

# 基于应用商店数据挖掘的旅游导航软件用户行为研究报告

## 1. 研究动机

随着近些年来科学技术的进步，导航软件逐步发展进步，普及范围也越来越大，并且逐渐对于民众的生活产生各种各样的影响，极大地便利了民众的出行，尤其是在外出旅游等场合，导航软件的使用都可以帮助人们更好地提高生活质量。导航软件，首先可以提供电子地图，相比传统地图，电子地图管理方便，实现计算机数据库管理，可以对地理信息及时更新。如今的导航软件往往可以进行矢量地图渲染，显示高质量地图效果，同时结合 AR 虚拟实景功能，给用户的感官提供了很高的舒适度，同时也保障了准确度。然后还有在线导航功能，导航软件通过实时在线导航，提供偏航判定，偏航重导功能，给用户规划了行程路线，给用户提供了单纯从地图上无法得知的路况信息，同时，语言导航功能等的应用，同样提高了用户的使用体验。



图 1.1 导航软件极大地便利了人们的生活

同样，如今旅游行业的兴盛，也带动着旅游软件功能的完善与发展。如今较为成熟的旅游软件就有很多。由图 1.2 旅游软件的发展：民众对于旅游软件的接受度也已经达到了较高的水准。由图 1.3 各旅游软件的活跃度：旅游软件大多提供机票，酒店的预定，以及旅游攻略来满足游客的旅游需求。目前的旅游软件 APP 数量繁多，但是功能又各有局限，有关于旅游攻略类的，有关于车票酒店预订类的，有关于行程记录分享。为了一次出行，或许需要下载很多 APP。

也就是说，如今在旅游导航领域，并没有一个很好的软件提供综合性的服务来满足民众对于旅游以及导航的综合需求。于是为了给这类软件的研发提供思路，我们试图整合旅游软件以及导航软件的评论，进行一个经验研究，讨论并给出设计这类软件的建议。



数据来源：艾媒咨询，《2018上半年中国在线旅游行业研究报告》

图 1.2 旅游软件的发展

2018年8月中国旅游出行APP月活跃人数排行榜TOP10			
排名	应用	活跃人数 (万)	环比增幅 (%)
1	携程旅行	6,122.16	+2.78%
2	去哪儿旅行	3,664.77	+2.44%
3	同程旅游	1,315.12	-19.0%
4	飞猪旅行	1,133.53	+2.56%
5	蚂蜂窝自由行	980.56	-1.95%
6	途牛旅游	780.37	-2.74%
7	艺龙旅行	289.76	+3.10%
8	TripAdvisor猫途鹰	112.79	+0.41%
9	百度旅游	74.95	-4.93%
10	驴妈妈旅游	34.34	-2.75%

数据来源：艾媒北极星系统(bjx.iimedia.cn)

图 1.3 各旅游软件的活跃度

## 2. 相关工作

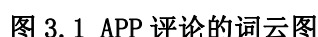
我们首先搜索了相关研究工作以确定我们的研究方法和研究角度。当前，获取移动应用用户需求的方法主要有访谈、调查<sup>[1]</sup>、分析用户活动日志<sup>[2][3]</sup>、挖掘应用商店数据<sup>[4]</sup>等等。访谈、调查容易获取用户的行为特点和背后的原因，却难以达到比较大的数据量；分析用户活动日志可以获得大量用户行为数据，却难以分析用户行为背后的原因。而挖掘应用商店数据则能比较好地兼顾这两方面的需求。

已经有很多研究者利用挖掘应用商店数据的方法来研究用户的行为特性。Pagano 和 Maalej 挖掘了 iOS 应用商店中排行前 25 的应用的评分和评论，用统计学方法调查用户如何进行评分和评论，以及何时进行评分和评论的问题<sup>[5]</sup>。他们的研究指出，用户评论集中出现在版本刚更新的时候，然后评论速度随着时间逐渐下降。用户的评论常常不是针对单一问题的，而是综合多个方面的综合评价，比如汇报 bug、用户体验、功能建议等。Harman, Jia, Zhang 等人挖掘了黑莓应用市场中付费应用的数据，包括应用描述、类别、用户评分、价格、



旅游、导航类 App 的用户特性也有一些研究者做了研究,但大部分研究者用的都是调查的方法,鲜有挖掘应用商店数据的研究方法。地图导航类 App 方面,吴婷通过访谈和问卷调研的方法对目标用户进行研究与分析,总结出获取型、分享型、社交型、自娱型四类用户及用户的个性化需求,并在调研结果的基础上,从本能、行为、反思三个层面概括了手机地图出行应用的用户体验目标,并提出了应用功能个性化的设想<sup>[8]</sup>。王永瑞从用户角度出发,结合用户使用百度地图的行为作为研究对象,以用户画像的分析方法对百度地图用户的用户属性、行为进行多维度的分析研究<sup>[9]</sup>。旅游类 App 方面,刘岚通过问卷调查、数据分析的方法,分析了影响地图类旅游 App 用户粘性的因素,发现感知易用性、感知有用性可以通过技术满意度对持续使用意愿产生正向的影响作用,且感知易用性作用大于感知有用性的作用<sup>[10]</sup>。潘美娟运用意向调查的方法调查了大学生的使用在线旅游 App 的意向,发现,旅游 APP 的使用并不局限于旅游活动;支付方式是学术关注的重点;服务质量和售后评价的影响非常大<sup>[11]</sup>。施香君,许明月,万绪才运用问卷调查和统计分析法,对旅游网站的用户信任度进行了评价,并进一步分析了信任度的影响因子及其结构关系。该研究表明,网站使用频次、在线评论信息、个性化定制服务、服务保障、网站信息可靠性和价格等对网站的信任度有着不同程度的影响,并且这些影响表现出一定的人口学差异<sup>[12]</sup>。

- (1) 用 python 爬虫爬取 IOS 应用商店中多个 APP 评论信息，每个应用爬取 500 条评论。
- (2) 发现评论中有广告，并且广告内容基本都相同，涉及色情内容。利用敏感词对评论进行清洗。
- (3) 信息量太大，对于多个 app 无法准确评价，先制作词云图，观察评论中出现频率较高的一些词语，观察大致评论信息。



(4) 编写 python 脚本，对评论长度进行统计，下图为概率密度函数，纵轴为对应评论长度

的评论数，横轴为评论长度。

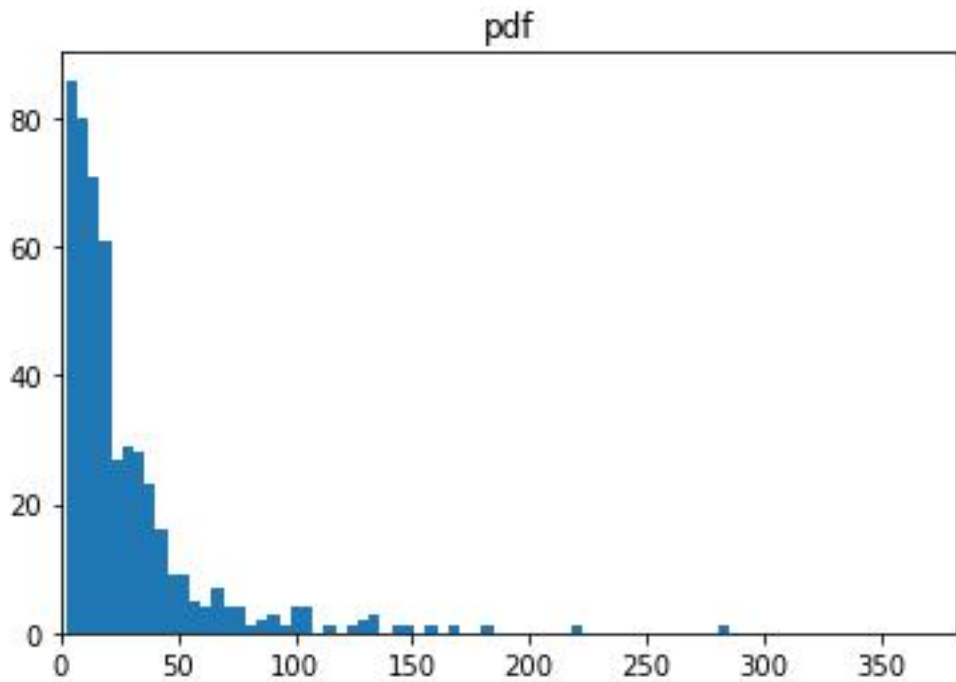


图 3.2 评论长度的概率分布

(5) 编写 python 脚本，将评论按用户评分进行分类，并分别统计各分数段评论的数量。

评分次序从左至右由 1-5 的评论数分布如下：

高德：[115, 19, 32, 34, 290]

百度地图：[135, 22, 40, 47, 256]

携程旅游：[53, 3, 5, 10, 429]

去哪儿：[11, 0, 2, 0, 487]

通过观察我们可以发现大部分应用的评分十分极端，不是 1 分就是 5 分，这是意料之外情理之中，一般去应用商城评论 app 的用户，不是因为对 app 意见很大来吐槽发泄情绪，就是因为这个 app 很好，如果一个 app 中规中矩用户一般不会去评论。不过评论的两极分化并不影响我们研究应用的需要改进的地方与受用户欢迎的地方。

(6) 观察到不同评分中评论数目的差异，将评论按评分分类，分别分析，对高德地图评分为 1 的词云图，将 topic 大致分成：导航错误、定位不准、推送垃圾、信息更新不及时、导航声音不合适、其他功能错误等多个类别。



图 3.3 高德地图评分为 1 的词云图

(7) 通过人工+关键词词频统计，统计不同 topic 的个数，得到统计结果表格如下：

表 3.1 通过人工关键词分类统计得到的评论主题分布结果

打分	1	2	3	4	5
多收费/欺骗消费者	29	2	0	0	0
抢不到票	3	1	0	0	0
信息不准确	2	0	0	0	0
应用卡顿/不稳定	7	2	4	1	0
其他功能错误或建议	8	0	0	0	0
功能好用方便	0	0	0	0	693
客服态度以及找不到	10	0	0	0	0
总数	59	5	4	1	693

(8) 数据量较小时可以人工分类，但是效率极低，样本量还不够，在已经人工标记的基础上，我们思考如何基于机器学习，对于更大规模评论进行自动分类。文本分类的机器学习算法有支持向量机、朴素贝叶斯等多个方法，我们分别进行了尝试。最终选择正确率为 76.9% 的朴素贝叶斯分类作为机器学习算法。

## 4. 技术难点与问题

### 4.1 数据来源

#### 4.1.1 介绍

评论分析，数据来源是基础。客观高质量的原始数据能保证结论的可靠正确。主要的评论信息来自于 python 网页爬虫，爬取了 IOS 应用商店中有关导航类和旅游类多个 APP 评论信息，每个应用爬取 500 条评论。网络爬虫是按照一定的规则，自动地抓取万维网信息的一种程序或者脚本。python 爬虫架构可以分成 URL 管理器、网页下载器、网页解析器等几部分，依赖不同的库完成相应操作。

#### 4.1.2 技术细节

(1) 利用 request 库向目标网站发送一个网页访问请求 request，服务器等到请求之后，反馈相应的 response，这个 response 就包含了我们平时访问网页所浏览的全部信息。

(2) response 中的信息经过 utf-8 解码之后得到 html 文件，利用 BeautifulSoup 库进行解析。然而由于包含了网页中的全部信息，不相关信息太多，我们需要筛选出我们需要的评论信息。

(3) 我们利用 re 库通过正则表达式，高效地筛选出我们想要的评论内容与评分，将这些爬虫得到的评论写入 txt 文件。

(4) 自动更换页面爬取下一页数据，如果网站有反爬虫措施，休眠一秒并更换网页请求的 header 以进行伪装。

### 4.2 评论分类分类

#### 4.2.1 动机

数据量较小时可以人工分类，但是效率极低，样本量还不够，在已经人工标记的基础上，我们思考如果基于机器学习，对于更大规模评论进行自动分类。文本分类的机器学习算法有支持向量机、朴素贝叶斯等多个方法，我们分别进行了尝试。

#### 4.2.2 模型介绍

##### (1) 朴素贝叶斯分类 (Nave Bayes)



朴素贝叶斯分类（Nave Bayes）是基于贝叶斯定理，被广泛应用于互联网新闻的分类、垃圾邮件的筛选等分类任务。

如果假设  $X$  是一个未标注的测试样本， $H$  是一个假设，如假设  $X$  数据类  $C$ ，那么  $P(H|X)$  表示当确定  $X$  条件下  $H$  成立的概率， $X$  是给定的观测样本。这其实就是  $H$  的后验概率。利用贝叶斯定理，

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

其中， $P(H)$  表示  $H$  的先验概率， $P(X)$  表示  $X$  的先验概率， $P(X|H)$  表示  $H$  条件下  $X$  成立的概率，即后验概率。

当样本属性是多个维度时，此时假设各个属性是相互独立，即

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} = \frac{P(H)}{P(X)} \prod P(x_i|H)$$

对于给定的待分类样本，求解在此样本出现的条件下各个类别出现的概率，哪个最大，就认为此待分类样本属于哪个类别。

(2) 支持向量机

支持向量机的目的是寻找一个超平面对样本进行分割，分割原则是间隔最大化，最终转化成一个凸二次规划问题来求解。

- 当训练样本线性可分时，通过硬间隔最大化，学习一个线性可分支持向量机；
- 当训练样本近似线性可分时，通过软间隔最大化，学习一个线性支持向量机；
- 当训练样本线性不可分时，通过核技巧和软间隔最大化，学习一个非线性支持向量机。

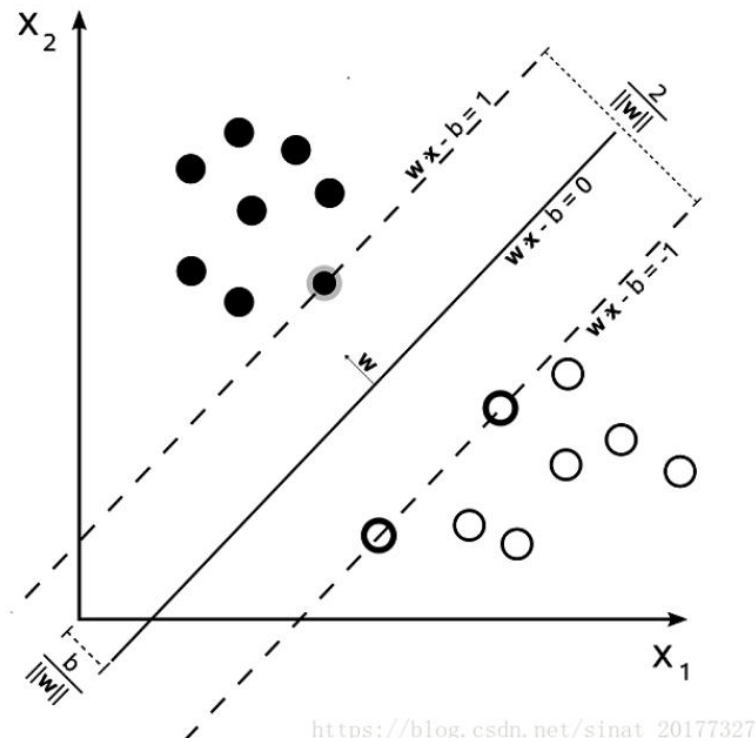


图 4.1 支持向量机中最优分类面示意图

4.3 技术细节

4.3.1 数据预处理

(1) 清洗广告：

利用敏感词对评论进行清洗，对评论进行便利，如果评论中含有“美女”、“薇信”等敏感词汇，就将其视为广告，去掉这个评论。对于导航类 APP，我们高德地图，去掉约 2% 的涉嫌广告评论。

自然语言中有大量 stopword 像“的”、“啊”、“了”，这种词语出现频率高且对文本分类没有帮助。利用 jieba 库将这种 stopword 清洗掉。

利用 sklearn 的 CountVectorizer 函数，将文本映射到特征向量。

### 4.3.2 模型搭建

```
In [3]: runfile('G:/zhangjinming/C++/python/20181206/svm.py', wdir='G:/
zhangjinming/C++/python/20181206')
['不走捷径' '一点都不准绕来绕去的' '定位不准确老是掉线' ... '太实用了谢谢' '好评'
'不错哦']
[0 0 1 ... 6 6 6]
G:\zhangjinming\Anaconda\lib\site-packages\sklearn\cross_validation.py:41:
DeprecationWarning: This module was deprecated in version 0.18 in favor of
the model_selection module into which all the refactored classes and
functions are moved. Also note that the interface of the new CV iterators
are different from that of this module. This module will be removed in 0.20.
"This module will be removed in 0.20.", DeprecationWarning)
0.6206896551724138

In [4]:
```

图 4.2 利用支持向量机模型分类的程序分析结果

因为 SVM 分类结果似乎不太理想，我们尝试用朴素贝叶斯分类器进行分类，建立贝叶斯分类器，用数据集训练。

```
20 pattern = re.compile(r'(\w{4,6})(\w{5,7})')
21 filterdata = re.findall(pattern,line)
22 comment = ''.join(filterdata)
23 X.append(comment)
24 Y.append(i)
25 infile.close()
26
27 infile=open("%s/%s.txt"%(dir,file),"w+",encoding='utf-8')
28 for line in infile:
29 pattern = re.compile(r'(\w{4,6})(\w{5,7})')
30 filterdata = re.findall(pattern,line)
31 comment = ''.join(filterdata)
32 X.append(comment)
33 Y.append(i)
34
35 new_X=np.array(X)
36 new_Y=np.array(Y)
37 per = np.random.permutation(new_X.shape[0])
38 new_X = new_X[per]
39 new_Y = new_Y[per]
40 print(new_X)
41 print(new_Y)
42 from sklearn.cross_validation import train_test_split
43 X_train,X_test,Y_train,Y_test = train_test_split(X,Y, test_size=0.15,random_state=42)
44
45 from sklearn.feature_extraction.text import CountVecorizer
46 vec = CountVecorizer()
47 X_train = vec.fit_transform(X_train)
48 X_test = vec.transform(X_test)
49 # 使用平滑处理初始化的朴素贝叶斯模型
50 mnb = MultinomialNB(alpha=1.0)
51 # 利用训练数据对模型参数进行估计
52 mnb.fit(X_train,Y_train)
53
54 # 对测试样本进行类别预测,结果存储在变量y_predict中
55 y_predict = mnb.predict(X_test)
56 from sklearn.metrics import accuracy_score
57 result=accuracy_score(Y_test,y_predict)
58 print(result)
```

图 4.3 利用朴素贝叶斯分类器的程序代码

将总评论的 80%用于训练，20%数据用了测试，统计正确率。

```

In [25]: runfile('G:/zhangjinming/C++/python/20181206/bayes.py', wdir='G:/
zhangjinming/C++/python/20181206')
['后台一群死人从不更新路况' '就是好不知道为什么' '不走捷径' ...
'一直使用并信赖高德地图也从来没让我失望过希望越做越好而不是像滴滴自己把自己玩儿
坏了' '有版的吗' '满意']
[4 6 6 ... 6 6 6]
0.7692307692307693

In [26]: runfile('G:/zhangjinming/C++/python/20181206/bayes.py', wdir='G:/
zhangjinming/C++/python/20181206')
['路口没有详细图'
'高德比百度地图好用而且准但是有一点让我只能用百度就是高德在开启导航时总会说一声
什么什么开始导航然后才来得及手动关声音而且再启动还要手动静音第二退出后台总要说高
德提醒您吧啦啦啦的很烦希望增加永久静音功能望采纳'
'方向准确导航无差错' ...
'先前挺信任高德的可是都好几次了明明说开始导航了我都走了几百米了导航的米数丝毫不
变今天更奇葩明明右手边就没路非要我右拐'
'导航状态信号弱根本不能导航' '在线查找速度快路况播报及时挺好用的']
[6 6 6 ... 0 6 6]
0.7692307692307693

In [27]:

```

图 4.4 利用朴素贝叶斯分类器的程序分析结果

最终得到其正确率为 76.9%。

## 4.4 对于分类器性能的评价

### 4.4.1 正确率没那么高的原因：

- (1) 朴素贝叶斯假定各属性之间相互独立，在训练时没考虑各个特征之间的联系，模型本身有局限性。
- (2) 数据集还不够，四百余条太少，增加标注好的数据集会提高精度。
- (3) 有的 topic 数目偏少，影响分类效果。

### 4.4.2 未来提高的地方：

- (1) 改进模型，利用 LSTM 模型分析文本。
- (2) 增加标注数据集规模。

## 5. 分析

通过数据采集部分，我们收集到了 iOS、华为、小米三个应用商店的评论整合，其中由于 iOS 应用商店爬取评论非常方便我们一共采集了 500 条评论数据，而 Android 系统的华为、小米应用商店无法爬取数据也无法复制评论，因此我们只好人工分析数据，每个应用商店一共提取了 100 条评论数据。其中导航 APP 我们选取的是高德地图和百度地图；旅游 APP 我们选取的是携程和去哪儿。最终按照评论分类标准的不同我们进行了如下分析：

### 5.1 评分分布

通过各 app 的评分分布，可以发现，评分主要都呈现一个两级分化的趋势，通过分析，可以理解，在给出评分时，用户一般会根据自身的使用体验，通过较为极端的评分表示赞同欣赏或者讨厌，除了 5 星用户一般没有差评，在 1，2，3，4 星评论中都有用户给出的各类建议。而 5 星部分，所占比例异常高，一般源于网络水军的干扰。所以，在后面的分析中，也会将评分异常的 app 的 5 星评论忽略。



各个应用的评分分布情况

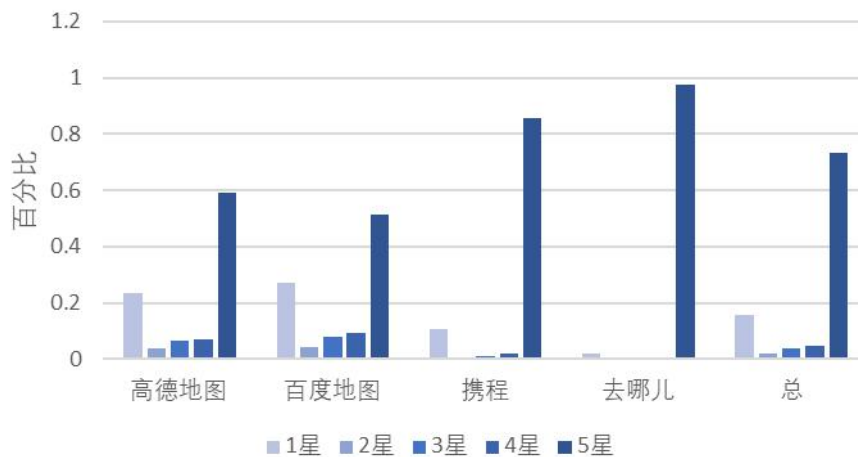


图 5.1 四个 APP 评分分布图

## 5.2 每个 APP 按不同星级分类后各评分星级下各种评论的比例

### 5.2.1 导航 APP：高德地图与百度地图

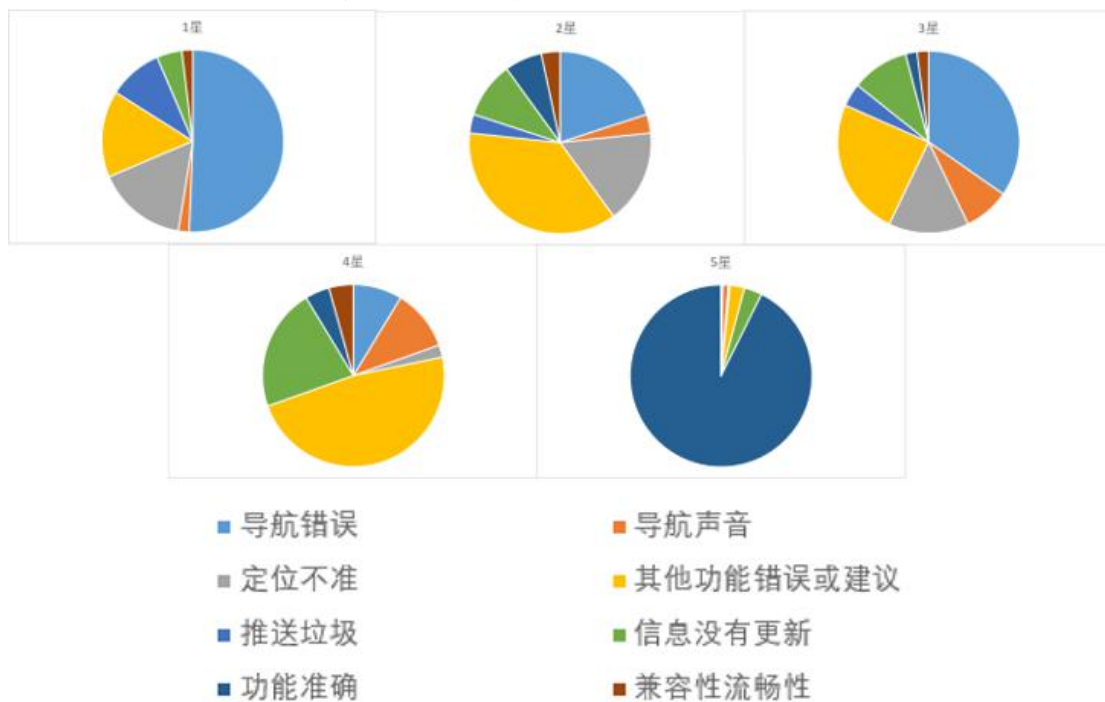


图 5.2 高德地图不同星级下各种评论类别的占比图

对于导航 app，在 1 星评论中，共同点是最主要原因因为导航错误，说明人们给出差评的原因还是出自导航软件最基本的考虑，希望导航软件更加准确可靠。主要原因是其他功能错误，说明人们给差评的最主要原因还是看功能。此外，可以看出，百度地图主要是信息更新不足，以及导航声音让人们不够满意，而高德地图则是，定位不准。

在 2, 3, 4 星评论中，相同点是，导航错误、信息没有更新和其他功能错误或建议呈现三足鼎立。说明设计导航 APP 最基本的功能的三个要素一直都是最基本的考察标准。两个 APP 中随着评定星级的升高，导航声音的比重越来越大。说明在满足了基本考察之后，导航软件的声音是非常重要的提分项，并且越喜欢使用软件的客户越希望声音好听。在这些评论中，两个软件并没有太大区别。

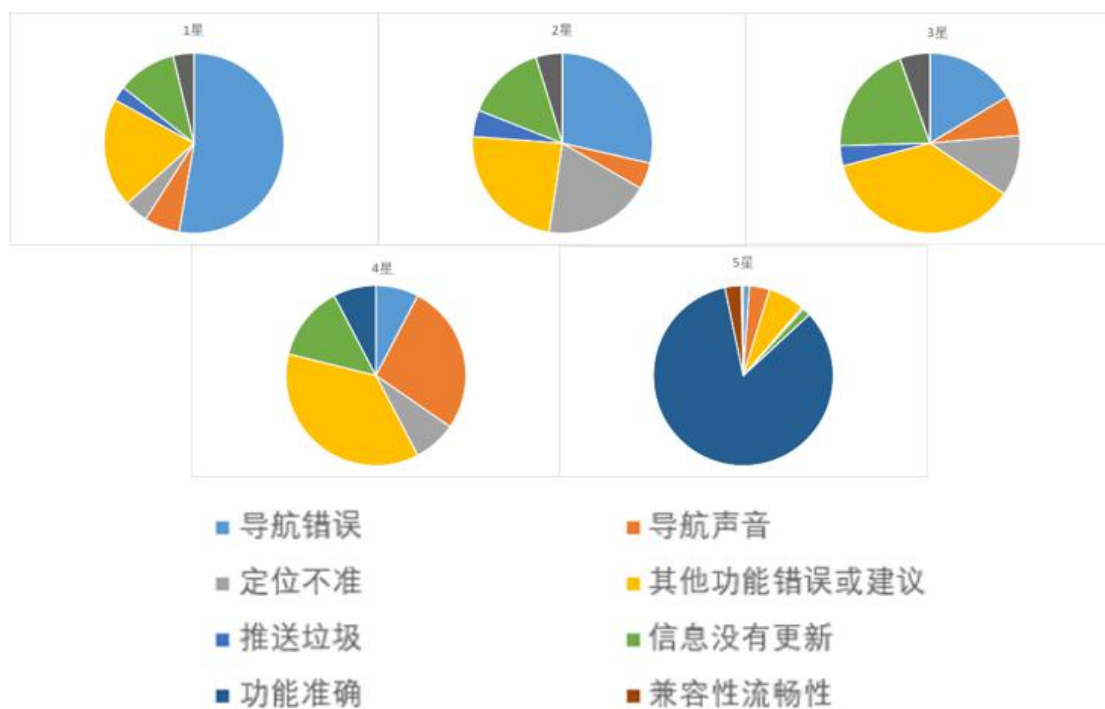


图 5.3 百度地图不同星级下各种评论类别的占比图

在 5 星评论中，相同点是，功能准确。说明人们给出好评和给出差评的原因基本一样，就是功能好不好。同时也会出现一些关于其他功能错误的评价，但是影响不大。对于百度地图主要是，表扬导航声音和表扬语音输入。

#### 5.2.2 旅游 APP：去哪儿 APP 和携程 APP

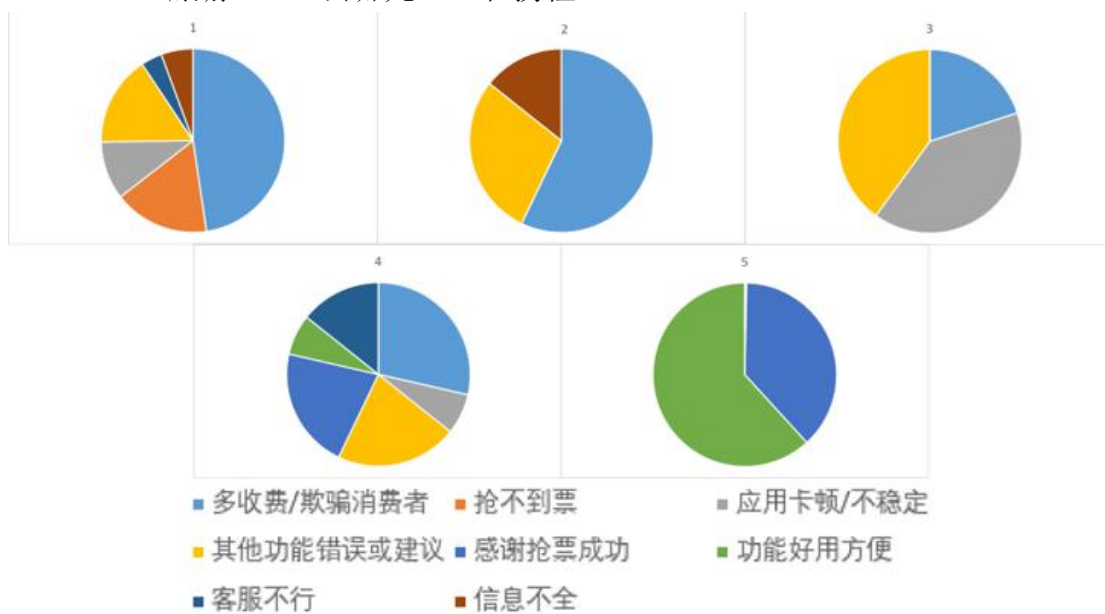


图 5.4 携程 APP 不同星级下各种评论类别的占比图

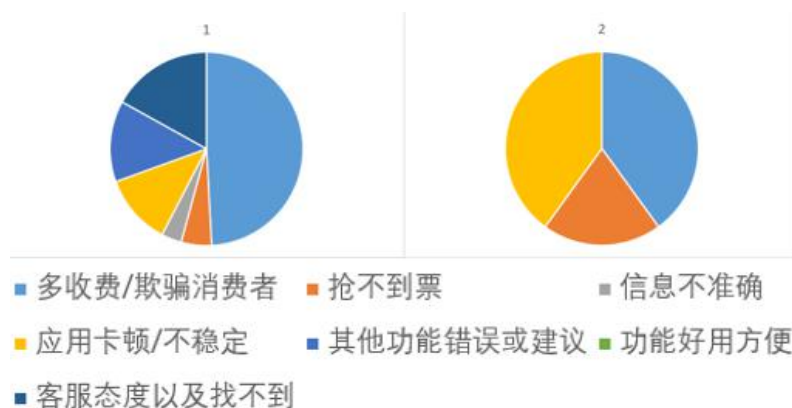


图 5.5 去哪儿 APP 不同星级下各种评论类别的占比图

对于旅游 app，1 星评论中，相同点主要在于，多收费、欺骗消费者。说明人们使用旅游 APP 最在意的就是诚信和价格是否实惠。同样，其他功能或建议与应用卡顿也有影响。说明人们使用 APP 最看重的还是功能和 APP 的流畅性，很可能与旅游时旅客非常看重旅游的时间安排，因为旅游往往非常赶时间，不流畅的 APP 很容易影响使用者的心情。对于携程，侧重点在抢不到票，对于去哪儿，侧重点在客服态度和无合适方案。

对于 2, 3, 4 星评论，携程主要存在多收费、欺骗消费者+其他功能错误或建议是不变的话题，还是基本功能。在 4 星评价及 5 星中才出现了关于抢票成功的表扬，说明很可能只有喜欢使用携程的用户才发现了抢票功能的好处，因此很可能说明抢票功能的 APP 竞争优势没有被很好的发挥出来。而去哪儿主要是应用卡顿。

对于 5 星评论，功能好是共同点。

另外，去哪儿的 3, 4, 5 星水军太多，无法分析。

### 5.3 针对每个 APP 不同评论类别观察每个类别下星级的分布

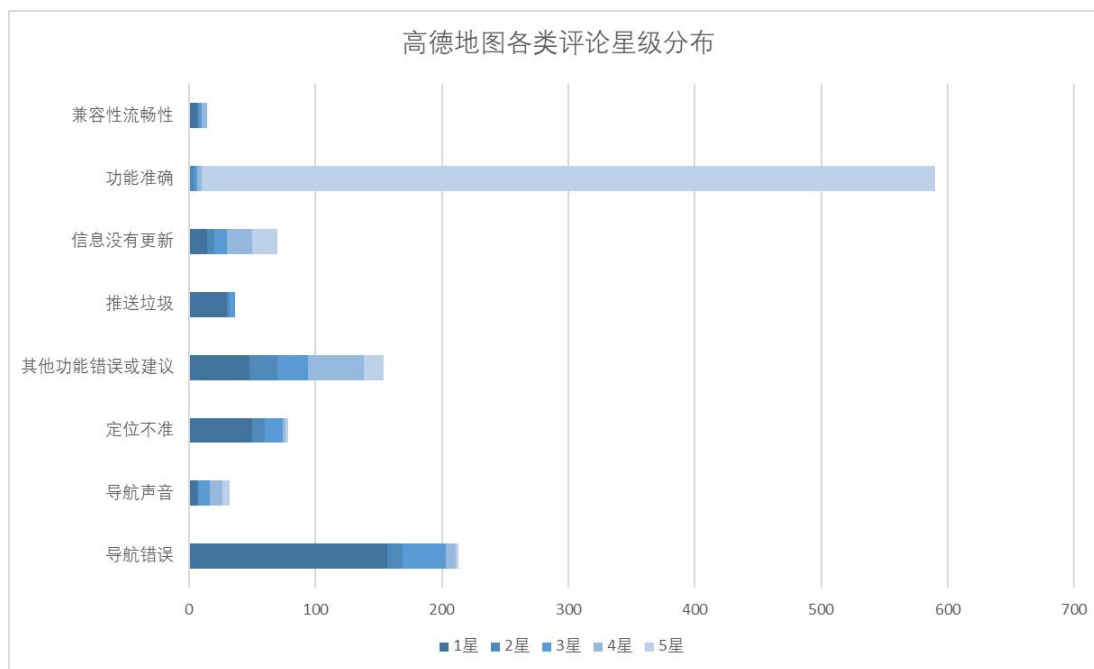


图 5.6 高德地图不同评论类别下星级的分布图



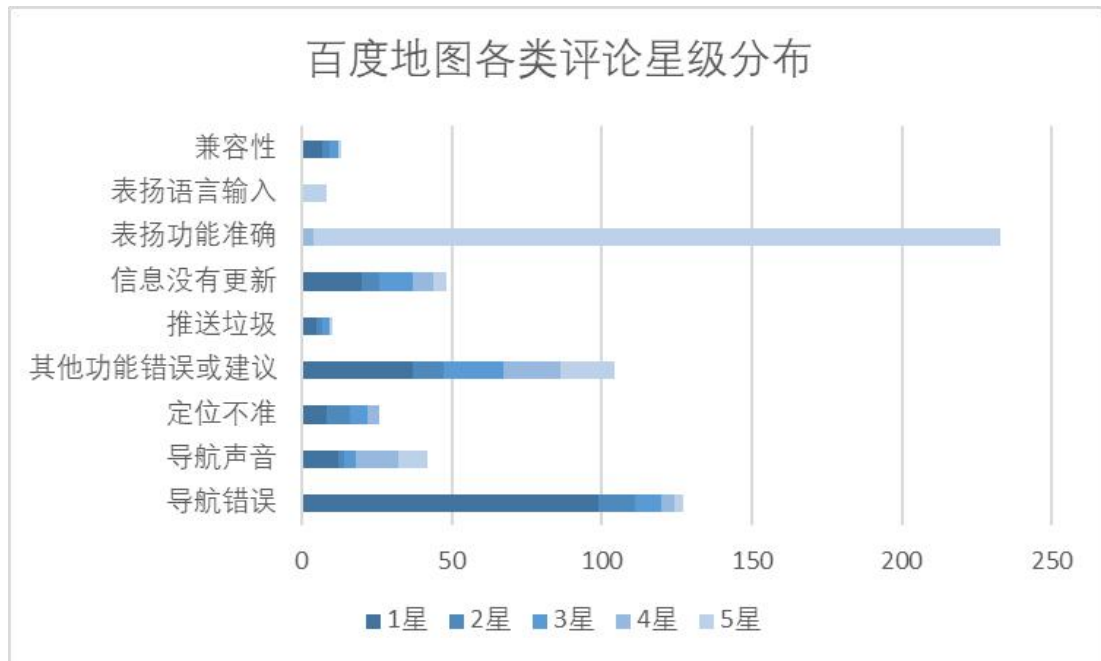


图 5.7 百度地图不同评论类别下星级的分布图

对于导航软件，根据每一部分评论内容中星级的分布（用户体验感受的分布），我们可以发现遇到：导航错误、定位不准、信息没有及时更新、推送垃圾、兼容性的用户更倾向于打 1 星，可见这些都是影响导航体验的主要因素。而导航声音和其他因素（比如，骑行导航等建议）都不会太多地影响用户的使用态度。对于导航错误遇到的用户基本都打了 1 星，也存在打 5 星的用户，这可以看出导航软件如果出现导航错误很大程度上可以抵消其他部分的优点。然后是导航声音，这个算是一个比较体验性的功能，评论在 1-5 星均有分布，说明这一话题根据个人感受的影响比较大，提高多样性以及播报信息的完整度都是可以提高的策略。对于定位不准，类似导航错误，但是往往和自身使用状况有关 所以 1 星比例比 导航错误较少。对于功能准确则是体现了实用性的重要。而推送垃圾，兼容流畅都是在说明一个 app 使用体验的影响，虽然所占比例较少，但是 1 星比例很高说明一旦受到这类问题的干扰，很大程度会影响对 app 的整体评价。信息更新这一问题所占比例也较低各星段都有，说明，信息更新速度的影响度较为复杂，有强有弱。在其他功能方面，比如骑行跟随，以及骑行路线选择，虽然所占比例较小，但是也是需要完善的一个地方，而卡车等选择路线，对路段信息的提示的完整度也极大地影响了卡车司机的用户体验。

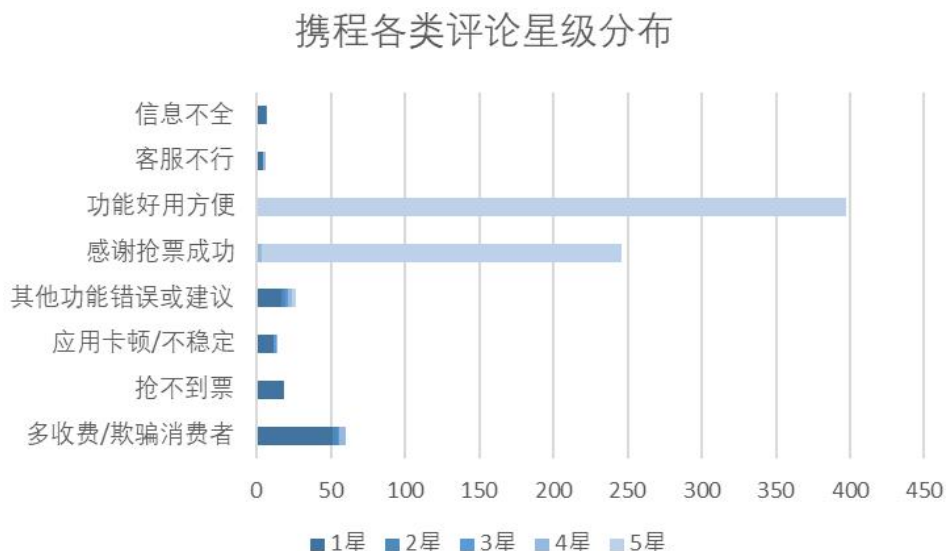


图 5.8 携程 APP 地图不同评论类别下星级的分布图

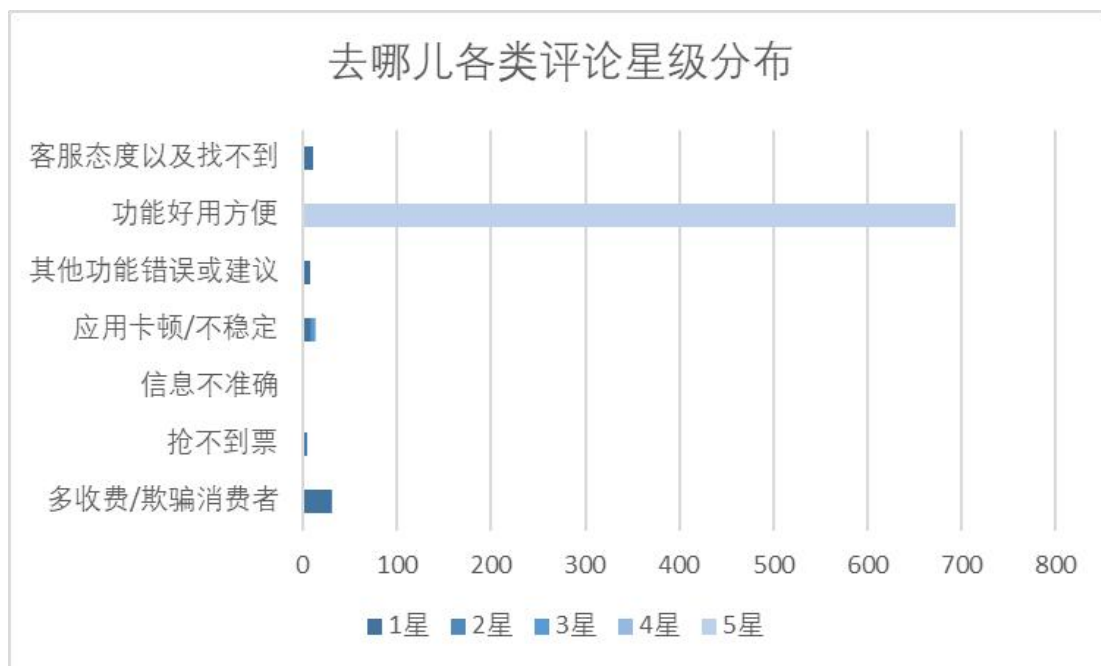


图 5.9 去哪儿 APP 德地图不同评论类别下星级的分布图

而旅游软件的分析，可以发现，抢票，收费以及客服反馈都是比较重要的环节。多收费/欺诈的 1 星比例很高，是很合理的，因为与用户利益相关。抢票，是大家关注的一个点，也是一个重要功能。关于抢票而引发的客服沟通问题，在知乎，贴吧上也十分热门。与客服的沟通问题，基本集中出现在 1 星，与客服沟通一般都是用户遇到问题才会去寻求客服的帮助，所以分布也很合理，这说明客服处理问题的态度能力都需要格外注意。

#### 5.4 关于评论星级和评论数量相关性的分析：

从挖掘到的数据来看，不同软件的不同评论内容在不同星级评论中的分布是不一样的。为了研究这种分布的差异性，更好地了解用户对于不同软件特性的敏感程度，我们计算了不同软件的各种评论话题在各个星级评论中的数量和星级的相关系数  $\rho$ ，其计算公式如下：

$$\rho = \frac{\sum_{i=1}^5 (i-3)(n_i - \bar{n})}{\sqrt{\sum_{i=1}^5 (i-3)^2} \sqrt{\sum_{i=1}^5 (n_i - \bar{n})^2}}$$

其中  $i$  表示评分星级，取值为 1 到 5， $n_i$  表示某个评论内容星级为  $i$  的评论中出现的次数。

在我们挖掘到的数据中，旅游类 App 的评论出现严重的两级分化现象，绝大部分分布在 1 星和 5 星，疑似存在刷评分的“水军”和竞争对手的商业攻击，研究其相关性系数意义不大，所以这里我们只考虑导航类 App。而导航类 App 中夸赞性评论几乎全部集中于 5 星，同样疑似有水军现象，研究相关性系数意义不大，故这里只考虑其中的缺陷性评论。相关系数的负号表示评分越低，用户反映该缺陷的倾向越大，正号表示相反的趋势，绝对值表示这种倾向的相关性大小。该相关系数越接近 -1，表明越敏感，容易因为此缺陷而打低分；越接近 1，表明用户对此缺陷不敏感，提出该缺陷只是对软件的改进建议。

由此，我们计算得到的高德地图、百度地图的各种软件缺陷的评论数和星级的相关系数如下表：

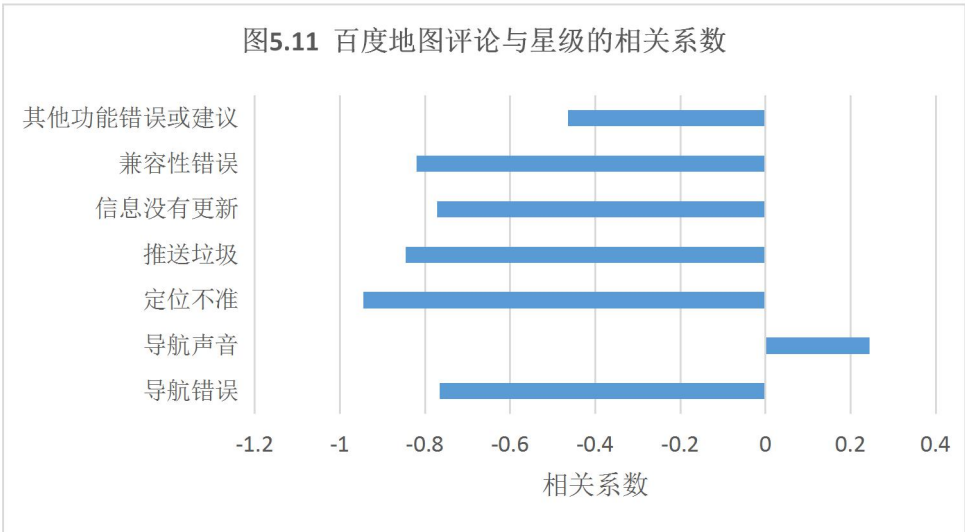
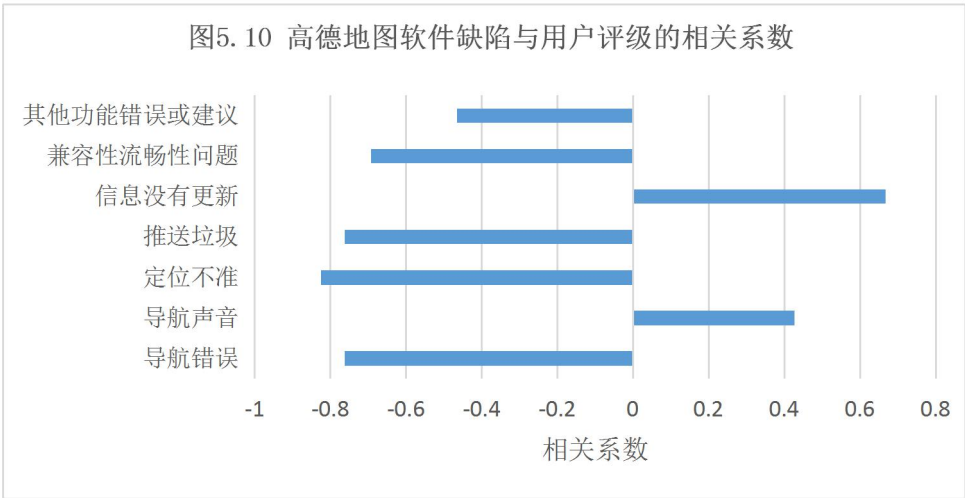
表 5.1 高德地图软件缺陷的评论数和星级的相关系数

评论内容	相关系数
导航错误	-0.76255
导航声音	0.426401
定位不准	-0.8255
推送垃圾	-0.76271
信息没有更新	0.666886
兼容性流畅性问题	-0.69338
其他功能错误或建议	-0.46586

表 5.2 百度地图软件缺陷的评论数和星级的相关系数

评论内容	相关系数
导航错误	-0.76555
导航声音	0.244339
定位不准	-0.94491
推送垃圾	-0.84515
信息没有更新	-0.77211
兼容性错误	-0.81929
其他功能错误或建议	-0.4639

绘制两个软件的各项相关系数条形统计图如下：





两个软件各项缺陷评论数与星级的相关系数大小排序如下:

高德地图: -1<定位不准<推送垃圾<导航错误<兼容性流畅性问题<其他功能错误或建议<导航声音<信息没有更新<1

百度地图: -1 定位不准<推送垃圾<兼容性错误<信息没有更新<导航错误<其他功能错误或建议<导航声音<1

可以看到两款软件在相关系数的排序上既有相似性又有差异性。定位不准都是用户最为敏感最无法容忍的问题,推送垃圾都是紧随其后;导航声音问题都是相关系数大于 0 的软件缺陷,两款软件的用户都对此不敏感,提出此问题的主要是打分较高的用户,他们希望软件能在此问题上更加优化。高德地图的信息更新问题与评分的相关系数为正并且为各项中的最大值,而百度地图此项为负值,说明高德地图的信息更新问题处理得更好,用户对此没有太大意见。

## 6. 结论

通过 6.2 和 6.3 节的分析我们可以得出一个相同的结论:不论是导航 APP 还是旅游 APP,最主要考虑的因素还是 APP 本身最主要的功能。因为在软件开发人员在开发过程中,常常有很多用户需求需要解决。对软件改进方向做决策时应该考虑到用户对各种问题的敏感性,用户敏感性最高的问题应该成为优先解决的问题。从上面的分析可以看到,定位、导航功能的开发和改进必须是导航 APP 开发人员首要考虑的问题;旅游 APP 对于游客利益的保障即为旅游 APP 开发最先要考虑的问题。

其次,软件运营商可以积极开发各种营利渠道,避免过多通过推送广告进行营利,另用户感到反感。同时由于导航和旅游 APP 的应用场景往往是用户处于非常急迫需要的情况,软件的流畅度是非常重要的问题,这往往需要开发者在功能的丰富程度和操作的流畅程度之间找一个平衡。

最后,对于导航 APP 语音问题是一个最个性化的因素,应该作为软件的润色方面,同时应该提供给用户尽可能多的选择,才能较好地实现个性化。

另外,对于水军控评对本研究造成的影响,我们建议软件开发者对给出评论的用户进行一定的筛查:只允许哪些熟悉软件功能并已经有一定的软件使用年龄的用户评价,同时可以针对本软件的用户群体开设信用额度,选取那些信用度高遵守网络文明的用户参与评价,防止来自于竞争对手等的水军通过重复创建账号进行灌水或控评。

## 7. 评价

### 7.1 研究优点:

- (1) 本研究分析了用户对各种软件缺陷的敏感程度,对软件开发人员的决策有很大的参考价值。
- (2) 本研究对采集到的评论数据采用朴素贝叶斯和支持向量机分类器进行了分类处理,并最终确定选取了自动分类准确率较高的朴素贝叶斯分类器作为我们的爬取评论数据的自动分类器,减轻了人工操作的负担。
- (3) 采用各种图表从不同角度分析了提取得到的数据,一定程度上结合了研究者生活中的经验分析处理了评论信息,为旅游导航类的软件开发提供了指导与建议。

### 7.2 研究不足与发展方向

- (1) 由于评论极大地受到了水军的干扰,无法分析旅游软件的评论与星级的相关系数。并且已经分析的数据中也不排除水军的存在,数据真实性难以验证;在今后的工作中能够找

出更有效的鉴别水军的自动鉴别程序,同时也希望软件开发者能注意到软件评论中的水军对用户带来的影响。

- (2) 评论分类比较单一,“其他功能错误或建议”仍有较多问题可以挖掘。关于评论的分类老师也给出了相关建议,认为我们不应该根据我们已有的经验自行人工地分类,因为得到的类别是我们熟悉的评论类别,而最好通过评论数据发掘新的评论角度和分类方法,这一点我们希望在未来能得到改进。
- (3) 提取的评论数据太少。真正用于最终评论分析的评论数只有 700 条。主要的原因在于 Android 系统下的应用商店无法爬取评论,只能使用人工方式进行评论的提取与分类,故数据提取的效率大大降低;而对于 iOS 系统由于研究时间的限制,提取的评论数量也不够。我们希望在未来的工作中能够提取更大信息量的数据,发掘更多有价值的意见与建议。

## 8. 分工

### 8.1 研究报告的分工:

分析部分中关于评论星级和评论数量相关性的分析、相关工作部分由陈沛欣(516051910009)撰写;工作流程与技术难点部分由张津铭(516021910177)撰写;研究动机部分、分析中评分分布部分、按照星级分类和按照评论类别分类部分由苗秋实(516051910020)撰写;总结和评价部分以报告的总体格式与构架由曾乐蔚(516051910054)负责。

### 8.2 实验研究中的分工:

前期文献查找主要由陈沛欣、苗秋实、张津铭三人完成;数据查找与提取主要是由张津铭编写的 Python 程序完成的 iOS 应用商店的评论提取;由苗秋实和陈沛欣同学人工完成的小米和华为应用商店的评论提取;分析部分主要由苗秋实、曾乐蔚、陈沛欣三人完成;展示部分主要由苗秋实制作展示 PPT,曾乐蔚负责现场展示与介绍部分。

## 参考文献

- [1] Kuniavsky M. Observing the user experience : a practitioner's guide to user research[M]. Morgan Kaufmann, 2013.
- [2] Liu X , Li H , Lu X , et al. Understanding Diverse Smartphone Usage Patterns from Large-Scale Appstore-Service Profiles[J]. IEEE Transactions on Software Engineering, 2018, 44(4):384-411.
- [3] Minh T , Do T , Blom J , et al. Smartphone Usage in the Wild : a Large-Scale Analysis of Applications and Context[C]// International Conference on Multimodal Interfaces. ACM, 2011.
- [4] Martin W , Sarro F , Jia Y , et al. A Survey of App Store Analysis for Software Engineering[J]. IEEE Transactions on Software Engineering, 2017, 43(9):817-847.
- [5] Pagano D, Maalej W. User feedback in the appstore: An empirical study[J]. ieee international conference on requirements engineering, 2013: 125-134.
- [6] Harman M, Jia Y, Zhang Y, et al. App store mining and analysis: MSR for app stores[C]. mining software repositories, 2012: 108-111.
- [7] Chen M, Liu X. Predicting popularity of online distributed applications: iTunes app store case analysis[C]// Iconference. ACM, 2011.

- [8] 吴婷. 手机地图出行应用的个性化用户体验和创新研究[D]. 浙江工业大学, 2014.
- [9] 王永瑞. 百度地图的用户信息多维度分析研究[D]. 武汉邮电科学研究院, 2018.
- [10] 刘岚. 地图类旅游 APP 用户粘性影响因素研究[D]. 西南大学, 2017.
- [11] 潘美娟. 大学生在线旅游手机应用软件 (APP) 使用意向研究——以新疆师范大学为例[J]. 特区经济, 2018 (04) :101-105.
- [12] 施香君, 许明月, 万绪才. 旅游网站信任度评价及其影响因素分析——以携程网为例[J]. 南京财经大学学报, 2018 (04) :100-108.