仍存在不足。Lin et al. (2013) 研究指出贷款人主要依据借款人在平台上披露的信息内容及数量来做投资决策, 所以贷款人倾向于更多地掌握借款人的有效信息。但 Steelmann and Aaron (2006) 实证发现, 与线下借贷市场相比, P2P 网络借贷市场中的借贷双方信息不对称现象更加突出。因为在线上市场中, 交易大多是在假名或匿名状态下进行的, 借款人可能会隐瞒自身的不利信息

收稿日期: 2015-06-10

作者简介: 彭红枫, 金融学博士, 教授, 武汉大学经济与管理学院, Email: fhpeng@whu.edu.cn.

赵海燕, 硕士研究生, 武汉大学经济与管理学院, Email: zhaohaiyanwhu@163.com.

周 洋, 金融学博士, 助理教授, 武汉大学经济与管理学院, Email: y.zhou@whu.edu.cn.

\* 作者感谢匿名审稿人的宝贵意见。文责自负。

以增加其获得贷款的可能性并降低借款成本,而贷款人的投资行为是透明的,且网络借贷不需要抵押担保,所以贷款人承担了主要风险。由此观之,深入研究 P2P 网络借贷平台上借款信息对借款利率和借款成功率的影响,对于降低信息不对称程度、控制网络借贷风险以及完善借贷平台的建设具有重大的现实意义。

国内外学者对借款信息与借款利率和借款成功率的关系进行了大量的理论和实证研究,研究所关注的信息大致可分为财务信息(如 Klafft, 2008; 王会娟和廖理, 2014)、社会资本信息(如 Guiso et al., 2004; Freeman and Jin, 2008)和人口特征信息(如 Beller et al., 1994; 李悦雷等, 2013; 廖理等, 2015)这三大类,但鲜有学者从信息的文本挖掘角度去研究借款信息对借款利率和借款成功率的影响。

心理学研究表明,自然语言中运用的词汇能在一定程度上反映出作者的自我认知和社会地位(Pennebaker et al., 2003),由此推断,借款人对借款订单的详细描述或许能为贷款人提供借款人的额外信息。结合互联网借贷的实际情况可知,P2P 借贷交易中存在较强的信息不对称,因而贷款人会关注借款人提供的借款陈述等文本信息,以降低其投资风险。但不同的读者对同一个文本内容的理解和认知是不同的,所以对文本信息进行精确地测量是非常有必要的。

虽然“文本的组织模式和语言特征可携带与作者相关的有效信息”(Ghose et al., 2012; Hancock et al., 2007; 等)这一结论在新闻传播、股票发行(IPO)等领域都已得到证明,但是学者们尚未深入探究文本信息在经济交易中的作用(Herzenstein et al., 2011)。而 P2P 网络借贷作为一种新兴且发展迅速的资金交换手段,现有文献中涉及与借款陈述相关的文本特征研究并不多,且大部分都只是将借款人对借款申请的描述作为控制变量,并没有对其进行系统分析(Larrimore et al., 2011),例如 Pope and Sydnor(2011)从借款文本中提炼出借款用途,并将其作为控制变量,以探究由借款人提供的图片所传达出的性别、种族等因素对网络借贷的影响;Duarte et al. (2012)将陈述文本的总单词数、单词的平均音节数等加入控制变量中,以研究外貌对借款的影响。此外,在仅有的少量较为系统地探究借款陈述如何影响网络借贷行为的研究中,几乎所有文献都将其视角锁定在文本内容的挖掘上(如 Michels 2012; 李焰等 2014)。

值得注意的是,虽然 P2P 网络借贷平台允许借款人提供借款陈述,但是这些文本不经过平台验证,不具备法律约束力,这意味着借款人不必为自己的“言行不一”承担责任,而不符合事实的借款描述可能会引导贷款人做出错误的投资决策,所以仅从内容层面上去研究借款陈述是不充分的。相对于内容而言,文本的语言特征更加难以掩饰或歪曲,作为文本语言特征的突出标志,可读性是衡量作者能力、受教育程度甚至其财务状况的一个简单有效的指标(Rennekamp, 2012)。因此文本可读性颇具研究价值,但在 P2P 领域的此类研究却近乎空白。

迷雾指数是目前最流行的测量文本可读性的工具,其在金融等多个领域的有效性都已得到证明(如 Li, 2008),同时 Loughran and McDonald(2014)研究发现文本长度也能够对金融文件的可读性进行有效测量。因此,本文对 P2P 借贷中的借款陈述进行文本可读

性分析,不仅研究借款陈述的有无对借款成功率和借款成本的影响,而且用借款陈述的迷雾指数和文本长度两个指标来从可读性层面探究借款陈述在互联网借贷中的作用,以期更加全面地考察借款人信息中的借款陈述对 P2P 网络借贷行为的影响。

此外,伴随着互联网金融的发展和全球经济形势的改变,P2P 网络借贷的利率形成机制发生了明显变化。以 Prosper 平台为例,该平台最初采用竞拍模式确定借款成本,但拍卖系统的操作难度大,加上金融危机等事件的冲击,Prosper 平台的贷款违约率和坏账率都大幅上升,所以 Prosper 平台经过多次调整并于 2010 年 12 月 19 日正式将利率形成机制改为固定利率制<sup>①</sup>。由于不同利率形成机制下的利率所包含的信息大有不同——竞拍模式下的利率更多地反映出贷款人对借款标的的市场评价,而固定利率模式下的利率则是平台对借款人违约风险的科学量化,因此本文将全样本按照利率形成机制变动时间点划分为两组,对比分析不同利率形成机制下借款成本的影响因素。

本文旨在不同利率模式下对比探究以下五个问题:第一,在 P2P 借贷中哪些借款人更倾向于提供借款陈述;第二,借款人提供借款陈述能否提高其借款成功率并降低其借款利率;第三,如果提供借款陈述可以影响网络借贷,那么借款陈述的迷雾指数能否影响借款成功率和借款利率;第四,如果提供借款陈述能影响网络借贷,那么借款陈述长度是否会影响借款成功率和借款利率;第五,在上述问题成立的前提下,提供借款陈述对哪部分借款人的作用更大。

本文的研究贡献主要体现在以下几个方面:首先,本文将在新闻学、IPO 等领域广泛使用的文本可读性分析(尤其是迷雾指数)引入到 P2P 网络借贷影响因素的研究中,并加入非线性分析,对比探究不同利率形成机制下网络借贷行为影响因素的差异性,丰富了研究视角,弥补了文本可读性分析在网络借贷领域的研究空缺;其次,基于 Lounghran and McDonald(2014)的研究,本文创新性地将“迷雾指数和文本长度的比较”延伸到 P2P 网络借贷的研究中,并得到较有价值的结论;最后,相对于小样本并手工搜集信息的文本研究(如 Herzenstein et al., 2011; 李焰等 2014),本文借助计算机编程对指标齐全的大样本进行系统的文本分析,避免了实证结果的偶然性。

## 二、假设提出

在利率竞拍模式下,贷款人会通过各种途径收集借款人的信息,并对借款人的还款能力进行判断进而做出投资决策;在固定利率模式下,平台会先审核借款人的借款申请,并对通过审核的借款申请进行定价,所以为了增加贷款人和平台对其的信任度,借款人会选择在提供平台要求填写的信息之外,增加额外的信息,例如借款人提供借款陈述来具体阐述自己的经济情况、借款目的等,以展示其具有良好的还款能力。Klaft(2008)研究发现,借款人的信用评级对网络借贷的影响最大。据此提出第一个假设:

<sup>①</sup> [https://en.wikipedia.org/wiki/Prosper\\_Marketplace](https://en.wikipedia.org/wiki/Prosper_Marketplace).

假设 1: 信用等级越低的借款人越倾向于提供借款陈述, 且该结论在利率竞拍模式和固定利率模式下均成立。

根据 Pennebaker et al. (2003) 的研究可知, 借款陈述是借款人对借款原因、财务状况等信息的自我描述, 能够反映出借款人对自身的评价及其社会地位, 也能传达出借款人对此次借款的重视度、迫切度等信息。可见, 借款人提供借款陈述可为贷款人的投资决策和平台的审核与定价决策提供额外信息, 有助于其申请贷款。据此提出第二个假设:

假设 2: 在两种利率模式中, 提供借款陈述均可提高借款成功率并降低借款利率。

迷雾指数是句子的平均长度和复杂单词所占比例的线性组合, 用于度量借款陈述的阅读难度。迷雾指数的值越小, 说明借款陈述的可读性越强。可读性过强的文本虽然生动易懂, 但是在语言表达的精确性、理论的严密性等方面相对不足 (阎达五和孙蔓莉, 2002), 所以迷雾指数过小的借款陈述不能准确地表达借款人的意图, 这不利于借款人申请贷款; 而当借款陈述的迷雾指数过大时, 贷款人和平台很可能会因读不懂而放弃认真阅读文本, 这会限制甚至阻碍贷款人和平台从中获得额外信息。据此提出第三个假设:

假设 3: 借款陈述的迷雾指数与借款成功率存在倒 U 型关系, 而与借款利率存在 U 型关系, 且该结论不受利率形成机制变动的影响。

借款陈述的长度能反映出借款人的还款能力及其对贷款的重视度。若借款陈述过短, 贷款人和平台也许 would 认为借款人对自己的还款能力不够自信或者是其对此次借款不够重视, 从而不愿为其提供资金或会要求较高的借款利率以补偿潜在的违约损失; 若借款陈述过长, 由于借款陈述具有较大的主观性, 客观信息处于弱势的借款人很可能用更多的正向文字去掩饰不足, 以增加其获得贷款的概率 (Tan et al., 2007), 因此过长的借款陈述更容易引起贷款人和平台对文本真实性的质疑。据此提出第四个假设:

假设 4: 借款陈述的长度与借款成功率存在倒 U 型关系, 而与借款利率存在 U 型关系, 且该结论不受利率形成机制变动的影响。

信用等级是对借款人信用状况的客观反映, 也是贷款人和平台评估借款人还款能力的重要指标 (Lin et al., 2013; Klafft, 2008)。当借款人的信用等级较低时, 借款人可以提供借款陈述以增加贷款人和平台对其的青睐度, 从而能更明显地提高借款成功率并降低交易成本; 而对于信用等级较高的借款人, 其借款成功的可能性本来就大, 再额外提供借款陈述的作用相对较小。据此提出第五个假设:

假设 5: 信用等级较低的借款人提供借款陈述更能提高借款成功率并降低借款成本, 且该结论在利率竞拍模式和固定利率模式下均成立。

### 三、研究设计

#### (一) 样本来源及其处理

本文选取美国 Prosper 网络借贷交易平台上 2007 年 12 月 3 日至 2012 年 2 月 1 日期间的借贷数据为样本, 所有观测都是竞标处于完成状态的借贷信息。首先采用随机抽样

选取 80000 个观测作为初始样本, 然后对初始样本做预处理: 删除信息不全(如缺失信用等级、学历等)的 20257 个观测, 去除带有噪音数据(如借款陈述为“NA”或“.”)的 16776 个观测, 最终整理得到 42967 个有效交易数据。由于 Prosper 平台在 2010 年 12 月 19 日将其利率形成机制从原来的竞拍模式变为固定利率模式, 因此本文将预处理得到的 42967 个观测按照利率机制变动时间点划分为两组: 一组是利率竞价机制下的样本(简称“分样本 1”), 共有 17355 个有效交易数据, 其中 13405 个观测提供借款陈述, 占比 77%, 3950 个观测未提供借款陈述, 占比 23%; 另一组是固定利率机制下的样本(简称“分样本 2”), 共有 25612 个有效交易数据, 其中 14972 个观测提供借款陈述, 占比 58%, 10640 个观测未提供借款陈述, 占比 42%。

表 1 变量定义表

类型	变量	变量名称	变量定义
核心变量 (Core Variables)	Success	借款成功率	投标结束时, 借到的资金/申请的总资金
	Interest	实际借款利率	用借款利率减去借款时的无风险利率来表示
	Description	是否提供借款陈述	当借款人提供借款陈述时取 1, 否则取 0
	FogIndex	迷雾指数	计算公式为: $0.4 * (\text{句子的平均长度} + 100 * \text{复杂单词所占比例})$
	Des_Length	借款陈述的长度	借款陈述中单词的个数
借款基本特征 (Basic Characteristics)	Money	借款金额	借款人申请借款的总金额, 以千美元为单位
	Term	借款期限	借款人借款的期限, 以月为单位
	Purpose	借款用途	按借款人借款的轻重缓急赋值: 创业投资取 0, 个体或公司用款取 1, 资金周转取 2, 个人消费取 3, 其他为 4
借款人信用特征 (Credit Characteristics)	Grade	信用等级	信用等级从高到低依次为 AA 至 HR 共 7 类, 分别赋值 0 至 6
	FICO	平台信用评价	Prosper 平台对借款人的信用评价, 共分为 780 + , 750 - 779 , 714 - 749 , 670 - 713 和 660 - 678 这五类, 依次赋值为 0 至 4
	Total_CRBA	总信用额度	借款人能够在平台上申请的最高资金额度, 以千美元为单位
借款人经济特征 (Economic Characteristics)	DI_Ratio	负债收入比	借款人的负债占其收入的比重
	House	住房情况	当借款人拥有住房时取 0, 当借款人抵押贷款买房时取 1, 当借款人租房时取 2
	Income	月收入	借款人每个月的收入, 以千美元为单位
	Worktime	工作年限	借款人从事工作的时间长度, 以年为单位
借款人历史借贷特征 (History Characteristics)	Inquiries	请款次数	借款人近 6 个月的请款次数
	Delinquencies	违约次数	借款人近 2 年的违约次数
宏观经济特征 (Macroeconomic Characteristics)	Year	年份	为控制经济波动和平台制度变动等宏观因素的影响, 设置年份虚拟变量, 两组分样本都以 2010 年为基准, 分别为 Year_2007, Year_2008 和 Year_2009, Year_2011 和 Year_2012

## (二) 研究模型和变量定义

为了检验前四个假设, 本文基于 Klafft(2008), Michels(2012) 等研究分别构建模型 (1) 和 (2), 并结合研究目的, 使用模型 (3) 研究借款陈述对网络借贷的非线性影响。

$$Description = c + \beta_1 Grade + \beta_2 Controls + \varepsilon \tag{1}$$

$$Y = c + \beta_1 Description + \beta_2 Controls + \varepsilon \tag{2}$$

$$Y = c + \beta_1 X + \beta_2 X^2 + \beta_3 Controls + \varepsilon \tag{3}$$

模型(1)中的被解释变量 *Description* 为借款人是否提供借款陈述,提供时取 1,否则取 0。模型(2)和(3)中的被解释变量 *Y* 为借款成功率(*Success*)和实际借款利率(*Interest*),分别是指借款人实际借到的金额与借款总金额的比值、借款利率与无风险利率的差值。模型(1)的核心解释变量 *Grade* 为借款人的信用等级,共有 7 个等级,从高到低依次赋值为 0 至 6。模型(2)的核心解释变量为 *Description*。模型(3)的核心解释变量 *X* 为迷雾指数(*FogIndex*)和借款陈述长度(*Des\_Length*),迷雾指数的计算公式为:  $FogIndex = 0.4 * (ASL + 100 * ACW)$ ,其中 *ASL* 是句子的平均长度,由总单词个数除以句子个数得到,*ACW* 是复杂单词的比例,由复杂单词(即音节大于 2 的单词)个数除以总单词个数得到;借款陈述长度是指陈述文本中的单词个数。同时,综合已有文献对 P2P 网络借贷行为影响因素的研究成果,本文选取借款金额、借款期限、月收入等变量作为控制变量,变量定义如表 1 所示。

为了验证假设 5,本文在模型(2)中加入变量  $Description * Grade$ ,通过考察该交叉项来检验不同信用等级的借款人提供借款陈述对借款成功率和借款利率影响的差异性。

表 2 变量的描述性统计

变量	模型(1)和(2)中变量的描述性统计				模型(3)中变量的描述性统计			
	分样本 1 (17355 个观测)		分样本 2 (25612 个观测)		分样本 1 (13045 个观测)		分样本 2 (14972 个观测)	
	均值	标准差	均值	标准差	均值	标准差	均值	标准差
<i>Success</i>	0.9100	0.2160	0.9841	0.0845	0.8894	0.2391	0.9819	0.0888
<i>Interest</i>	0.1153	0.0323	0.1227	0.0414	0.1153	0.0316	0.1232	0.0418
<i>Description</i>	0.7724	0.4193	0.5846	0.4928				
<i>FogIndex</i>					11.7880	3.4372	11.7649	3.4860
<i>Des_Length</i>					87.2211	91.9745	65.4329	64.8308
<i>Money</i>	10.2934	6.3937	12.3884	8.1870	10.4420	6.3132	13.0873	8.2736
<i>Term</i>	40.0339	8.9747	43.8732	11.2683	39.3999	8.3692	44.6786	11.5316
<i>Purpose</i>	1.7515	1.1112	1.6297	1.0171	1.7443	1.1166	1.6011	1.0085
<i>Grade</i>	1.5689	1.3126	1.5895	1.4343	1.5764	1.3021	1.6103	1.4562
<i>FICO</i>	2.4027	1.0961	2.4461	1.0626	2.4053	1.0909	2.4130	1.0531
<i>Total_CRBA</i>	97.2385	1014	99.5863	686	86.7684	426	100.7857	713.1128

续表

	模型(1)和(2)中变量的描述性统计				模型(3)中变量的描述性统计			
<i>DI_Ratio</i>	0.1257	0.0654	0.1385	0.0661	0.1257	0.0653	0.1403	0.0660
<i>House</i>	1.4224	0.6336	1.3997	0.6178	1.4375	0.6267	1.3935	0.6153
<i>Income</i>	5.8420	7.3273	5.8516	3.9611	5.8224	6.5799	5.8817	3.9791
<i>Worktime</i>	4.6029	3.5469	5.3147	3.5410	4.4427	3.5467	5.2890	3.5311
<i>Inquiries</i>	0.9211	1.1053	0.8325	1.0311	0.9414	1.1242	0.8278	1.0348
<i>Delinquencies</i>	0.1428	0.4665	0.1434	0.5013	0.1430	0.4688	0.1371	0.4801

注: 由于篇幅有限, 年份变量的描述性统计结果没有列出, 如有需要可以向作者索取。

### (三) 描述性统计分析

表 2 汇总了所有变量的描述性统计结果。对于模型(1)和(2)中的变量而言, 分样本 1 中 77% 的借款人提供了借款陈述, 91% 的借款人成功获得贷款, 平均实际借款利率约为 11.5%, 借款金额的均值约为 10293 美元, 借款期限的均值为 40.0339 个月, 而分样本 2 中只有 58% 的借款人提供了借款陈述, 但其借款成功率高达 98%, 平均实际借款利率、平均借款金额和平均借款期限与分样本 1 相比都偏大, 这说明在固定利率模式下借款人成功获得资金的可能性、资金量以及资金使用期限都会增加, 但也会付出相对较高的借款成本。两组样本中借款人的信用等级和平台信用评价的均值都较小, 说明两组样本中借款人的信用状况均处于中高地位。两组分样本中借款人的月收入均值都在 5850 美元左右, 工作年限均在 5 年左右, 表明在两组分样本中大多数借款人都是属于中低收入且参与工作时间较短的群体。

对于模型(3)中的变量而言, 在提供借款陈述的观测中, 两组分样本的迷雾指数均值分别为 11.7880 和 11.7649, 说明两组分样本中借款陈述的阅读难度相近; 而借款陈述长度的均值分别为 87.2211 和 65.4329 个单词, 可见在利率竞价机制下借款人提供的陈述长度更长。此外, 在依据是否提供借款陈述进行筛选后, 分样本间的变量差异有所改变, 例如在提供借款陈述的观测中, 借款成功率在分样本 1 和分样本 2 间的差距更大。

## 四、实证结果及分析

### (一) 验证假设 1——P2P 网络借贷中哪些借款人更倾向于提供借款陈述

本文采用 Probit 响应变量模型对模型(1)进行回归, 其回归结果如表 3<sup>②</sup> 所示。

② 由于篇幅所限, 控制变量的回归结果没有汇报在表中, 如有需要可以向作者索取。

表 3 信用等级对借款人是否提供借款陈述的影响

变量	模型( 1) 被解释变量: <i>Description</i>			
	分样本 1		分样本 2	
<i>Constant</i>	0. 1894 *** ( <. 0001)	0. 6107 *** ( <. 0001)	0. 7293 *** ( <. 0001)	-0. 2756 *** ( 0. 0011)
<i>Grade</i>	0. 0152 *** ( 0. 0058)	0. 1701 *** ( <. 0001)	0. 0112 ( 0. 1639)	0. 0268 ( 0. 1811)
<i>Interest</i>		-0. 8639 *** ( 0. 0040)		-3. 9969 *** ( <. 0001)
Control Variables	<i>NO</i>	<i>YES</i>	<i>NO</i>	<i>YES</i>
Log likelihood	- 17380. 95	- 8298. 402	- 9307. 562	- 17152. 55
N	17355	17355	25612	25612

注: *N* 代表样本量; \*\*、\*、\* 分别代表在 1%、5% 和 10% 的显著水平,小括号内为对应系数的 *P* 值; Control Variables 包括 *Credit Characteristics*、*Economic Characteristics*、*History Characteristics* 和 *Macroeconomic Characteristics* 这五类控制变量; *YES* 和 *NO* 分别代表回归中加入和未加入控制变量; 后续表格中对应标识的含义均与表 3 相同。

由表 3 可知,对于分样本 1,当模型( 1) 中只加入信用等级( *Grade*) 时, *Grade* 的系数为 0. 0152,且在 1% 的水平上显著,表明信用等级越低的借款人越倾向于提供借款陈述。这主要是由于为了增加贷款人的信任度以如期获得贷款,信用等级越低的借款人越倾向于具体描述自己的借款原因、经济情况等,以吸引贷款人投标。在模型( 1) 中加入控制变量后, *Grade* 的系数增大至 0. 1701,且仍显著,说明在控制其他因素后,信用等级的变化对是否提供借款陈述的影响增大,即信用等级低的借款人更可能提供借款陈述。然而,在分样本 2 中无论是否加入控制变量,信用等级( *Grade*) 对是否提供借款陈述( *Description*) 的回归系数均不显著,表明利率形成机制变动后,借款人的信用等级与其是否提供借款陈述之间不存在线性关系。

综上所述,在利率竞拍机制下,信用等级越低的借款人越倾向于提供借款陈述;在固定利率机制下,借款人的信用等级与是否提供借款陈述无显著关系。

( 二) 验证假设 2——提供借款陈述是否能提高借款成功率并降低借款利率

本文分别以借款成功率和借款利率为被解释变量,以是否提供借款陈述为核心解释变量,探究借款陈述对借贷行为的影响,同时加入控制变量以确保结果的准确性。由于被解释变量均是连续变量,因此本文对模型( 2) 采用 OLS 回归,其回归结果如表 4 所示。

由表 4 中第二、三列可知,对于分样本 1,在加入控制变量的情况下,是否提供借款陈述( *Description*) 的回归系数在 1% 的水平上显著,且为负数,说明提供借款陈述会降低借款成功率,即提供额外信息的借款人获得贷款的难度反而增加。这主要是因为一方面借款陈述带有较大的主观性,贷款人对内容的真实性有所顾虑,另一方面借款陈述可读性的强弱直接影响到贷款人从中捕捉到的有效信息的多少,并非所有的陈述都可以增加贷款人的信任。这点将会在对假设 3 和假设 4 的验证中具体展开分析。但是在分样本 2 中,



是否提供借款陈述( *Description*) 的回归系数并不显著,即提供借款陈述并不能增加借款人获得资金的可能性。两组样本回归结果的差异性可能是由利率模式变动造成的,在利率竞拍模式下借款成本由贷款人竞价投标形成,借款人存在用文本描述来吸引投资专业度不高的贷款人并降低借款利率的机会,但是在固定利率模式下,借款成本由相对专业的 *Prosper* 平台考核确定,借款人依靠主观的借款陈述来弥补客观信息不足的行为已被预期到,因而借款陈述在网络借贷中的影响力降低,所以借款陈述和借款成功率的关系变得不显著了。因此对于分样本 2,本文将只探究借款陈述的阅读难度和长度在固定利率模式下与借款成本之间的关系,而不再关注其对借款成功率的影响。

表 4 借款陈述对借款成功率和借款利率的影响

变量	模型(2) 被解释变量: <i>Success</i>		模型(2) 被解释变量: <i>Interest</i>	
	分样本 1	分样本 2	分样本 1	分样本 2
<i>Constant</i>	0.9199 *** ( <.0001)	1.0283 *** ( <.0001)	0.0647 *** ( <.0001)	0.0295 *** ( <.0001)
<i>Description</i>	-0.0620 *** ( <.0001)	0.0007 ( 0.5155)	-0.0006 *** ( 0.0040)	-0.0006 *** ( <.0001)
<i>Interest</i>	3.6323 *** ( <.0001)	0.3027 *** ( <.0001)		
Control Variables	YES	YES	YES	YES
Log likelihood	7587.397	28291.64	55532.98	82390.61
N	17355	25612	17355	25612

由表 4 中第四、五列可知,在引入控制变量后,两组分样本中是否提供借款陈述( *Description*) 的系数均在 1% 的水平上显著为负,这说明提供借款陈述有助于降低借款利率,且这种关系不受利率形成机制变动的影响。借款陈述虽不能客观反映借款人的真实情况,但在某种程度上可作为贷款人了解借款人的渠道,具有诚意的借款陈述有助于降低借款人的借款成本。此外,对于模型(2)的控制变量,两组分样本中的大部分变量都显著,表明虽然利率形成机制的变动会使得借贷行为的影响因素发生改变,但大部分因素,如借款金额、信用等级等,在两种利率模式下均与网络借贷行为保持显著的相关性。可见,贷款人和平台在对借款人的还款能力和违约风险进行评估时考虑的影响因素大致相同。

综上所述,无论是在利率竞拍模式还是固定利率模式下,借款人提供借款陈述都能降低借款成本,但是不一定能增加借款成功率。

### (三) 验证假设 3——借款陈述的迷雾指数能否影响借款成功率和实际借款利率

本文运用模型(3)对借款陈述的迷雾指数进行非线性回归,结果如表 5 所示。根据表 5 中第二列可知,在分样本 1 中控制其他变量后,迷雾指数( *FogIndex*) 的一次项系数和

二次项系数在 1% 的水平上显著 ,且一次项系数为正 ,二次项系数为负 ,说明借款陈述的迷雾指数和借款成功率之间呈现出显著的“倒 U 型”关系 ,意味着借款陈述的迷雾指数过大或过小都会阻碍贷款人从借款陈述中更加深入地了解借款人 ,增加了借款人的借款难度 ,只有当借款陈述的迷雾指数在某个适宜的范围内 ,提供借款陈述才会显著提高借款人的借款成功率。从表 5 中第三、四列可知 ,在两组分样本中 ,迷雾指数的二次项系数均不显著 ,说明借款陈述的迷雾指数和实际借款利率的“U 型”关系不成立。

表 5 借款陈述的迷雾指数对借款成功率和实际借款利率的影响

变量	模型(3) 被解释 变量: <i>Success</i>	模型(3) 被解释 变量: <i>Interest</i>		令模型(3) 中的 $\beta_2 = 0$ , 被解释变量: <i>Interest</i>	
	分样本 1	分样本 1	分样本 2	分样本 1	分样本 2
<i>Constant</i>	0.6221 *** ( < .0001)	0.0506 *** ( < .0001)	0.0397 *** ( < .0001)	0.0492 *** ( < .0001)	0.0429 *** ( < .0001)
<i>FogIndex</i>	0.0178 *** ( < .0001)	-0.0002 ( 0.2880)	0.0007* ( 0.0958)	7.16E-05 ** ( 0.0399)	0.0001 *** ( 0.0018)
<i>FogIndex_Square</i>	-0.0006 *** ( < .0001)	1.04E-05 ( 0.1321)	-2.41E-05 ( 0.6362)		
<i>Interest</i>	3.8396 *** ( < .0001)				
Control Variables	YES	YES	YES	YES	YES
Log likelihood	5171.471	32411.06	34203.03	32409.93	34197.02
N	13405	13405	14972	13405	14972

为了进一步探究迷雾指数与实际借款利率的关系 ,本文去掉模型(3) 中的二次项以进行线性回归。根据表 5 中最后两列可知 ,在两组分样本中 ,迷雾指数的回归系数在 5% 的水平上均显著为正 ,表明利率形成机制的变动并没有改变文本阅读难度和借款成本之间的关系 ,具体表现为实际借款利率会随着迷雾指数的增大而增大。这主要是因为借款陈述的迷雾指数越大 ,贷款人对借款陈述的理解难度越大 ,也就越难准确地从陈述文本中提取出有效信息 ,而信息不对称程度加大会导致更高的风险溢价 ,因此贷款人会向借款人索取更高的收益;对于已经获得贷款的借款人而言 ,迷雾指数越小 ,越有利于借款人向贷款人传达自己的想法和意图 ,增加贷款人对其的信任度 ,进而达到降低借款成本的目的。由此可见 ,借款陈述的迷雾指数与借款利率之间存在显著的正相关关系。

综上所述 ,在利率竞拍模式下 ,借款陈述的迷雾指数与借款成功率呈现“倒 U 型”关系 ,与借款利率表现为线性正相关 ,且该正相关性在固定利率模式下仍显著。

(四) 验证假设 4——借款陈述的长度能否影响借款成功率和实际借款利率

本文用与验证假设 3 相同的模型来检验假设 4 ,保持其他变量不变 ,只是将核心解释变量换作借款陈述长度 ,以观察借款陈述长度对网络借贷行为是否具有预测功能。基于

假设分析,本文对借款陈述长度进行非线性回归,其回归结果如表 6 所示。

由表 6 中第二列可知,在加入控制变量的情况下,分样本 1 中的借款陈述长度(*Des\_Length*)的一次项系数显著为正,二次项系数显著为负,这说明借款陈述长度和借款成功率之间的确呈现出显著的“倒 U 型”关系,即只有当借款陈述的长度在某个范围内时,提供借款陈述才会显著提高借款人的借款成功率。从表 6 中第三、四列可知,在控制其他变量的影响后,对于分样本 1,借款陈述长度的二次项系数不显著,说明在利率竞拍模式下,借款陈述长度和实际借款利率的“U 型”关系不存在;而在分样本 2 中,借款陈述长度的一次项系数和二次项系数在 5% 的水平上均显著,且一次项系数为负,二次项系数为正,即在固定利率模式下,借款陈述长度与借款成功率表现出显著的“U 型”关系,表明借款陈述过长或过短都将降低贷款人从借款陈述文本中获得有效信息的程度,长度适宜的借款陈述才能降低借款人的借款成本。

表 6 借款陈述长度对借款成功率和实际借款利率的影响

变量	模型(3) 被解释 变量: <i>Success</i>	模型(3) 被解释 变量: <i>Interest</i>		令模型(3) 中的 $\beta_2 = 0$ , 被解释变量: <i>Interest</i>
	分样本 1	分样本 1	分样本 2	分样本 1
<i>Constant</i>	0.4130 *** ( < .0001)	0.0648 *** ( < .0001)	0.0278 *** ( < .0001)	0.0649 *** ( < .0001)
<i>Des_Length</i>	0.0003 *** ( < .0001)	$3.68E-06^*$ ( 0.0932)	$-1.72E-05^{***}$ ( < .0001)	$2.30E-06^{**}$ ( 0.0173)
<i>Des_Length_Square</i>	$-3.77E-07^{***}$ ( 0.0001)	$-3.38E-09$ ( 0.4842)	$1.50E-08^{**}$ ( 0.0373)	
<i>Interest</i>	7.6662 *** ( < .0001)			
Control Variables	YES	YES	YES	YES
Log likelihood	2259.588	42678.38	48124.64	42678.14
N	13405	13405	14972	13405

为了进一步探究分样本 1 中借款陈述长度与实际借款利率的关系,本文去掉模型(3)中的二次项以进行线性回归。根据表 6 的最后一列可知,在利率竞拍阶段,借款陈述长度的回归系数在 5% 的置信水平上显著为正,表明借款人提供的借款陈述越长,其借款利率反而越高。由此观之,当借款成本的决定权在贷款人手中时,相对于冗长复杂的借款陈述文本而言,言简意赅的借款陈述更具有说服力,更有助于增加贷款人的有效信息量,所以借款成本与借款陈述长度的正相关关系会呈现在真实网络借贷交易中。

综上所述,在利率竞拍模式下,借款陈述长度与借款成功率呈现“倒 U 型”关系,与借款利率表现为线性正相关,而在固定利率机制下,借款陈述长度与借款利率的关系表现为“U 型”。可见,在利率形成机制变动前,借款陈述的长度和迷雾指数对网络借贷行为的

影响大致相同，即均与借款成功率存在二次关系，与借款利率成正比，意味着当文本的长度和阅读难度都在适宜范围内时，借款人提供的借款陈述可提高借款成功率；阅读难度较低且长度较短的借款陈述能够显著降低借款成本；而在固定利率模式下，借款陈述的长度和迷雾指数对实际借款利率的影响并不相同，具体表现为借款利率与借款陈述长度存在显著的“U 型”关系，而与迷雾指数只表现出线性正相关。

然而，Loughran and McDonald(2014)通过对上市公司披露的公开文件进行可读性分析，发现相对于迷雾指数，用文本的长度来度量可读性更加有效，这主要是因为上市公司定期披露的公开文件的阅读对象一般都是金融或经济领域的专业人员，他们能够轻松理解大多数常用但复杂的专业词汇，而迷雾指数只是从词语的音节数来定义复杂单词，因此迷雾指数在度量上市公司文件的可读性上存在较大的误差。而本文发现在利率竞拍模式下，迷雾指数和文本长度均可有效度量借款陈述的可读性，且用二者度量文本可读性得到的实证结论一致，这强有力地说明了 P2P 网络借贷交易的大众化程度较高，大多数贷款人都不具备专业的金融素养，对专业的金融词汇不完全熟悉，所以用迷雾指数去度量文本可读性依然有效。对于固定利率模式下的样本而言，借款陈述的长度和迷雾指数与实际借款利率的关系存在较大的差异，这很可能是因为利率形成机制变动后，借款成本的最终决定权也从普通的贷款人手中转移到 Prosper 平台的专业人员手中，虽然不同的定价主体对借款人的还款能力进行判断的评估因素相似，但这种利率决定权的转移使得定价更加科学化、专业化，因而借款陈述长度和迷雾指数对文本内容的度量产生了偏差。此外，因为互联网投资具有较大的风险，所以若要促进 P2P 网络借贷平台的持续发展，平台有义务为贷款人提供与金融常识有关的培训，加强其风险投资意识，提高其理财能力，进而缓解贷款人在互联网匿名交易中的不利局势。

(五) 验证假设 5——提供借款陈述对哪部分借款人作用更大

本文在模型(2)的基础上，通过加入交叉项  $Description * Grade$  来检验不同信用等级的借款人提供借款陈述对借款成功率和利率影响的差异性，回归结果如表 7 所示。

表 7 借款陈述和信用等级对借款成功率和借款利率的交叉影响

变量	被解释变量: <i>Success</i>	被解释变量: <i>Interest</i>	
	分样本 1	分样本 1	分样本 2
<i>Constant</i>	0.5698 *** ( < .0001)	0.0627 *** ( < .0001)	0.0292 *** ( < .0001)
<i>Description</i>	-0.0737 *** ( < .0001)	0.0030 *** ( < .0001)	0.0001 ** ( 0.0371)
<i>Description * Grade</i>	0.0076 *** ( 0.0033)	-0.0023 *** ( < .0001)	-0.0003 *** ( 0.0002)
<i>Interest</i>	7.0918 *** ( < .0001)		

续表

变量	被解释变量: <i>Success</i>	被解释变量: <i>Interest</i>	
	分样本 1	分样本 1	分样本 2
Control Variables	YES	YES	YES
Log likelihood	7588.318	55672.84	82397.64
N	17355	17355	25612

表 7 中的第二列是在被解释变量为借款成功率( *Success*) 时模型( 2) 的基础上加入交叉项 *Description \* Grade* 后的回归结果。在分样本 1 中,是否提供借款陈述 *Description* 的回归系数在 1% 的水平上显著为负,但交叉项 *Description \* Grade* 的系数在 1% 的水平显著为正,这说明信用等级越低的借款人提供借款陈述越可以增加其借款成功率。表 7 中的第三、四列是在被解释变量 *Y* 为实际借款利率( *Interest*) 时模型( 3) 的基础上加入交叉项 *Description \* Grade* 后的回归结果。对于两组分样本而言,是否提供借款陈述 *Description* 的回归系数在 5% 的水平上均显著为正,而交叉项的系数在 1% 的水平上显著为负,这说明信用等级越低的借款人提供借款陈述可以降低其借款利率。这可能是由于贷款人和 P2P 平台都认为信用等级低的借款人的违约可能性较大,所以信用等级低的借款人额外提供借款陈述更能弥补其客观信息的不足,增加贷款人或 P2P 平台对其的信任度,从而更有利于其参与网络借贷交易。

综上所述,提供借款陈述对信用等级低的借款人作用更大,且此结论不受利率形成机制变动的影响。

## 五、稳健性检验

为了保证研究结论的可靠性,本文运用两种方法进行稳健性检验:第一种是重选样本法,即选取借款成功的观测作为样本,研究结论保持不变;第二种是关键变量替换法,即用借款陈述的字符个数替代借款陈述的单词个数来度量文本长度,研究结论保持不变。<sup>③</sup>

## 六、研究结论

本文选取 Prosper 网络借贷平台上 2007 年 12 月至 2012 年 2 月的交易数据为样本,从文本可读性分析的视角对比探究 Prosper 平台借贷利率形成机制变动前后借款陈述对 P2P 网络借贷交易中的借款成功率和实际借款利率的影响,为 P2P 网络平台借贷交易的发展和完善提供了理论依据。本文的主要结论和建议归纳如下:

③ 由于篇幅所限,稳健性检验中的回归结果没有列出,如有需要可以向作者索取。

首先,在利率竞拍机制下,信用等级越低的借款人越倾向于提供借款陈述;在固定利率机制下,借款人的信用等级与是否提供借款陈述无显著关系。在利率竞拍机制下借款成本由贷款人投标决定,信用等级低的借款人更可能违约,因而贷款人对此类借款人投资的可能性较小,即使投资也会要求较高的利率回报,所以信用等级低的借款人更倾向于提供借款陈述以赢得贷款人青睐。但利率形成机制变动后,借款利率由较为专业的P2P平台决定,信用等级低的借款人试图用主观的软信息去弥补硬信息不足的做法在专业的信用风险评估模式下难以发挥作用。

其次,在两种利率模式中,借款人提供借款陈述都能降低借款成本,但是不一定能增加借款成功率。借款陈述虽然带有较大的主观性,但能在一定程度上为贷款人提供额外信息,尤其是具有诚意的借款陈述更有利于借款人获得较低利率的资金,而有无借款陈述对借款成功率的影响并不显著。这说明在P2P借贷交易中借款人有必要如实提供借款陈述以降低借款成本。此外,决定借款成本的不同主体(即贷款人和平台)在对借款人还款能力和违约风险进行评估时考虑的影响因素大致相同,差别在于平台的评估相对较专业。

再者,在利率竞拍模式下,借款陈述的长度和迷雾指数对网络借贷行为的影响大致是相同的,即均与借款成功率存在“倒U型”关系,与借款利率成正比;而在固定利率模式下,借款陈述的长度和迷雾指数对借款利率的影响并不相同,即借款利率与借款陈述长度存在显著的“U型”关系,而与迷雾指数只表现出线性正相关。利率形成机制变动前,只有当文本的长度和迷雾指数在适宜范围内时,提供借款陈述才可提高借款成功率,而迷雾指数较低且文本长度较短的借款陈述能显著降低借款成本,这说明借款人提供的借款陈述应当尽量精炼准确。但在利率形成机制变动后,借款陈述长度与借款利率的关系转变为“U型”,说明借款人仅靠增加文本长度而不注重质量来博取评估人信任的手段在固定利率模式下已难以奏效,只有长度适中的借款陈述才有助于借款人降低其借款成本。

同时,P2P借贷的大众化程度较高,大多数贷款人都不具备专业的金融素养,所以将借款利率的最终决定权由贷款人转移到相对专业的P2P平台是明智之举,这对于科学合理地评估借贷风险、保障投资收益以刺激贷款人的投资热情以及降低平台的坏账率等都具有重要作用。此外,由于互联网投资具有较大风险,因此P2P平台有义务为贷款人提供金融常识培训,加强其风险意识,以促进网络借贷平台的持续发展。

最后,信用等级较低的借款人提供借款陈述对借贷行为的影响更大,且此结论不受利率形成机制变动的影响。信用等级低的借款人额外提供借款陈述更能弥补自己客观信息的不足,增加定价主体对其的了解度和信任度,交易中的信息不对称程度降低,从而更有利于其参与网络借贷活动。

## 参考文献

- [1] 李焰、高弋君、李珍妮、才子豪、王冰婷和杨宇轩 2014,《借款人描述性信息对投资人决策的影响——基于 P2P 网络借贷平台的分析》,《经济研究》增 1 期 第 143 ~ 155 页。
- [2] 李悦雷、郭阳和张维 2013,《中国 P2P 小额贷款市场借贷成功率影响因素分析》,《金融研究》第 7 期 第 126 ~ 138 页。
- [3] 廖理、吉霖和张伟强 2015,《借贷市场能准确识别学历的价值吗? ——来自 P2P 平台的经验证据》,《金融研究》第 3 期 第 146 ~ 159 页。
- [4] 王会娟和廖理 2014,《中国 P2P 网络借贷平台信用认证机制研究——来自“人人贷”的经验证据》,《中国工业经济》第 4 期 第 136 ~ 147 页。
- [5] 阎达五和孙蔓莉 2002,《深市 B 股发行公司年度报告可读性特征研究》,《会计研究》第 5 期 第 10 ~ 17 页。
- [6] Beller A. , G. Borjas, M. Tienda, D. Bloom and G. Grenier. 1994. “Beauty and the Labor Market”, *The American Economic Review* ,84( 5) : 1174 ~ 1194.
- [7] Duarte J. , S. Siegel and L. Young. 2012. “Trust and Credit: the Role of Appearance in Peer – to – Peer Lending” , *Review of Financial Studies* ,25( 8) : 2455 ~ 2484.
- [8] Freedman, S. , and G. Z. Jin. 2008. “Do Social Networks Solve Information Problems for Peer – to – Peer Lending? Evidence from Prosper. com” , Working Papers , College Park , MD: NET Institute.
- [9] Ghose A. , P. G. Ipeirotis and B. Li. 2012. “Designing Ranking Systems for Hotels on Travel Search Engines by Mining User – Generated and Crowdsourced Content” , *Marketing Science* ,31( 3) : 493 ~ 520.
- [10] Guiso, L. , P. Sapienza and L. Zingales. 2004. “The Role of Social Capital in Financial Development” , *American Economic Review* ,94: 526 ~ 556.
- [11] Hancock J. T. , L. E. Curry, S. Goorha and M. Woodworth. 2007. “On Lying and Being Lied To: A Linguistic Analysis of Deception in Computer – Mediated Communication” , *Discourse Processes* ,45( 1) : 1 ~ 23.
- [12] Herzenstein M. , S. Sonenshein and U. M. Dholakia. 2011. “Tell Me a Good Story and I May Lend You Money: the Role of Narratives in Peer – to – Peer Lending Decisions” , *Journal of Marketing Research* ,48( SPL) : S138 ~ S149.
- [13] Klafft M. . 2008. “Peer to Peer Lending: Auctioning Micro Credits over the Internet,” Proceedings of the 2008 International Conference on Information Systems , Technology and Management( ICISTM 08) , March , Dubai , United Arab Emirates.
- [14] Larrimore L. , L. Jiang , J. Larrimore , M. David and G. Scott. 2011. “Peer to Peer Lending: the Relationship between Language Features , Trustworthiness and Persuasion Success” , *Journal of Applied Communication Research* ,39( 1) : 19 ~ 37.
- [15] Loughran T. , and B. McDonald. 2014. “Measuring Readability in Financial Disclosures” , *The Journal of Finance* ,69( 4) : 1643 ~ 1671.
- [16] Li F. . 2008. “Annual Report Readability , Current Earnings and Earnings Persistence” , *Journal of Accounting and Economics* ,45( 2) : 221 ~ 247.
- [17] Lin M. , N. R. Prabhala and S. Viswanathan. 2013. “Judging Borrowers by the Company They Keep: Social Networks and Adverse Selection in Online Peer – to – Peer Lending” , *Management Science* ,59( 1) : 17 ~ 35.
- [18] Michels J. . 2012. “Do Unverifiable Disclosures Matter? Evidence from Peer – to – Peer Lending” , *The Accounting Review* ,87( 4) : 1385 ~ 1413.
- [19] Pennebaker J. W. , M. R. Mehl and K. Niederhoffer. 2003. “Psychological Aspects of Natural Language Use: Our Words , Our Selves” , *Annual Review of Psychology* ,54( 1) : 547 ~ 577.

- [20]Pope D. G. and J. R. Sydnor. 2011. "What's in a Picture? Evidence of Discrimination from Prosper. com", *Journal of Human Resources*, 46( 1) : 53 ~92.
- [21]Rennekamp K. . 2012. "Processing Fluency and Investors' Reactions to Disclosure Readability", *Journal of Accounting Research*, 50( 5) : 1319 ~1354.
- [22]Steelmann J. , and B. Aaron. 2006. "Bypassing Banks", *Region Focus*, *Federal Reserve Bank of Richmond*, 10( 3) : 37 ~40.
- [23]Tan , K. W. P. , D. Swee , C. Lim , B. H. Detenber and L. Alsagoff. 2007. "The Impact of Language Variety and Expertise on Perceptions of Online Political Discussions", *Journal of Computer - Mediated Communication*, 13( 1) : 76 ~99.

## Could the Loan Description Affect the Loan Cost and the Loan Success Rate? Textual Analysis Based on the Loan Description

PENG Hongfeng ZHAO Haiyan ZHOU Yang

( Economics and Management School ,Wuhan University)

**Abstract:** Whether the loan descriptions have effects on P2P lending behaviors in different formation mechanisms of interest rate is tested from the perspective of textual analysis , using the transaction data on Prosper between December 2007 and February 2012. The results show , in the auction mode of interest rates , borrowers whose credit ratings are lower are more inclined to provide the loan description; both of fog indexes and lengths of loan descriptions form an inverted U-shaped relation to the loan success rate and have a linear positive correlation with the interest rate. Second , the interest rates form a U-shaped relation to the lengths of loan descriptions and are positively correlated with the fog indexes in the pre-set rates mode. Third , providing the loan description could decrease the interest rate , but fails to increase the success rate , and the impacts of providing the loan description are more obvious for borrowers with a lower credit rating in both mechanisms.

**Key words:** Loan Description ,P2P Lending ,Textual Analysis

( 责任编辑: 林梦瑶) ( 校对: ZL)