

SnowNLP | 2 | 爬取评论情感分析

1. 基本环境

程序名称	getComment.py getModel.py pieChart.py snowanalysis.py iphone_wordcloud.py
硬件环境	Win
软件环境	Python3, Anaconda

2. 核心问题：爬取京东某商品评论，并用开源中文情感分析库 SnowNLP 进行情感分析

3. 解决过程：

3.1. 爬虫

利用 python requests + re 爬取京东 iphone8 plus 商品下的共计 500 页，以及其 iphone8plus 的型号（颜色，存储大小）1000 组，存入 txt。

对评论数据进行了简单的清洗，去除了带有图片的评论中带有的 html 代码：

```
18 第一次用苹果，以前用的三星NOTE2，五年了，现在各种卡，内存不足，别笑2G网，流量是省了，但是各种不方便，所以综合各种原因，决定换个实用性强，耐用，五六年不会出大毛病的。这个定单
.22下的单，说是要等30号才有货，没想到25号货就发了，26号收到。从各种信息了解到可能苹果8系比7系升级效果不明显，用户不认账，有些供大于求了。不像以前那样，出现购买狂潮，东西才这么
快到货了。不管什么原因，还是比预计的到手早几天，给你们点赞！本来对手机外观从图片和视频都看到过，但是打开后现实手拿着看视觉上前后双玻璃还是很漂亮，周边做开也很精细。由于第
一次用苹果，一个人搞了一下午才搞定，这里对第一次安卓换苹果的用户提两点，SIM卡要剪成最小的，在剪卡前记得把原来手机上的号码，图片一类对自己有用的信息用qq同步软件，同步一下，以
防像我的情况新手机读不出来手机卡上的号码，各种麻烦。手机到今天也用3天了，对操作系统的慢慢熟悉，用着感觉确实比安卓系统更方便，简捷，流畅，很是满意！其他功能性东西慢慢去了解，
不作评价，因为自己还没搞明白。很好的一次购物。毕竟是双玻璃，东西到了就给它加上防护了。不漏后壳谁知道我用的啥！哈哈！低调！<div class="uploadingdiv"></div><div class="uploadingdiv"></div><div
class="uploadingdiv"></div><div class="uploadingdiv"></div><div class="uploadingdiv"></div><div class="uploadingdiv"></div><div
class="uploadingdiv"></div><div class="uploadingdiv"></div>
```

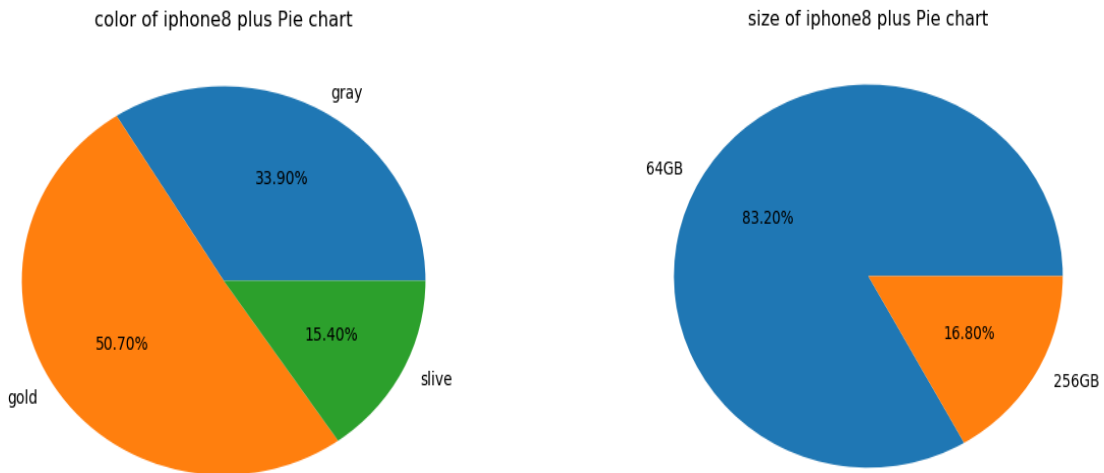
图 1 原图

```
第一次用苹果，以前用的三星NOTE2，五年了，现在各种卡，内存不足，别笑2G网，流量是省了，但是各种不方便，所以综合各种原因，决定换个实用性强，耐用，五六年不会出大毛病的。这个定单
9.22下的单，说是要等30号才有货，没想到25号货就发了，26号收到。从各种信息了解到可能苹果8系比7系升级效果不明显，用户不认账，有些供大于求了。不像以前那样，出现购买狂潮，东西才
这么快就到货了。不管什么原因，还是比预计的到手早几天，给你们点赞！本来对手机外观从图片和视频都看到过，但是打开后现实手拿着看视觉上前后双玻璃还是很漂亮，周边做开也很精细。由于
第一次用苹果，一个人搞了一下午才搞定，这里对第一次安卓换苹果的用户提两点，SIM卡要剪成最小的，在剪卡前记得把原来手机上的号码，图片一类对自己有用的信息用qq同步软件，同步一下，
以防像我的情况新手机读不出来手机卡上的号码，各种麻烦。手机到今天也用3天了，对操作系统的慢慢熟悉，用着感觉确实比安卓系统更方便，简捷，流畅，很是满意！其他功能性东西慢慢去了解，
不作评价，因为自己还没搞明白。很好的一次购物。毕竟是双玻璃，东西到了就给它加上防护了。不漏后壳谁知道我用的啥！哈哈！低调！
```

图 2 清洗后数据

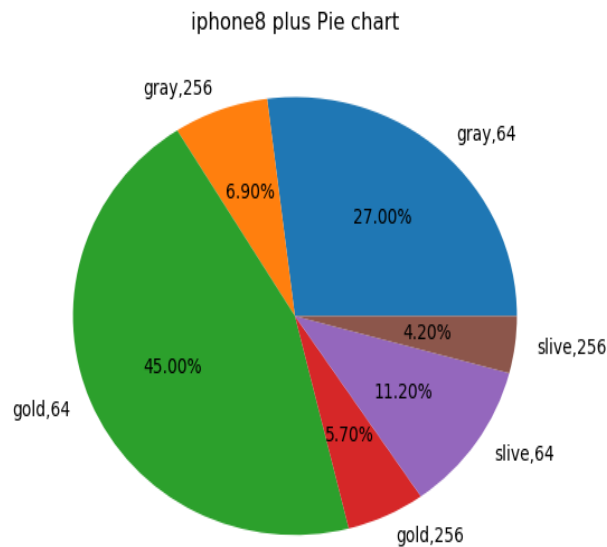
3.2. iphone8 plus 各个颜色、型号的受欢迎程度

利用 matplotlib 画饼图进行分析：



可以看出金色是最热销的颜色，占到销量的一半以上，而银色仅占到整体的 15.4%。

从存储容量来看，64GB 版本是 80% 以上消费者的选择，而只有 16.8% 的消费者选择显得容量过大的 256GB 版本。



综合各个型号的饼图，金色 64GB 版本是最热销的版本的 iPhone 8 Plus。

3.3. 对评论制作词云和进行情感分析

3.3.1. 情感分析

利用开源中文情感分析库 snowNLP (<https://github.com/isnowfy/snownlp>) 对评论进行情感分析。

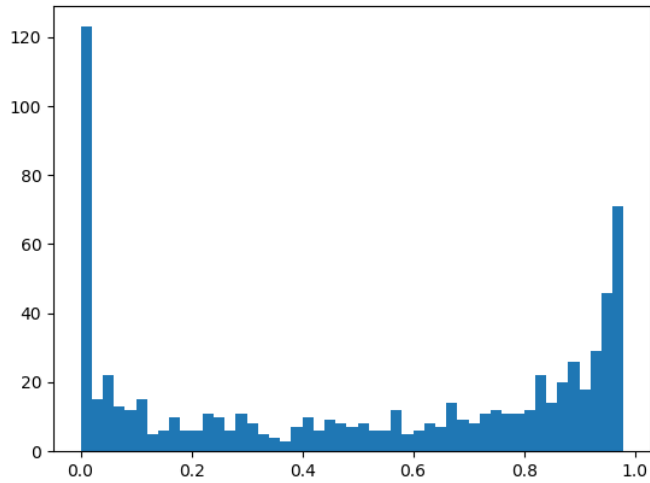


图 3 情感分析结果

注：0-1 的区间，越靠近 1 越积极，靠近 0 为消极。

图中出现较为明显的两极分化的评论，选取部分区间评论查看他们的评价。

[0-0.2]展示部分：

内容呢机首质的二直跳出来，控制不了
0.1975713684063622
知道是月份才买的，买量了三百多
0.175109321864439
夫之前就网上看到各种边框有缝隙，摄像头有灰尘，果然到手后边框有缝能插纸，后面两个摄像头都有灰尘，中框上还有一点不明显的掉色，估计追求完美的人会受不了，我自己觉得在接受范围内，懒得换了，麻烦，第一次用苹果系统，各方面还在摸索中。
1.5880991573685733e-06
响的刚买第二天就降价
0.11660871731351874
物流还是很快，早上5点抢到，第二天早上就收到，开箱没什么感觉，和7差不多，就这样吧。
0.0072911320757093
实话实说，手机刚到手也担心出现网上说的那些各种问题导致拆封前犹豫不决迟迟不敢下手，然后心一横就给拆了，拿出来顺利激活查了序列码等一系列流程，把该下的app都装上，玩了整整一晚发现这玩意儿真的没问题，流畅到不行，没等到X确实挺遗憾的，但真心讲，8p的体验不会比X差多少，也算个心安理得吧，但不管怎样，的确各方面都满意，黑色好看的不行，就仅仅是指纹真的太明显了这也没办法毕竟玻璃，总体满意，再使用一段时间吧，没什么问题会继续追评给大家真实的使用体验。
0.09883819237562402
实话实说，手机刚到手也担心出现网上说的那些各种问题导致拆封前犹豫不决迟迟不敢下手，然后心一横就给拆了，拿出来顺利激活查了序列码等一系列流程，把该下的app都装上，玩了整整一晚发现这玩意儿真的没问题，流畅到不行，没等到X确实挺遗憾的，但真心讲，8p的体验不会比X差多少，也算个心安理得吧，但不管怎样，的确各方面都满意，黑色好看的不行，就仅仅是指纹真的太明显了这也没办法毕竟玻璃，总体满意，再使用一段时间吧，没什么问题会继续追评给大家真实的使用体验。
0.09069572700833993
原装正品，玻璃机身手感比7P好太多，这个颜色看着好像粉色?有点后悔没上金色，之前7就是亮黑，粘指纹，所以这次直接pass。本来打算等双十一再下手，但是还是没忍住。
0.024537137871857184
上午下单，下午到货，有专人送货，并且需要验证码确认才能收货，对机主审美疲劳了，没有什么感觉，一年多没有用过ios系统，有点不习惯，发现一个问题，登录ID下载安装百度地图，然后打开地图下载离线地图，暂时无法不上网络，其他App也连不上网络，包括wifi和数据都连不上，进入网络设置界面后就卡住了，按返回键可以退出，卸载百度地图后其他一切正常。是系统不兼容呢？还是个案？
6.431983612387171e-09
开拆第二天早上8点一直等，总是没有货，等了一天，失望了一天，幸亏设置了到货提醒，等到半夜12点突然通知有货了，但是只有5分钟抢货时间，我是立马下单付款，隔了一天第三天到的，整体上都不错，挺好的，很舒服，只是苹果ID注册和登录花了我一下午时间，不能下载软件，总是提示错误，联系iTunes工作人员，打电话排队一个半小时，终于整好了，真是一路坎坷，整体还是挺满意的
9.776362430713448e-08

[0.4-0.6]：

早上八点抢购，次日收到的货，由于区域没有京东店，但是京东物流在业内是出了名的快，就是这么666，全新iOS11系统，不打算升级，手上苹果6也是在京东买的，系统升到8.41，依旧给力，用iOS系统只要不升级，用三年绰绰有余，这个是安卓无法媲美的，苹果就是数据线恶心，如果不是原装线充电不上，不管买多贵的线依旧如此，原装线容易开裂，硬伤。
0.4758833273536131
新电话，搞了挺久的还是自己慢慢琢磨吧？昨天凌晨一点钟，下的单，今天下午就到了挺快的快递。
0.5353229958318392
好.....
0.5072640160946941
刚拆包装，第一眼过去质感非常不错，后面板是玻璃的，简直就是一件艺术品！
0.4941173766393959
感觉还可以，而且还有手机壳虽然不怎么喜欢，不过质量蛮好的，还没开用，等下再说一下
0.41444053130641466
没想到那么快到了，希望没有问题毕竟首发
0.4696863964678234
没想到那么快到了，希望没有问题hellip;hellip;毕竟首发
0.44363290938971524
感觉很不错，手感都很好，喜欢它的指纹支付，刚开始去应用商店下载app的时候有点麻烦，之后就很好多了，再就是在修改苹果ID账户信息的网络不好的话会有点卡，其他的都很不错。
0.5066677568730278
物流确实给力，手机颜色好评
0.564103324915191
开箱检查还不错，整体外观没有瑕疵，开机小试，没有激活，还没查代码的一致性。
0.4392779156016342

[0.8-1.0]：

东西收到了，家人都很喜欢，京东物流真快。
0.9281349444944961
昨天刚到，用着一般，系统优化没到位，总感觉不是很流畅
0.972636783290846
非常不错，很喜欢。买新不买旧，我几乎用过所有苹果产品，8比7还是有很多不同的，支持京东。
0.9840090802015523
第一次在京东上买苹果手机，用惯了这个系统，还是想接着用，这次的物流不算快，没有之前的速度了，可能是其他因素，手机是没一点问题，一贯的漂亮好用，我会好好珍惜的
0.99924135693099046
非常感谢京东给予的优质服务，从仓储管理、物流配送等各方面都是做的非常好的。送货及时，配送员也非常的热情，有时候不方便收件的时候，也安排时间另行配送，同时京东商城在售后管理上也非常好的，以解客户忧愁，排除万难，给予我们非常好的购物体验。顺祝！thank you very much for the excellent service provided by Jingdong mall, and it is very good to do in warehouse management, logistics distribution and so on. Delivery in a timely manner, distribution staff is also very enthusiastic, and sometimes inconvenient to receive the time, but also arranged for time to be delivered. At the same time in the mall management Jingdong customer service is also very good, to solve customer suffering, overcome all difficulties. Give us a very good shopping experience. Shu h. Kei. kei!
0.9723430431184501
手机手感很不错，颜色也很漂亮，速度流畅，之前6s更换过来的，上手好。
0.9998354169511996
很漂亮，第二天早上就拿到了，送货上门，开心
0.884164885270224
京东送货给力，手机很好
0.8403253915807718
第一次买苹果，感觉还不错，毕竟是最新款，到手先验证了一下真伪，果然没问题，相信京东。

从部分实例看出，在[0.8-1.0]及[0.4-0.6]及中立及积极区间内的评价较为准确，但在消极[0-0.2]区间内的评论有比较严重的失准情况出现。在对手机某些特征如手感、外观、材质等情感分析准确度较低。

查看 snowNLP 的原始训练集发现，两端数据集中没有关于手机等消费电子产品的数据，可能是上述情况产生的原因。

3.3.2. 词云

使用 jieba 分词 和 wordcloud 建立词云：



“京东”、“自营”、“快递小哥”出现得最多，但这些词汇与 iPhone8Plus 无关，而是对购物平台京东的评价。原因可能是在数据处理时没有处理反水军评论，无法确定这些好评中的水分。但这个现象还是印证了生活中大家对于京东的评价甚高，如自营的正品保障以及快递速度高等等。

寻找关于 iPhone8plus 的词汇可以看到“系统流畅”、“玻璃材质”、“无线充电”等，这些都是 iPhone8plus 的重要卖点。

当然字体最大的苹果手机和苹果可以看出，消费者对于苹果这个品牌的认同感。