**淘宝评论情感分析系统需求分析文档**

2019/11/18

崔少峰 赵奕翔 王一帆 张舒翔 刘一博

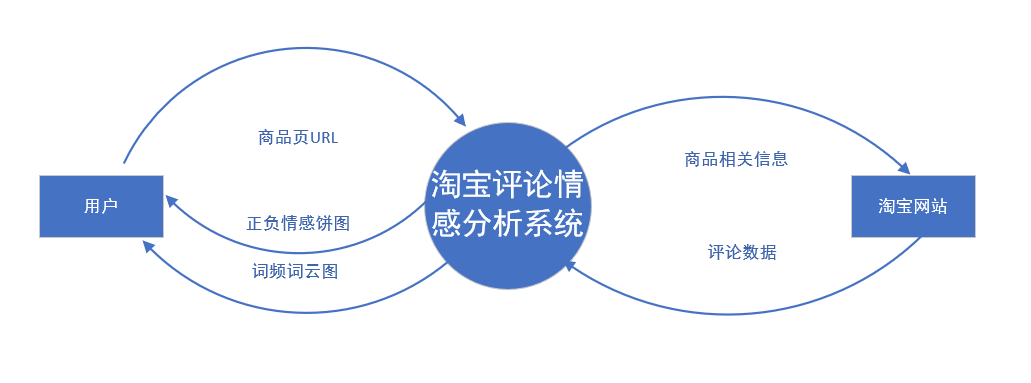
**一、系统介绍**

该系统是一个淘宝商品评论情感分析工具，实现的功能有：

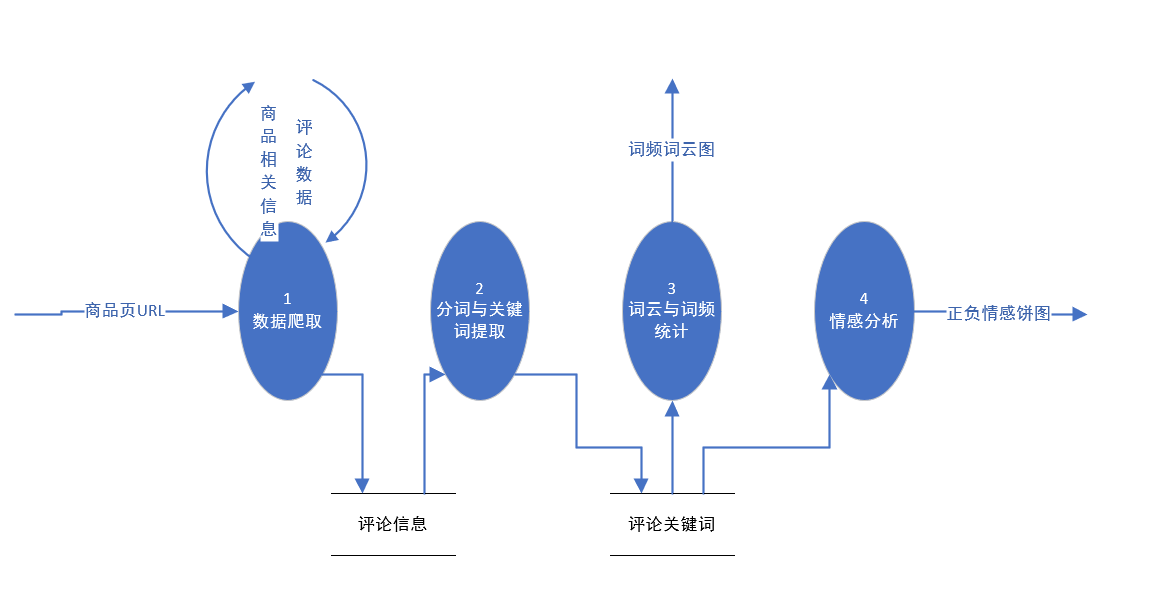
1. 淘宝某商品评论数据爬取
2. 分词与关键词提取
3. 词云与词频统计
4. 情感分析

**二、系统的分层数据流图**

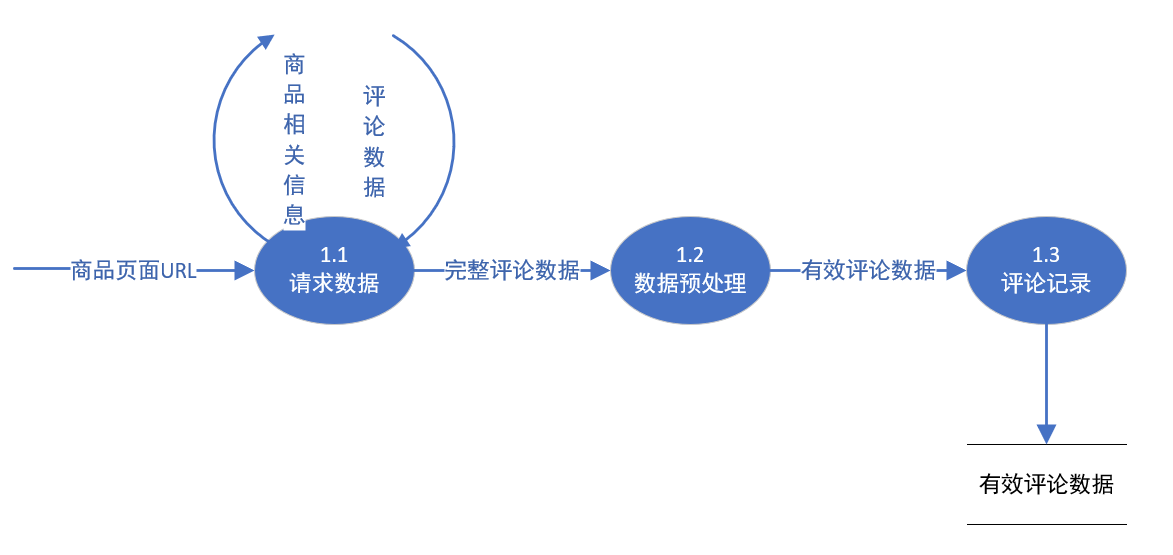
## 0层图



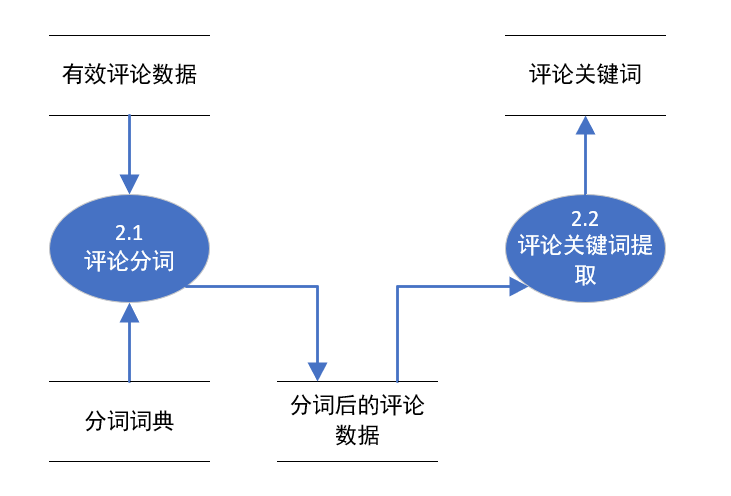
## 1层图



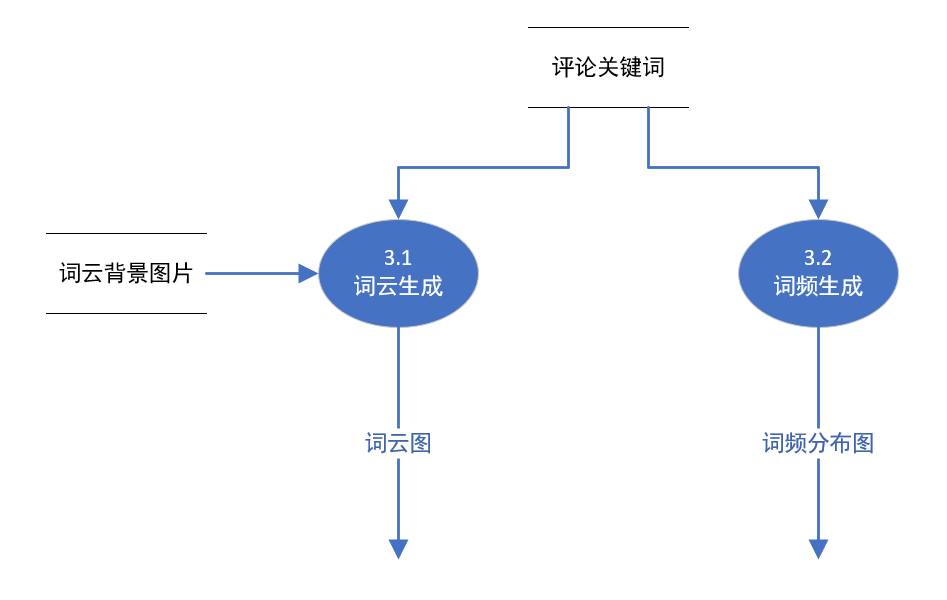
## 数据爬取子图



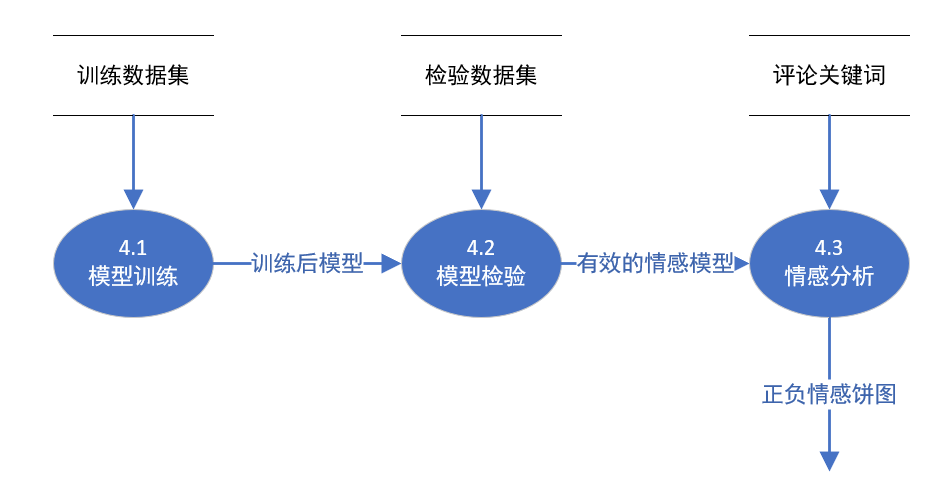
## 分词与关键词提取子图



## 词云与词频统计子图



## 情感分析子图



**五、数据字典**

**1）数据流**

商品页面URL=URL字符串

商品相关信息=商品id+spuId+卖家id+（爬取评论页数）

原始评论数据=来源+数据项数目+标签+{买家用户名+评论时间+头像地址+评论内容+买家id+评分}

完整评论数据={评论内容}

有效评论数据={有效评论内容}

词云图=词云图片

词频图=词频图片

训练后模型=marshal

有效的情感模型= marshal

正负情感饼图=情感分布饼图

**2）文件**

有效评论数据={有效评论内容}

分词词典={[分词 | 标点符号]}

分词后的评论数据={分词后的评论内容}

评论关键词={每条评论的关键词}

词云背景图片=用作词云背景的图片

训练数据集={正向情感词}+{负向情感词}

测试数据集={评论关键词 + [1 | -1]}

**3）加工**

**数据爬取**

根据爬取的页数执行下列程序:

设置请求头headers

根据itemId、spuId、sellerId拼接请求url

发起网络请求获取返回数据

将返回数据转换为json类型

进行下一步处理

**情感分析**

加载训练模型

读取评论关键词文件逐行分析

如果情感超过一半则判断为积极情绪

否则判断为消极情绪

画图