

# 系统概述

观点挖掘系统是一种自动挖掘产品评论中的评论观点,并输出评论观点标签及评论观点极性的系统. 下图展示的就是JD做的产品评论的挖掘功能,最上面矩形框里面的就是挖掘出来的观点



观点挖掘系统的商业用途:

## 1.产品营销分析和竞品对比分析

我司作为一家有2C业务的公司,市场人员可以通过这个系统追踪消费者对自家产品的态度,如消费者是否喜欢这款产品,消费者不喜欢这款产品的哪方面,消费者的主要抱怨是什么,我们应该增加哪些特征

同时我们也可以分析竞品的情况,如别家产品的弱点在哪里,我们该怎样和他们竞争,消费者是否喜欢别家产品的所有特性

## 2.用户意图识别和用户画像

我司ToB业务的主要发力点在客服机器人,在客服系统中,系统挖掘出来的用户观点可以作为特征来辅助客服系统

可参考百度AI开放平台的观点抽取功能:[baidu AI](#)

# 系统设计

观点挖掘的这个任务的过程可分为以下几步:

- 1.确定产品对象
- 2.确定目标对象的特征
- 3.挖掘评论中与对象或对象特征相关的观点词
- 4.评论观点的极性
- 5.总结归纳所有观点

观点的形式化定义

➤ 一个观点表示为五元组

$$(o_j, a_{jk}, so_{ijkl}, h_i, t_l),$$

其中

- $o_j$  为目标对象.
- $a_{jk}$  是对象  $o_j$  的特征
- $so_{ijkl}$  为观点所表达的情感值 (如倾向性分类)
- $h_i$  为观点持有者
- $t_l$  为观点表达的时间

每步详解:

1.确定产品对象

手表?音箱?耳机?一般由需求方给出

2.确定目标对象的特征

对象的特征一般先通过需求方给出一些,然后再通过挖掘的方法发现更多的特征

挖掘特征分为挖掘频繁特征和挖掘非频繁特征

频繁特征可使用关联挖掘方法

非频繁特征待研究

注意两个问题,一个是名词短语,一个是特征聚类,也就是同义特征可进行聚类,每个特征类都表示该对象独一无二的某个特征

可以调研这个方法:

<https://www.ijcai.org/Proceedings/09/Papers/202.pdf>

[https://www.mitpressjournals.org/doi/pdf/10.1162/coli\\_a\\_00034](https://www.mitpressjournals.org/doi/pdf/10.1162/coli_a_00034)

3.挖掘评论中与对象或对象特征相关的观点词

观点词可通过抽取距离特征最近的形容词或名词等来寻找,然后根据情感词典确定观点词的极性

注意否定词

注意观点词的指代方向

#### 4. 评论观点的极性

根据特征,观点词,否定词等确定整个观点的情感极性,参考情感分类的方法

#### 5. 总结归纳所有观点

进行观点聚类,合并意思相近的观点,得到整个产品观点的规范化表达

这一步很难跨领域,而且需要人工参与的成分较多

相关资料:

<https://wenku.baidu.com/view/4c84257704a1b0717fd5ddf7.html>

<https://wenku.baidu.com/view/7152237033d4b14e84246878.html>

<https://blog.csdn.net/u014313009/article/details/38958113>

英文语料:

<https://www.gelbukh.com/resources/implicit-aspect-extraction-corpus/>

<http://alt.qcri.org/semeval2014/task4/>

<http://alt.qcri.org/semeval2015/task12/>

<http://alt.qcri.org/semeval2016/task5/>

<https://github.com/uclmr/jack/tree/master/data/sentihood>

<http://jmcauley.ucsd.edu/data/amazon/>

<https://cseweb.ucsd.edu/~jmcauley/datasets.html>