

# 基于文本挖掘的丁香医生 APP 用户评论分析\*

白雪艳 陈超亿 冯占春

(华中科技大学同济医学院医药卫生管理学院 武汉 430000)

〔摘要〕 介绍医疗 APP 发展概况及消费者满意度研究现状, 基于丁香医生 APP 在线评论文本分析用户满意度影响因素并提出建议, 包括强化核心形象塑造、提高品牌辨识度、加强回复时效性等。

〔关键词〕 丁香医生 APP; ROST 数据挖掘系统; 形象感知; 内容分析

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2022.08.010

**Analysis of User Reviews of Dingxiang Doctor APP Based on Text Mining** BAI Xueyan, CHEN Chaoyi, FENG Zhanchun, School of Medical and Health Management, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430000, China

〔Abstract〕 The paper introduces the development situation of medical APP and the research status of consumer satisfaction, analyzes the influencing factors of user satisfaction based on the online review text of Dingxiang Doctor APP, and puts forward suggestions, including improving core image building and brand recognition, strengthening response timeliness, etc.

〔Keywords〕 Dingxiang Doctor APP; ROST data mining system; image perception; content analysis

## 1 引言

伴随信息通信技术的进步, 线上问诊成为患者获得医疗服务的重要方式, 基于移动健康的方法, 即通过移动电话和个人数字助理等移动技术提供健康服务和信息的电子健康模式在卫生保健领域快速发展。根据医疗干预的阶段不同, 医疗应用程序可划分为以下 6 种类型: 预约及缴费, 患者可通过平

台预约医生、缴付诊疗费用; 医疗咨询及评估, 注册患者可通过平台向医生咨询, 注册医生可回答患者问题; 医学教育, 通过网站或应用程序为患者提供与疾病相关的科普和教育文章, 如日常生活思考、自我监控、自我管理和对某些疾病的应急反应; 药品和医疗器械销售, 患者在此类网站上核实有效处方后可以购买药物; 电子病历, 面向医生提供制作医疗电子记录的相关服务; 慢性病管理, 通常分为医生和患者版本。目前有研究表明使用移动医疗应用开展行为干预在疾病管理中具有潜在有效性<sup>[1]</sup>。国家对于“互联网+医疗健康”的支持推动了在线医疗服务的发展。丁香医生 APP 由医学网站丁香园团队研发, 是一款面向大众的药品信息查询及日常安全用药辅助工具, 具有较强代表性, 对于丁香医生 APP 用户评论的分析具有现实意义<sup>[2]</sup>。

〔修回日期〕 2021-10-04

〔作者简介〕 白雪艳, 硕士研究生; 通信作者: 冯占春, 院长, 教授。

〔基金项目〕 国家卫生健康委员会委托课题“县域医共体在新冠肺炎防治中发挥作用与影响研究”。

## 2 相关研究

### 2.1 基于在线评论中数值型信息的消费者满意度研究

主要针对评论数量、言论质量、评论效价对顾客价值创造的影响开展相关研究。研究表明，数值型在线评论的效价参考价值逐渐降低，主要表现在以下两个方面：第一，消费者对于在线评论关注点发生转移。在商品同质化严重的背景下，用户对信息细致程度有更高要求。作为简化版信息的数值型在线评论，内容详细程度和信息量不足，导致其影响力消减。第二，数值型在线评论的效价之间没有显著差异，使其参考价值降低。有学者通过对 Airbnb 和 TripAdvisor 在线评论的评分比较，发现大多数评分都集中在 4.5~5 分之间（总分为 5）。综上所述，数值型信息研究参考价值降低，有较大局限性。

### 2.2 基于用户评论文本的消费者满意度研究

此类研究提取评论中的文本并通过情感分析方法判断情感倾向<sup>[3]</sup>。国内学者对用户服务满意度研究起步较晚，提取用户评论文本进行满意度影响因素研究主要集中在酒店、旅游、电商领域。移动医疗作为新型产业，该领域相关研究数量较少。有学者从医疗论坛文本入手，分析脑血管病论坛用户关注的热点主题以及论坛成员的真实需求，以提供更精准服务；有学者提出在信任中介作用下，评论文本的数量及其时效性、信任倾向和评论文本接收者的参与程度都会影响消费者的认知信任和情感信任，从而影响购买意愿；有学者探讨影响移动医疗 APP 用户接受行为的因素并通过问卷调查收集数据进行分析，提出感知有用性、感知易用性、信息质量等皆为移动医疗 APP 接受行为的影响因素；有学者利用问卷调查方法分析用户对在线医疗服务满意度，研究表明挂号和查询等医疗服务满意程度较高；有学者探讨问诊类移动 APP 用户持续使用影响因素，提出医生沟通、关怀行为会显著提升用户对医生的信任进而提升对 APP 的信任度和满意度；另有学者以在线评论为切入点，提出用户服务满意度比较重要的影响因素有“在线诊疗”“健康管理”“其他服务”及“平台活动”，用户最关注的因素是

“在线诊疗”；有学者通过文本挖掘研究得出，正面评论占大多数得益于医疗健康类 APP 的使用便捷性，用户对大部分因素表示满意，但对服务质量、可优化性、价格的满意度较低。

### 2.3 针对医疗 APP 产品满意度相关研究

目前针对医疗 APP 产品满意度的研究较少，且对用户满意度的研究多以调查问卷为主要数据收集方式，从文本挖掘角度进行分析的研究相对较少；问卷设计局限在于有效率较低、开放性问题偏少，挖掘更深层、隐蔽的信息难度较大。因此本文在总结现有文献的基础上，使用文本挖掘法对丁香医生 APP 的在线评论进行研究，将文本主题分析、情感分析相结合，深入挖掘在线评论文本中所蕴含的满意度影响因素，并针对医疗 APP 发展提出相关建议，以期为促进移动医疗市场发展提供参考<sup>[4]</sup>。

## 3 数据采集

### 3.1 对象选取

本研究选取国内具有代表性的平台之一丁香医生 APP。使用八爪鱼采集器抓取丁香医生 APP 在 Apple Store 的评论文本，采集的用户评论文本发布时间为 2018 年 8 月 - 2021 年 8 月。最终爬取了超过 2 000 条的用户评论信息，内容包括服务反馈、改进意见等，利用 Excel 的自定义排序、筛选等功能去除无用评论，如空格、符号、表情、广告等，初步整理出有效用户评论文本<sup>[5]</sup>。

### 3.2 研究思路（图 1）

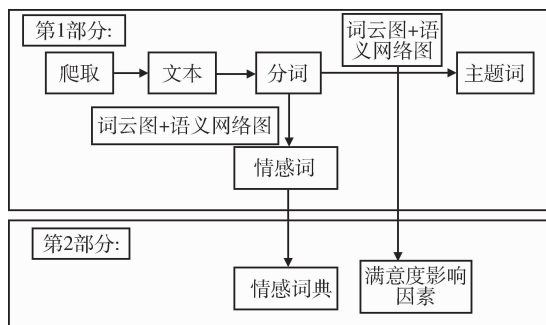


图 1 研究思路

基于八爪鱼采集器爬取丁香医生 APP 在 Apple Store 店固定时间段在线评论文本，对原始文本进行格式统一化处理，去除广告、表情包、无意义符号等；通过 ROST - CM6.0 软件对处理后评论文本进行分词；为直观地展现客户评论关注焦点，本研究通过生成高频词表、词云图和网络分析图实现可视化，以便进一步分析。用户情感倾向隐含在用户评论信息中，在分析影响用户满意度因素之前，用基于情感词典的 ROST - CM6.0 软件挖掘和分析，对

用户评论信息进行情感分析。最后总结研究结论、提出改进建议。

## 4 结果

### 4.1 概述

运用八爪鱼采集器爬取评论并拆分为发表时间、作者、评级、标题、内容 5 个维度，见表 1。

表 1 采集评论样本

发表时间	作者	评级	标题	内容
2020 年 11 月 25 日 14: 39: 53	Ameng233	5	请修复推送的文章点开无法跳转这个问题	每次点开通知中心的文章，进入主界面，没有任何反应，无法跳转到点开的文章
2020 年 11 月 12 日 09: 47: 28	ak7107	5	很实用	挺有用的，很方便地可以查到生活中医生的建议，希望以后能改变看病难的现状
2020 年 11 月 10 日 15: 28: 10	zjz8381	5	一点好评	实用、方便、专业、可信
2020 年 11 月 4 日 14: 13: 06	faye3332	5	推荐丁香医生线上问诊平台	使用感受非常好！已推荐给身边的人
2020 年 10 月 31 日 07: 22: 46	悠悠子 123	5	好评如下	丁香医生是最实用的 APP 了，有什么问题在这里都可以知道
2020 年 10 月 28 日 13: 34: 49	狗仔包	5	非常好的一个软件	这个软件不需要添加问诊人信息便可以获取网络资源，非常方便

### 4.2 词频分析



图 2 词云

4.2.1 用户满意度影响因素 可视化词云图可直观反应评论主题，可知影响用户满意度最重要的因素是“医生”，见图 2，这说明用户更关注服务主体提供者医生的行为。用户对 APP 总体评价“方

便”。其他影响用户满意度的因素包括“医院”“开发者”“科普”“平台”“咨询”“问诊”等。词云图不说明各主题之间联系，因此采用网络语义分析对各主题关系进行分析<sup>[6]</sup>。

4.2.2 APP 认知高频词主题分类 分析排名前 60 位的高频词可知：用户对于 APP 整体印象评价趋于正面，如“方便”“实用”“感谢”“耐心”等高频词汇；高频词中对 APP 整体形象认知主要分为 4 个类型，即服务主体类、设施类、患者反馈类、用户感知类<sup>[7]</sup>，见表 2。

表 2 丁香园 APP 认知高频词

序号	高频词	序号	高频词	序号	高频词	序号	高频词
1	医生	16	回答	31	好评	46	点击
2	删除	17	使用	32	处理	47	良心
3	丁香	18	知识	33	医疗	48	疫情
4	问题	19	建议	34	科学	49	抱歉

续表 2

5	医院	20	联系	35	功能	50	身体
6	方便	21	时间	36	医学	51	信号
7	开发者	22	健康	37	值得	52	上面
8	问诊	23	实用	38	解答	53	百度
9	平台	24	患者	39	小时	54	申请
10	感谢	25	详细	40	帮助	55	三甲
11	软件	26	提问	41	看病	56	每次
12	科普	27	体验	42	垃圾	57	数据
13	咨询	28	及时	43	态度	58	手机
14	反馈	29	服务	44	信赖	59	平时
15	您好	30	耐心	45	设置	60	程序

4.2.3 服务主体类高频词分析 此类高频词汇包括“医院”“医生”“开发者”等，反映 APP 服务主要提供者包括专业医者、技术支持人员等，其为 APP 运营提供专业知识、技术支持。

4.2.4 设施类高频词分析 此类高频词汇包括“平台”“回答”“使用”“程序”“手机”“反馈”等，反映 APP 使用方式主要是依托移动电子平台进行查询、咨询、提问、反馈等<sup>[8]</sup>。

4.2.5 患者反馈类高频词分析 此类高频词汇包括“健康”“体验”“科学”“看病”“健康”等，表明 APP 不仅日常提供患者健康咨询和健康相关服务，还成为了健康知识辟谣的窗口。

4.2.6 用户感知类高频词分析 此类高频词汇包括“感谢”“详细”“服务”“耐心”“信赖”“垃圾”等。对于 APP 提供服务的认知多数为积极情绪，占比 67.52%，中性情感占比 18.87%，负面情感占比 13.61%，感知高频词分布符合这一结论。

#### 4.3 语义网络分析

语义网络分析直观展现要素联系，词语距中心节点（核心节点）距离越近关联性越强。线条指向越密集则共现次数越多、关联越强<sup>[9]</sup>。使用 ROST-CM 6.0 内容挖掘系统的“社会网络和语义网络分析”功能，导入文本生成共现语义网络。与“医生”关联形容词有“耐心”“方便”“及时”等，

表明丁香医生 APP 中医生的评价感知整体正面；与“医生”联系密切的词汇有“建议”“服务”“态度”“知识”“问诊”“科普”等，表明医生起到了科普和服务等作用。“问题”一词与“联系”“信号”“时间”“解答”等形成语义分析网络，说明影响用户满意度的主要是问答质量、信息设备和时效。“医院”与“挂号”“咨询”“解答”“排队”密切相关，体现了丁香园 APP 提供传统医院服务，能够实际解决问题。

#### 4.4 情感分析

通过利用 ROST - CM6.0 软件，基于软件自定义情感词典对丁香园 APP 的使用感知情况进行分析。整体情绪分段统计中，积极情绪占比 67.52%，中性情绪占比 18.87%，消极情绪占比 13.61%。

### 5 建议

#### 5.1 加强核心形象塑造，吸引潜在用户

研究发现大部分用户对于 APP 的情感倾向是正面的，认为其是值得感谢、服务耐心、值得信赖的。根据网络语义图可知用户所看重的核心服务提供者是医生和开发者。同时词频分析和语义网络分析显示“医生”和“平台”两个高频词与反馈、问题等因素联系密切，说明 APP 服务要重视技术平台和专业人才两方面的支持；分析评论可知一部分用户是从知乎等平台引流到丁香医生 APP，因此应充分借助电视广告、网站视频、社交网络等宣传推广手段，消除对于医疗 APP 专业资格的顾虑，有效吸引潜在顾客。

#### 5.2 提高品牌辨识度，增强市场竞争力

医疗 APP 市场竞争激烈，应树立品牌特色提升竞争力。本文发现部分用户通过权健事件提升了对丁香医生 APP 的好感度，部分用户为分辨关于健康的“事实”和“谣言”下载了 APP，提升了丁香医生 APP 在医疗 APP 市场的知名度和辨识度。因此医疗 APP 要借助日常科普和重大社会事

件加强品牌形象建设, 让用户在使用中感受到品牌特色。

### 5.3 加强回复时效性, 提升用户黏性

分析发现用户点评中咨询、问诊、科普、回复等词出现频率较高, 部分用户评论反映回复时效性差、回复没有针对性等问题。医疗 APP 主营线上服务, 提供日常健康维护、答疑解惑、科普等服务, 主要依靠问答、咨询等形式完成, 所以要针对回复时效性和人员专业性完善服务。另外评论中“小孩”“老人”提到次数较多, 因此应把握老幼医疗市场的巨大潜力, 增加针对性服务。同时尽量减少重复空泛的回答, 为用户提供个性化的专业服务, 改善用户服务体验、提升用户黏性。

### 5.4 注重产品更新迭代, 完善系统漏洞

促进 APP 升级换代, 及时修复系统漏洞。根据评论分析可知部分用户对丁香医生 APP 软件闪退和卡顿问题表示不满, 而且不合理的操作步骤给用户带来较差体验。应注重用户体验反馈, 及时优化和完善产品功能, 如优化登记页面、简化操作步骤等。

## 6 结语

近年来在线医疗发展迅速, 满足了公众对便捷和非接触式医疗服务的需求。丁香医生 APP 是具有代表性的移动医疗服务 APP, 在全国范围内提供便捷医疗服务。用户评论中核心层词汇反映出 APP 服务主体, 即医疗 APP 是医生、医院、开发者等共同合力支撑的“不出门的医院”。通过情感分析和满意度分析进一步发现, 部分用户对回复时效性和 APP 技术支持表现出负面情绪, 对 APP 总体表现出正面的情感意愿和较高的满意度, 如认为丁香医生 APP 是使用方便、值得感激、可以起到科普作用的 APP。大多数用户评论文本收

到的积极反馈表明, 在特殊情况下在线保健服务是用户首选。根据词频分析和语义网络分析, 在点评词汇中存在部分消极评论, 究其原因, 回复时效性是主要影响因素, 此外部分用户未能完全接受线上问诊方式, 对于线上问诊认可度较低, 对此类产品有抵触心理<sup>[10]</sup>。本文从用户体验、情感分析等角度分析影响丁香医生 APP 满意度的要素并提出改善建议, 相较问卷调查研究方式更容易获取到深层信息。但本文未在评论文本信息中针对治疗效果进行科学的调查和评估, 未来将针对此问题进行进一步研究。

### 参考文献

- 1 Doraiswamy S, Abraham A, Mamtani R, et al. Use of Telehealth During the COVID - 19 Pandemic: Scoping Review [J]. *Med Internet Res*, 2020, 22 (12): e24087.
- 2 肖雅婧, 邵圣文. 移动医疗信息平台在延续护理中的应用现状 [J]. *护理研究*, 2021, 35 (7): 1229 - 1231.
- 3 马英才, 李春月. 消费者对电子商务平台销售生鲜产品满意度测算研究——基于京东生鲜在线评论的分析 [J]. *价格理论与实践*, 2020 (5): 117 - 120.
- 4 由雨萍, 何玲玲. 基于框架语义的在线医疗评论情感分析 [J]. *现代情报*, 2020, 40 (3): 111 - 116, 125.
- 5 黄仁东, 韩静, 崔乃雪. 公众对“网约护士”的态度和关注点: 基于微博评论的文本挖掘研究 [J]. *护理研究*, 2021, 35 (17): 3021 - 3025.
- 6 王梦洽, 方卫华. 互联网医疗服务平台公众使用意愿影响因素研究 [J]. *北京航空航天大学学报 (社会科学版)*, 2020, 33 (3): 150 - 156.
- 7 邓君, 胡明乐. 基于用户感知的在线医疗社区信息服务质量影响因素研究 [J]. *情报科学*, 2019, 37 (10): 40 - 45.
- 8 王若佳, 张璐, 王继民. 基于扎根理论的在线问诊用户满意度影响因素研究 [J]. *情报理论与实践*, 2019, 42 (10): 117 - 123.
- 9 张公让, 鲍超, 王晓玉, 等. 基于评论数据的文本语义挖掘与情感分析 [J]. *情报科学*, 2021, 39 (5): 53 - 61.
- 10 张绍婵, 冯国忠. 移动医疗 APP 监管问题探究 [J]. *现代商贸工业*, 2019, 40 (11): 90 - 91.