

项目简介-文献组

双十一资料整理

庄建伟

厦门大学 经济学院

2016年12月10日

目录

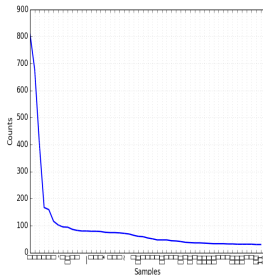
- ① 简介
- ② 消费者（郝泽栋）
- ③ 零售商（刘晓曼, 庄建伟）
- ④ 平台（张伟贤）
- ⑤ 小结

为项目组收集资料

- 任务：搜集关于双十一的文本消息，利用自然语言处理的方法分析其对消费者的影响。
- 初步思路：爬取知乎等网站上相关的文本数据，对所得文本进行处理。
- 处理方法：1. 分词；2. 词频统计；3. 分词性词频统计；4. 情感分析。

- 上周进度：抓取了知乎上四个排名靠前的与双十一相关的问题的回答，对这些答案进行了汇总，并进行了词频统计；记录了出现频率最高的50 个词并绘制了频数分布图。

- 目前的结果：由于未进行词性标注处理，且样本较小，目前所得top 50的词基本没什么意义，其中甚至包含“分割线”这样的词。



- 下一步工作：抓取更大容量的样本；处理过程中加入词性标注，分词性进行词频统计，进一步观察结果。

- 天猫双十一商家的各种活动,例如：万店同庆、价格竞猜、集猫活动、店铺优惠券等
- 网店的各种Cheating，例如：超低价商品（秒杀商品）、刷单等

