社会化网络的用户数据分析

赵卫东 博士

复旦大学软件学院

wdzhao@fudan.edu.cn







议程

- 社会化网络的发展
- 电子零售网站顾客商品评论分析
- 微博用户的情感分析





社会化网络的发展

- 增强商家与顾客、顾客之间互动
- 建立口碑并增强信任
- 社会化营销是一种新的营销方式





美丽说一典型的社会化网络

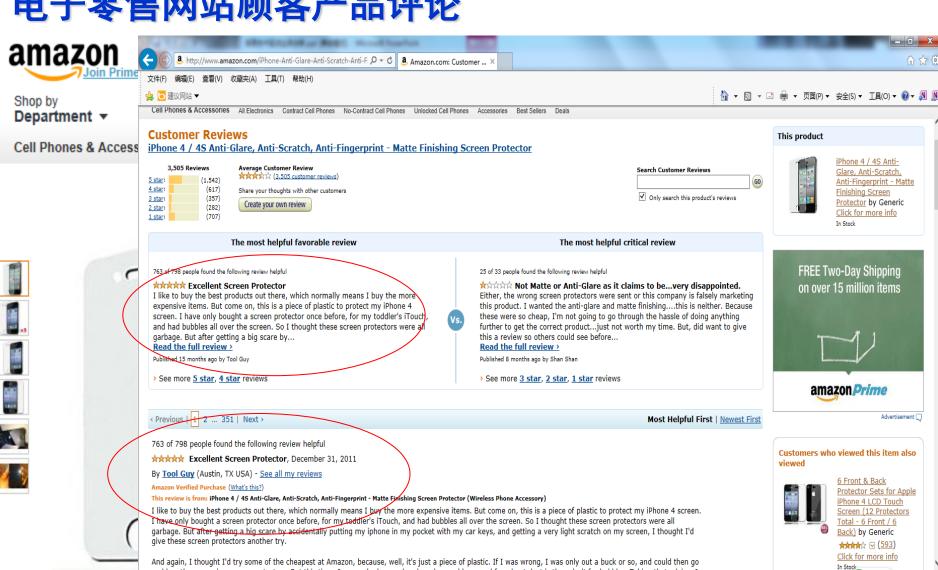






↓ 6.1K/5 € 100% ▼

电子零售网站顾客产品评论



and buy the expensive screen protectors. But this time, I researched around and read some blogs, and found out dust is the culprit for bubbles. Taking that advice, I

l***0

ABBANICATION TOWN A MARCHANIC

iPhone5

搜索

热门搜索: 尼康S5100 华为W1 台电A11四核 三星 i939 RCA播放器 华为G520 三星

前页 服装城 搭配购 团购 夺宝岛 在线游戏 全部商品分类 **手机数码** > 时尚影音 > MP3/MP4 > 台电 > 台电X19+ 全部评价(10247) 好评(9222) 中评(742) 差评(283) 有晒单(70) 台电X19 ★★★★★ 2013-04-03 13:13 点:续航时间长 小巧精致 足:有点容易按到朗读文字功能。。。 毅術家 得: 不错 续航时间长 小巧 银牌会员 色: X19+ 购买日期: 2013-03-26 回复(0) 有用(0) 物有所值 ★★★★★ 2013-04-03 11:15 点: 待机时间够长 不 足: 播放顺序不能按照文件名排列 piggy0122 得:待机时间够长,音质也可以,就是播放顺序很乱,不是按照文件名的顺序播放,想用它听连续的音频就会很麻烦。 铁牌会员 辽宁 购买日期: 2013-03-23 有用(0) 回复(0) 台电科技 ★★★★★ 2013-04-03 08:45 点:续航能力强 足:暂时还没发现缺点哦!





议程

- 社会化网络的发展
- 电子零售网站顾客商品评论分析
- 微博用户的情感分析

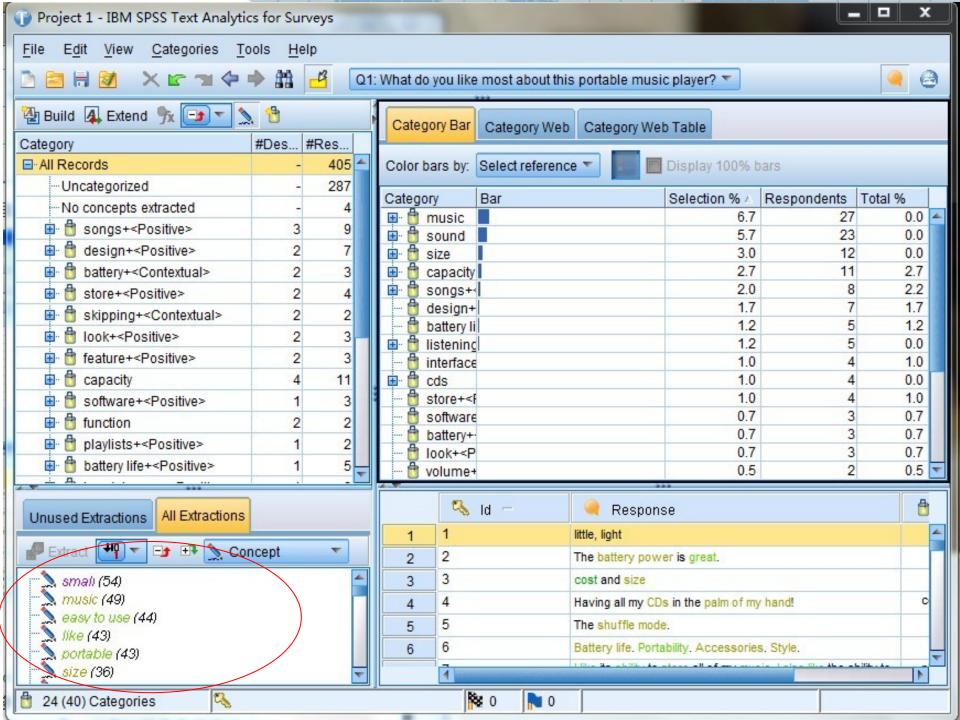


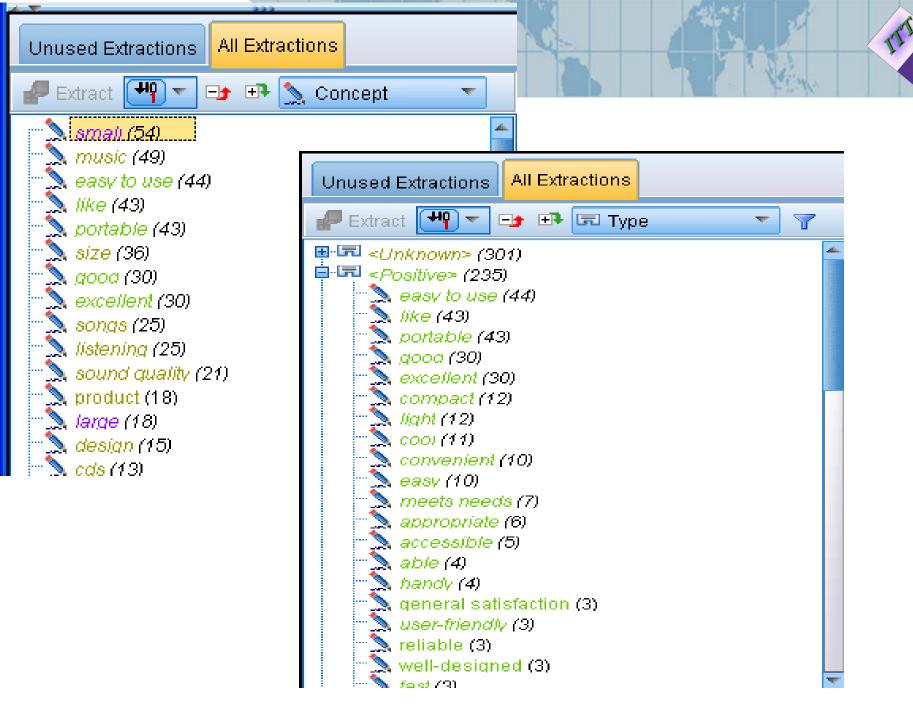


在线产品评论分析过程(重点)

- 数据抓取
- 分词
- 关键词聚类(语义识别)
- 情感分析(兴趣主题、情感极性和强度) 难点!
- 如何应用?

| 文 | 开始 开始 | 插入 页面布局 公式 数据 | 居 审阅 视图 加裁项 Ac | | |
|----------|-----------------------------|---|--|--|--|
| | 3 数 剪切 3 复制 → | Arial - 10 - 1 | A A = = >>- = = | | |
| 料贴 《 格式刷 | | B I U - 3 - 4 - | · 慢 ■ ■ ■ ● ■ ● ● ● ● ● ● ● ● ● ● ● ● ● ● | | |
| 剪贴板 5 | | 字体 | 5 对齐方式 | | |
| | F6 | √ 🦢 🏂 Male | | | |
| .al | Α | В | С | | |
| | Respondent | Q1: What do you like most about | | | |
| 1 | ID . | this portable music player? | this portable music player? | | |
| 2 | 1 | little, light | expensive | | |
| | 2 | The battery power is great. | The screen is hard to see when | | |
| 3 | | , | outside. | | |
| 4 | 3 | cost and size | difficult software | | |
| 5 | 4 | Having all my CDs in the palm of my hand! | Nothing, I love it! | | |
| 6 | 5 | The shuffle mode. | Battery life seems shorter than advertised. | | |
| 7 | 6 | Battery life. Portability. Accessories. Style. | Ubiquitousness; everyone has one. | | |
| | 7 | I like its ability to store all of my | I wish the 40GB model was still | | |
| | | music. I also like the ability to | available. I have a 20GB model and | | |
| 8 | | create playlists. | need more memory. | | |
| 9 | 8 | portability, capacity, sound quality, durability | it doesn't have a light. | | |
| 10 | 9 | Small, great sound, capacity. | Nothing. I love it. | | |
| | 10 | Able to hold all of my songs in one | It is in the shop due to a hardware | | |
| 11 | | place. | failure. | | |
| | 11 | It's portable! I can take it anywhere. | smudges on the display | | |
| 13 | 12 | Living in my own little world | Battery life | | |
| 14 | 13 | mobility | Technical difficulties setting it up initially and managing the library of songs on my PC. | | |
| | 14 | I like that Product A has a lot of | It is a little heavy, and the battery | | |









案例讨论

- 顾客对某种具体的mp3关注哪些主题?顾客对这些主题的具体偏好是什么?
- 对于购物网站,可了解哪些商品、哪些属性受顾客群欢迎?安排采购和商品展示。
- 对于商品的制造厂商,利用上述分析,改善设计和产品销售。





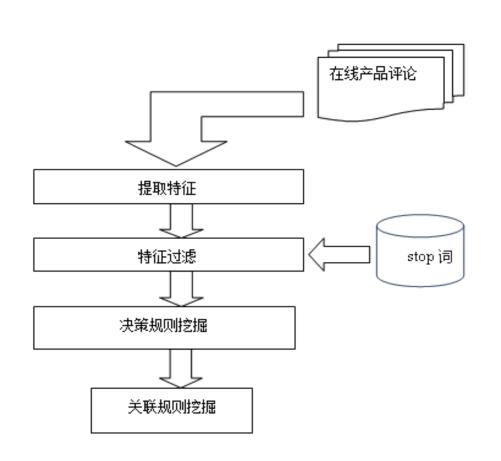
研讨(1)

- 阅读下面论文,讨论顾客对商品的评论与评分之间的关联规则应用价值:
- Wingyan Chung, Tzu-Liang Tseng. Discovering business intelligence from online product reviews: a rule-induction framework. Expert Systems with Applications, 2012, 39(15):11870-11879





研讨(2)



• 挖掘出的关联规则格式如下:"文本关键词1、文本关键词2、....、文本关键词n=>评论等级",例如对于商品"scooter"的评论,可以挖掘出关联规则"fun,

kids(7)=>5(6,0.857)", "fun, easy, great (12) =>5(12,1)",从中可以看 出该商品操作简单,给小 孩子带来乐趣,商家就可 以把商品推荐给有小孩的 顾客。





思考

- 一些学术问题:
- 社会化网络的信任传播模型
- 社会化网络对顾客购买心理的影响机理研究





实验

IDM

Solutions *

Services +

Products +

Support & downloads -

My IBM +

Welcome [IBM Sign in] [Register]

Search

SPSS Text Analytics for Surveys

System requirements

Library

Home

Success stories

Training and certification

Services

Trials and demos

How to buy

Support

IBM Software > Business analytics > SPSS > SPSS Statistics > SPSS Text Analytics for Surveys > Trials and demos >

United States [change]

SPSS Text Analytics for Surveys

Trial download for version 4.0.1

IBM® SPSS® Text Analytics for Surveys uses powerful natural language processing technologies specifically designed for survey text. It leads the way in unlocking open-ended responses for better insight and statistical analysis. The evaluation period begins on the date that you agree to the terms of the agreement.

- · Review systems requirements
- · Learn more about the product

SPSS Text Analytics for Surveys, V4.0.1 available packages

Operating systems Languages Min. sizes

Windows 7 Home Premium, Windows 7 Professional, Windows
Vista Business, Windows Vista Enterprise, Windows Vista
Home Basic, Windows Vista Home Premium, Windows XP
Home, Windows XP Professional, Windows 7 Enterprise: IBM
SPSS Text Analytics for Surveys for Trial 4.0.1 Windows Multilingual
eAssembly

After you have read the <u>systems requirements</u>, please select 'Continue' to log in to your IBM account or to register to download the package.



Information Management

Ready to buy?

⇒ View pricing & buy

Related trials

→ SPSS Data Collection

Related resources

- → SPSS Text Analytics for Surveys Product Information
- → SPSS Statistics Products
- SPSS Text Analytics for Surveys Data Sheet
- → Analyzing Survey Text: A Brief Overview
- → How to get more value from your survey data
- Get Adobe® Reader®

About IBM

Privacy

Contact

Terms of use

se

IBM Feeds J

Jobs





软件和实验指导书

- 软件SPSS Text Analytics for Surveys下载地址:
 http://www14.software.ibm.com/download/data/web/en US/trialprograms/P254879Q17958R48.html
- 实验指导书: http://61.129.34.202/Blweb/SPSS Text OAnalytics.pdf





议程

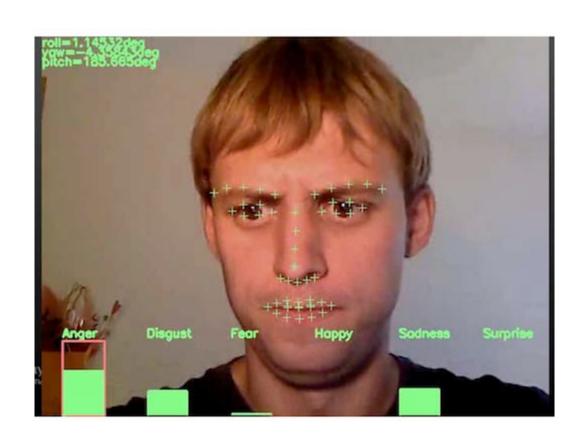
- 社会化网络的发展
- 电子零售网站顾客商品评论分析
- 微博用户的情感分析





情感分析技术

- 通过"电脑的眼睛"来 阅读人的面部表情,然 后通过计算分析出人的 情感(情绪反应)
- 商业价值?
- http://docs.realeyesit. com/demo/?v=emotio ns







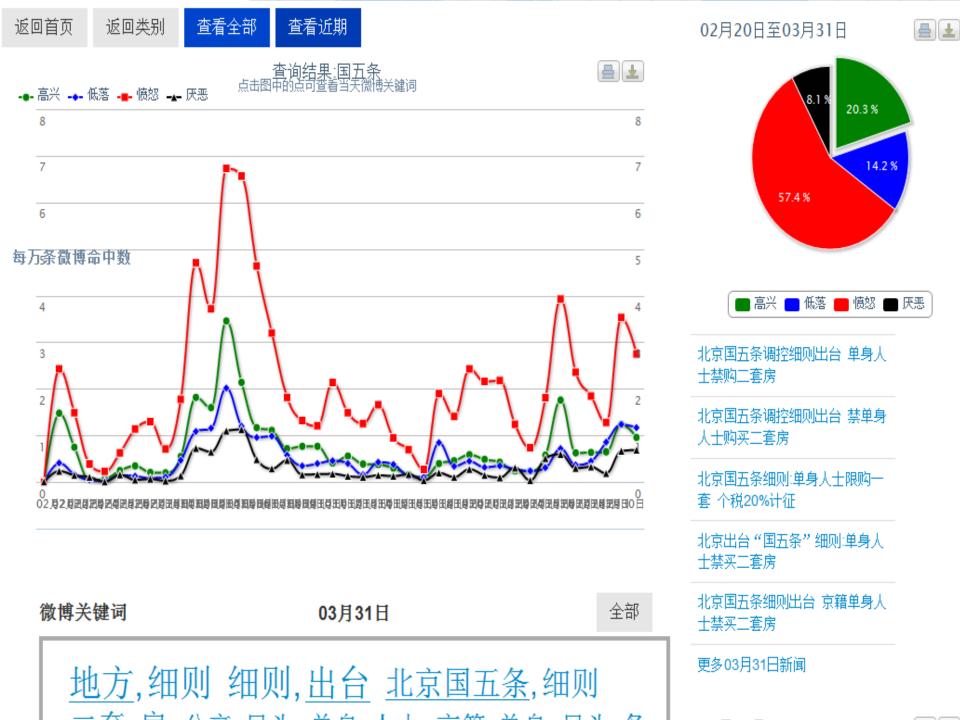
心情学搜索

- 分析新浪微博用户对某 些事件或新闻的情感: 高兴、低落、愤怒和厌 恶等
- 商业价值:與情监控、客户体验分析还有?
- http://xinqings.nlsde. buaa.edu.cn/



国五条 近期 全部

| 实时 | 意甲 | 拜仁 | 英超 | 颈椎 黑帮 | 更多〉〉 |
|----|----|-----|----|------------|------|
| 财经 | 淘宝 | 地产 | 房产 | VC IPO | 更多〉〉 |
| 时尚 | 宝马 | 奔驰 | 奥迪 | 欧莱雅 大宝 | 更多〉〉 |
| 生活 | 美的 | 青春 | 减肥 | 苹果 美团网 | 更多〉〉 |
| 游戏 | 梦想 | 我们的 | 奥特 | ·曼 剑网3 eve | 更多〉〉 |
| 体育 | 卡卡 | 法国 | 德国 | 意大利 俄罗斯 | 更多>> |







研讨1:识别社会化网络的意见领袖

在社会化网络中, 意见领袖能够在真实的社会中影响用户。可以用下面的公式识别具有意见领袖:

$$U_{s} = \frac{\sum_{i \in A} R_{U}(i) - R_{\mu}(i)}{|\operatorname{Con}|}$$

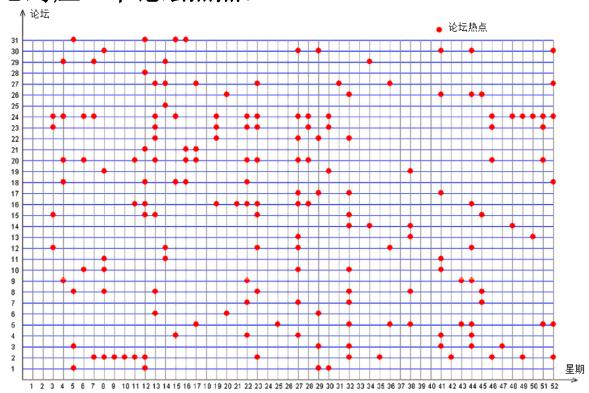
- 其中Con是内容(例如音乐、电影等)集合,U是用户集合,|Con|是Con的数量,是U中的用户对于内容i的评价,是用户对i的平均评价值。上式说明每个用户评价和平均评价的近似程度,选择与平均评价最不近似的用户作为意见领袖。通过对Yahoo! 音乐数据库(包括音乐、用户和评论等数据)和GroupLens电影数据库(包括电影、用户和评论等数据)的实验。
- 意见领袖有很多价值,其中一个应用是网络营销。意见领袖的微博很容易引导人们关注,因此可以在意见领袖的微博上投放广告。





研讨2:检测在线论坛热点

 首先,分析每个论坛帖子的情感倾向(emotional polarity)。然后利用依据帖子情感倾向聚类,每个聚类 中心对应一个论坛热点。







谢谢

