

# APP用户差评行为影响因素研究

丁文强, 苗虹

(江苏科技大学经济管理学院, 江苏镇江 212003)

**摘要:** 差评行为会对APP应用下载量产生影响。通过对抓取的APP用户差评进行文本分析, 探究开发者推出的增值服务是否会对消费者差评行为产生影响, 分析导致消费者差评行为的因素, 为APP开发者提供建议。利用Python爬取评论, 采用逻辑回归模型进行分类, 挖掘评论中的差评信息; 利用LDA模型对预处理后的数据进行主题分析, 并对结果进行可视化解读。研究表明, 开发者上架试用版本一段时间后推送增值服务的行为会引发用户的差评行为, 增值服务的质量、功能和开发者推送付费版本的频率对差评行为起到一定的调节作用。基于研究结论, 建议开发者在推出增值服务时采取合理的方式, 并注意推送频率。

**关键词:** 移动APP; 文本挖掘; LDA模型; 差评行为

**DOI:** 10.11907/rjdk.211240

**中图分类号:** TP391

**文献标识码:** A

**开放科学(资源服务)标识码(OSID):**

**文章编号:** 1672-7800(2022)001-0233-05



## Research on the Influencing Factors of APP Users' Negative Evaluation Behavior

DING Wen-qiang, MIAO Hong

(School of Economics and Management, Jiangsu University of Science and Technology, Zhenjiang 212003, China)

**Abstract:** Bad comments will have an impact on APP downloads. Through the text analysis of the bad evaluation of the captured APP users, explores whether the value-added services launched by developers will have an impact on the consumer's bad evaluation behavior, and further finds out the factors that cause the consumer's bad evaluation behavior, providing advice to developers. Using Python crawling comments, using logical regression model to classify, mining bad information in the comments, using LDA model to pre-processed data subject analysis, and the results of visualization and interpretation. Research results show that after a period of time, the behavior of pushing value-added service will cause users to make bad evaluation behavior, and the quality, function and frequency of developers pushing paid version of value-added service will play a certain role in regulating its influence process. Developers need to take a reasonable way to pay attention to push frequency when introducing value-added behavior.

**Key Words:** APP; text mining; LDA model; negative evaluation behavior

## 0 引言

随着信息技术的不断发展, 应用程序(APP)如雨后春笋般出现在各大应用市场中, APP开发者之间的竞争异常激烈。应用市场允许开发者发布以及用户购买、下载、评论APP, 用户可以根据使用体验给APP打出最低1星、最高5星的评分。Palomba等<sup>[1]</sup>研究表明, 星级评分和用户评论对开发团队的盈利有重大影响, 低星级的APP往往很难在应用市场中生存下来, 大量优质评论能使APP获得更高曝光度, 负面评论则会使开发者面临损失。研究表明, 当评论数量达到一定规模时, 低认知需求用户受评论数量的影

响较大, 高认知需求用户受评论质量的影响较大<sup>[2]</sup>; 而对于体验型产品而言, 评论数量对用户的影响更为明显<sup>[3]</sup>。因此, 开发者通常会通过推出试用版本的方式积累原始用户, 通过扩充功能、进阶增值服务等方式获得盈利。

在互联网时代, 越来越多的用户愿意通过在线评论的方式分享自身使用体验, 这些评论包含大量有价值的信息。潜在用户除了通过开发者介绍的信息了解APP外, 还会借助用户评分进行决策, 在线评论成为其获取信息的重要来源<sup>[4-5]</sup>。然而, 不少消费者都有过给出好评等级却写出负面评论的经历, 消费者担心给出差评会受到卖家的报复, 从而违心给出正面评价<sup>[6]</sup>。评论星级与文本之间的一致性使得高星级评论中也有着差评的存在, 用户的识别

收稿日期: 2021-02-26

基金项目: 国家自然科学基金重点项目(71972090)

作者简介: 丁文强(1994-), 男, 江苏科技大学经济管理学院硕士研究生, 研究方向为数据分析; 苗虹(1979-), 女, 江苏科技大学经济管理学院副教授, 研究方向为信息系统云化迁移、信息系统建模。

难度大大增加<sup>[7]</sup>。鉴于此,本文拟探究用户是否会对开发者在试用版本后推出增值收费服务等营利行为作出负面评价,以及新版本软件的功能、质量与开发团队推送频率等对用户差评意愿的影响。

## 1 相关研究

文本挖掘是从文本中提取出高质量信息的过程,其应用多集中在商品以及企业产品评价方面。目前,有许多学者面向APP评论进行文本挖掘研究,例如胡甜媛等<sup>[8]</sup>将评论体现的反馈定义为软件满足的需求、存在的问题以及未达到的期望,通过构建评价对象与评价观点抽取规则,有效挖掘体现使用反馈的APP用户评论;吕宏玉等<sup>[9]</sup>针对APP用户特性,提出基于挖掘范围识别和关联规则的APP评论特征方法;张莉曼等<sup>[10]</sup>从APP用户需求表达状态和表达过程出发,剖析用户需求的3种表达形式,从用户生成数据中聚合需求信息,生成高价需求情报,然后再推送给需求者;Zhou等<sup>[11]</sup>以亚马逊生态系统为例,利用LDA模型提取与客户需求相关的主题,并使用KANO模型对客户需求进行定量分类;Kim等<sup>[12]</sup>利用LDA模型对Twitter用户发布的信息进行挖掘,从而更加高效地对用户进行好友推荐和内容推送;王欣研等<sup>[13]</sup>提出学术APP在线评论主题语义关联研究方法,基于词向量相关性获得主题语义关联,构建语义关联图谱,为APP平台运营者完善功能提供了参考依据。以上研究多从用户需求和软件特征的角度进行文本挖掘。本文尝试从开发者的角度分析评论文本,通过开发者将应用免费上架后再向用户推送付费版本这一行为,研究用户作出负面评价的影响因素。

## 2 研究框架与方法

### 2.1 研究框架

本文研究框架如图1所示。首先采集APP的用户评论,对数据进行预处理;然后从用户评论集中人工筛选出500条好评和500条差评,分为训练集和测试集,采用逻辑回归模型进行分类,筛选出差评文本。利用LDA模型对差评集合进行主题提取,生成文档-主题矩阵词汇-主题矩阵,并展示可视化结果。通过上述实验结果,验证开发者在试用版本后推出增值收费服务等营利行为是否会得到用户差评,以及是否存在某些因素对此起到调节作用。

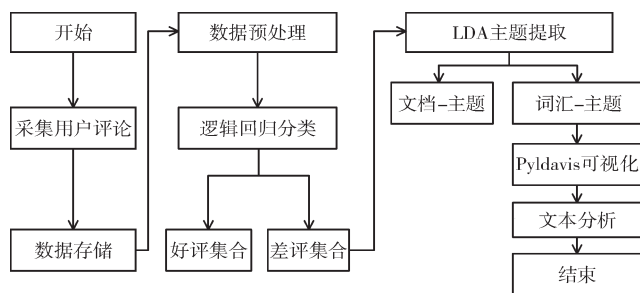


Fig. 1 Research framework

图1 研究框架

### 2.2 LDA模型

LDA主题模型是包含文档、主题、词项3种结构的3层贝叶斯概率模型<sup>[14]</sup>,其利用概率统计思想对文档进行建模,能将文档集中每篇文档的主题以概率分布的形式给出,通过判断文本的相似性进行更深层次的语义挖掘,将文本中隐藏的主题挖掘出来。

LDA模型的原理如图2所示。首先定义一个包含m个文档的文档集合 $D(d_1, d_2 \dots d_m)$ 、主题集合T,字典中有V个词汇。已知每个文档包含n个词汇,则 $N_m$ 表示第m篇文章中的词汇量, $W_{m,n}$ 表示第m个文档中的第n个词。文本集中隐藏了k个主题( $T_1, T_2 \dots T_k$ ), $Z_{m,n}$ 表示第m个文档中第n个词的主题。每个文档都有各自的主体,服从Dirichlet分布,参数为 $\alpha$ ,则对于第m个文档,其主题分布 $\theta_m$ 为Dirichlet( $\alpha$ )。主题中的词汇服从Dirichlet分布,参数为 $\beta$ ,则对于任意一个主题K,其词分布 $\phi_k$ 为Dirichlet( $\beta$ )。 $\alpha, \beta$ 为模型中的超参数,影响着模型主题分布与词汇分布的平滑度。两个隐含变量 $\theta$ 和 $\varphi$ 分别表示文档集D中d文档对应到不同主题的概率和主题集T中k主题生成不同单词的概率,其中 $\theta$ 为K维向量, $\varphi$ 为v维向量。

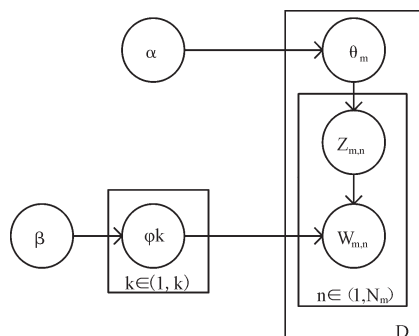


Fig. 2 LDA model

图2 LDA模型

LDA的联合概率计算方法如式(1)所示,式中w为已知变量,先验参数可参考前人经验确定;z为某篇文章的所有主题,可由 $\theta$ 确定。LDA的训练过程主要为确定 $\theta$ 和 $\varphi$ 值的过程,这些参数一旦确定,对于新加入的文档,可以认为主题词-词分布的 $\varphi$ 是稳定的,参照Gibbs Sampling公式确定 $\theta$ 和 $\varphi$ 。

$$p(\theta, z, w | \alpha, \beta) = p(\theta | \alpha) \prod_{n=1}^N p(z | \theta) p(w | z, \beta) \quad (1)$$

LDA生成文本的步骤见表1。

Table 1 LDA model test generation steps

表1 LDA模型生成文本步骤

步骤	参数分布	结果
1	Dirichlet 分布参数 $\alpha$	文档d的主题分布参数 $\theta$
2	多项式分布参数 $\theta$	文本中某个词的主题z
3	Dirichlet 分布参数 $\beta$	z主题对应的词分布参数 $\varphi$
4	多项式分布参数 $\varphi$	生成词w
5	重复步骤1、2、3、4	生成文本

### 2.3 逻辑回归模型

逻辑回归模型<sup>[15]</sup>是一种经典的分类算法,适用于对用户评论数据进行分类。其训练速度较快,且具有较好的分

类效果,是使用最为广泛的分类方法之一。该模型可根据一组自变量输入计算出归属每种类别的概率,常使用二项逻辑回归方法,即类别只有 0 和 1 两种,其条件概率分布表示为:

$$p(Y = 1|x, w) = \frac{e^{w^T x + b}}{1 + e^{w^T x + b}} = \frac{1}{1 + e^{-(w^T x + b)}} \quad (2)$$

$$p(Y = 0|x, w) = \frac{1}{1 + e^{w^T x + b}} \quad (3)$$

式中,  $x \in R^n$  为输入,称为实例特征;  $Y \in (0, 1)$  为输出,两个分类可以对评论是好评还是差评进行判断;  $w \in R^n$  和  $b \in R$  为参数;  $w$  为权值向量,对应每个输入特征的权重;  $b$  为偏置。

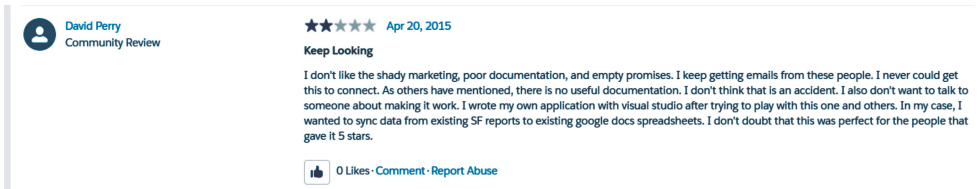


Fig. 3 Partial test data display

图 3 部分试验数据展示

### 3.2 数据预处理

由于用户在平台中的评论比较随意,不可避免地会出现很多与研究主题无关的无效评论,因此对数据进行预处理有助于提升模型效率,处理方法主要分为以下 3 种:①分词。进行数据处理时,计算机理解的最小处理单位为单词,因此需要将语句拆分成有意义的单词;②去除停用词。分词结果中通常会出现很多无意义的符号,使用停用词表删除文本中的连词、量词、介词等无意义词语,剔除无用标点,可以提高检索效率、优化内存,文本处理时若遇到停用表中的词,系统会停止处理,将其扔掉;③次干提取,词形还原。去除相似单词的词缀,得到单词词根,否则主题模型会将这些单词当作不同的实体进行处理,词形还原则是利用上下文语境将单词还原成词根。

### 3.3 逻辑回归分类

对采集到的评论信息进行人工标注,得到好评和差评各 500 条。做好标签,正面评论用 1 表示,负面评论用 0 表示,采用空间向量表示预处理后的词组,组成语料库。将语料库以 7:3 的比例划分为训练集和测试集,利用逻辑回归算法对训练矩阵进行分类器训练,将训练好的分类器运用到测试矩阵中进行预测,然后对模型效果进行评估,最后调用训练好的逻辑回归模型对采集到的评论进行分类。

经过训练,得到如图 4 所示的试验结果,可以看出各项模型评价指标均大于 0.9,说明预测效果较为理想。使用训练好的分类器对用户评论进行分类,可形成用户差评集合。

	precision	recall	f1-score	support
0	0.95	0.95	0.95	144
1	0.96	0.96	0.96	156
accuracy			0.95	300
macro avg	0.95	0.95	0.95	300
weighted avg	0.95	0.95	0.95	300

Fig. 4 Logical regression model evaluation

图 4 逻辑回归模型评估

## 3 数据采集与预处理

### 3.1 数据来源

APPexchange (<https://APPexchange.salesforce.com/>) 是 salesforce 推出的应用平台,现今已上架 3 000 多种 APP,拥有上百万在线用户,旨在为用户提供大量优质 APP。利用 Python+Selenium 的方法采集 APPexchange 平台中免费 APP 的用户评论数据,共有 14 290 条,删除非英文评论、无意义评论(“Waste Time”“Highly Recommended”)以及显示异常评论等,得到 14 068 条有效评论。爬取的部分数据内容如图 3 所示。

### 3.4 主题个数确定

LDA 算法需要确定超参数  $\alpha$  和  $\beta$ ,以及主题数量  $k$ 。实践中常将超参数  $\alpha$  和  $\beta$  值设置为 0.1 和 0.01<sup>[16]</sup>,本文亦是如此。

确定主题数量是构建 LDA 模型的难点,主题数量影响着整个模型性能的优劣。当主题数目过多时,会产生很多无明显意义的主题;当主题数目过少时,主题粒度过大,一个主题包含多层语义。目前主题数量常基于困惑度和主题连贯性确定<sup>[17]</sup>,由于困惑度依赖概率和词频,没有考虑到词汇在用户评论中的上下文联系,因此本文选择后者确定用户评论集的最佳主题数量<sup>[18]</sup>。通过 Python 自带的模型方法确定主题数量,根据 LDA 提取的主题相似度计算相应得分,得分越高,模型质量越好。将主题数量分别设置为 10、15、20、25、30,通过网格搜索法(Grid Search)进行调参。由图 5(彩图扫 OSID 码可见)可以看出,当主题数量为 10 时效果最好,当主题数量继续缩小时,评分仍会继续上升,但主题数目过少时会造成一个评论包含多个主题的情况,不利于分析提取的主题。经过多次试验,最终设定主题数量为 10。

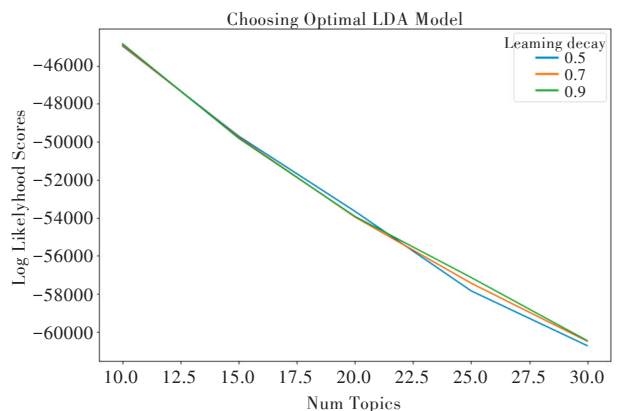


Fig. 5 Number of themes

图 5 主题数量

### 4 主题分析

通过 LDA 模型对用户差评集合进行分析,得到文档—主题矩阵和主题—词汇矩阵。表 2 为运行 LDA 模型后得到的主题—词汇矩阵,展示了评论集中出现的与 10 个主题

相关的特征词。图 6 为主题 4 包含的前 30 个主题词,特征词“APP、free、pay”共现在部分软件版本升级的差评集中,说明确实存在开发者先将 APP 免费上架,在限免一段时间后推送增值服务的情况,最终有用户针对该现象给出了负面评价。由此证实,将付费 APP 伪装成免费型上架,后期再推送付费版本会对用户差评行为产生影响。

Table 2 Theme-vocabulary matrix

表 2 主题—词汇矩阵

Topic1	support	issue	salesforce	team	time	customer	product	service	integration	company
Topic2	object	custom	record	datum	time	field	tool	easy	file	standard
Topic3	salesforce	APP	package	solution	support	install	team	instal	project	production
Topic4	email	free	version	send	salesforce	APP	support	pay	address	tool
Topic5	user	APP	survey	set	question	issue	test	easy	field	support
Topic6	salesforce	task	user	event	update	activity	track	create	time	object
Topic7	product	account	easy	user	business	recommend	sale	integration	salesforce	team
Topic8	create	lead	lightning	salesforce	file	site	link	account	box	record
Topic9	field	lead	account	APP	time	datum	tool	issue	chatter	salesforce
Topic10	error	message	instal	package	fail	install	component	class	apex	limit

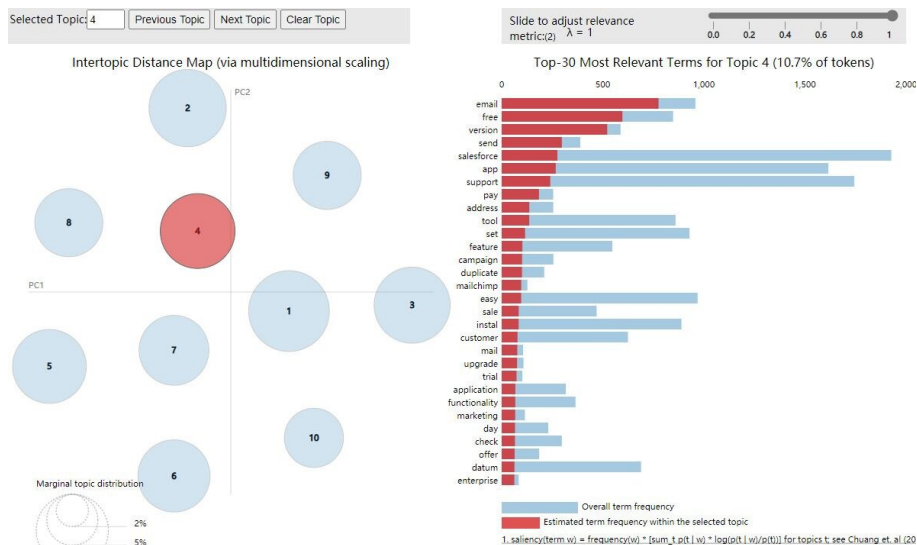


Fig. 6 PyLDAvis visualization

图 6 PyLDAvis 可视化图

通过 pyLDAvis 对主题进行可视化,将图 6 中的关键词导入到主题词关联图中(见图 7),发现了关键词“email、address、day”,说明开发者可能会通过电子邮件的方式向用户推送营销信息,且较为频繁。关键词“pay、version、feature、upgrade”联系紧密,说明更新版本可能会增加新的功能适配用户需求,如果用户对新版本感到满意可能会接受付费行为,但当升级版本没有达到预期效果时会给出差评。

团队的支持与服务、APP 的适配程度和程序包解决方案等问题。后几个主题解读效果不是很好,主要为 APP 的账户、插件、设置等出现的问题以及对开发团队的反馈。总体来说,开发者将 APP 上架于免费专区,后期进行增值服务的行为容易造成用户不满,开发者需提升更新后版本的质量,达到物超所值的效果。

从提取的主题中可以看出,导致用户差评行为的因素有很多,需要对这些因素的重要程度进行比较。Pyldavis 左侧面板展示了 LDA 模型分类出的不同主题,主题气泡大小可显示主题强度,气泡越大说明该主题的重要程度越高,气泡之间的距离表示两个主题之间的相似程度。根据以上信息绘制图 8,可以看出,主题 4 占比为 10.7%,说明付费 APP 伪装成免费 APP 上架,后期进行收费的主题较为靠前,且与主题 1 的 12.5% 较为接近。前 3 个主题反映的是开发

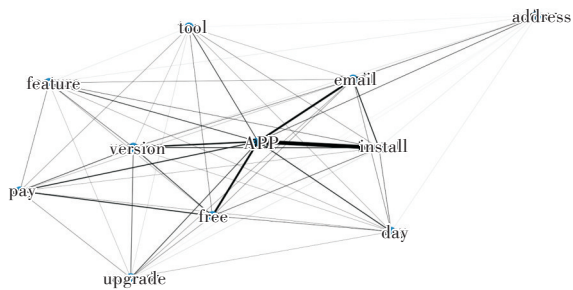


Fig. 7 Keywords association chart

图 7 关键词关联图谱

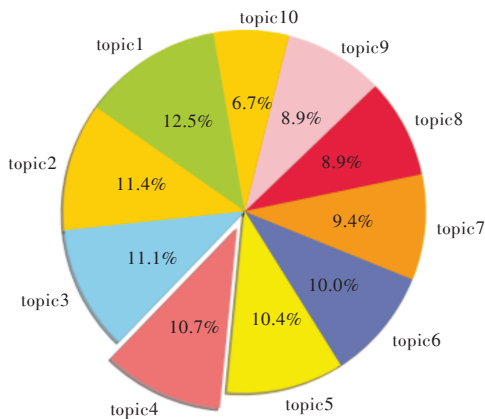


Fig. 8 Proportion of themes

图 8 各主题占比

## 5 结语

本文基于逻辑回归算法对 APP 的评论集进行分类, 筛选出差评数据, 然后利用 LDA 模型分析评论数据, 探究是否存在开发者在试用版本后推出进阶收费服务等营利行为, 以及这一现象与用户差评行为的关系。实验结果表明, 开发者将应用免费上架后再向用户推送付费版本的行为是用户作出差评的原因之一, 尤其是当开发团队频繁地对用户推送付费版本信息时, 更有可能导致用户给予差评; 付费版本的质量和功能可能会对用户差评行为起到调节作用。

当使用免费版本的用户达到一定规模时, 开发团队为获取更高利润, 往往会推出付费或高级版本进行营利行为, 如果未做好营销推广工作可能会导致用户给予差评, 影响 APP 下载量。开发者应合理制定价格策略, 注重与用户的沟通。如果付费版本的功能与免费版相比得到明显提升, 且价格合理, 相信用户乐意进行付费版本的尝试。

本文研究方法仍存在一定局限性, 根据实际需求调整 pyldavis 中参数  $\lambda$  的值可以得到更有意义的主题词。此外, LDA 模型超参数的设定会影响主题模型的结果, 因此在今后研究中将考虑更多参数对试验结果的影响, 而不是参照其他文献的推荐值, 以便模型更加拟合试验数据。

### 参考文献:

- [1] PALOMBA F, LINARES-VASQUEZ M, BAVOTA G, et al. User reviews matter! tracking crowdsourced reviews to support evolution of successful APPs[C]//Bremen: 2015 IEEE International Conference on Software Maintenance and Evolution (ICSME), 2015.
- [2] CHEN X H, ZENG P. An experimental study on the effects of mobile shopping reviews on consumer purchasing intention[J]. Research on Economic and Management, 2016, 37(6): 122-129.  
陈晓红, 曾平. 移动购物评价对消费者购买意愿影响的实验研究[J]. 经济与管理研究, 2016, 37(6): 122-129.
- [3] LI Z W, ZHANG Y H, LUAN D Q. What factors affect consumers' online purchase decisions? the driven role of customer perceived value[J]. Management Review, 2017, 29(8): 136-146.  
李宗伟, 张艳辉, 梁东庆. 哪些因素影响消费者的在线购买决策?

- 顾客感知价值的驱动作用[J]. 管理评论, 2017, 29(8): 136-146.
- [4] HU M, LIU B. Mining opinion features in customer reviews[J]. International Journal of Computer Engineering in Research Trends, 2016, 3(9): 535-539.
  - [5] CHEN R, WANG Q, XU W. Mining user requirements to facilitate mobile APP quality upgrades with big data[J]. Electronic Commerce Research and Applications, 2019, 38: 100889.
  - [6] XU N. Analysis of the problem of infringement's rights and interests resulting from online evaluation[D]. Kunming: Yunnan University, 2017.  
徐娜. 因网评引起的侵犯消费者权益问题分析[D]. 昆明: 云南大学, 2017.
  - [7] RAN M, JIANG Y, XIANG Q X, et al. Consistency judgment method between user's comment and user's mark for APP software[J]. Computer and Digital Engineering, 2017, 45(4): 692-695.  
冉猛, 姜瑛, 向祺鑫, 等. APP 软件用户评论与评分星级一致性判断方法[J]. 计算机与数字工程, 2017, 45(4): 692-695.
  - [8] HU T Y, JIANG Y. Mining of user's comments reflecting usage feedback for APP software[J]. Journal of Software, 2019, 30(10): 3168-3185.  
胡甜媛, 姜瑛. 体现使用反馈的 APP 软件用户评论挖掘[J]. 软件学报, 2019, 30(10): 3168-3185.
  - [9] LV H Y, FAN K, YANG J L. Research on software feature mining for APP user reviews[J]. Library Theory and Practice, 2019(7): 106-112.  
吕宏玉, 樊坤, 杨建林. 面向 APP 用户评论的软件特征挖掘研究[J]. 图书馆理论与实践, 2019(7): 106-112.
  - [10] ZHANG L M, ZHANG X X, LU H, et al. Research on user demand intelligence push model of academic APP based on demand aggregation[J]. Journal of Intelligence, 2020, 39(7): 1-8.  
张莉曼, 张向先, 卢恒, 等. 基于需求聚合的学术 APP 用户需求情报推送模型研究[J]. 情报杂志, 2020, 39(7): 1-8.
  - [11] ZHOU F, AYOUB J, XU Q, et al. A machine learning approach to customer needs analysis for product ecosystems[J]. Journal of Mechanical Design, 2020, 142(1): 1-56.
  - [12] KIM Y, SHIM K. TWILITE: A recommendation system for Twitter using a probabilistic model based on latent Dirichlet allocation[J]. Information Systems, 2014, 42(6): 59-77.
  - [13] WANG X Y, ZHANG X X, ZHANG L M. Semantic association of online comment topics of academic APP users[J]. Information Science, 2020, 38(6): 25-31.  
王欣研, 张向先, 张莉曼. 学术 APP 用户在线评论主题语义关联研究[J]. 情报科学, 2020, 38(6): 25-31.
  - [14] BLEI D M, NG A Y, JORDAN M I. Latent Dirichlet allocation[J]. Journal of Machine Learning Research, 2003, 3: 993-1022.
  - [15] HOSMER D W, LEMESHOW S. Applied logistic regression[M]. New York: Wiley Inc, 2000.
  - [16] HAO H, ZHANG K, WANG W, et al. A tale of two countries: International comparison of online doctor reviews between China and the United States[J]. International Journal of Medical Informatics, 2017, 99: 37-44.
  - [17] MIMNO D, WALLACH H, TALLEY E, et al. Optimizing semantic coherence in topic models[C]//Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, 2011: 262-272.
  - [18] XIE T, QIN P, ZHU L. Study on the topic mining and dynamic visualization in view of LDA model[J]. Modern Applied Science, 2018, 13(1): 204.