华南理工大学

基于云的P2P网贷信息监控与分析系统

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？

- 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。

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产品口碑的分析，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

数据来源

数据分为三部分，包括新闻数据、品牌评论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。
- 品牌评论数据从网贷之家爬取，总共23566条品牌评论数据，包含1695个P2P品牌。
- 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

通过P2P资讯网站爬取到的新闻数据，数据格式比较规范，大部分数据具有很高的可用性，但是有小部分的数据存在乱码问题，我们通过筛选掉这些乱码数据，留下高质量的数据，以json的形式存储在mongodb数据库中。

通过网贷之家爬取的品牌评论数据，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，品牌评论数据存储在mysql中。

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，将处理后的数据存储在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，有效地发现热点实体的抽取。
- 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

可视化展现


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
新闻舆情
更多
【每日指数】11月24日全国P2P网贷成交额54.17亿元 (0/169)
年末温州民间融资活跃度小幅提升 (0/140)
6月24日互联网理财产品收益播报 (0/136)
互联网金融能为《中国制造2025》贡献什么？ (0/101)
监管层要求P2P“小额化” 平台两极分化加剧 (0/92)
“一带一路”对P2P的影响 (0/89)
P2P平台的企业级别大额借贷业务投资风险更大 (0/76)
黄金价格继续受压 短期内难以翻身 (0/69)
美元理财年末迎来机遇期 (0/64)
天眼： 继农行、招行之后 交通银行也暂停P2P充值了 (0/61)
微博舆情
更多
农村“两权”抵押贷款，温江成试点区 (0/1)
我就说说当下的逻辑，不一定对： 因为自上周四以互... (0/1)
马云竟然上了胡润艺术榜 一幅画拍出3469万高价 (0/1)
郑重声明：本人微博所展示的个人实盘操作，并未收费... (0/1)
信息量很大，值得多看几遍。 □中国互联网金融协会在... (0/1)
看看互联网金服联盟到底几家P2P挤进去了，答案是16... (0/1)
断崖人生：泛亚投资者群像：//22万人中，并不是都没... (0/1)
东凯中学7月竣工9月招生 (0/1)
大BOSS在 #梦想合伙人# 里截了一些顾巧音的造型，... (0/1)
下午好，打开账户，准备开战！！下午建仓：传媒龙头... (0/1)
热点话题
互联网金融有助于打破金融垄断 新闻
年末温州民间融资活跃度小幅提升 新闻
典当行开分号:反映“冷热不均” 瓶颈在人才 新闻
红岭创投关于担保标自动投标比例调整的公告 新闻
一周热点回顾：网金社“拼群爹”叫板陆金所 新闻
谈玩P2P的期限配置：投几个月的标比较好？ 新闻
深圳首付贷存量超20亿 多地监管摸排“合围” 新闻
中国P2P公司“合法”前 三种可试的上市路径 新闻
数据看懂：宜人贷与Lending Club的上市路PK 新闻
美元理财年末迎来机遇期 新闻
爬虫数据
更多
深扒这2家平台就是想说：国资系不值得迷恋
恒生电子澄清借壳传闻 蚂蚁金服：无此计划
非法集资案登记平台开通 首问“e租宝”案
网贷之家新春拜年第八弹：平台君们送祝福
博弈新焦点：民资系网贷引人关注
从农发贷获得融资，看已至风口的农村金融
征信数据共享会是一门好生意吗？
“校园贷”到了该整顿的时候
唐小僧：行业迎来大洗牌，我们要静下心来做合规
互联网保险受资本青睐 近20家平台完成融资
的响应...

可视化展现


美元理财年末迎来机遇期
新闻
互联网保险受资本青睐 近20家平台完成融资
标签云图
热点人名
朱明春
庞氏
黄震
王岩岫
黄诗樵
周小川
计葵生
马云
马骏
郭田勇
徐红伟
信息岛图
开发团队: 华南理工大学Paddle团队
关于我们 | 联系我们 | 网站地图
Copyright 2016 All Rights Reserved

可视化展现


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
话题热度趋势图
200
150
100
50
0
01-01
03-26
话题报道正负面趋势图
200
150
100
50

可视化展现

- 【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

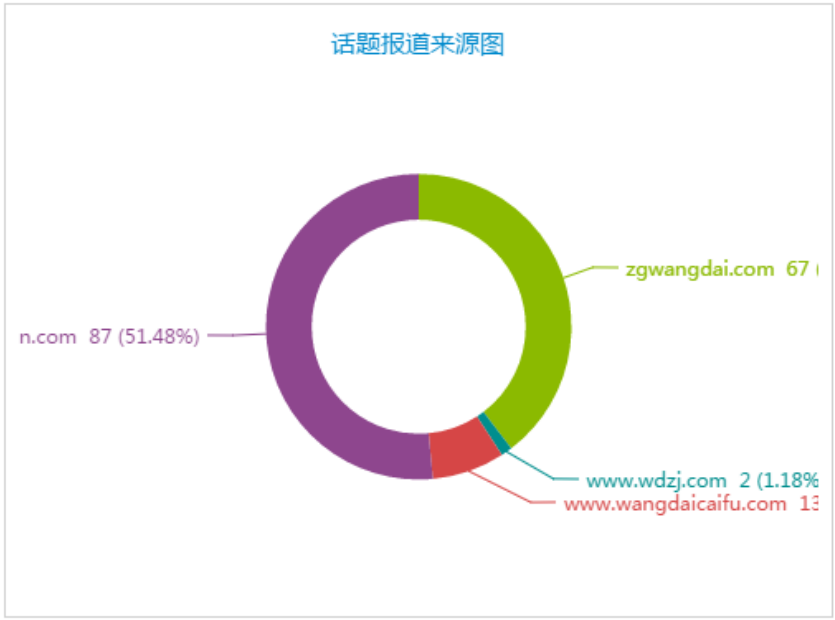
随着某宝事件爆发以及征求意见稿的出台，很多投资者担心手里的金融类理财产品也会出现违约风险。而近期，也有大量金融类理财产品的P2P机构开始转型...

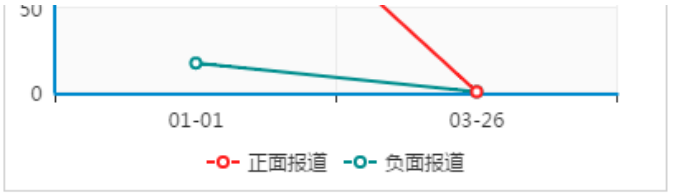
强度值：0.6 极性：褒义

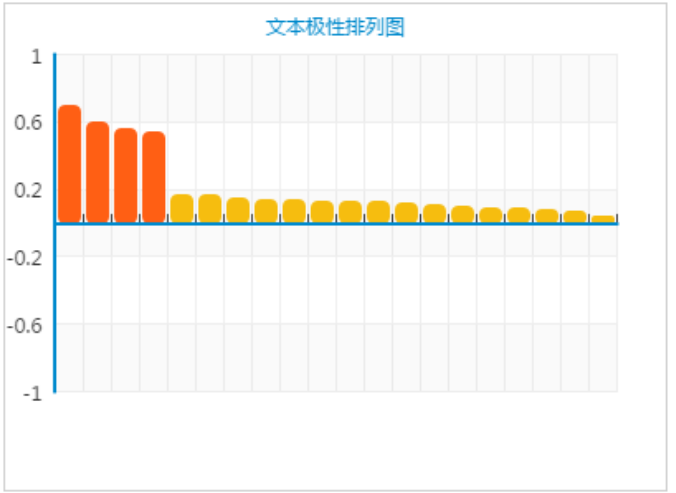
- 2月深圳网贷利率创新低 贷款期限创新长

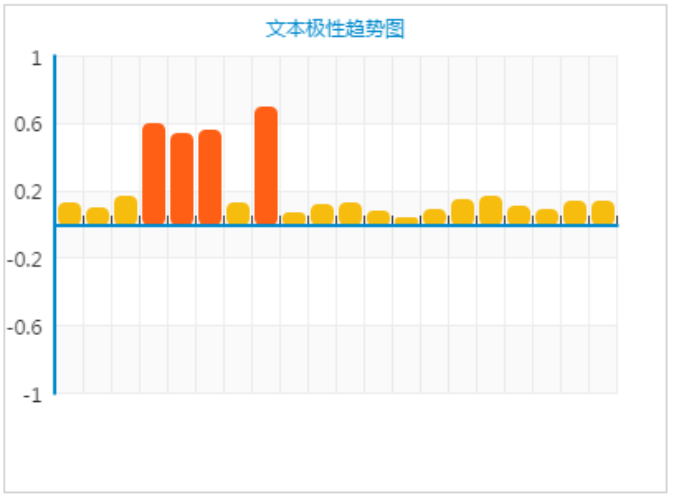
记者昨天获悉，深圳2014年2月份P2P网贷月成交额、日成交额同创历史新高，平均综合年利率创历史新低，平均网贷期限创历史新长，人气...

强度值：0.54 极性：褒义









数据产品交互

数据产品交互

- 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，系统包含七大模块，包括首页，新闻舆情，微博舆情，来源统计，热点实体，品牌分析，爬虫数据。数据统计分析结果通过图表的形式进行可视化展示，主要技术采用echart，用户可以直观地发现数据背后的规律，方便用户做决策。

产品使用价值

产品使用价值

- 本产品通过及时地爬取P2P舆情网站以及新浪微博平台的热点资讯，发现这些新闻的话题，跟踪话题的趋势以及舆论变化，有效地把握P2P行业以及企业的主要舆情热点。
- 本产品通过热点实体的发现，发现最近的热点人名，热点品牌和热点机构名。
- 本系统通过爬取P2P品牌的评论，对评论信息进行挖掘分析，计算P2P品牌的口碑，根据口碑值对P2P品牌进行排序展示，根据展示结果可以指导用户选择合适的P2P平台进行稳健投资。

创新性

- 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/微博平台的正文和评论，并有效过滤噪音信息。
- 对P2P舆情信息进行情感极性分析，分析信息的情感趋势。
- 对P2P舆情信息进行自动聚类，实时发现舆情热点并追踪热点发展动态和舆论变化趋势。
- 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。
- 对P2P舆情热点话题、事件进行时间、空间等多维度分析。
- 数据的展示与交互，我们采用echart进行可视化展示。

成熟性

- 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，爬虫数据七大模块，功能全面，开发人员对于以上技术比较熟悉，对于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。同时数据分析统计结果用图表的形式进行可视化展示，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。

系统代码链接

百度云：

http://pan.baidu.com/s/1bNGkFk

提取密码： ei2y

谢谢大家