



华南理工大学

基于云的P2P网贷信息监控与分析系统



- 背景
- 数据来源
- 相关技术
- 可视化展现
- 数据产品交互
- 产品使用价值



华南理工大学

目录

背景



→ 什么是P2P？

- P2P指是一种个人与个人之间的借贷，而P2P理财是指以互联网平台为中介机构，把这借贷双方对接起来在网站上实现各自的借贷需求。借款方可以是无抵押贷款或是有抵押贷款。出借方可以为给人。而中介一般是收取双方或单方的手续费为盈利目的或者是赚取一定息差为盈利目的的新型理财模式。

→ 为什么出现P2P？

- 由于传统银行无法满足个人信贷和小微企业贷款的需求，P2P平台的出现弥补了传统金融的不足，拓宽了中小企业融资渠道，并与互联网信息技术相结合，作为一种金融创新，它对促进金融市场发展和普惠金融体系的建设都起到了巨大的促进作用。



→ P2P出现的问题

- P2P平台出现的问题也是五花八门，有“跑路”、诈骗、老板失联、倒闭、网站关闭这类注定损失本金的问题，也有提现困难、经侦介入、平台资质造假等让人提心吊胆的问题。

→ 为什么开发P2P网贷信息舆情监控系统？

- 通过P2P舆情监控，可以发掘P2P行业以及企业的主要话题，及时地把握P2P的发展动态。同时对P2P平台进行多维度的分析，包括利率，成交量，投资人数，借款人数以及口碑等，帮助投资人选择合适的P2P平台进行稳健投资，实现利益最大化。



数据来源



数据分为四部分，包括新闻数据、P2P平台评论数据、P2P平台基本信息数据和新浪微博数据。

- 新闻数据从P2P资讯平台爬取，包括网贷时代，网贷财富，网贷之家，网贷新闻，新闻数据包含2013-7-12 11:05:03至2016-03-22 21:10的新闻资讯，总共28647条。
- P2P平台评论数据从网贷之家爬取，总共56721条品牌评论数据，包含3120个P2P平台。
- P2P平台基本信息数据从网贷之家爬取，包括利率，成交量，投资人数，借款人数等，包含3120个P2P平台的基本信息数据。
- 新浪微博数据包含2015-11-30至2016-3-26的部分P2P微博，总共2148条。



相关技术



■ 爬虫方法与工具应用

爬虫采用java语言开发，支持去重，断点重爬，持续更新等功能，具有很高的可用性。目前P2P资讯网站反爬虫策略不够完善，通过http协议模拟浏览器的方式可以获取国内主要P2P资讯网站的数据，对于新浪微博的数据，我们通过分析网络协议，抓取微博平台上的热点新闻。在我们的系统中，数据从网贷时代，网贷财富，网贷之家，网贷新闻四个P2P资讯网站和新浪微博获取。为了避免造成对方网站服务器的压力，我们实现了爬虫参数可配置化，包括爬虫更新时间间隔，爬虫速率等。



■ 数据清洗方法

(1) 通过P2P资讯网站爬取到的新闻数据，格式规范，大部分数据具有很高的可用性，但是存在小部分数据存在乱码问题，我们通过筛选掉这些乱码数据，留下高质量的数据，以JSON格式存储在MongoDB中。

(2) 通过网贷之家爬取的P2P平台评论数据，存在人为随意性，我们通过句法分析，语义分析以及领域知识挖掘出评论中的主要内容，挖掘结果存储在MYSQL中。

(3) 通过网贷之家爬取的P2P平台基本信息，格式规范，以JSON格式存储在MongoDB中。

(4) 通过新浪微博爬取的数据，数据格式不太规范，我们通过去除正文中的噪音，包括图片链接地址、@等，筛选出正文中有价值的信息，筛选后的数据存储在MYSQL中。



■ 文本分析与数据分析

- 文本分析包含新闻文本的分析，微博文本的分析以及评论口碑文本的分析，主要包括对
- 文本进行细粒度的情感分析，情感分析主要包括基于依存句法分析的文本分词和句子结构、构造结合情感字典的舆情/话题正负面极性计算。
- 关键词云的抽取，关键词云主要通过词频-反文档频率统计，筛选一段时间内的关键词。
- 品牌口碑分析，基于依存句法结构分析、结合情感字典的对品牌评论进行正负面极性计算，结合品牌的全部正负面极性，计算品牌的口碑，根据好评率进行排序展示。
- 热点实体抽取，基于词频统计，同时通过人为降低热门词的权重，有效地对热点实体的进行抽取。
- 主题句生成，通过话题内的文章自动生成话题的标题。
- 话题发现以及跟踪，实时发现舆情热点并追踪热点发展动态和舆论变化趋势。



可视化展现



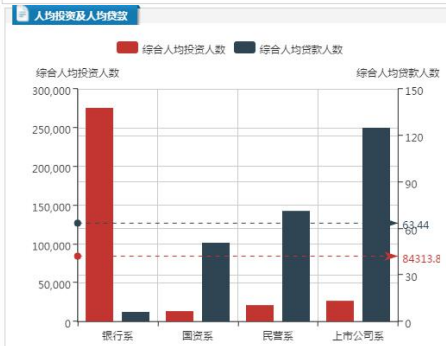
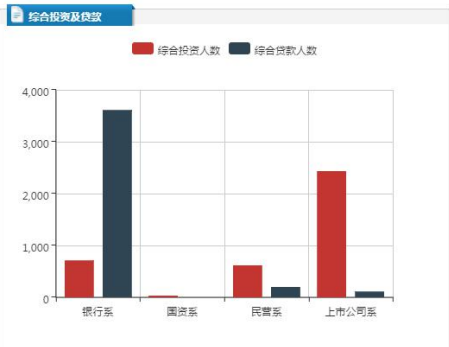
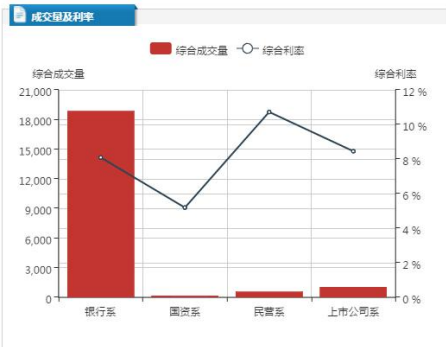
→ 系统展示链接

<http://120.27.109.170:8080/P2PSystem>



基于云的新一代P2P网贷信息监测与分析系统

- 首页
- 品牌分析
- 新闻舆情
- 微博舆情
- 来源统计
- 热点实体
- 品牌区域分析
- 爬虫数据



品牌排行榜

品牌	利率/%	成交量/元	投资人数/人	贷款人数/人	期限/天
1 幸福贷					24.49
2 九融金融					23.95
3 乐惠财富					22.37
4 智融财富					21.81
5 瑞融网贷					21.04
6 拍拍贷					20.96
7 微易贷					19.9
8 量子贷					19.69
9 途汇通					19.6



新闻舆情

- 【每日指数】11月24日全国P2P网贷成交额54.17亿元 (0/169)
- 年末温州民间借贷活跃度小幅提升 (0/140)
- 6月24日互联网金融产品收益率播报 (0/136)
- 互联网金融能为《中国制造2025》贡献什么? (0/1011)
- 监管要求P2P“小额化”平台等级化加剧 (0/92)
- “一带一路”对P2P的影响 (0/89)
- P2P平台的企业级大额借贷业务投资风险更大 (0/76)
- 黄金价格继续承压 短期内难以回升 (0/69)
- 美元理财将迎来平价周期 (0/64)
- 天翼：继农行、招行之后 交通银行也暂停P2P充值了 (0/61)
- 中国P2P公司“合法”前 三种可行的上市路径 (0/61)
- 股权众筹“原上”新三板 或成新退出机制 (0/54)
- 谈玩P2P的期限配置：几个月的标比放好? (0/51)
- 央行征信局局长：征信市场千亿规模不容忽视 (0/50)
- 红岭创投关于担保资产自动标标比例调整的公告 (0/45)

热点话题

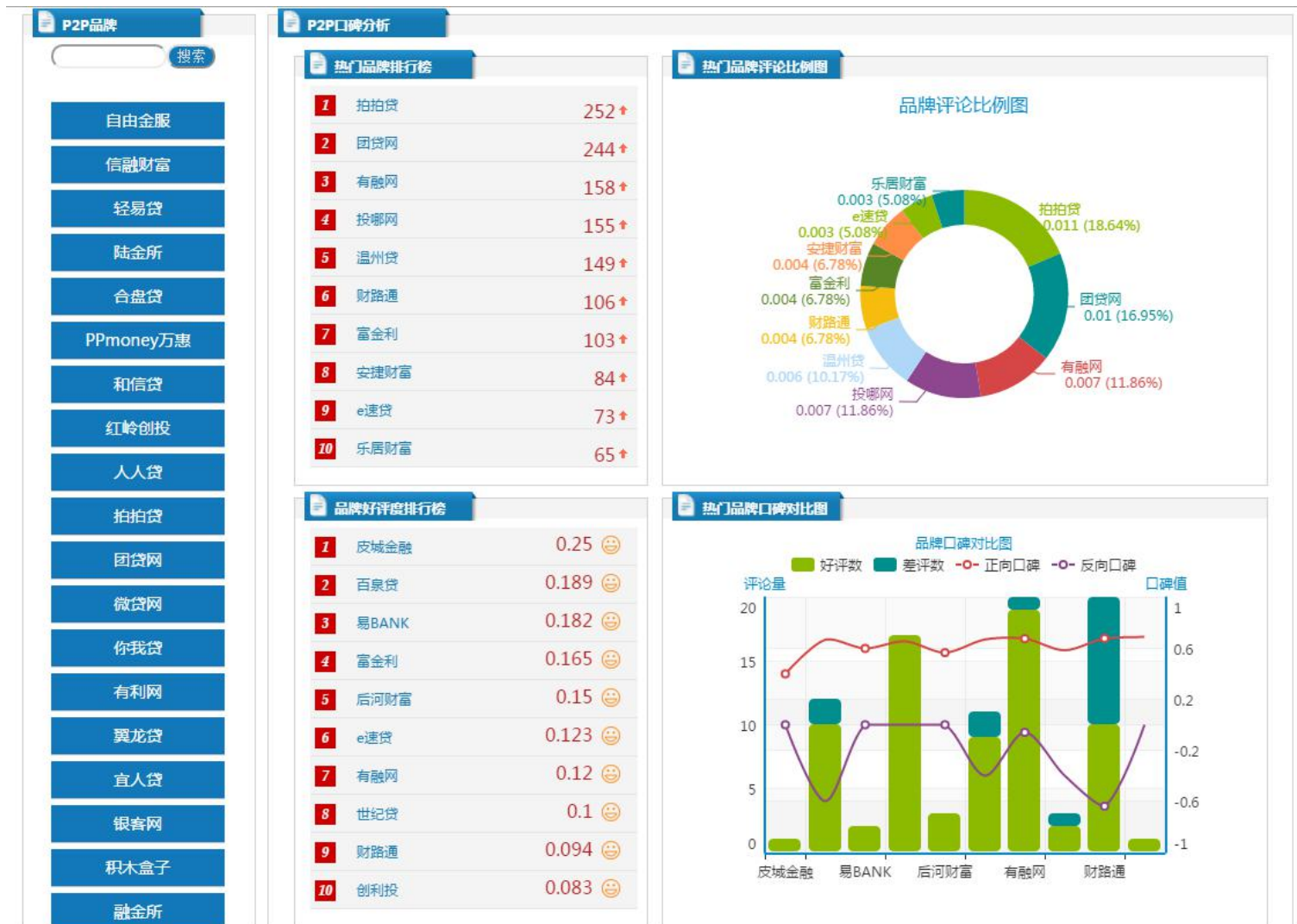
- 动力煤期货首单交割 (新闻)
- 互联网金融能为《中国制造2025》贡献什么? (新闻)
- 互联网金融有助于打破金融垄断 (新闻)
- 年末温州民间借贷活跃度小幅提升 (新闻)
- 典当行开分号:反映“冷热不均”瓶颈在人才 (新闻)
- 红岭创投关于担保资产自动标标比例调整的公告 (新闻)
- “一厘贷”问题，网金社“拼数据”叫板陆金所 (新闻)
- 谈玩P2P的期限配置：几个月的标比放好? (新闻)
- 深圳首付贷存案超20亿 多地监管喊停“套套” (新闻)
- 中国P2P公司“合法”前 三种可行的上市路径 (新闻)
- 数据看谁：宜人贷与Lending Club的上市路PK (新闻)
- 美元理财将迎来平价周期 (新闻)
- 微普公司试水互联网金融“越网”卖不良 (新闻)
- P2P平台的企业级大额借贷业务投资风险更大 (新闻)
- 探析：哪些网贷投资人能“玩”股权投资? (新闻)



标签云

热点人名

黄诗樵, 郭田勇, 李兆雄, 张俊, 徐马云, 计葵生, 马化腾, 李彦宏





P2P品牌

搜索

- 自由金服
- 信融财富
- 轻易贷
- 陆金所
- 合盘贷
- PPmoney万惠
- 和信贷
- 红岭创投
- 人人贷
- 拍拍贷
- 网贷网
- 微贷网
- 你我贷
- 有利网
- 翼龙贷
- 宜人贷
- 银客网
- 积木盒子
- 融金所
- 理想宝

P2P口碑分析

品牌名：拍拍贷

品牌信息

注册资本	10000 万元
平均年化	21.46%
资金托管	无托管
投标保障	投标50个列表且满足一定条件，100%本金保障
保障模式	
担保机构	-

您可能感兴趣的品牌

关键词

评论信息

评论总数:252

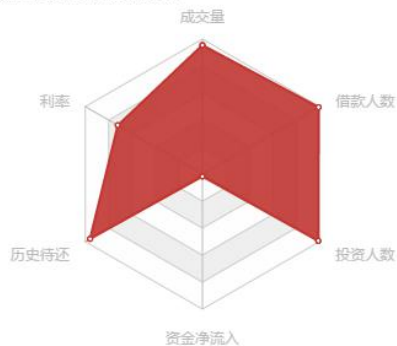
- 拍拍贷平台回款散乱，无法大额提现是软肋，发布安...
时间：2016-03-25
- 客服的态度还算不错。。标谁难抢的
时间：2016-03-23
- 总的来说，安全性不错，充值1000块以上不需要...
时间：2016-03-22
- 发现在之家的论坛上还是投资人偏多，我也是个投...
时间：2016-03-21
- 总体还是不错的，主要针对个人借贷，模式是主流的...
时间：2016-03-19
- 当初看拍拍贷利息高投了6万，但逾期几个月了，无...
时间：2016-03-19

【首页】 【上一页】 【下一页】 【尾页】



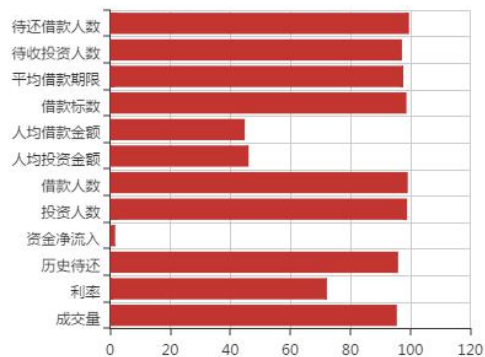
投资偏好分析

拍拍贷投资偏好分析



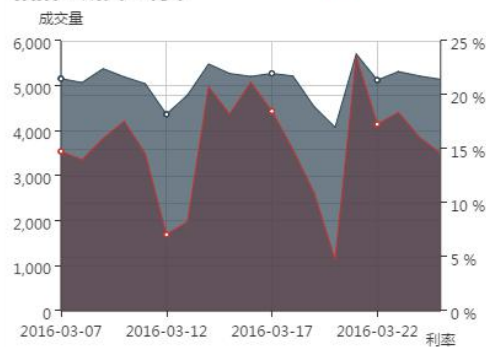
核心指标

拍拍贷核心指标



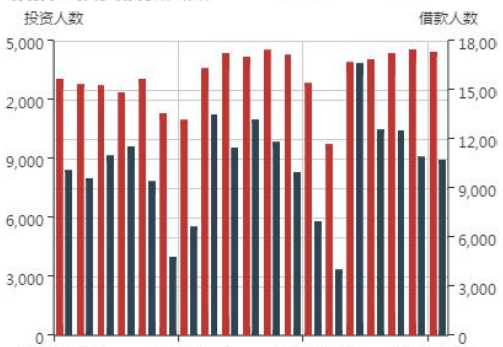
成交量利率分析

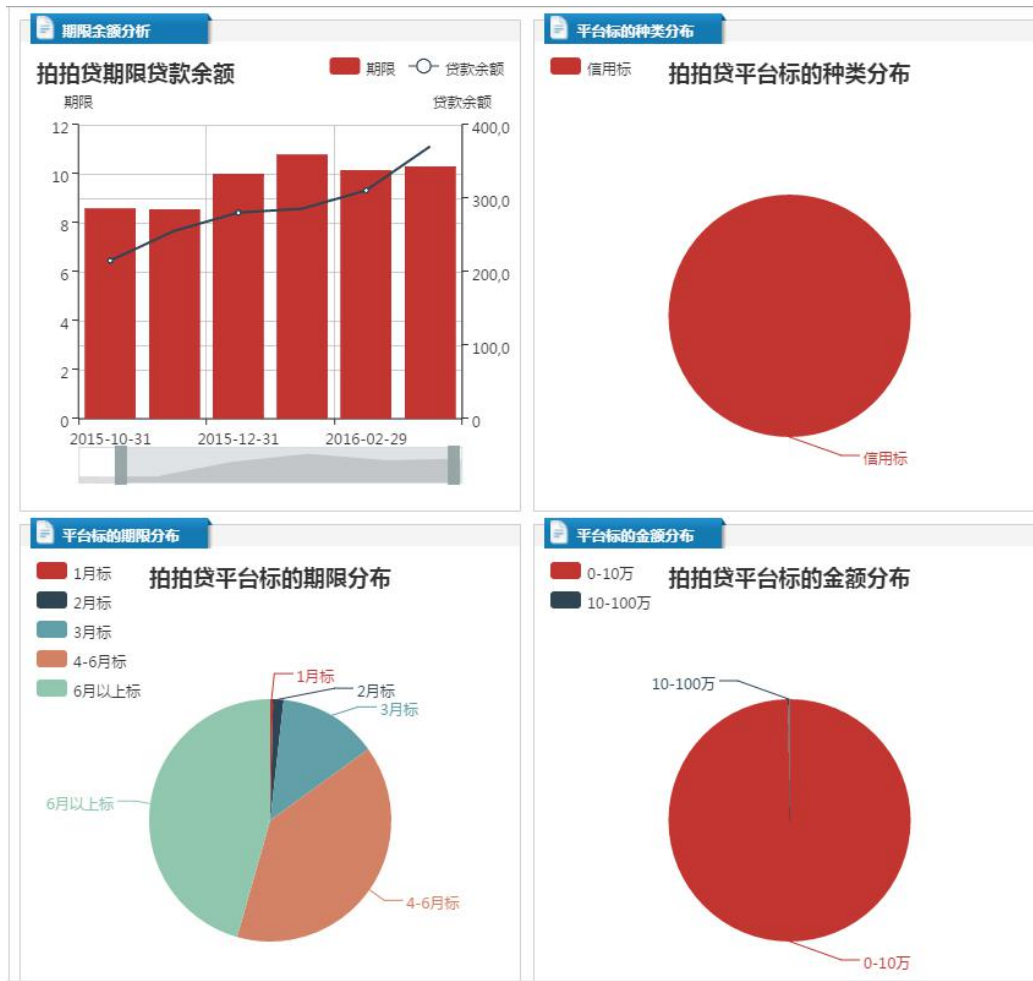
拍拍贷成交量利率



投资借款人数分析

拍拍贷投资借款人数







热点人名

李克强

庞氏

马云

徐红伟

王岩岫

马骏

周世平

张俊

朱明春

计葵生

石鹏峰

黄震

周小川

黄诗樵

郭田勇

热点人名相关信息

· **李克强**：股市汇市政策是正确的 避免了系统性金融风险

作者: 依凡 “猴年一开年, 全球经济、特别是各国股市急剧下跌, 也给我国经济带来很大的挑战和新的不确定因素。因为今天的中国经济已经与世界经济深度融合。” 在春节假期后首日的国务院常务会议上, **李克强**总理严肃地说。过去一周, 当中国人沉浸在自己...

李克强：严惩金融领域各种违法违规行为

据新华社电 1月13日, 中共中央政治局常委、国务院总理、党组书记**李克强**主持召开国务院党组会议, 部署深入推进政府系统党风廉政建设和反腐败工作。会议强调, 要多措并举维护财政金融安全, 对财政专项资金实行清单管理, 对沉淀闲置资金一律收回调整使...

盘点：**李克强**总理都在哪些场合力挺了P2P

随着行业的迅猛发展, 互联网金融一直备受关注与追捧, 亦不乏质疑。支持者强调, 互联网金融信息透明度高, 操作过程简单, 降低了借贷门槛和资金成本, 能够更好的服务于实体经济, 尤其可以满足小微企业和个人的金融需求, 是大势所趋。质疑者认为, 互联网金...

李克强出席双创周 发布国家级创业服务平台

10月19日, 中共中央政治局常委、国务院总理**李克强**在北京出席首届“全国大众创业万众创新活动周”, 并考察主题展区。 为推动实施创新驱动发展战略、展示“双创”成果、提供平台促进创业创新要素聚集对接, 在全社会营造良好氛围, 经国务院批准, ...

李克强：鼓励互联网金融依托实体经济发展

10月16日, 国务院总理**李克强**在主持召开的金融企业座谈会上提出, 加大对实体经济支持;鼓励互联网金融依托实体经济规范有序发展;坚守不发生区域性系统性金融风险底线。 互联网金融监管的一只“靴子”已经落地, 另一只细则“靴子”落地进入读秒...

[【首页】](#)

[【上一页】](#)

[【下一页】](#)

[【尾页】](#)

热点实体 (**李克强**) 相关报道共有25条/5页。这是第1页。 转到第 页



P2P平台区域分布





基于云的新一代P2P网贷信息监测与分析系统

- 首页
- 新闻舆情
- 微博舆情
- 来源统计
- 热点实体
- 品牌分析
- 爬虫数据

话题名：【每日指数】11月24日全国P2P网贷成交额54.17亿元

分类：新闻 创建时间：1970-01-01 文本数：0/169 热度：0.0 极性：0.171

今日更新

今日没有与该话题有关的报道。

往日报道

更多

· 12月25日全国P2P网贷成交额57.60亿元

P2P业内权威数据机构12月27日讯,周五(25日)中国P2P网贷指数新闻日报显示,该日全国P2P网贷总成交额为57.60亿元、利率...

强度值：0.13 极性：褒义

· 1月24日全国P2P网贷成交额18.59亿元

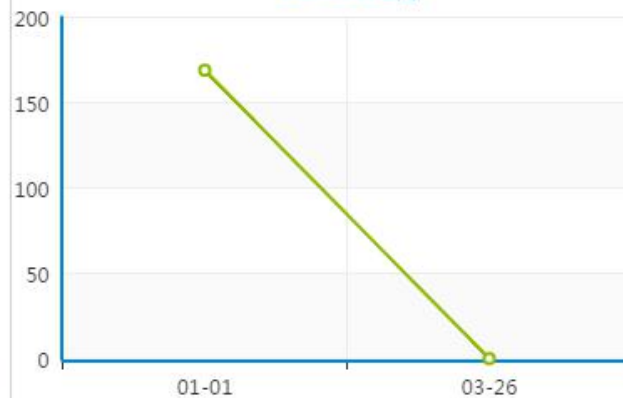
中国经济网深圳1月26日讯,布的周日(24日)中国P2P网贷指数日报显示,该日全国P2P网贷成交额18.59亿元、利率11.71%、期限4....

强度值：0.1 极性：褒义

· 【互金指数播报】招商银行小企业e贷极速上位,冲入三甲

第一网贷(深圳钱诚)编制发布的中国互联网金融指数由中国P2P网贷指数、中国民

话题热度趋势图



话题报道正负面趋势图





- 【互金指数播报】招商银行小企业e贷极速上位，冲入三甲

第一网贷（深圳钱诚）编制发布的中国互联网金融指数由中国P2P网贷指数、中国民间借贷市场利率指数、理财指数、中国众筹指数四大部分组成。 ●中国...

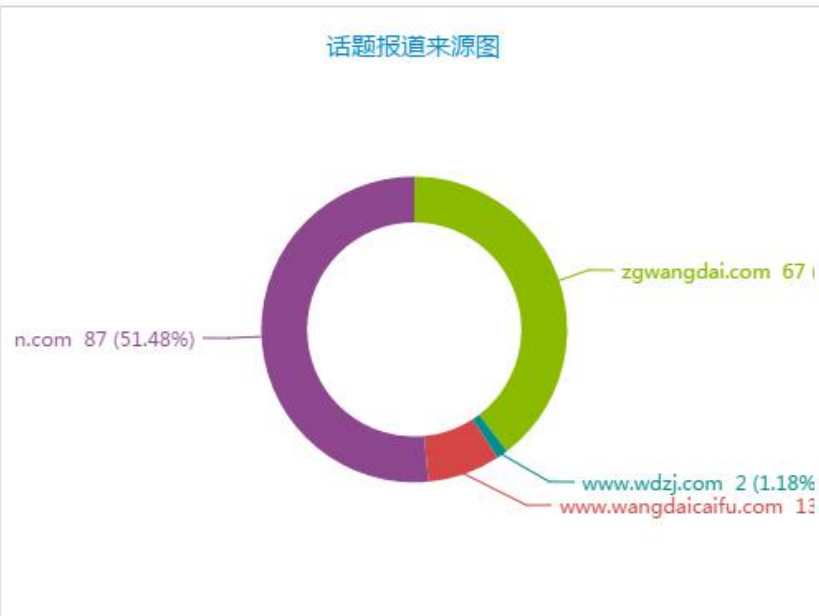
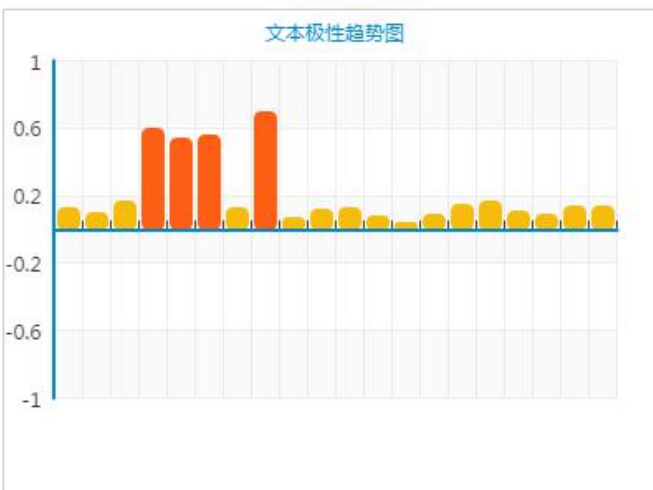
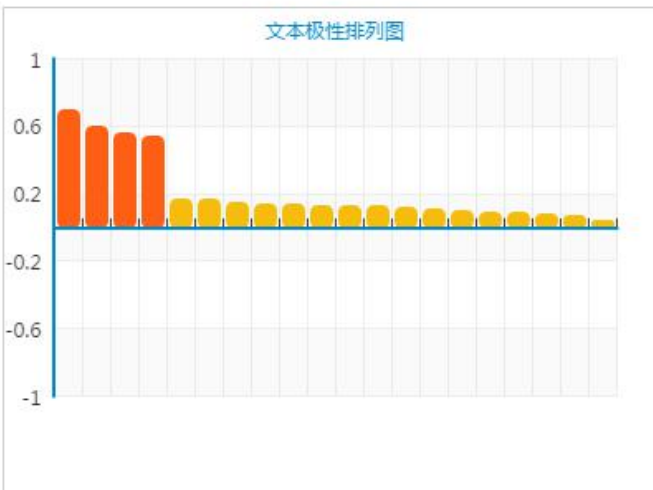
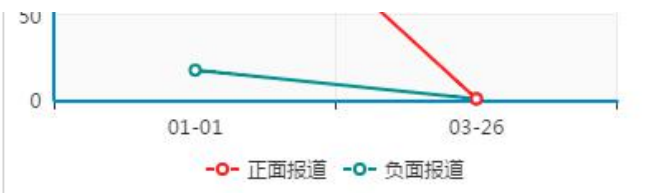
强度值：0.17 极性：褒义
- 走在监管前面的P2P平台才笑的更久

随着某宝事件爆发以及征求意见稿的出台，很多投资者担心手里的金融类理财产品也会出现违约风险。而近期，也有大量金融类理财产品的P2P机构开始转型...

强度值：0.6 极性：褒义
- 2月深圳网贷利率创新低 贷款期限创最长

记者昨天获悉，深圳2014年2月份P2P网贷月成交额、日成交额同创历史新高，平均综合年利率创历史新低，平均网贷期限创历史最长，人气...

强度值：0.54 极性：褒义





数据产品交互



- (1) 本系统主页采用上左右的分栏风格，上部是导航栏，用户通过点击导航栏，可以直接定位到系统的不同模块，系统包含八大模块，包括首页，品牌分析，新闻舆情，微博舆情，来源统计，热点实体，品牌区域分析、爬虫数据。
- (2) P2P网贷监控系统首页包含图表以及舆情信息，通过图表展示P2P平台的综合指数、排行榜、主要话题等，让用户对P2P行业有个整体性的认识，实现用户认知从“面”到“点”的过渡，符合用户体验的认知规律。
- (3) 数据统计分析结果，根据数据特征通过柱状图，折线图以及扇图等形式进行可视化展示，比如利率适合用折线图，投资借款人数适合用柱状图，可视化工具采用Echarts3和Fresh，通过可视化，用户可以直观地发现数据背后的规律，方便用户做决策。
- (4) 本产品包含3000多个P2P平台的基本信息以及5w多评论信息，考虑平台数量比较多，本产品提供了搜索功能，同时提供评论信息的关键搭配对，用户可以通过评论搭配对定位到感兴趣的评论内容。
- (5) 对于可视化图表，均添加动画效果，从信息量的角度，使用户更快地定位到有价值的信息，从动态视觉的角度上，提高了用户体验。



产品使用价值



- (1) 本产品通过及时地爬取P2P平台舆情信息、新浪微博的热点资讯，发现以及跟踪话题的演化以及舆论变化，有效地把握P2P行业的舆情趋势。
- (2) 本产品通过爬取P2P平台的基本信息，对基本信息进行数据分析，将分析结果进行可视化展示，让用户对P2P平台有具体的认识，有效地指导用户投资实现利益最大化。
- (3) 本产品能够发现最近的热点人名，热点品牌和热点机构名，让用户对于舆情热点有基本的把握。
- (4) 本产品通过爬取P2P平台的评论，对评论信息进行文本挖掘，分析文本的极性、抽取P2P平台评论关键搭配对以及计算P2P平台的口碑，根据挖掘结果对P2P平台进行可视化展示，根据展示结果指导用户选择口碑好的P2P平台进行稳健投资。



→ 创新性

- (1) 实现爬虫，自动采集P2P舆情信息，能够实时监控并抓取P2P舆情站点/微博平台的正文和P2P平台基本信息以及P2P平台的评论信息，并有效地进行数据清洗，将数据存储在高可用、可扩展存储介质中。
- (2) 对P2P平台的基本信息进行整体性分析，再具体到每个P2P平台，从“面”到“点”。
- (2) 对P2P平台舆情信息进行情感极性分析，分析信息的情感趋势。
- (3) 对P2P平台舆情信息进行周期性地多次聚类，发现舆情热点并追踪热点发展动态和舆论变化趋势。
- (4) 对P2P舆情一些重点关注实体（例如热点人物、品牌、机构等）进行舆论正负面分析和监测追踪。
- (5) 对P2P舆情热点话题、事件进行时间、空间等多维度分析。
- (6) 对P2P平台的评论信息进行关键搭配对抽取，分析P2P平台的口碑。
- (7) 数据的展示与交互，我们采用Echarts3进行可视化展示。



→ 成熟性

- (1) 该舆情分析产品包括首页，新闻舆情，微博舆情，来源统计，热点实体，品牌分析，品牌区域分析和爬虫数据八大模块，从多方位分析P2P平台，为用户提供全面而有价值的信息。
- (2) 涉及的主要技术有情感分析，口碑分析，聚类，推荐系统以及可视化展示等，开发人员对于以上技术比较熟悉，聚类方面考虑长文本以及短文本特征，推荐系统主要采用P2P平台的基本信息进行相似度计算，特别是情感计算以及口碑分析，具有较高的准确度，准确率达到90%以上，技术完成度较好。
- (3) 数据分析统计结果用柱状图、折线图以及扇图等图表的形式进行可视化展示，同时从产品设计上，考虑TOC产品化，服务人群主要是P2P行业投资人群，具有较高的产品可用性。系统已经完成初步测试，具有较高的稳健性和可靠性。



华南理工大学

致谢

谢谢大家