社会科学与计算研究
ISSN: 2789-553X(print)
雙清學術出版社
www.qingpress.com

四十年（1980-2020）来个人借款领域的研究主题变迁
-基于文本挖掘 LDA 算法的主题发现和可视化

陈媛先
中央财经大学，金融学院 （10098）
蒂尔堡大学(荷兰)

摘要：伴随中国经济的快速发展，个人借款业务发展快速，涌现了不少热点和难点问题，这些都需要学者们进行深入研究。回顾研究历史，能更好地帮助研究人员掌握研究的发展脉络，也利于学者找到当前研究的空白，从而推动领域内的研究不断向前。基于此目的，本文研究了 40 年来(1980-2020 年)我国在“个人借款”研究论文涉及主题的变迁和重要主题的相关性。借助机器学习 LDA 主题分析算法，作者对 2000 多篇文章进行了主题挖掘，并进行了可视化演示。文章的重要发现包括：一、从 2013 年以后，个人借款领域的研究文章大幅增长，而文章的阅读和引用次数从 2010 年开始增长加快。二、从 2000 年开始，人工智能相关在个人借款领域的文章占比增大。三、在个人借款领域，我国的研究具有比较鲜明的三个特征，其中，1980-2000 年，主要为探索期；2001-2010 年为个人住房贷款、企业相关的个人贷款重点研究期间，而 2011 年到 2020 年为个人借款研究深入的阶段。

关键词：个人借款，主题，主题可视化，40 年

1. 引言

个人金融是金融的重要组成部分，其中，个人的借款又是个人金融的重要组成部分。个人借款涉及链条较长，因此，个人借款的研究有很多领域，包括：金融机构、金融产品、借款人、出借人、合同与法律、风险管理、金融行为等。自改革开发以来，国内在个人借款方面的业务发展比较快速，相关的科研论文也较多，但是，对于该领域的论文回顾比较零散，因此，做一次系统的回顾和分析，有助于为该领域的研究往前推进。本文基于这个考虑，结

Cite this paper: 陈媛先. 四十年（1980-2020）来个人借款领域的研究主题变迁-基于文本挖掘 LDA 算法的主题发现和可视化. 社会科学与计算研究研究，2021，卷 1，第 4 期，1-32 页.

合机器学习的方法，用主题分析和主题可视化的技术，对国内近 40 年来个人借款领域的文献进行了系统的分析和讨论。

实际上，当前关于个人金融比较重要的综述性文章主要涉及于商业银行风险的研究综述，比如，商业银行操作风险相关文献的综述（蒋亚利. 2012）；关于个人金融隐私保护的研究（姚蔚子. 2016）；关于居民储蓄研究的综述（人寿. 1992），个人住房抵押贷款风险研究的综述（乔薇. 2012），个人金融客户保持的研究综述（于小亿. 2009），P2P 研究综述（莫易娴.2011; 王学龙, 张璟. 2010），农村民间借贷综述（周扬, 刘义杰. 2015）等。在回顾、综述性文章中，比较多的是汇总分析方法进行，对研究脉络、国内外差异、方法差异进行个人梳理和总结。但是，当前关于个人借款领域研究主题的发现和时期变迁的研究较少，对该领域的研究推进和未来的研究重点缺乏有效的分析数据，这凸显了本文研究的必要性和重要性。本文采用机器辅助分析，利用 LDA 主题发现技术，对个人借款领域最近 40 年（1980-2020 年）研究论文进行主题挖掘，分析研究的推进和主题变迁，将弥补这个空白。

在本文中，我们将重点分析这个领域的研究涉及到哪些重要主题，在不同的时期具有哪些不同，内在的迁徙逻辑和中国经济的发展是否具有内在的关联？最近关于个人借款领域的热点研究主题有哪些？个人借款领域的论文整体的产出和引用的趋势如何，哪些论文获得最多的引用，这些论文的研究主题涉及到哪些内容？在具体的研究中，涉及到重要的论文，我们将采用列表方式进行展示，列出作者和引用等数据，方便查询；涉及到主题相关性，我们将采用可视化的方式进行展示，方便读者在四个象限中直接查看主题之间的接近程度。

我们的重要发现包括：一、从 2013 年以后，个人借款领域的研究文章大幅增长，而文章的阅读和引用次数从 2010 年开始增长加快。二、从 2000 年开始，人工智能相关在个人借款领域的文章占比增大，这个国内外人工智能逐渐深入到社会各个领域的应用趋势是一致的。三，在不同的时间区间，关于个人借款的研究主题呈现出明显差异，其中，在 1980-2000 年这 20 年中，个人借款相关的研究主题主要和国家、形式、探讨、运营、会计等相关，显示在这个区间内，研究处于探索的阶段，国家的整个关于个人借款的发展还在早期；在 2001-2010 年这 10 年内，个人借款和住房贷款紧密结合的研究主题开始凸显重要性，反映出房地产市场的快速发展也带来了大量与此相关的研究。与此同时，在这个阶段，个人创业相关借款的研究开始凸显，企业借款中涉及个人信用部分的研究开始凸显，比如，企业负责人的信用问题。伴随着贷款业务的发展，关于风险和保险的探讨也变得重要。信用卡和消费信贷、汽车相关贷款的研究也开始逐渐大量涌现。凸显重要性。三、在 2011-2020 年这 10 年，个人借款的研究开始和“个人”进行了更多的结合，个贷、大学生贷款、创业贷款等研究增多，结合小微企业的研究也显著增多。在此期间，P2P 等网络贷款的研究是一个重要的研究方向，民间金融和相关法律的研究也显著增多。风险管理的相关研究更广泛，涉及商业银行信贷风险和个人征信等，而随着大数据和人工智能算法，借助模型和数据进行信用评估等的研究也逐渐结合更多应用案例而呈现显著增加。

本文的创新点在于：一、通过分析最近 40 年来（1980-2020）个人借款领域的文献主题变迁，涉及 2225 篇和个人借款相关的论文，通过分析论文的主题、不同时期的主题变迁、重点主题的相关研究文献，对个人借款领域的研究主题历史变迁进行了一个系统的回顾，为总结该领域的研究空白提供了一个分析的基础；二、本文通过收集个人借款相关的博文和新闻报道，分析近期普通民众关注的重点话题和新闻报道的重要事件，为个人借款相关研究未来

Cite this paper: 陈媛先. 四十年（1980-2020）来个人借款领域的研究主题变迁-基于文本挖掘 LDA 算法的主题发现和可视化. 社会科学与计算研究研究，2021，卷 1，第 4 期，1-32 页.

的研究方向提供了一些思路和启发。三、本文将 LDA 主题挖掘算法引用个人金融主题研究领域，这是一个创新，也是本文的一个重要贡献。

我们认为研究尚存在的不足在于：一、为了不干扰分析过程，并尽量客观，本文在分析过程中只对基本的无意义的词汇进行过滤，并没有考虑足够的停用词，因此，在部分主题结果中会包含部分无意义的发现主题，尽管这并不影响主题的解读，但是会造成部分部分噪音。二、文本的分析并没有对重要的论文和观点进行解读，而只是从主题的角度进行分析，这在未来可能可以改进。三、本文所用数据基于整理百度学术网站搜索到的文章数据，因此，文献的收集局限于该数据来源，如果百度学术尚未收录的文章会被遗漏掉，即使是非常重要的文章；另外，文章的收集是基于作者定义的词语所触发的搜索结果，因此，不免会遗漏没有使用到的词语对应的文章；最后，文章数据的整理过程未免会有所遗漏。

2. 研究方法、思路和数据

本文的研究方法包括基于机器学习的 LDA 主题发现和基于可视化技术的主题可视化技术。LDA（Latent Dirichlet Allocation）主题发现技术属于机器学习领域的重要算法，该技术基于三层贝叶斯算法，建立文档、主题、关键词三层概率关系。在其他的学术领域，LDA 的技术已经开始较为广泛的应用。比如，在社会研究领域，探讨话题的演变（单斌, 李芳. 2010），微博事件提取（高永兵, 熊振华. 2015），知识流动（宋凯等. 2017）。在金融相关领域，LDA 主题挖掘的技术也开始应用在股票分析等相关领域，如分析舆情对股指的影响（徐翔等. 2018)，移动支付风险分析（封思贤, 袁圣兰. 2018），消费者信用评估（向晖, 杨胜刚. 2010），商业银行操作风险（丰吉闯等. 2011）。

LDA 算法属于文本挖掘算法，通过构建文档到主题、主题到词语的多项式分布，支持如下分析：文档中出现指定数量主题的概率，特定主题先出现特定词语的概率。LDA 为非监督性性算法，主要通过机器学习的方法来对输入的语料进行主题挖掘，底层算法为三层贝叶斯概率模型。使用 LDA 进行主题，第一步是对语料进行分词，然后建立词向量。然后通过贝叶斯模型进行文档、主题和词的概率计算，详细如下图 1：

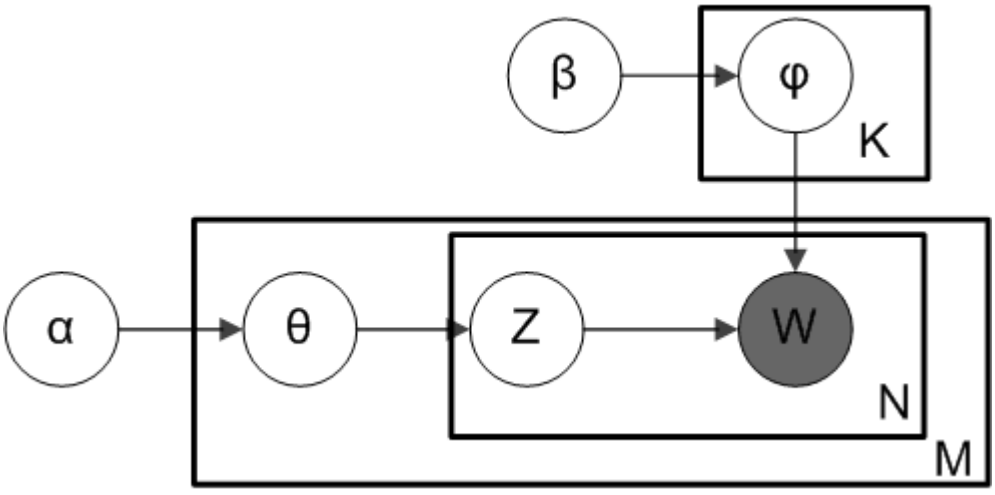

图 1：LDA 的实现逻辑

其中，M 为文档集合，在参数为 β 的 Dirichlet 分布中，通过采样主题（topic）生成词（word）的分布参数 φ；m 为 M 中的文档，通过参数为 α 的 Dirichlet 分布进行采样，形成文档（doc）

Cite this paper: 陈媛先. 四十年（1980-2020）来个人借款领域的研究主题变迁-基于文本挖掘 LDA 算法的主题发现和可视化. 社会科学与计算研究研究，2021，卷 1，第 4 期，1-32 页.

对于主题（topic）的分布参数 θ。基于文档 m 的第 n 个词语（W_mn），按照 θ 分布采用文档 m 的隐含主题（Z）。最后，在按照 φ 分布采样主题(Z_m）的一个词语（W_mn）。

模型的联合分布如下：

$$P(Z, W, \theta, \varphi|\alpha, \beta) = P(W|\varphi, Z) \cdot P(Z|\theta) \cdot P(\theta) \cdot P(\varphi)$$

主题可视化是指将 LDA 挖掘的主题以可视化的方式现在在四象限上，其中，LDAvis 是主题可视化的一种方案。通过将主题进行可视化展现，可以提供直观的图示。下图 2 为 LDAvis 的主界面。

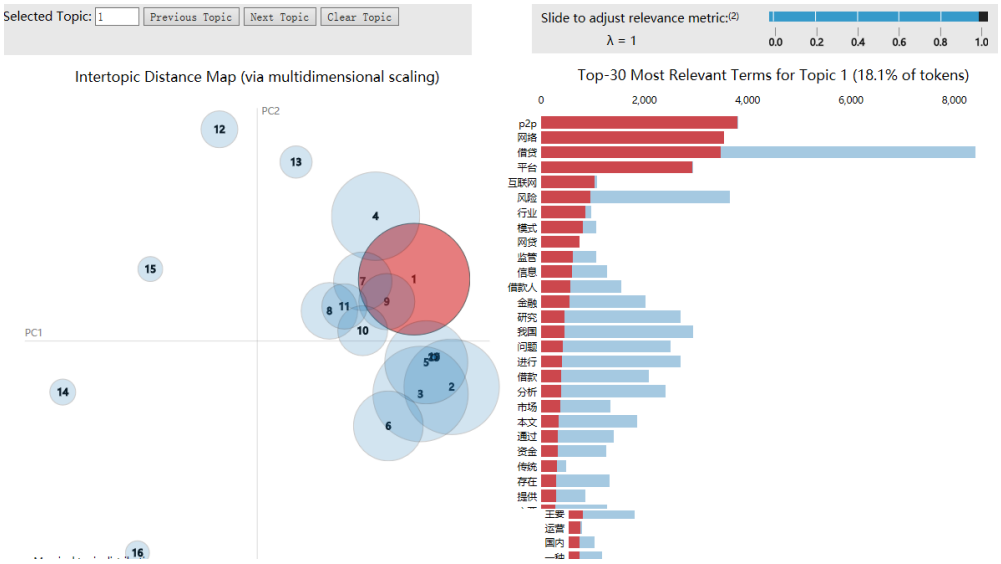

图 2：主题可视化呈现

在主界面左侧显示发现的主题，采用多维标度法（Multidimensional Scaling）进行展示。其中，气泡的大小代表主题出现的概率，气泡之间的距离代表主题之间的相似程度远近，两个气泡距离越远，代表他们的相似程度越低。在左侧上方面，点击按钮可以切换主题进行查看。

在主界面的右侧显示每个主题下最相关的前 30 个词汇。当点击左侧任意一个主题气泡的时候，右侧的词语会对应变化。在右侧最上方显示 lamda 设置标尺，可以拖动标尺对 lamda 进行切换。Lamda 作为一个系数，用于调整和主题相关词汇的展示。其中，lamda*p（w/t）度量主题下词语出现的频率，而 lamda*p(w/t)/p(w)度量了词汇在某个主题下的独特性。通过调整 lamda 来对这两个部分进行权重调整，当 lamda 为 1 的时候，只考虑整体出现次数的概率，而 lamda 为 0 的时候，只考虑独特性。因此，relevance(term w | topic t)的计算综合了词频和词语在主题下的独特性。

$$\text{relevance(term w | topic t)} = \lambda * p(w|t) + (1 - \lambda) * p(w|t)/p(w)$$

Cite this paper: 陈媛先. 四十年（1980-2020）来个人借款领域的研究主题变迁-基于文本挖掘 LDA 算法的主题发现和可视化. 社会科学与计算研究研究，2021，卷 1，第 4 期，1-32 页.

点击右侧的词语（鼠标悬停），左侧相关主题的气泡大小会有所变化，对应该词语在主题上的条件分布概率。

在本文中，我们将按照如下思路对文献进行研究：

第一步：分析 40 年文献的数量、阅读趋势、引用趋势；

第二步：研究个人借款领域重要问题的研究趋势和问题之间的相关性，重要问题我们用词语来定义，包括：融资可得性、消费观念、合规、监控、监管、人工智能、大数据、模型、法律、立法、证券和证券化、资产处置、投资和行为金融。

第三步：对研究进行历史阶段区分，分为 1980-2000 年、2000-2010 年、2010-2020 年三个历史阶段进行主题对比分析，研究主题的历史变迁规律。

第四步：对最近社交媒体、新闻涉及到的个人借款数据进行主题分析，挖掘领域内的热点问题，有助于启发后续研究的选题。

3. 数据来源、处理和模型参数设置

本文所使用的数据均基于自行收集和整理的论文数据。数据来自百度学术网站，数据的获取时间为 2020-5-20 至 2020-6-30 日，因此，文中后续所显示引用数据、阅读数据等均为该时间周期内某个时点的数据。通过搜索“个人借款”、“个人信贷”、“个人贷款”、“个人贷款”四个关键词，搜索到相关的论文。将论文的标题、摘要和关键词进行整理后作为分析的基础数据。在数据的处理上，主要包括四步：

第一步：汇总论文数据。

第二步：将重复的数据进行去重处理，去重逻辑为如下四项信息完全相同：摘要、作者、发表年份、标题。

论文数据详细信息见表 1：

表 1：论文数据

搜索关键词	原始论文篇数	去重后论文篇数
个人借贷	5887	2225
个人贷款	5361	
个人借款	7054	
个人信贷	618	

在数据挖掘上，本文主要借助 Python，采用结巴分词工具将内容进行分词后，采用 LDA 算法进行主题发现（支持 Python 运行的 gensim 包包含的 lda 算法模块），然后采用 LDAvis 进行可视化分析。在 LDA 算法进行主题发掘的时候，设置基本参数为：20 个主题、每个主题下 20 个词语，相关代码演示如附件 1。

4.1 分析结果

4.1 引用最多、阅读最多和文章发表数量趋势

在表 2 中显示了个人借款研究领用被引用最多的文章列表，列表中的文章引用次数均超过 30 次。从表 2 可以看到，最近 40 年来，引用数最多的文章为《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》，引用次数为 531 次。引用排在前列的 19 个

Cite this paper: 陈媛先. 四十年（1980-2020）来个人借款领域的研究主题变迁-基于文本挖掘 LDA 算法的主题发现和可视化. 社会科学与计算研究研究，2021，卷 1，第 4 期，1-32 页.

研究中，有 9 篇文章和 P2P 网络贷款研究相关，文章的发表时间从 2010 年到 2014 年，而这个期间也是 P2P 网络借贷的起步初期，这几篇文章从信用机制、中外比较、法律规制、风险、监管、反洗钱等方面进行探讨。我国互联网金融从 2013 年余额宝出现后得到了快速的发展，而互联网借贷更是在 2014 年-2015 年间开始快速起步，并在 2017 年后因为监管加强，行业的发展逐渐收缩。这些高引用文章的发展，反应了最近 10 年来，伴随中国网贷行业发展，研究的深入和传播。

表 2：个人借款领域引用最高的文章

标题	发表期刊	年份	引用	作者
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	2014	531	王会娟，廖理
P2P 在线借贷的中外比较分析——兼论对我国的启示	金融发展评论	2010	430	尤瑞章；张晓霞
P2P 网络借贷:金融创新中的问题和对策研究	科技信息	2011	343	陈静俊
网络平台借贷的法律规制研究	法学家	2013	208	姚海放，彭岳，肖建国等
民间借贷逾期行为研究——基于 P2P 网络借贷的实证分析	金融论坛	2013	135	陈霄，丁晓裕，王贝芬
中国个人消费信贷状况及风险防范研究	金融论坛	2005	130	杨大楷，俞艳
个人消费信贷的博奕分析	金融研究	2003	98	黄儒靖
商业性 P2P 网络借贷的风险与法律规制	人民司法	2013	88	茅建中
个人消费信贷:中美比较与借鉴	金融论坛	2007	84	郭慧，周伟民
我国民间借贷:现状、成因、影响与对策	金融理论与教学	2004	59	张大龙
我国 P2P 网络借贷发展现状及其监管思考	金融理论与实践	2014	57	田俊领
P2P 网络借贷中个人信息对借贷成功率影响的实证研究——以人人贷为例	财务与金融	2013	51	陈建中，宁欣
防范民间借贷风险的对策选择	经济理论与经济管理	2011	49	胡乃武，万晓芳
P2P 网络借贷洗钱风险剖析及策略选择	中国金融电脑	2014	40	马伟利，许井荣
个人信贷业务风险管理问题与对策	西部金融	2012	39	刘芬芳
网络信贷中的双重信任及其对借贷意愿的影响机制	福州大学学报(哲学社会科学版)	2013	35	陈冬宇，林漳希
商业银行个人信贷风险管理研究	河北金融	2012	31	张金兰
新一代个人信贷业务系统	中国金融电脑	2005	30	李继超
个人信贷中信用风险识别的信号博弈分析	湖南大学学报(社会科学版)	2010	30	晏艳阳，金鹏

在表 3 中列出了阅读量超过 150 次的文章，一共 18 篇。从阅读数和引用数来看，《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》 在引用和阅读数上均排在第一，但是，其他列表中的文章引用和阅读的排名有所差别。分析阅读最多的文章，有 9 篇文章和风险相关。对风险的关注不仅来自学术界，也会来自业界，而业界在面对快速发展的行业时候，也希望从学术研究成果中找到相关指导的依据，这或是引用和阅读排名差异背后体现的差异。

表 3：个人借款领域阅读最多的文章

标题	发表期刊	年份	阅读	作者
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	2014	600	王会娟，廖理
中国个人消费信贷状况及**风险**防范研究	金融论坛	2005	424	杨大楷，俞艳
大数据背景下商业银行的个人信贷**风险**控制--以工商银行为例	企业导报	2016	334	孙培耘，闻君，颉浩颖等
P2P 借贷**风险**管理案例分析——以人人贷为例	今日财富	2016	327	宋慧荣
我国 P2P 网络借贷发展现状及其监管思考	金融理论与实践	2014	322	田俊领
P2P 借贷中借款人的违约风险评估——基于“人人贷”数据的实证分析	经济问题	2017	320	阮素梅，何浩然，李敬明
论大数据背景下商业银行的个人信贷**风险**控制	科技经济导刊	2019	315	龙兴婷
商业银行个人贷款存在**风险**点剖析	时代金融	2017	304	高淑英
商业银行个人消费贷款**风险**防范分析	经营管理者	2017	275	高嘉璘，张鑫
影响借款人选择还款方式的因素分析	集团经济研究	2004	244	张祖平，钟菊
解析新个人所得税法中的专项附加扣除	湖南税务高等专科学校学报	2019	230	钟艺
商业银行个人信贷业务的现状及发展研究	湖南社会科学	2001	224	刘星
互联网金融背景下我国个人征信行业发展实践及展望	金融理论探索	2018	212	刘国刚
金融机构向小微企业贷款的利息收入免征增值税	税收征纳	2019	210	程辉
对挪用资金罪"归个人使用"和"借贷给他人"的考量	决策探索	2006	209	张云龙
农商行小微贷款**风险**评估及其预警——基于经济新常态背景的研究	农业技术经济	2017	203	葛永波，曹婷婷，陈磊
浅析商业银行个人信贷业务的潜在**风险**和防范	中国投资	2013	152	刘思文

为了分析个人借款相关文章的引用和阅读数趋势，本文将文章的引用和阅读数按照年度进行绘图，见图 3。可以看到，关于个人借款相关的引用和阅读从 2019 年后有显著的增加，但是，文章的引用数（蓝色点）和文章的阅读数（红色小叉）具有显著差异，这也反应了我们上面分析的，学术研究和业界关注可能存在差异。

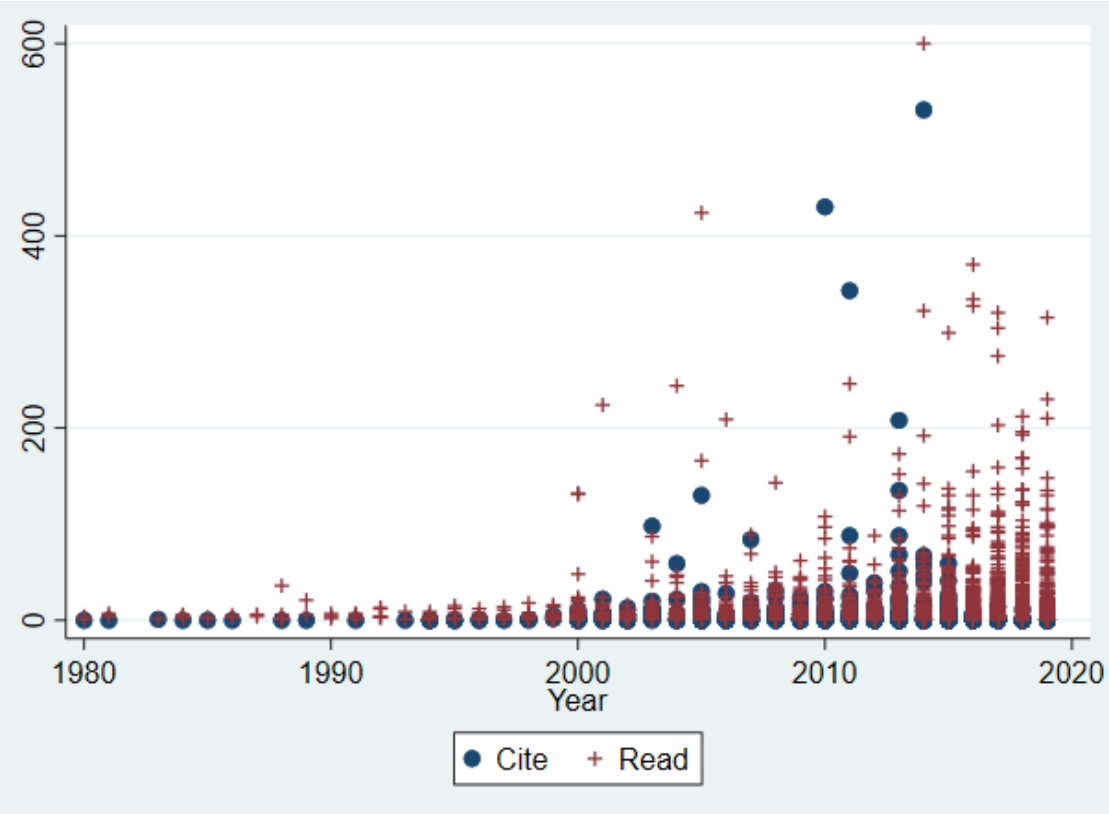

图 3：个人借款文章按年引用和阅读数

为了分析个人借款领域的历史研究成果数量，我们将历年发表的文章数和总文章数比例按年度进行作图，见图 4 灰色线条。随着大数据和 AI 技术的发展，AI 技术辅助金融业务成为个人金融领域的重要应用。那么，在研究上，AI 结合个人借款的研究有什么趋势？我们将 内容中包含如下关键字大数据、人工智能、AI、模型的文章定义为 AI 相关研究，然后将 AI 研究的历年发表文章数量和当年发表的文章数进行一个对比分析，显示在图 4 中黄色线条。可以看到，个人借款的相关文章在 2000 年后增多，而出现波发现增增从 2010 年后准便开始增多。与此同时，从 2015 年开始，AI 文章占当年的文章数也逐渐呈现增长的趋势。

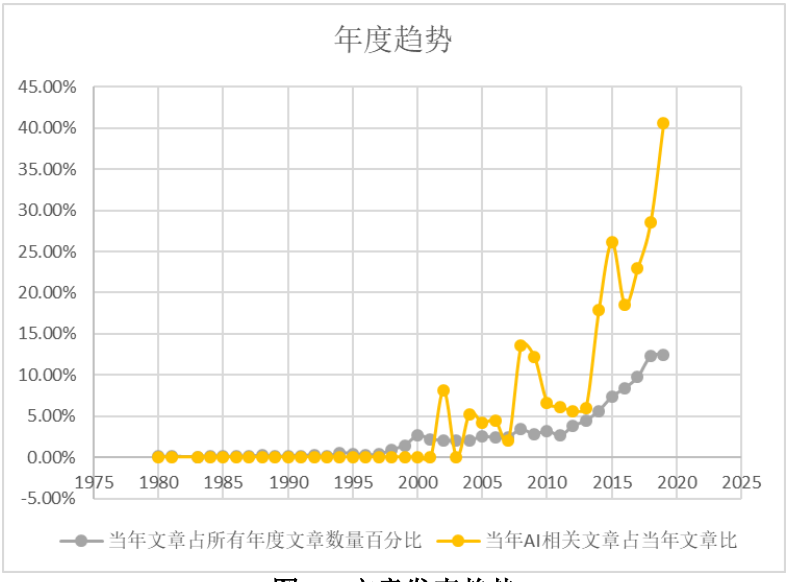

图 4：文章发表趋势

Cite this paper: 陈媛先. 四十年（1980-2020）来个人借款领域的研究主题变迁-基于文本挖掘 LDA 算法的主题发现和可视化. 社会科学与计算研究研究，2021，卷 1，第 4 期，1-32 页.

4.2 主题发现

记住 LDA 主题分析，和 LDAvis 可视化分析，我们得到 1980-2020 年 40 年来，我国个人借款领域相关研究的主题 16 个，见图 5 和表 4。主题按照重要性排序，主题为 20，每一个主题下词语为 20, 在可视化设置上，我们将 lamda 为 1。

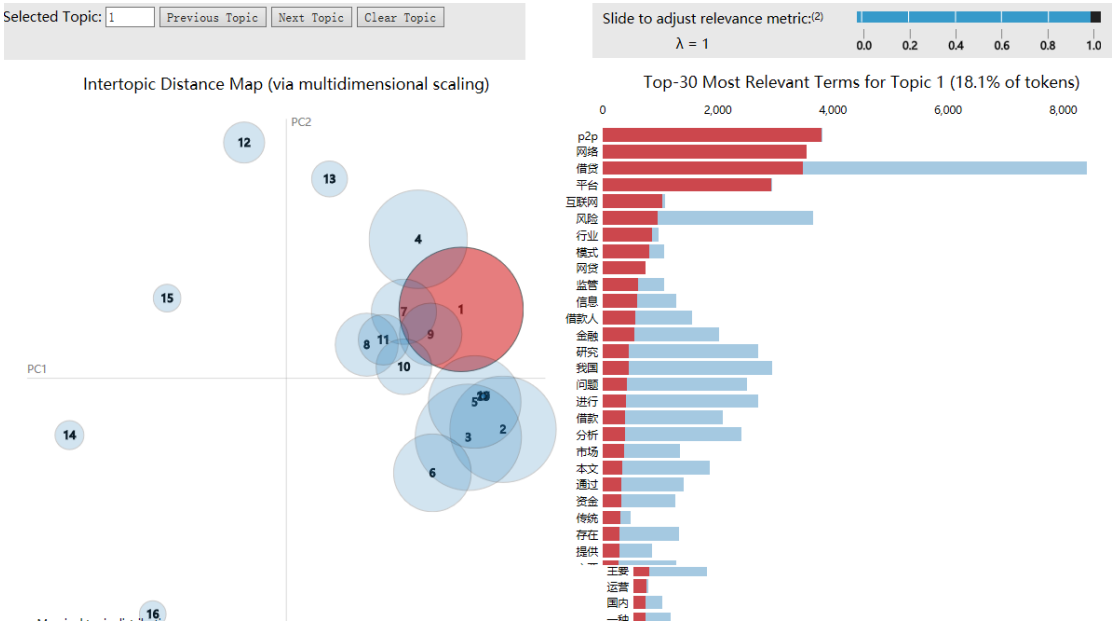

图 5：LDA 算法挖掘的 20 个重要主题可视化分布（lamda 为 1）

表 4：LDA 算法挖掘的 20 个重要主题

由大到小	核心词汇	内容倾向	研究象限
主题 1	p2p 网络 借贷 平台 互联网 风险 行业 模式 网贷 监管 信息 借款人 金融 研究 我国 问题 进行 借款 分析 市场	p2p	1
主题 2	风险 进行 管理 分析 研究 系统 银行 本文 业务 主要 以及 实现 提出 设计 理论 控制 方面 基础 问题 通过	风险	4
主题 3	商业银行 信贷业务 业务 银行 风险 我国 信贷风险 贷款 市场 问题 研究 分行 管理 随着 分析 风险管理 经济 防范 经营 成为	商业银行信贷	4
主题 4	借贷 民间 我国 法律 问题 资金 金融 经济 社会 监管 存在 中小企业 制度 企业 完善 规范 进行 规制 正规 以及	民间信贷	1
主题 5	模型 信用 数据 评估 信用风险 违约 个人信用 进行 借款人 研究 信息 方法 评价 基于 本文 客户 指标 预测 建立 构建	模型和预测	4
主题 6	贷款 抵押 住房 住房贷款 个人住房 借款人 房地产 公积金 银行 担保 按揭 贷款风险 方式 还款 风险 市场 房屋 房价 业务 管理	住房贷款	4
主题 7	法律 公司 合同 担保 问题 部分 保护 征信 小额贷款 分析 关系 认定 报告 进行 行为 笔者 案件 个人信息 机构 效力	法律和合同	1
主题 8	借款 企业 创业 债务 万元 经营 夫妻 合同 资金 自己 小微 共同 单位 利息 偿还 信用社 贷款 自然人 可以 是否	创业和小微贷	1
主题 9	影响 因素 研究 农村 农户 分析 行为 大学生 实证 进行 本文 校园 贷款 金融 意愿 调查 理论 地区 特征 显著	农户和大学生	1
主题 10	消费信贷 消费 个人消费 我国 汽车 经济 居民 需求 消费者 问题 增长 促进 中国 对策 城市 人们 信用 建立 政策 分析	消费贷	1
主题 11	投资者 投资 交易 行为 研究 效应 扣除 个人所得税 融资 进行 影响 偏好 过度 羊群 显著 分析 波动 专项 本文 通过	投资者	1

Cite this paper: 陈媛先. 四十年（1980-2020）来个人借款领域的研究主题变迁-基于文本挖掘 LDA 算法的主题发现和可视化. 社会科学与计算研究研究，2021，卷 1，第 4 期，1-32 页.

主题 12	融资 行为 企业 之间 公民 关系 交易 社会 活动 组织 金融机构 资金 法人 以及 经济 采用 由于 当事人 一定 金融	融资行为	2
主题 13	借贷 私人 利息 费用 国家 助学 农村 这种 教育 之间 农民 支出 债权人 一些 问题 政策 发生 自由 经济 方式	农村和助学	1
主题 14	信贷 资产 租赁 机构 方式 管理 直接 资金 运作 实行 具备 通过 本身 人员 相对 采取 最大 途径 一定 各种	信贷与租赁	3
主题 15	金融 证券化 我国 现金 证券 支持 发行 资产 政府 国际 金融市场 2017 处置 信用 以及 催收 包括 国家 金融机构 正式	证券化	2
主题 16	利率 渠道 存款 银行 货币政策 吸收 公众 超过 银行贷款 水平 通过 货币 人民银行 央行 调控 放贷 合法 已经 背后 最终	利率与渠道	3
主题	借贷 经济 民间 生产 个人消费 消费 增长 推动 部门 美国 金融 拓展 程度 企业 引发 信贷业务 万元 分析 金额 范围		
主题	借贷 民间 利率 我们 范围 信用 探讨 个人信用 国内 农村 理论 性质 形式 如果 来看 逾期 合作 首先 国家 应该		
主题	程度 分配 工作 掌握 法律法规 丰富 方面 评级 价格 分类 传统 应对 日益 信任 变量 依托 存在 市场 视角 模式		
主题	个人住房 住房贷款 信贷业务 国有 保证 商业银行 信贷 对策 资金 存在 担保 合同 银行 困难 住房 目前 国家 必须 许多 需求		

为了分析每个主题的特殊性，我们调整 lamda 为 0.2，强化每个主题下词和主题的独特性，然后再观察前 5 个主题的重要词汇。

在主题 1 中，主要的词汇：P2P、网络、平台、行业等，主要是涉及互联网金融行业的发展相关论文。

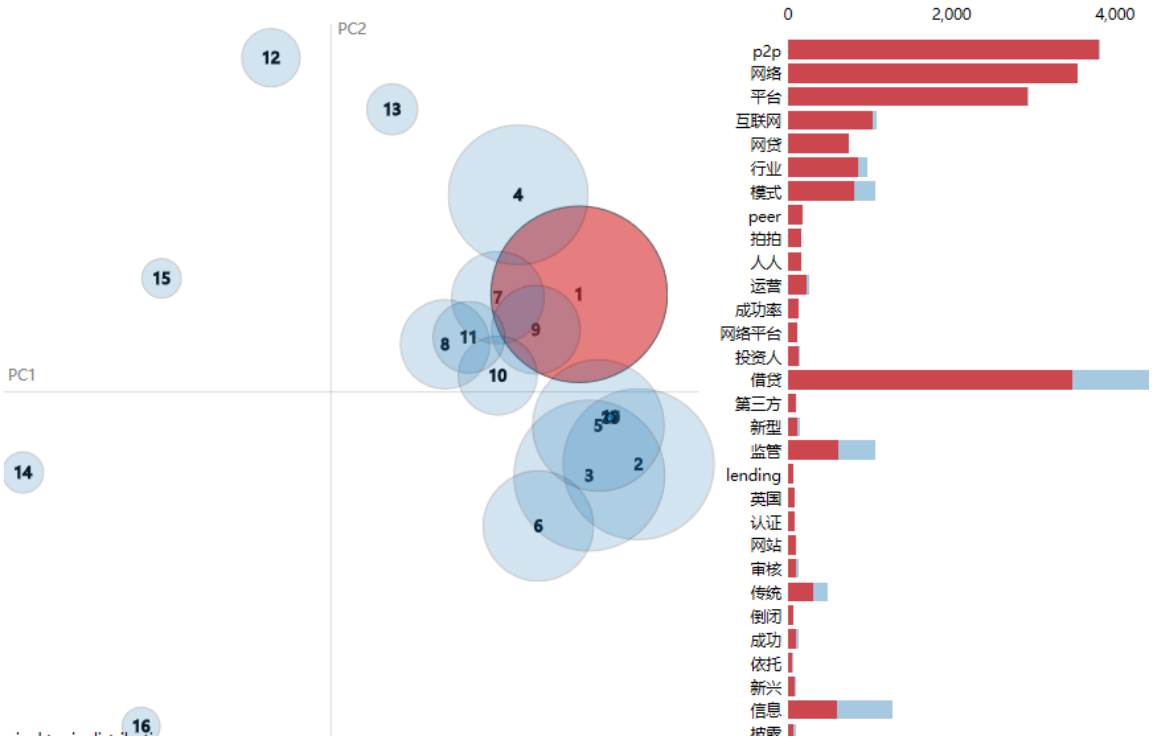

图 6：LDA 算法挖掘的主题 1 可视化分布（lamda 为 0.2）

在主题 2 中，主要的词汇：系统、模块、功能等，因此，可以看到位于象限 4 的主题 2，主要是倾向银行相关的业务系统建设。

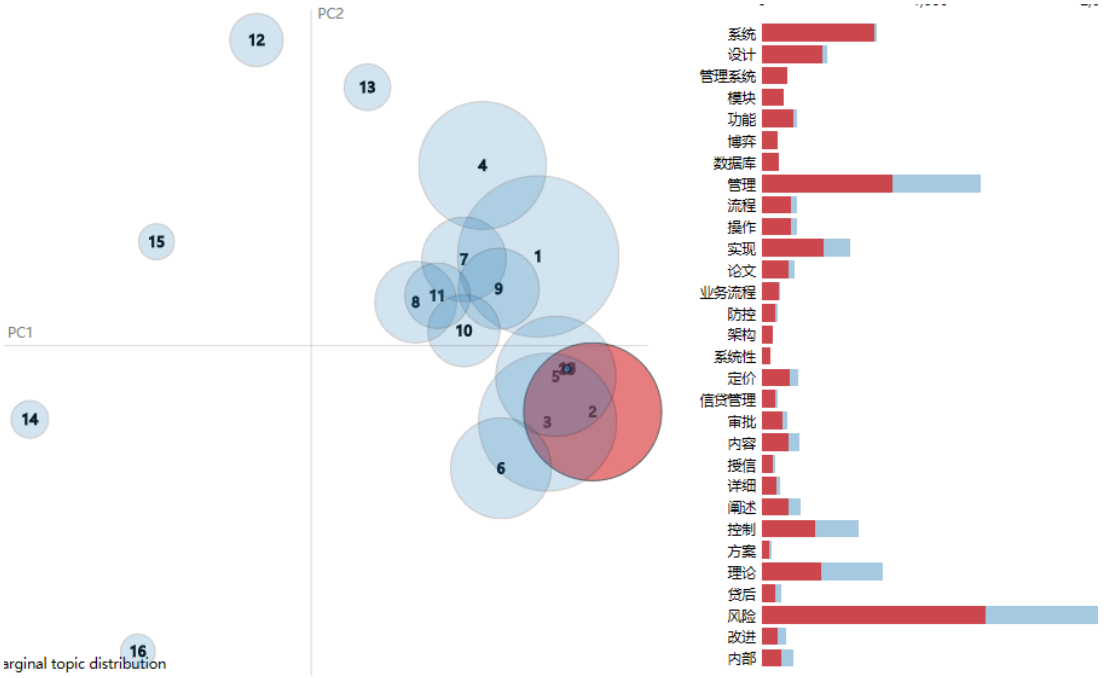

图 7：LDA 算法挖掘的主题 2 可视化分布（lamda 为 0.2）

在主题 3 中，特色词汇主要有：商业银行、信贷业务、信用风险、不良贷款等。该主题主要是研究商业银行信用不良和处置相关论文。

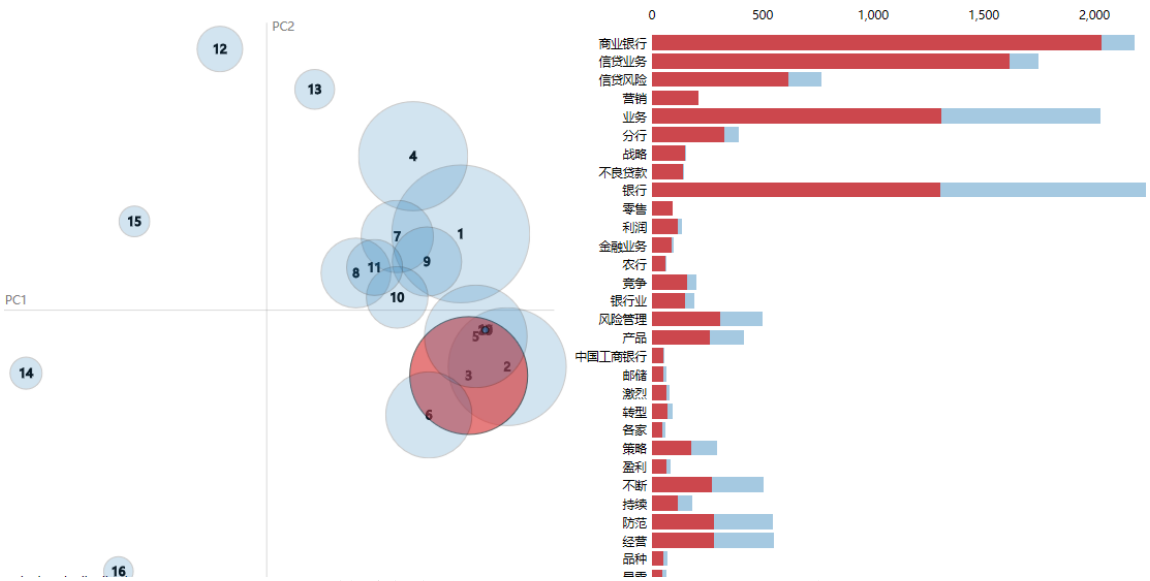

图 8：LDA 算法挖掘的主题 3 可视化分布（lamda 为 0.2）

在主题 4 中，特殊词汇主要是：民间、借贷、规划、中小企业、正规等，主要偏向探讨民间借款合规相关问题。

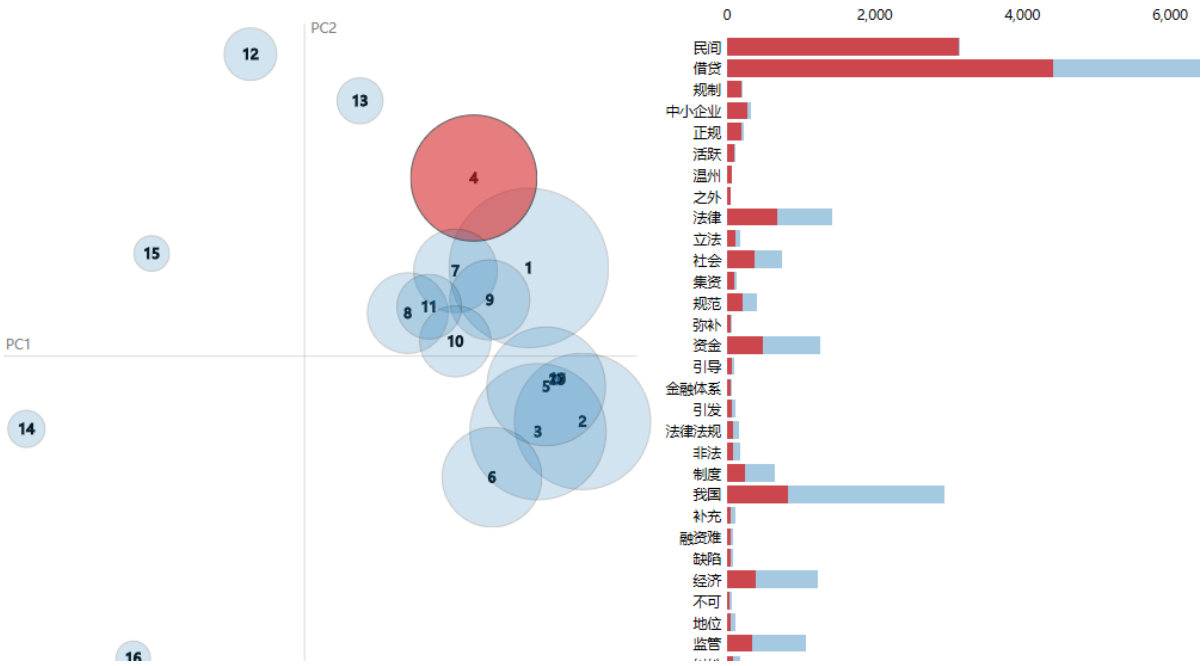

图 9：LDA 算法挖掘的主题 4 可视化分布（lamda 为 0.2）

在主题 5 的词汇中，可以发现，在模型应用中，使用到的模型主要有：回归、Logistics、随机森林、神经网络、决策树、向量等。相关应用大致涉及到信用评估、指标构建、评分和预测。

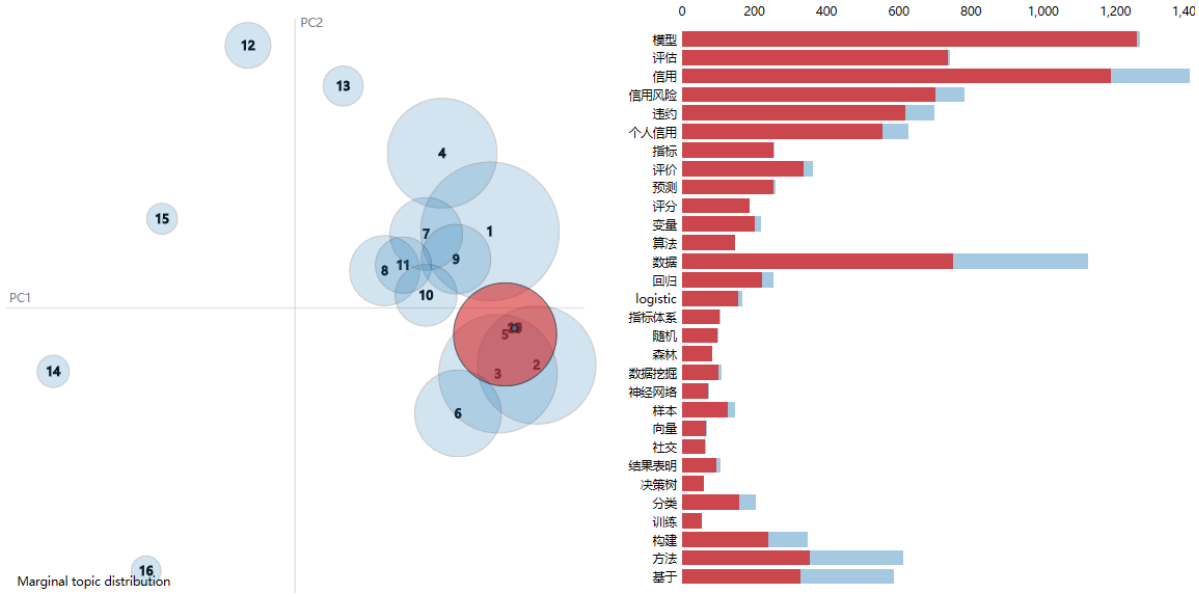

图 10：LDA 算法挖掘的主题 5 可视化分布（lamda 为 0.2）

4.3 按个人借款涉及领域分析

4.3.1 融资可得性

文献中涉及“中小企业”的文章主要分布在第一象限。近年来，中小企业融资难是一个突出的问题，金融如何扶持中小企业发展成为一个比较关键的课题。从数据可以看到，学者们对于通过支持中小企业的发展研究主要集中在创新金融领域，重点在 P2P 和民间金融，和 P2P（主题 1）、民间借贷（主题 4）和借款与创业（主题 8）相关。而在银行领域，相关的研究主要和风险相关（主题 3）。

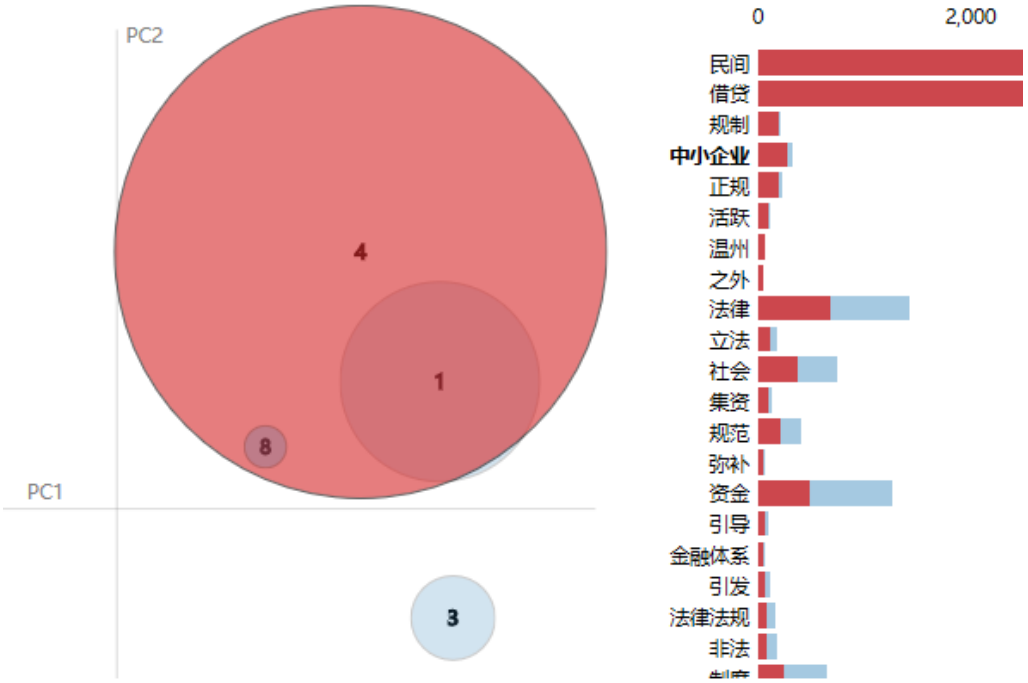

图 11：主题“中小企业”可视化分布

“小微”的相关研究主要分布在象限 1，和借款与创业（主题 8）、P2P（主题 1）相关。而在象限 4 中，主要研究和系统相关（主题 2）。

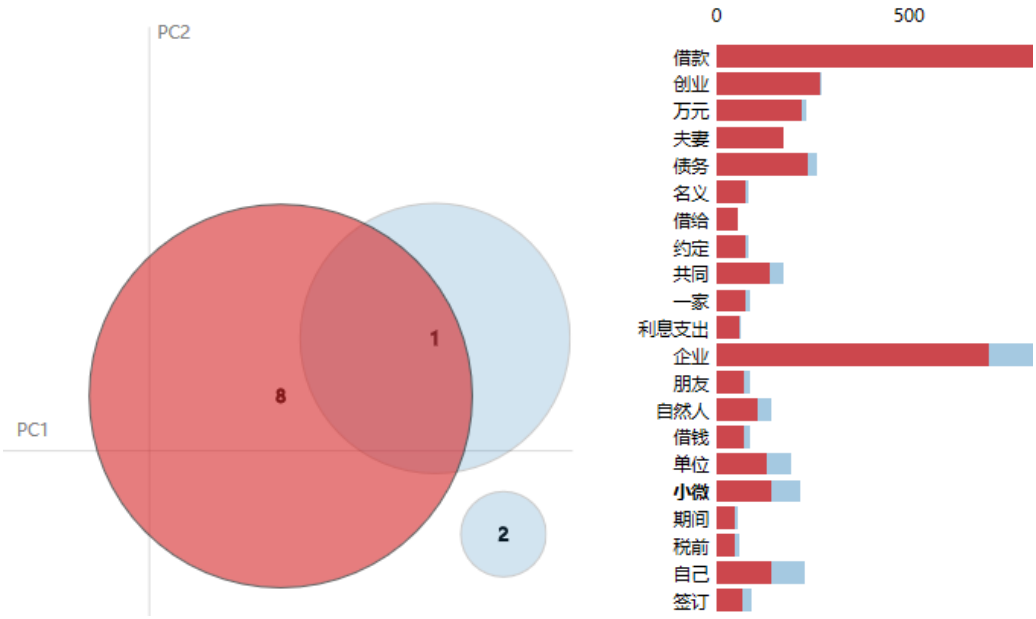

图 12：主题“小微”可视化分布

Cite this paper: 陈媛先. 四十年（1980-2020）来个人借款领域的研究主题变迁-基于文本挖掘 LDA 算法的主题发现和可视化. 社会科学与计算研究研究，2021，卷 1，第 4 期，1-32 页.

“融资难”出现在第 3 和第 2 象限，在第三象限主要和渠道相关（主题 14），而在第而象限主要和 P2P（主题 1）、民间借贷（主题 4）、合同和保护（主题 7）、农村和大学生（主题 9）相关。可以看到，农村和大学生存在融资难的问题，而在解决这个问题上，学者们探讨通过 P2P 和民间借贷作为解决渠道，但是相关的问题，比如合同和保护可能是比较重要的问题。

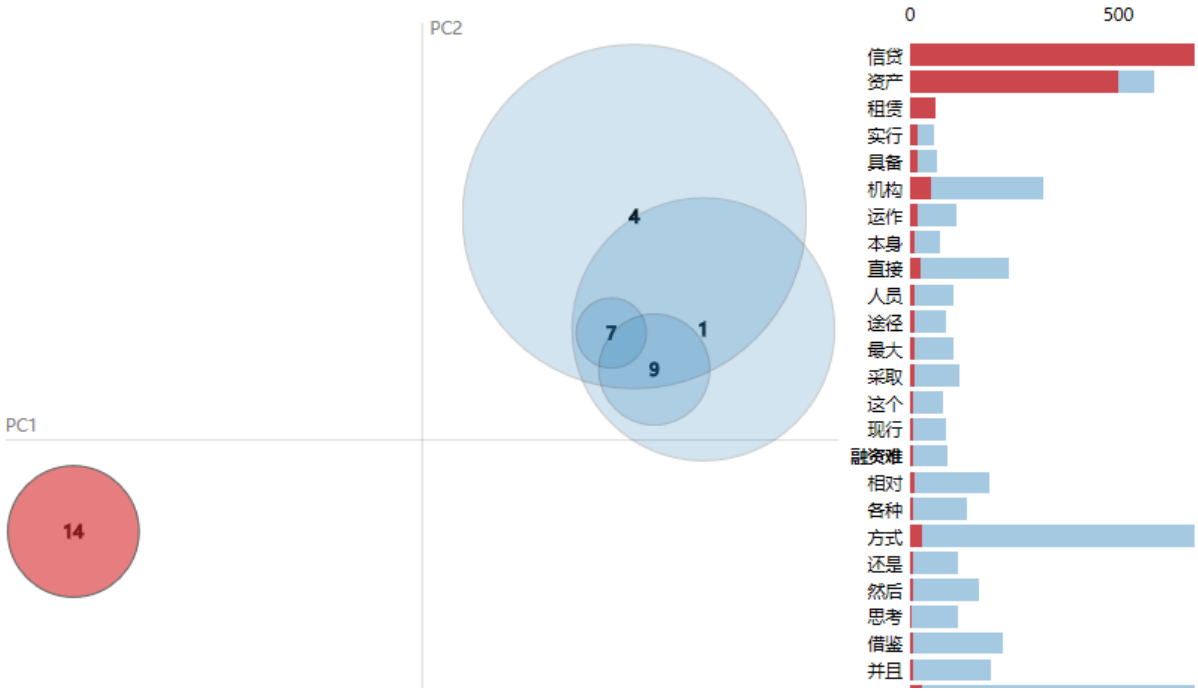

图 13：主题“融资难”可视化分布

“担保”的相关研究在第 1 象限主要是合同与保护（主题 1）、借款与创业（主题 8）和民间融资（主题 4）相关。在类似 P2P 等新的融资渠道借款用于支持创业与投资，如何提供担保和健全合同相关机制成为学者关注的重要课题。而在第 4 象限主要和房产抵押（主题 6）相关。

Cite this paper: 陈媛先. 四十年（1980-2020）来个人借款领域的研究主题变迁-基于文本挖掘 LDA 算法的主题发现和可视化. 社会科学与计算研究研究，2021，卷 1，第 4 期，1-32 页.

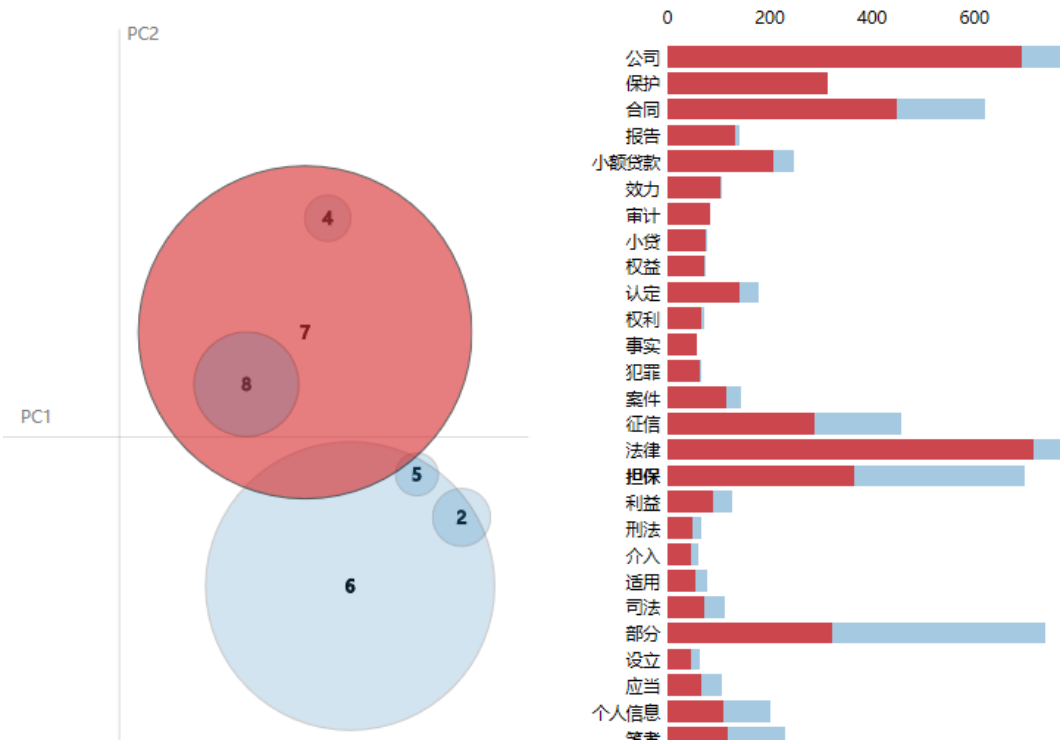

图 14：主题“担保”可视化分布

4.3.2 消费观念

“消费观念”相关的研究主要分布在第 1 和第 4 象限。在第 1 象限中，主要和消费信贷（主题 10）、农户和大学生（主题 9）相关。在象限 4，主要和商业银行信贷风险相关（主题 3）。

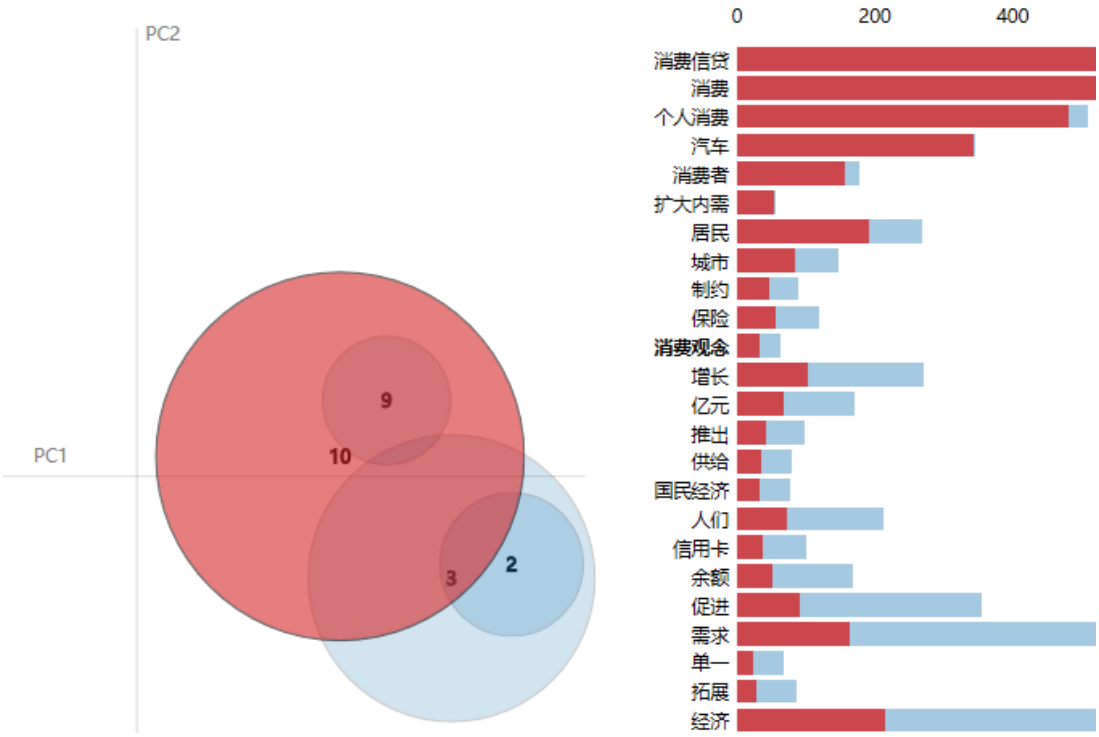

图 15：主题“消费观念”可视化分布

Cite this paper: 陈媛先. 四十年（1980-2020）来个人借款领域的研究主题变迁-基于文本挖掘 LDA 算法的主题发现和可视化. 社会科学与计算研究研究，2021，卷 1，第 4 期，1-32 页.

“转变”的主题分布大致和“消费观念”一致，但是研究涉及更广，还和民间借贷（主题 4）和 P2P（主题 1）相关。因此，可以看到，在消费观念转变上，大部分的研究倾向将消费观念转变的资金需求和民间借贷、P2P 进行关联研究。

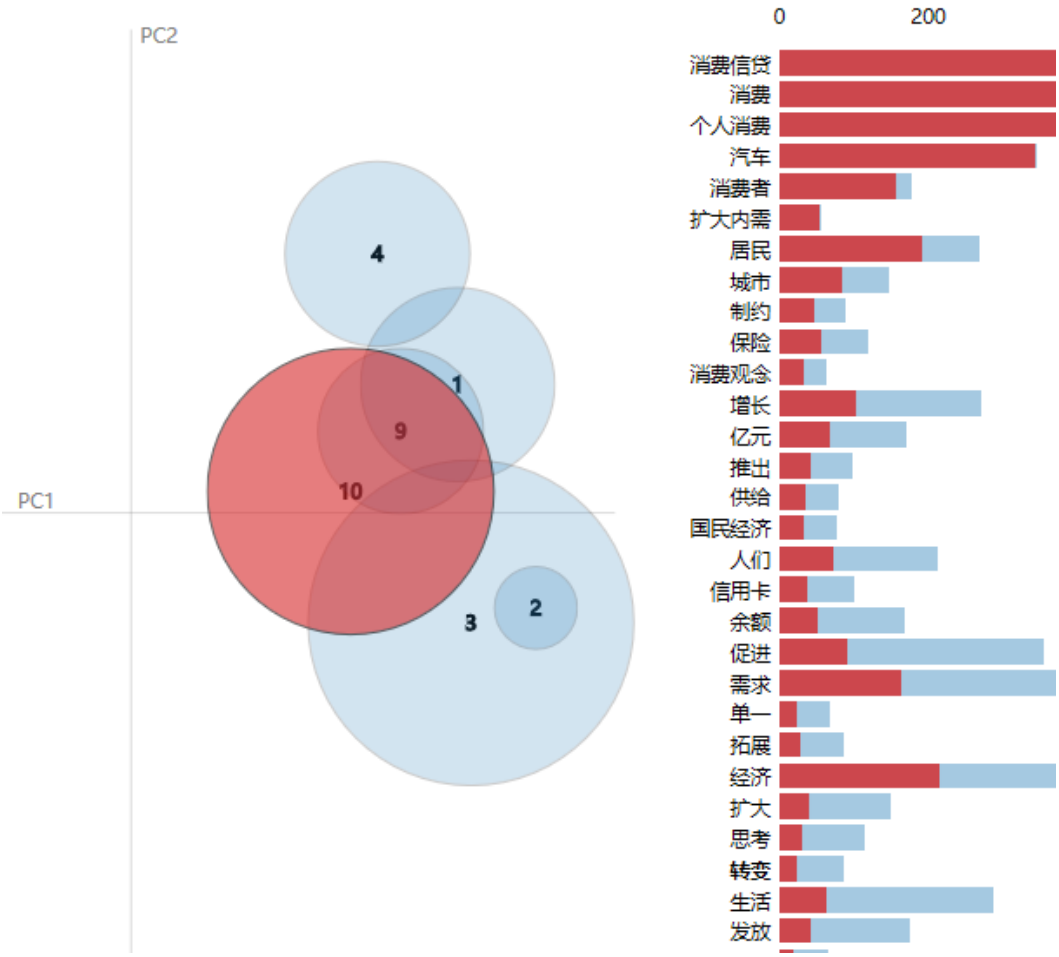

图 16：主题“转变”可视化分布

4.3.3 合规、调控、监管

关于“利率”的研究主要分布在象限 3，和渠道相关（主题 16）。此外，在第 1 象限中，P2P 的相关主题，比如行业、农民和农村、创业和投资也和利率相关。在象限 4 中，和利率相关的研究主要和房贷相关。

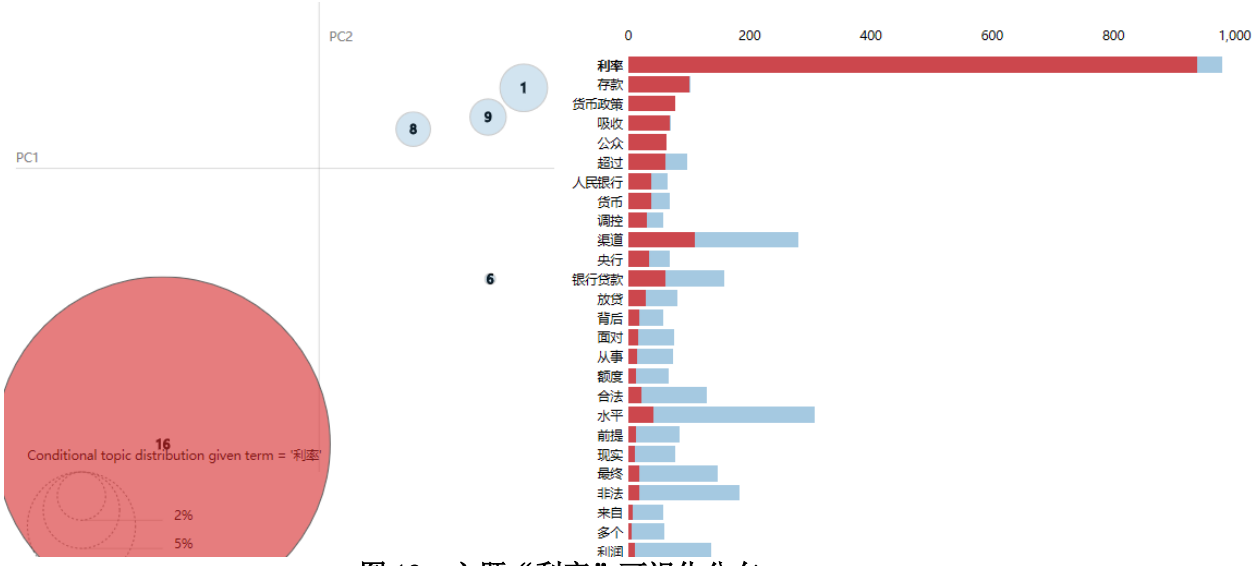

图 13：主题“利率”可视化分布

和“调控”相关的主题主要相关文章主要分布在渠道（主题 16）、民间金融（主题 4）相关、房贷（主题 6）、风险（主题 3）相关。

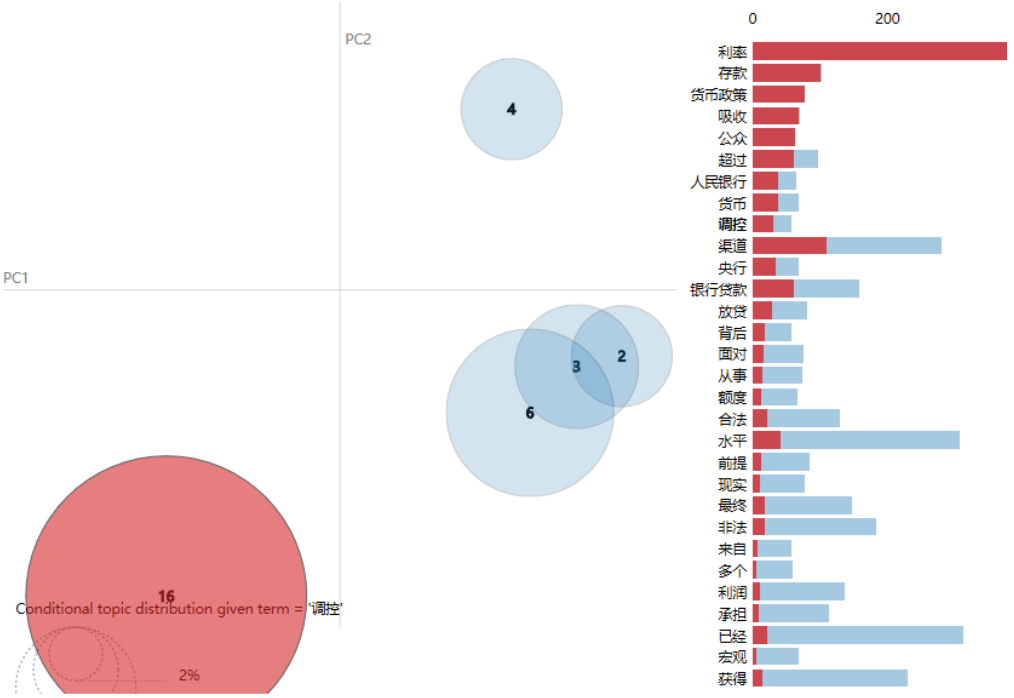

图 18：主题“调控”可视化分布

“规范”的词汇主要分布在象限 1，即研究法律和规范相关的文章主要和 P2P 相关，涉及行业、民间融资和合同关系等主题。涉及到象限 4 的主要和系统及抵押贷款主题相关，更倾向于内部管理和具体产品。

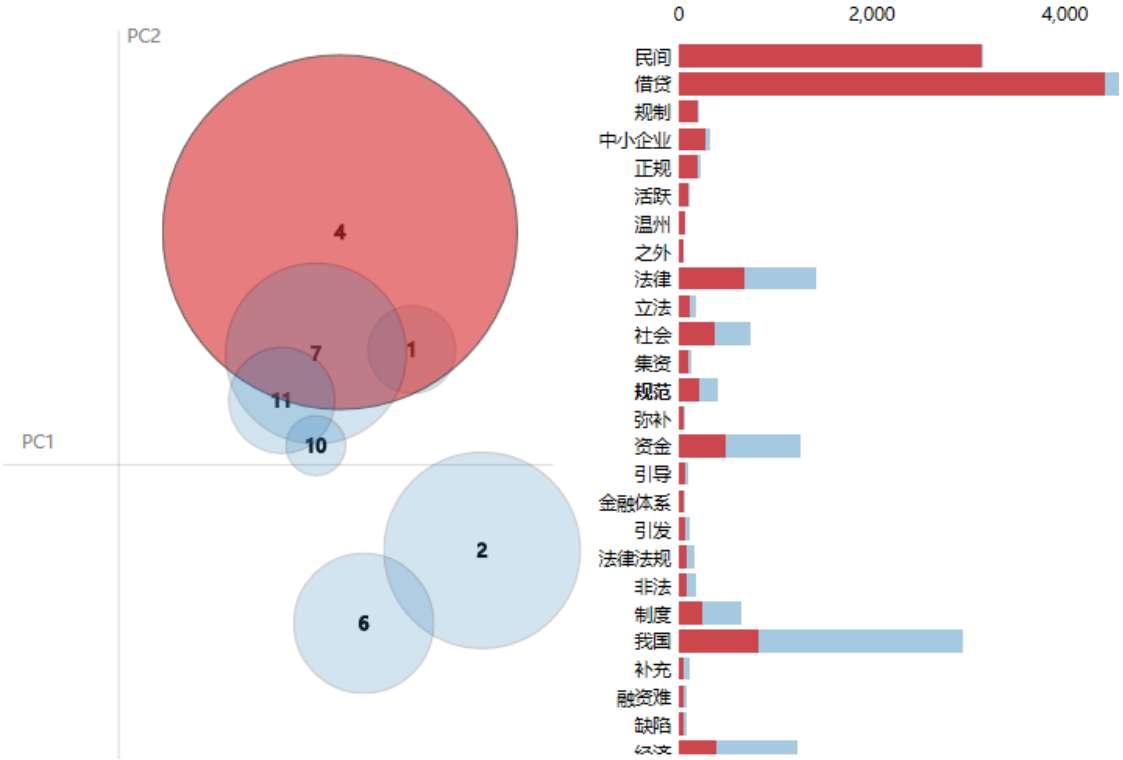

图 19：主题“规范”可视化分布

涉及“监管”主要分布在象限 1，主要和 P2P（主题 1）、民间借贷（主题 4）、合同与保护（主题 7）、投资者（主题 11）相关。

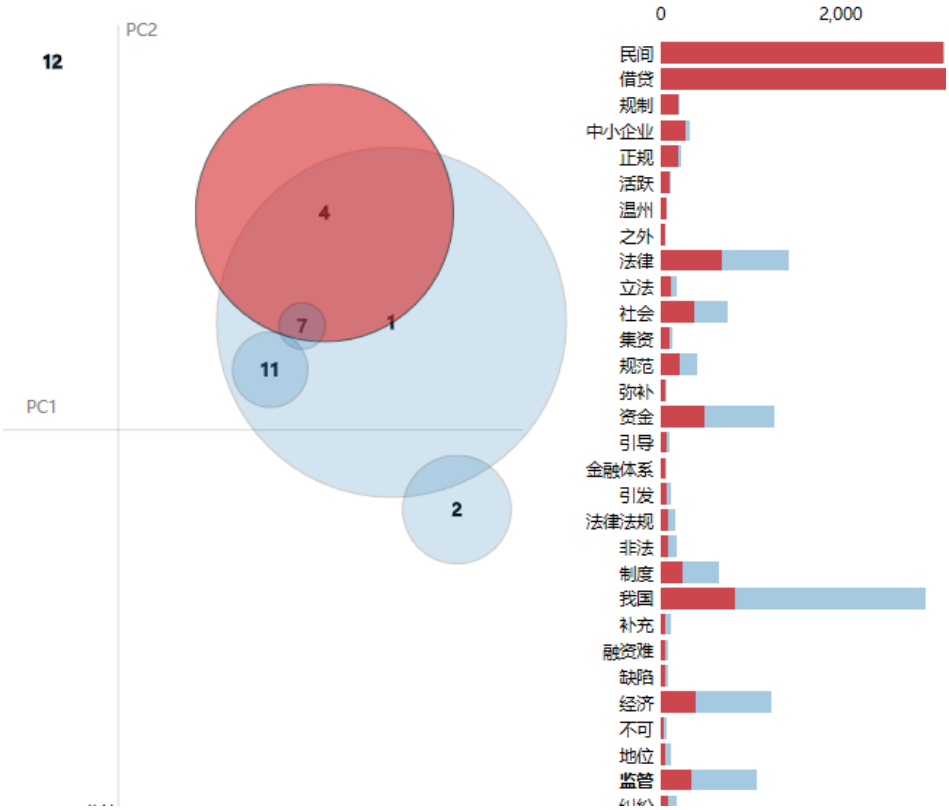

图 20：主题“监管”可视化分布

4.3.6 大数据、人工智能、模型

Cite this paper: 陈媛先. 四十年（1980-2020）来个人借款领域的研究主题变迁-基于文本挖掘 LDA 算法的主题发现和可视化. 社会科学与计算研究研究，2021，卷 1，第 4 期，1-32 页.

我们选择 P2P 词汇进行分析，可以看到 P2P 相关的研究主要落入象限 1，部分落入象限 4。

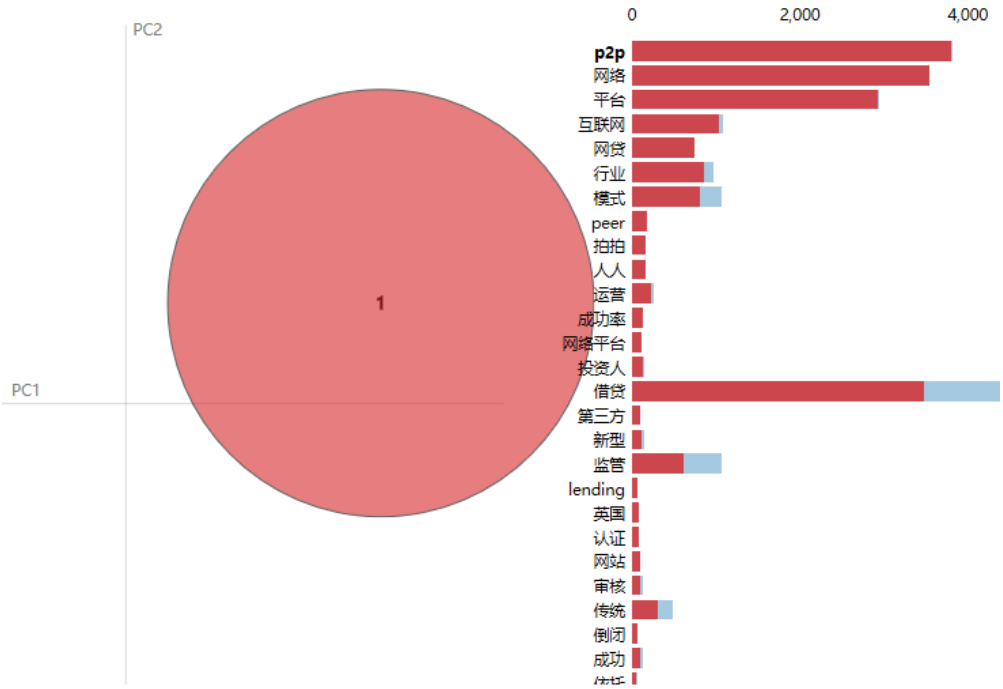

图 21：主题“P2P”可视化分布

选择“银行”词汇进行分析，可以看到该词主要如如象限 4，部分落入象限 1。

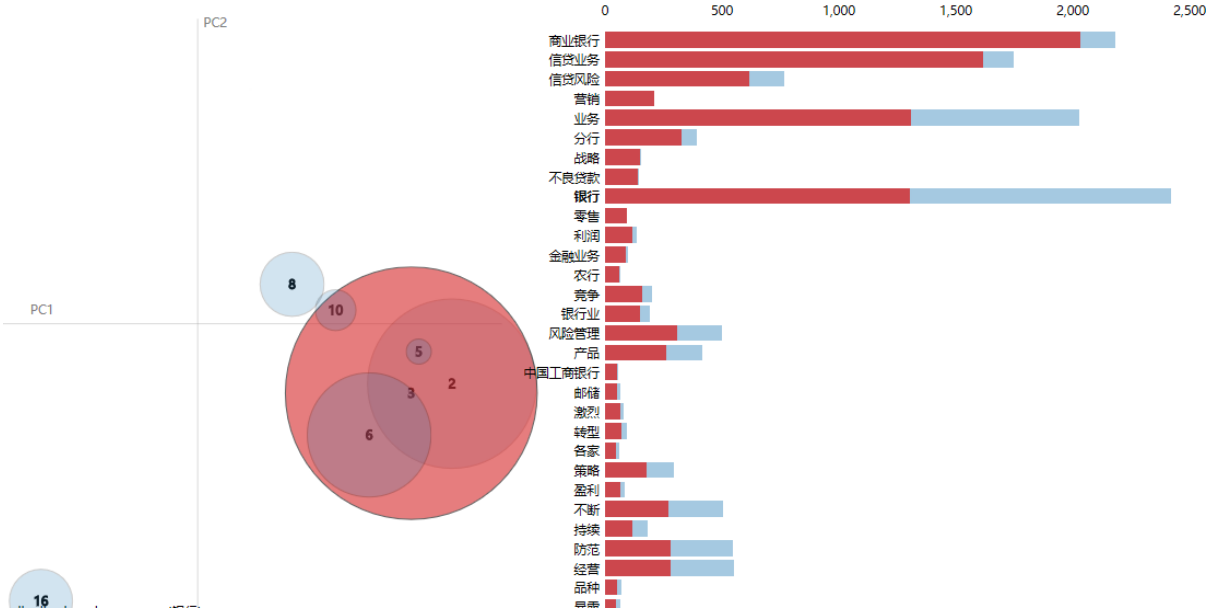

图 22：主题“银行”可视化分布

4.3.6 法律、立法

涉及“法律”的词汇主要分布在象限 1，即研究法律和规范相关的文章主要和 P2P 相关，涉及行业、民间融资和合同关系等主题。涉及到象限 4 的主要和系统及抵押贷款主题相关，更倾向于内部管理和具体产品。

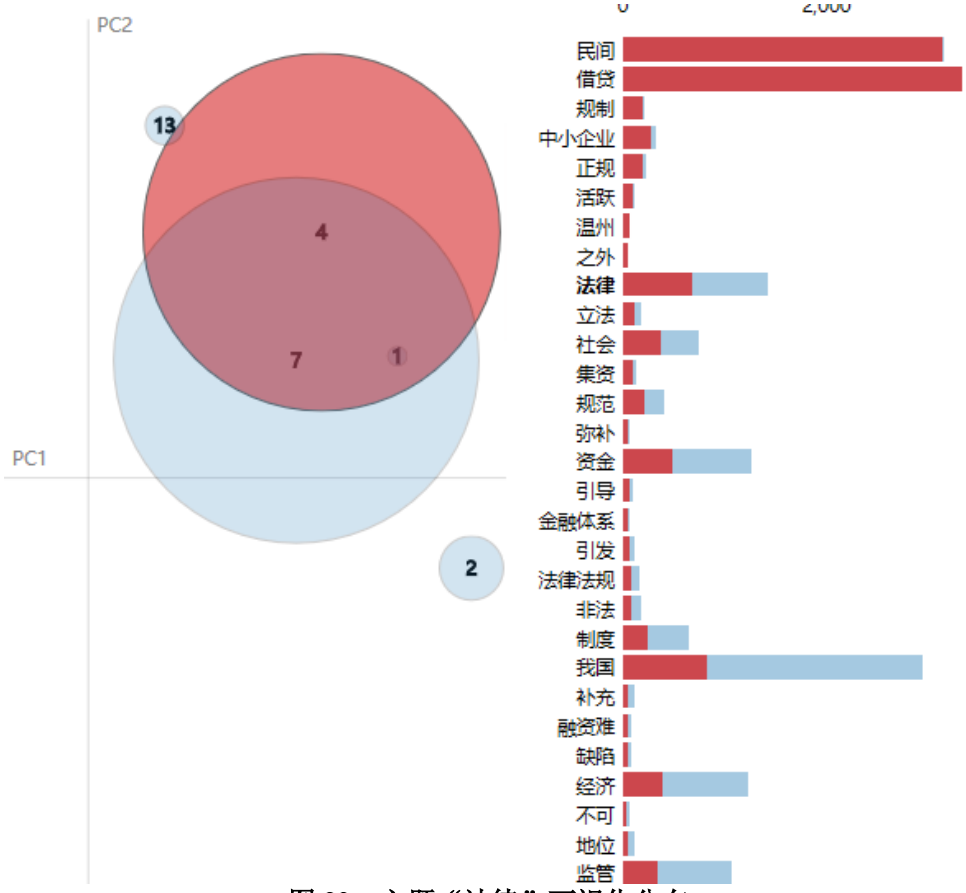

图 23：主题“法律”可视化分布

出现“争议”最多的研究分布在第 1 象限，主要涉及合同与保护（主题 7）、借款与创业（主题 8）、P2P（主题 1）。可以看到，出现“争议”最多的研究领域与涉及“担保”最多的研究大部分是重合的，也即涉及到复杂金融合同关系的领域，目前也是研究最多的领域。

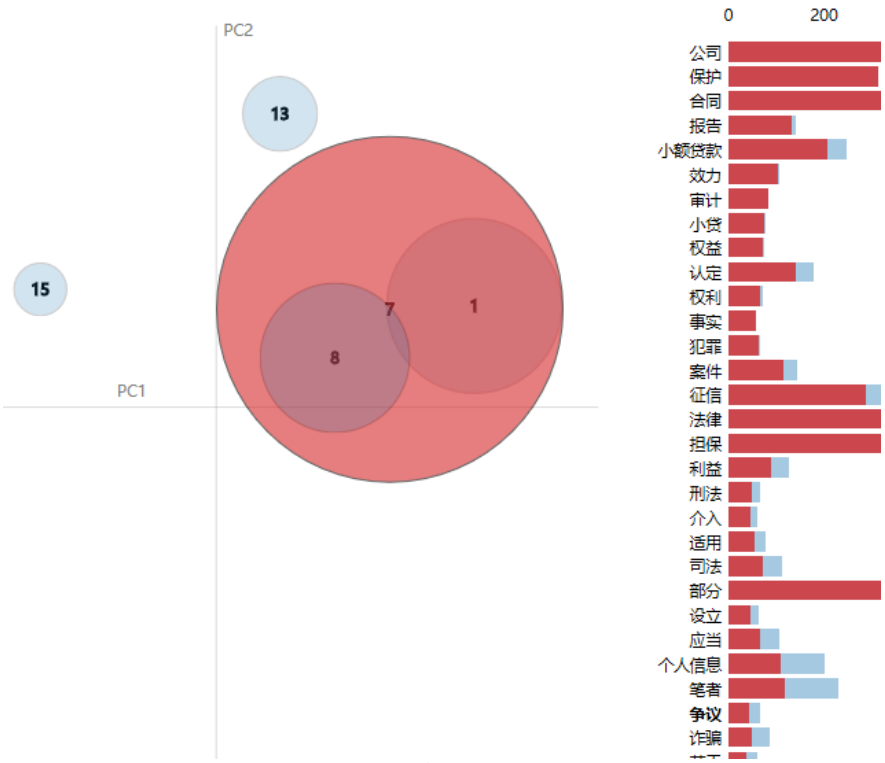

图 24：主题“争议”可视化分布

民间借贷、P2P 等容易游走在法律边缘，甚至滋生犯罪行为（如套路贷），涉及“诈骗”更多的研究也是分布在第 1 象限，和 P2P、民间借贷、合同与保护、农村和大学生主题相关。

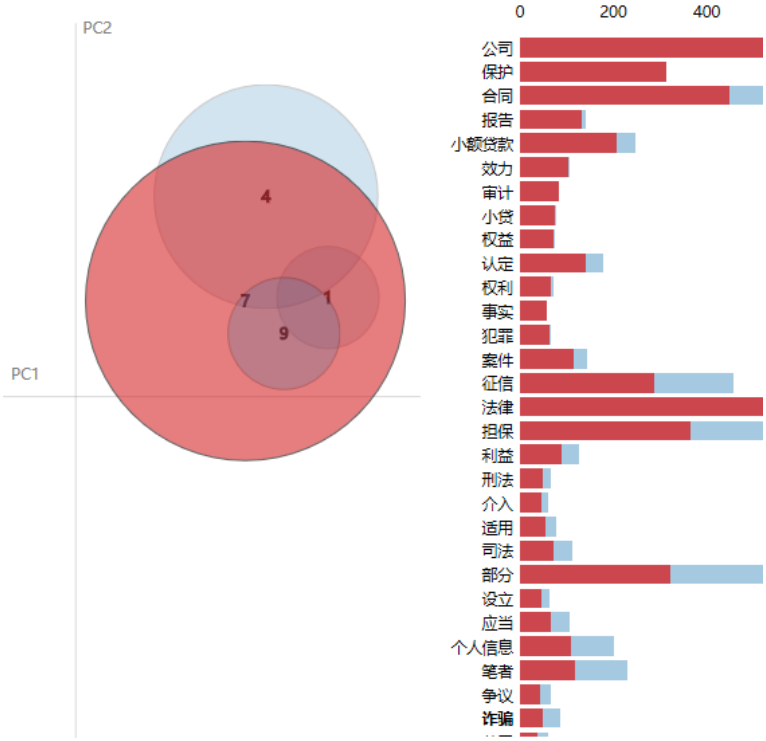

图 25：主题“诈骗”可视化分布

而涉及到“立法”具体诉求的文章主要分布在象限 1。

Cite this paper: 陈媛先. 四十年（1980-2020）来个人借款领域的研究主题变迁-基于文本挖掘 LDA 算法的主题发现和可视化. 社会科学与计算研究研究，2021，卷 1，第 4 期，1-32 页.

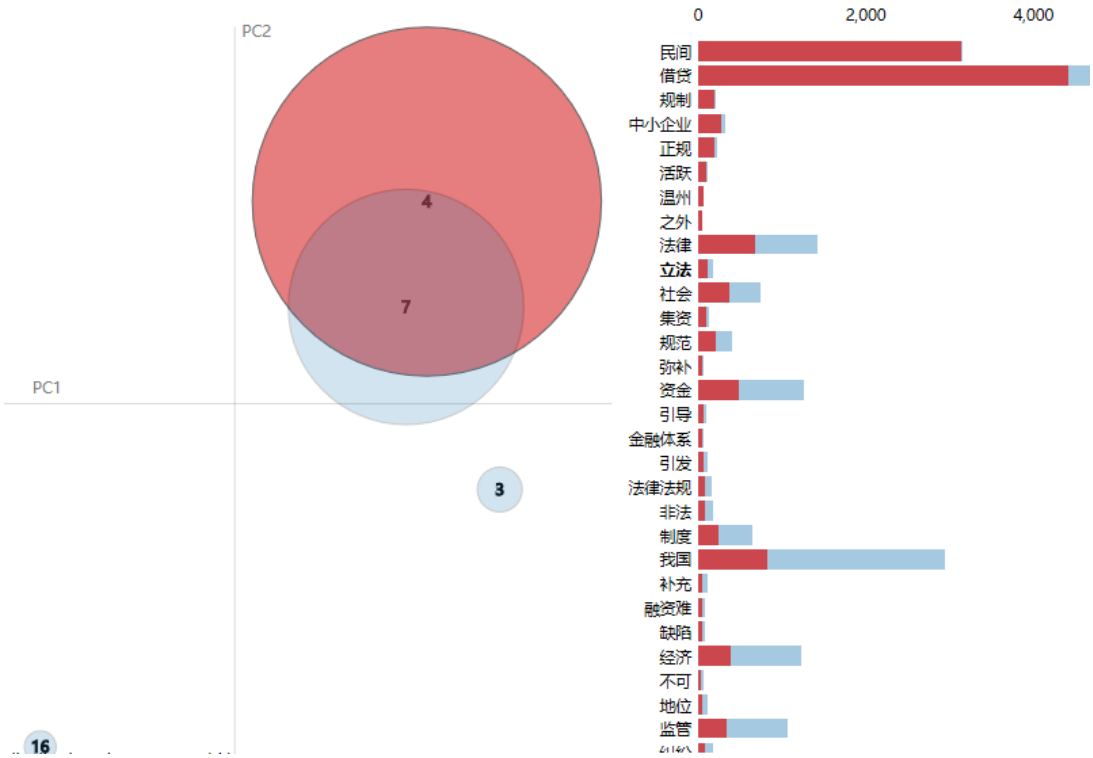

图 26：主题“立法”可视化分布

4.6.7 证券和证券化

涉及“证券”的主要分布在象限 2、象限 1 和部分分布在象限 4。

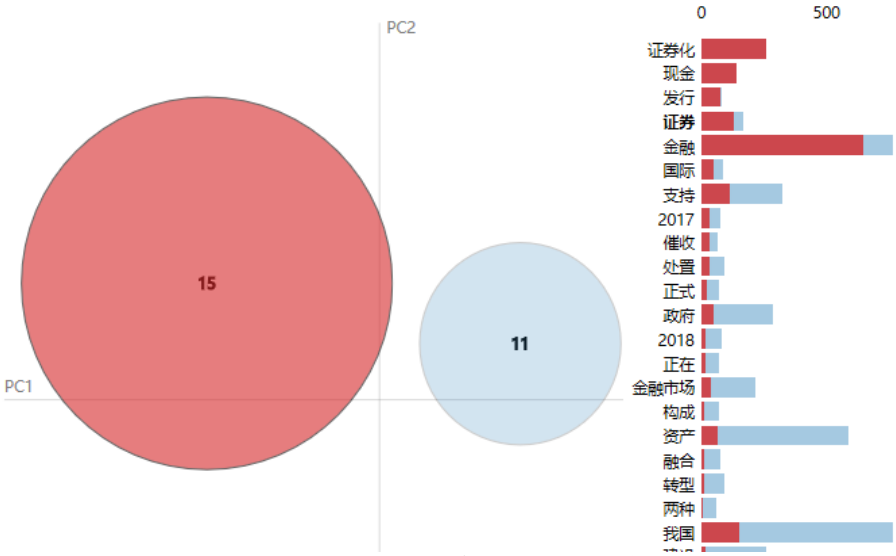

图 27：主题“证券”可视化分布

而“证券化”主题主要分布在象限 2、象限 1，象限 4 几乎不涉及。

Cite this paper: 陈媛先. 四十年（1980-2020）来个人借款领域的研究主题变迁-基于文本挖掘 LDA 算法的主题发现和可视化. 社会科学与计算研究研究，2021，卷 1，第 4 期，1-32 页.

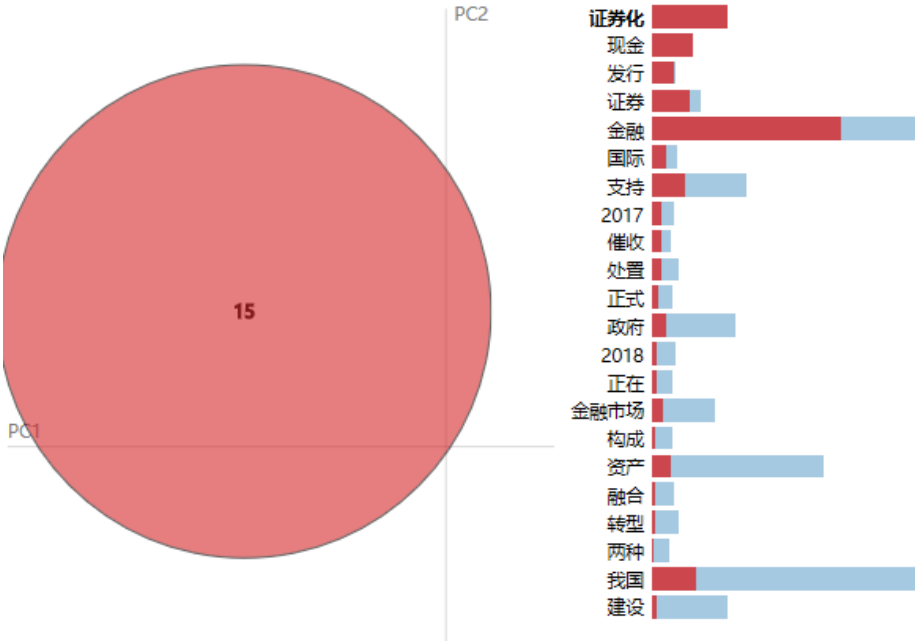

图 28：主题“证券化”可视化分布

4.3.4 资产处置

涉及“处置”主要分布在象限 1、象限 2 和象限 4. 其中，象限 1 关于 P2P 领域的处置主要和投资者相关，即投资者的保护和利益维护。而在象限 2 中，主要和证券化相关，即部分文章研究通过资产证券化的方式进行贷款处置。而在象限 4 中，主要涉及住房贷款和系统建设，即，在银行体系中，主要的资产处置和住房贷款不良相关，而系统如何支持不良资产的处置也是银行体系的一个重要研究课题。

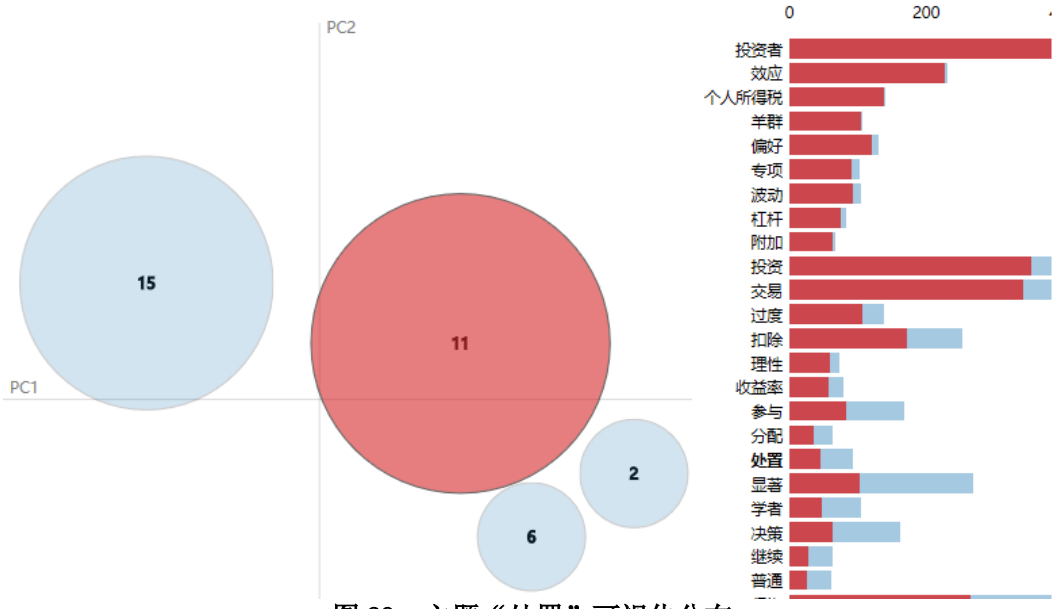

图 29：主题“处置”可视化分布

在“催收”方面，最相关的研究主题分布在第 2 象限证券化（主题 15）和象限 1 的公司合同保护（主题 7）。资产进入证券化后，催收工作和回款流程更复杂，而在 P2P 领域，随着暴力催收的出现，如何合理、合法催收，成为新金融领域的课题。

Cite this paper: 陈媛先. 四十年（1980-2020）来个人借款领域的研究主题变迁-基于文本挖掘 LDA 算法的主题发现和可视化. 社会科学与计算研究研究，2021，卷 1，第 4 期，1-32 页.

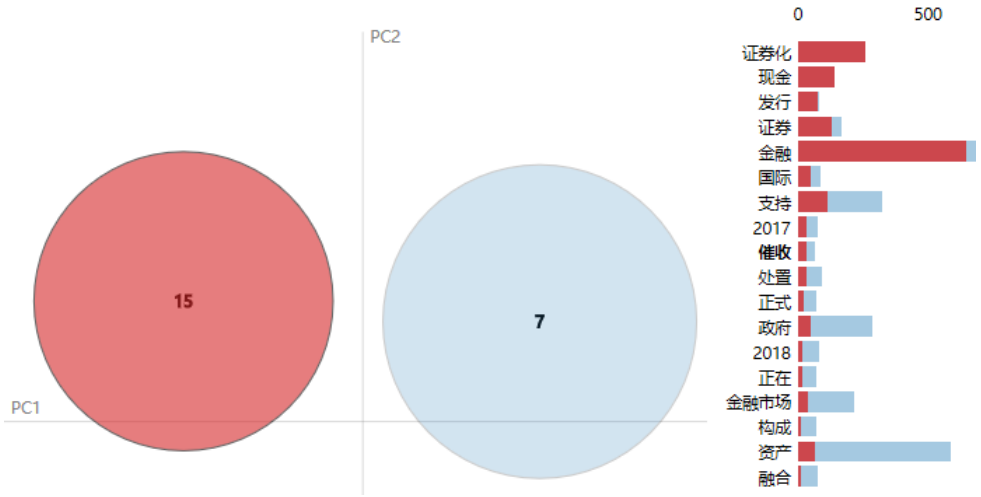

图 30：主题“催收”可视化分布

4.6.8 投资和行为金融

可以看到，“投资”这个词主要出现在主题 11 和主题 1。

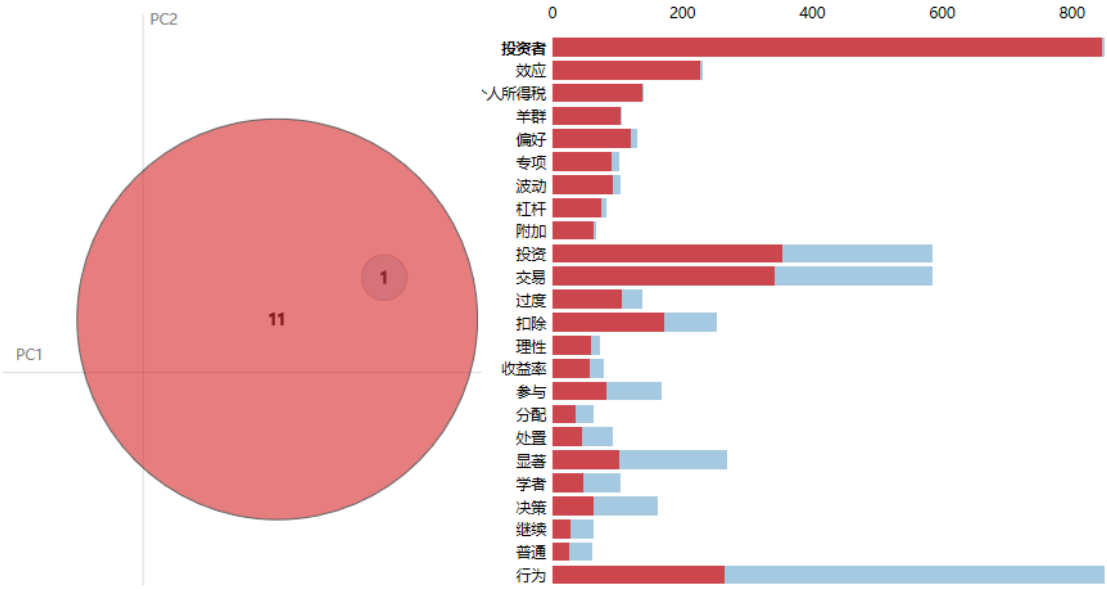

图 31：主题“投资”可视化分布

“行为”这个词主要出现在主题 11、12 和 9 中。

Cite this paper: 陈媛先. 四十年（1980-2020）来个人借款领域的研究主题变迁-基于文本挖掘 LDA 算法的主题发现和可视化. 社会科学与计算研究研究，2021，卷 1，第 4 期，1-32 页.

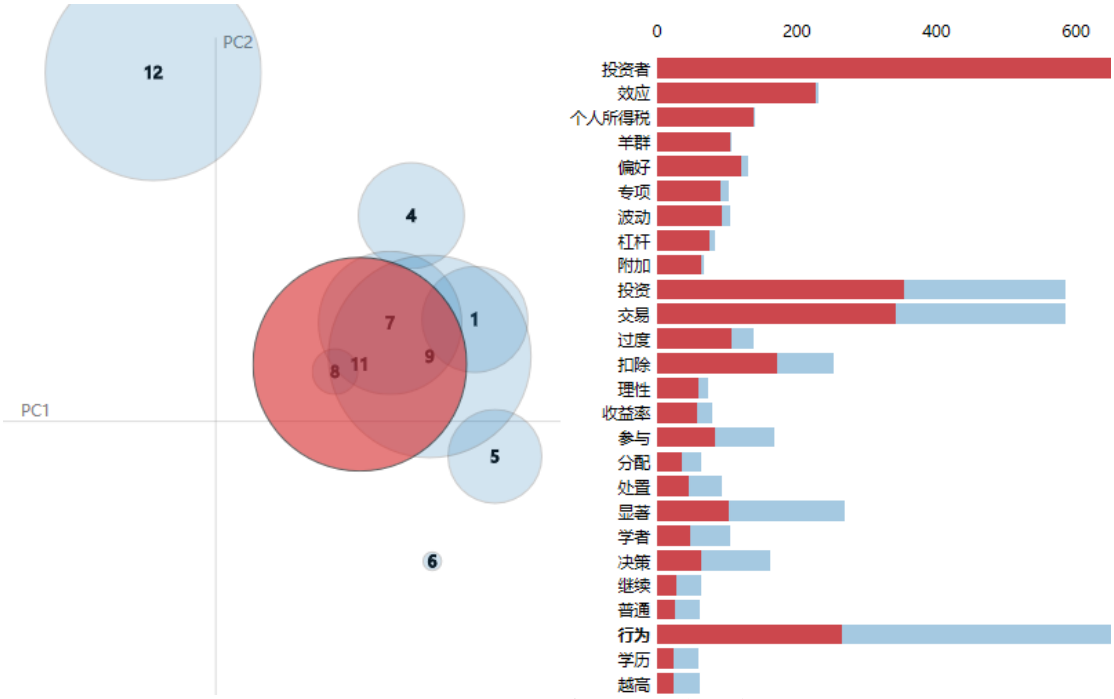

图 32：主题“行为”可视化分布

“偏好“主要出现在象限 1 和象限 4.，主要和投资者（主题 11）和农村及大学生（主题 9）相关。

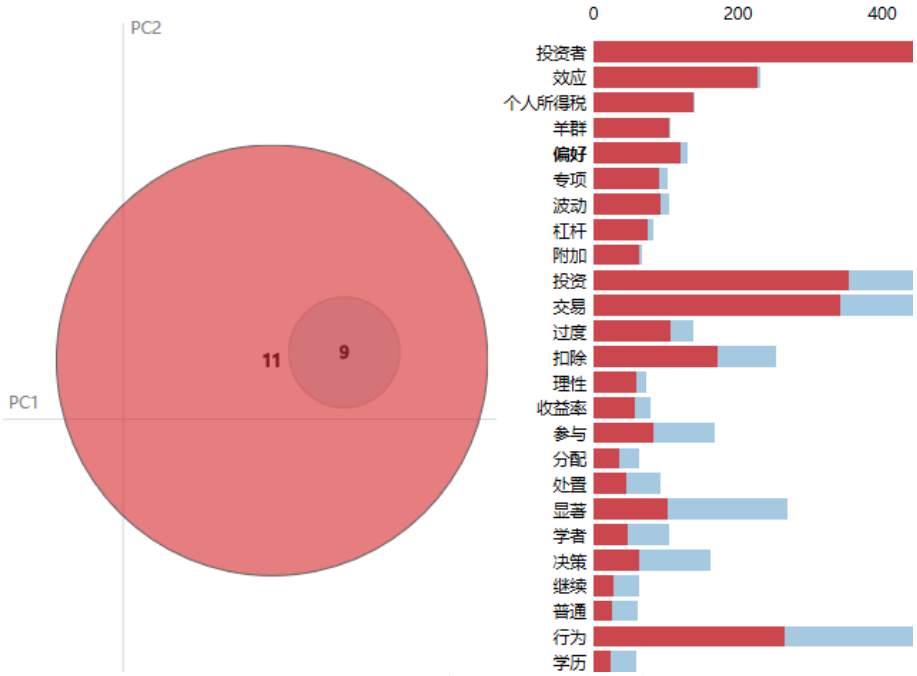

图 33：主题“偏好”可视化分布

“理性“词汇主要分布在象限 1 和 4.主要和投资者、民间借贷、农户及大学生、住房贷款、系统相关。

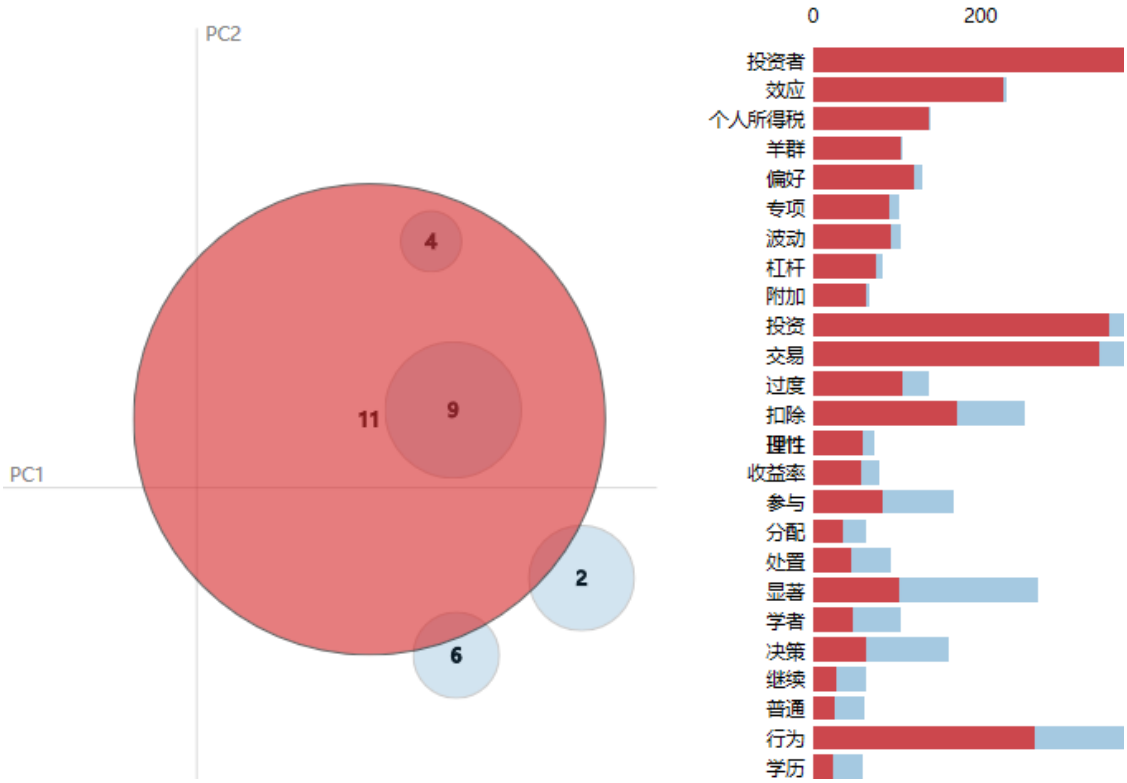

图 34：主题“理性”可视化分布

从以上“行为“、”理性“、”偏好“来看，行为金融的研究主要是在 P2P 的领域，其中，涉及到投资者和农村大学生。在 P2P 的发展中，校园贷曾经火热一时，大村大学生通过校园贷进行提前消费的方式成为研究的重点。

4.4 研究主题的变迁

为了研究个人借款领域研究主题的变迁，本文将论文按照三个不同的时期分别进行了归类，并去掉了没有采集到年份信息的论文，分类如下：

表 5：按时期统计发表论文

年度	论文数	占比
1980 年-2000 年	197 篇	9.19%
2001 年-2010 年	556 篇	25.94%
2011 年-2020 年	1390 篇	64.86%

因为进行分类后，每个时期的论文数量变少，为了更精准聚集主题，我们将分析的主题数量限定在 10 个，并且每个主题提炼词汇数也为 10 词语，按年度主题和对应词语如表 6。表 6 第一列按照主题概率大小排列。

如表 6 显示，在 1980-2000 年区间，个人借款相关的论文主题主要和国家、形式、探讨、运营、会计等相关（主题 6、7、8、9、10），显示在这个区间内，研究处于探索的阶段，国家的整个关于个人借款的发展还在早期。

Cite this paper: 陈媛先. 四十年（1980-2020）来个人借款领域的研究主题变迁-基于文本挖掘 LDA 算法的主题发现和可视化. 社会科学与计算研究研究，2021，卷 1，第 4 期，1-32 页.

如表 6 显示，在 2001-2010 年区间，在这个区间，住房贷款研究的主题（主题 7、9）开始凸显，这个我们这个时期内房地产市场的快速发展有紧密的关系。与此同时，和企业相关的主题开始占据主要部分（主题 2、4、5、6 均和企业相关）。在企业相关主题的研究中，创业借款开始凸显，企业借款中涉及个人信用部分的研究开始凸显，比如，企业负责人的信用问题。伴随着贷款业务的发展，关于风险和保险的探讨也变得重要（主题 1、8）。在这个阶段，信用卡和消费信贷、汽车相关贷款的研究也开始凸显重要性（主题 3、10）。

如表 6 显示，在 2011-2020 区间，个人借款的研究开始和“个人”进行了更多的结合，个贷、大学生贷款、创业贷款等研究增多（主题 8、10），结合小微企业的研究也增多（主题 9）。在这期间，P2P 等网络贷款的研究重要性提升（主题 1），民间金融和相关法律的研究也凸显（主题 4）。风险管理的相关研究更广泛，涉及商业银行信贷风险（主题 2）和个人征信（主题 7）等，而随着大数据和人工智能算法，借助模型和数据进行信用评估等的研究也逐渐增多（主题 3）。

三个时期研究主题和对应词语的变化揭示了随着我国经济发展，研究方向的变化趋势。

表 6：不同时期的主题

时期	1980-2000	2001-2010	2011-2020
主题 1	借贷 民间 **农村** 之间 经济 资金 发展 行为 **我国** 公民	信贷业务 商业银行 发展 我国 风险 消费信贷 信贷 业务 **信贷风险** 分析	借贷 **p2p** 网络 平台 风险 互联网 借款人 行业 信息 **网贷**
主题 2	借款 合同 企业 利息 **单位** 双方 借贷 贷款 万元 规定	借贷 **民间** 借款 **企业** 农村 法律 合同 私人 经济 资金	**风险 商业银行** 业务 信贷业务 我国 信贷 银行 问题 管理 分析
主题 3	消费信贷 消费 **发展** 信贷 经济 我国 信贷业务 金融 银行 中国	贷款 借款人 风险 个**人住房** 抵押 消费 银行 住房贷款 消费信贷 **汽车**	**模型 数据** 信用 进行 研究 评估 违约 融资 特征 评价
主题 4	贷款 借款人 **住房** 住房贷款 银行 个人住房 抵押 利率 **办法** 期限	融资 **企业 中小企业** 银行 公司 模式 金融 业务 市场 自然人	**借贷 民间 法律** 我国 金融 问题 资金 社会 企业 融资
主题 5	资金 **管理 经营** 企业 借款 私人 方式 发展 由于 信用社	贷款 **创业** 利率 **经营** 万元 自己 如果 融资 一家 资金	**抵押 住房贷款** 影响 资产 房地产 个人住房 农户 研究 因素 分析
主题 6	法律 借贷 保护 合法 **规定** 关系 合同 借款 私人 之间	借款 **企业** 评估 **费用** 贷款 处理 个人信用 问题 财务 信用	**银行** 业务 **系统** 客户 分行 信贷 进行 管理 实现 产品
主题 7	办理 抵押 房地产 住房贷款 手续 贷款 借款人 费用 对于 商业银行	信贷 消费 市场 中国 **住房** 城市 个人住房 金融 关系 信息	公司 住房 **公积金 征信 信用** 小额贷款 金融 融资 制度 资金
主题 8	**会计** 借贷 笔者 同志 要求 **探讨** 一个 比较 重要 情况	**保险** 支付 影响 很大 支持 还是 各项 巨大 借款	**担保 大学生 校园 创业** 债务 合同 夫妻 借款 房屋 法律
主题 9	借贷 贷款 住房贷款 利率 民间 行为 **农村** 消费信贷 借款人 经济	住房 办法 建设 管理 **住房贷款** 中国人民银行 有关 运用 偿还 满足	借款 企业 逾期 扣除 个人所得税 **小微 用途 经营** 费用 利息
主题 10	国家 包括 我国 社会 我们 实践 **形式 探讨** 目前 支付	**信用卡** 个人信用 银行 三个 信用 信贷 方面 自身 借款 办法	信贷业务 加快 业务 **个贷** 方面 二是 资产 积极 有着 促进

我们将三个时间的主题进行可视化，lamda 设置为 0.2，图示分别为图 35、图 36 和图 37。

从 1980-2000 年的研究主题看，研究主题比较分散、4 大主题之间彼此距离也很大。从 2001-2010 年的研究主题看，研究主题开始聚拢。而从 2011-2020 的研究主题看，研究主题进一步聚拢，几大主题形成一个彼此交互的研究群，即，研究的热点比较集中。

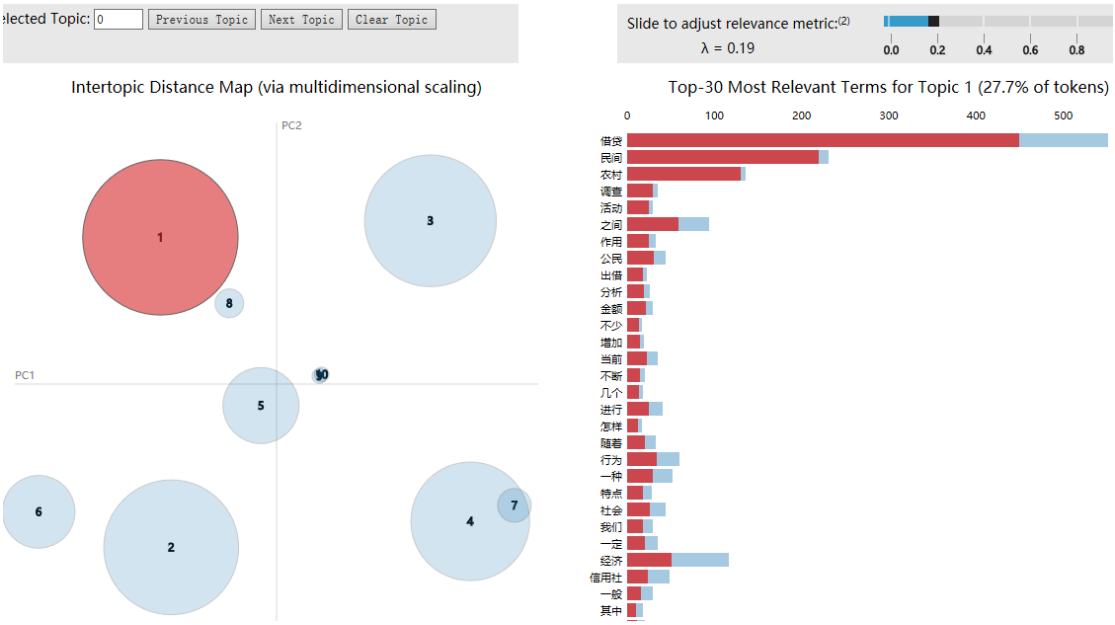

图 35：1980-2000 年区间重要研究主题可视化分布

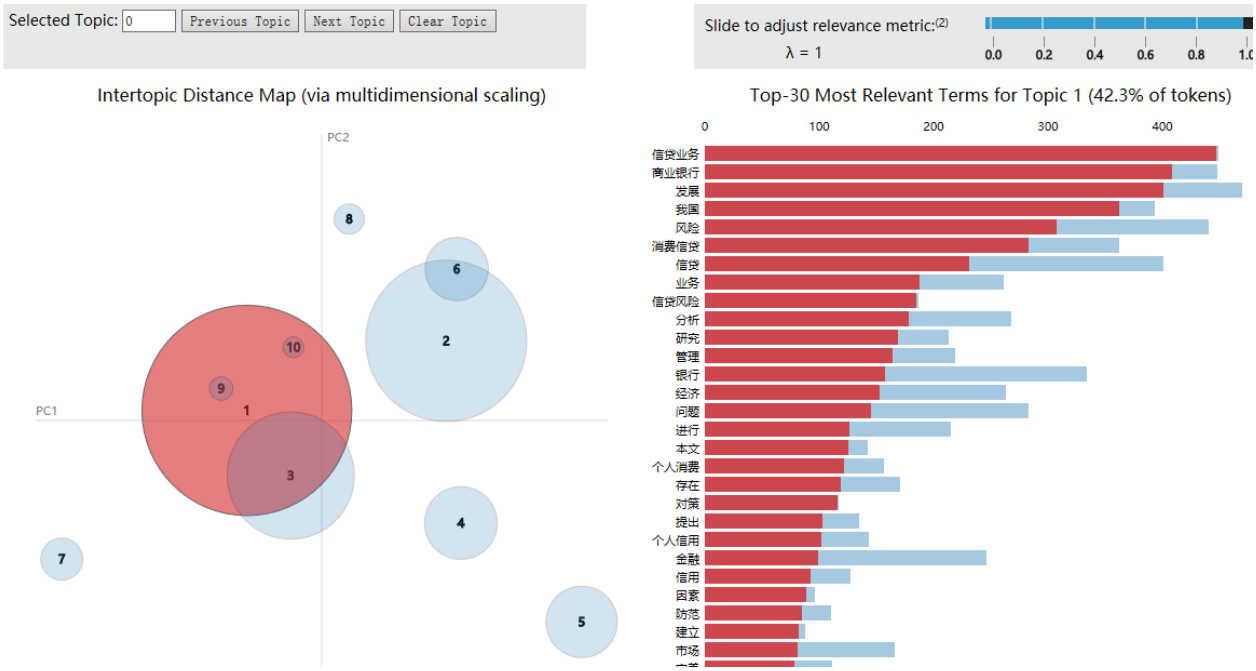

图 36：2001-2010 年区间重要研究主题可视化分布

Cite this paper: 陈媛先. 四十年（1980-2020）来个人借款领域的研究主题变迁-基于文本挖掘 LDA 算法的主题发现和可视化. 社会科学与计算研究研究，2021，卷 1，第 4 期，1-32 页.

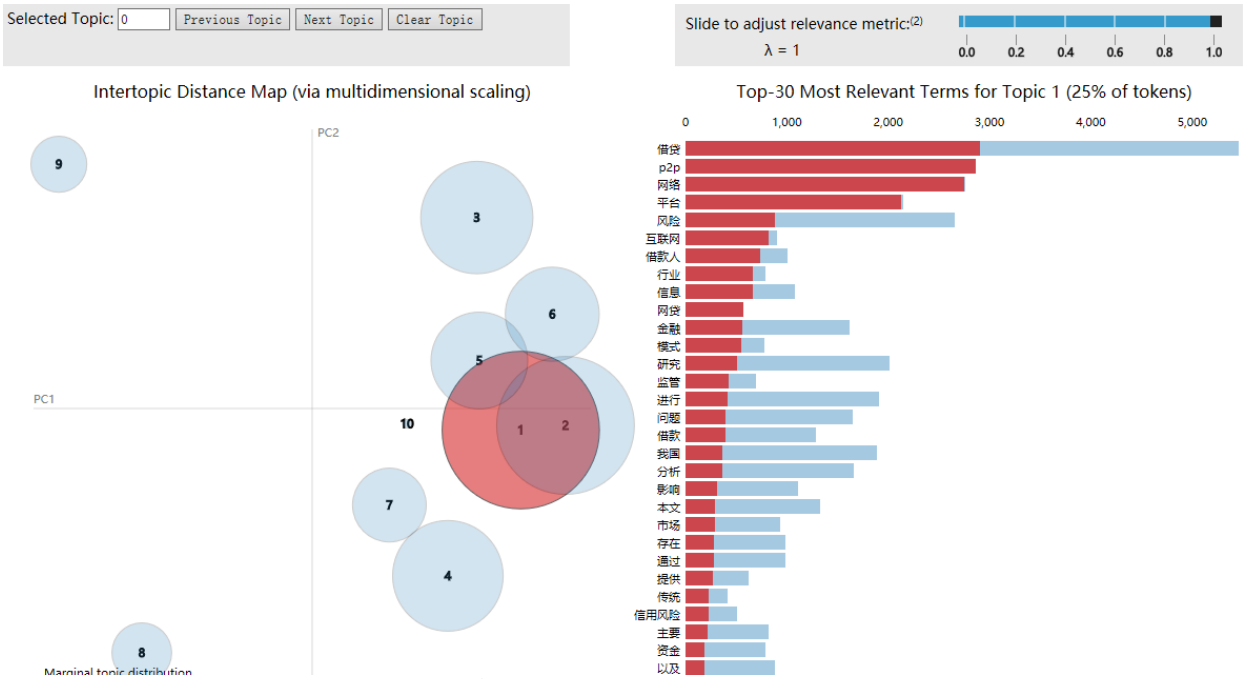

图 37：2011-2020 年区间重要研究主题可视化分布

4.5 未来的研究主题分析

社会的发展带来了个人借款领域更多的热点问题和继续解决的难题，为了分析未来个人借款的研究方向，本文使用 LDA 技术对今年社交领域、新闻领域提到的个人借款相关信息进行分析，提炼近期的探讨重要主题。

在这个部分，我们合并了两个来源的数据进行分析。数据一，来自社交媒体新浪微博；数据二，来自正式的新闻报道。通过两个数据源的分析，我们可以对比分析两个领域的主题差异。

首先，微博是更大众化的讨论平台，通过微博文章，可以看出大众当前更关心的问题。作者使用"个人借款"作为关键词，收集到微博最近几年文章数量 418 篇（时间从 2017 年 9 月 11 日到 2020 年 6 月 12 日）。使用 LDA 方式分析，设置关键词数据，并且将 lamda 设置为 0.2，更凸显每个主题下的词的特殊权重。在发现的 5 个主题，排列在前面的相关词和内容侧重点分别如下表 7 所示。

表 7：近期微博关于个人借款的讨论主题

主题	重要词
主题 1	公积金 深圳 自查 **违规 流入**
主题 2	征信 风险 能力 审核 放款
主题 3	**夫妻 共同 债务 贷款 一方**
主题 4	**诈骗 视频 约定 理财 收益**
主题 5	借贷 执行 民间 失信 纠纷

其中，主题 1 涉及借款借款相关的资金违规流入房地产市场的问题，比如，面向个体的大额消费贷实际流入房地产市场的问题。而主题 3 中，夫妻共同债务的讨论也是个人借款领

Cite this paper: 陈媛先. 四十年（1980-2020）来个人借款领域的研究主题变迁-基于文本挖掘 LDA 算法的主题发现和可视化. 社会科学与计算研究研究，2021，卷 1，第 4 期，1-32 页.

域一个新的问题。在主题 4 中，涉及个人借款资产和打包后的理财收益问题。这三个主题反应出最近 3 年民众关注的问题，但是，在之前的个人借款学术研究中涉及较少，可能是未来的研究方向。

其次，来自正式渠道的新闻报道更多反应了正式纳入媒体关注和政府关注的话题。作者采用“个人借款”为关键词，收集到最近媒体报道的新闻一共 223 篇（时间从 2019 年 10 月 19 到 2020 年 6 月 12 日），使用 LDA 方式分析，设置关键词数据，并且将 lamda 设置为 0.2，更凸显每个主题下的词的特殊权重。在发现的 5 个主题，排列在前面的相关词和内容侧重点分别如下表 8 所示。

表 8：近期新闻报道关于个人借款涉及的主题

主题	**重要词**
主题 1	**亿元　逾期　增长** 征信 **余额**
主题 2	企业　创业　就业　万元　最高
主题 3	利率　定价　基准　消费　业务
主题 4	公积金　住房 **调整　疫情　政策**
主题 5	**公司　质押　经营　股份　股东**

表 8 显示，在新闻报道中，位于主题 1 的是贷款余额和逾期的增长，显示了个人借款业务发展中，伴随业务增长，逾期问题的关注度获得很多的关注。主题 4 凸显了在新型冠状肺炎中，公积金政策的一些调整也是新闻关注的重点。而主题 5 中，可以看到关于公司经营过程中，股东质押贷款等相关问题也是重要的媒体关注重点。

从以上两个数据源的主题挖掘结果看，普通民众更关注和自己相关的法律问题、权益问题，而新闻报道中，更多的是关注和公司、行业相关的问题。

5．结论

伴随着中国经济的快速发展，个人金融领域的发展呈现蓬勃的态势。在这期间，个人房产抵押贷款、信用卡消费贷款、互联网金融、个人创业贷款等不同形式的借款涌现出来，在不断发展的同时，也出现了很多的问题，需要学者们进行关注和深入研究。回顾研究历史，能帮助研究人员更好的掌握研究的脉络，也能更好的协助领域内的学者找到研究的空白，从而推动领域内的研究不断向前。基于此目的，本文研究了 40 年来(1980-2020 年)我国在“个人借款”研究主题的变迁和重要主题的相关性。

借助机器学习 LDA 主题分析算法，作者对 2000 多篇文章进行了主题挖掘，并进行了可视化演示和分析。基于分析结果，作者认为：一、在个人借款领域，我国的研究具有比较鲜明的三个特征，其中，1980-2000 年，主要为探索期；2001-2010 年为个人住房贷款、企业相关的个人贷款重点研究期间，而 2011 年到 2020 年为个人借款研究深入的阶段。二、通过对近期微博和新闻报道的主题发现，在个人借款领域，涉及夫妻共同债务、个人借款打包理财相关研究可能是未来的研究重点。三、研究揭示，学术界、业界关注的内容可能差异，普通民众和新闻媒体关注的重点存在差异，关注这些差异，可能会为未来该领域研究主题的选择提供思路和启发。

Cite this paper: 陈媛先. 四十年（1980-2020）来个人借款领域的研究主题变迁-基于文本挖掘 LDA 算法的主题发现和可视化. 社会科学与计算研究研究，2021，卷 1，第 4 期，1-32 页.

人工智能的发展蓬勃向前，使用 AI 技术辅助进行学术研究将会发挥越来越重要的作用。本文通过 LDA 主题发现技术和主题可视化技术辅助，对个人借款领域的研究进行了分析和挖掘，无疑提供了一种对学术文献进行深度回归的思路，同样的方法可以用于研究更多的文献，相信该技术会在更多的文献研究中发挥重用。

附件

主题分析代码片段：

```
from sklearn.decomposition import LatentDirichletAllocation
n_topics = 20
lda = LatentDirichletAllocation(n_topics, max_iter=50,
                                learning_method='online',
                                learning_offset=50.,
                                random_state=0)
lda.fit(tf)
def print_top_words(model, feature_names, n_top_words):
    for topic_idx, topic in enumerate(model.components_):
        print("Topic #%d:" % topic_idx)
        print(" ".join([feature_names[i]
                        for i in topic.argsort()[:-n_top_words - 1:-1]]))
    print()
n_top_words =20
```

可视化代码片段：

```
import pyLDAvis
import pyLDAvis.sklearn
pyLDAvis.enable_notebook()
data=pyLDAvis.sklearn.prepare(lda, tf, tf_vectorizer)
pyLDAvis.show(data,open_browser=True)
```

参考文献

[1] 蒋亚利.中国商业银行个人金融业务操作风险防控机制文献综述[J]. 沿海企业与科技，2012, 000(001):6-9,5.

[2] 姚蔚子. 金融隐私保护与征信制度国内研究综述[J]. 时代金融(下旬), 2016, 000(007):58-59.

[3] 人寿. 居民储蓄存款适度增长与储蓄分流研究综述[J]. 浙江金融, 1992(1 期):28-29.

[4] 乔薇. 个人住房抵押贷款违约风险的研究综述及其启示[J]. 中国商界:上半月, 2012(7):27-28.

[5] 于小亿. 基于个人信用信息的金融业客户保持研究综述[C]// 信用经济与信用体系国际高峰论坛. 2009.

[6] 莫易娴. P2P 网络借贷国内外理论与实践研究文献综述[J]. 金融理论与实践, 2011(12):103-106.

[7] 王学龙, 张璟. P2P 关键技术研究综述[J]. 计算机应用研究, 2010, 027(003):801-805, 823.

[8] 周扬, 刘义杰. 农村民间借贷的经济效应及其监管研究文献综述[J]. 商情, 2015.

[9] 单斌, 李芳. 基于 LDA 话题演化研究方法综述[J]. 中文信息学报, 2010, 24(6):43-50.

[10] 高永兵，熊振华. 基于 LDA 的专业个人微博事件提取[J]. 内蒙古科技大学学报，2015, 034(003):257-261.

[11] 宋凯，李秀霞，赵思喆等. 基于 LDA 模型的国家间知识流动分析[J]. 情报杂志，2017,

Cite this paper: 陈媛先. 四十年（1980-2020）来个人借款领域的研究主题变迁-基于文本挖掘 LDA 算法的主题发现和可视化. 社会科学与计算研究研究，2021，卷 1，第 4 期，1-32 页.

036(006):55-60.

[12] 徐翔, 靳菁, 吕伟欣. 网络舆情作为社会传感器对股票指数的影响——基于 LDA 主题模型的挖掘分析[J]. 财务与金融, 2018, 176(06):5-13.

[13] 封思贤, 袁圣兰. 用户视角下的移动支付操作风险研究——基于行为经济学和 LDA 的分析[J]. 国际金融研究, 2018.

[14] 向晖, 杨胜刚. 消费者信用评估的 SVM-LDA 组合模型[J]. 消费经济, 2010, 026(001):54-56.

[15] 丰吉闯, 李建平, 高丽君. 商业银行操作风险度量模型选择分析[J]. 国际金融研究, 2011(8):88-96.

Cite this paper: 陈媛先. 四十年（1980-2020）来个人借款领域的研究主题变迁-基于文本挖掘 LDA 算法的主题发现和可视化. 社会科学与计算研究研究，2021，卷 1，第 4 期，1-32 页.