

柯洁, 杨婉, 黄桂玲, 等. 移动医疗 App 在线评论维度挖掘与情感分析[J]. 中华医学图书情报杂志, 2022, 31(6): 20-29.

DOI: 10.3969/j.issn.1671-3982.2022.06.003

· 研究与探讨 ·

## 移动医疗 App 在线评论维度挖掘与情感分析

柯洁<sup>1</sup>, 杨婉<sup>1</sup>, 黄桂玲<sup>2</sup>, 王璇<sup>3</sup>, 刘倩<sup>2</sup>

**[摘要]**目的: 对我国移动医疗 App 的在线评论进行维度挖掘与情感分析, 以便精准评价用户满意度。方法: 基于 App Store 中国应用市场中医疗类 App 的在线评论数据, 采用集成多策略的深度学习的方法, 首先应用 TF-IDF 算法、BERT 模型和 Canopy+K-means 聚类分析方法提取移动医疗 App 在线评论的主要维度, 然后通过计算各维度权重值, 明确不同维度对用户整体评价意见的重要性, 最后利用 LSTM-CNN 模型对各维度的用户评论进行细粒度情感分析。结果: 用户关注的移动医疗 App 在线评论主要维度依次为专业性、可靠性、交互性、易用性和特色性; 用户对移动医疗 App 的整体满意度不高, 在可靠性、交互性、特色性维度上的评论积极情感倾向率较低; 同时, 移动医疗各细分领域的 App 在不同维度上也存在明显不同的优势和劣势, 需要根据自身特色和发展目标进行优化与完善。结论: 集成多策略的深度学习的方法在移动医疗 App 在线评论维度挖掘和情感分类上具有很好的适用性、稳定性与可推广性, 可为 App 在线评论文本分析和用户满意度评价提供重要的方法支撑。

**[关键词]**移动医疗 App; 深度学习; 在线评论挖掘; 情感分析

**[中图分类号]**G252; R-05 **[文献标志码]**A **[文章编号]**1671-3982(2022)06-0020-10

### Dimension mining and sentiment analysis of online reviews of mobile medical App

KE Jie<sup>1</sup>, YANG Wan<sup>1</sup>, HUANG Gui-ling<sup>2</sup>, WANG Xuan<sup>3</sup>, LIU Qian<sup>2</sup>

(1. Department of Neurological Rehabilitation, Zhongnan Hospital of Wuhan University, Wuhan 430071, Hubei Province, China; 2. Department of Nursing, Zhongnan Hospital of Wuhan University, Wuhan 430071, Hubei Province, China; 3. Department of Spine and Bone Oncology, Zhongnan Hospital of Wuhan University, Wuhan 430071, Hubei Province, China)

Corresponding author: WANG Xuan

**[Abstract]** **Objective** Dimension mining and sentiment analysis of online reviews of mobile medical Apps in China were conducted, so as to accurately evaluate user satisfaction. **Methods** Based on the online review data of medical Apps in the App Store China App Market, the deep learning method of integrated multi-strategy was adopted, and the TF-IDF algorithm, BERT model and Canopy+K-means clustering analysis method were first used to extract the main dimensions of online reviews of mobile medical Apps, and then the importance of different dimensions to the

**[基金项目]**武汉大学中南医院护理学学科培育项目“基于信息化平台的骨科患者出院准备服务系统的构建及实践”(2NXKPY2022009)

**[作者单位]**1. 武汉大学中南医院神经康复科, 湖北 武汉 430071; 2. 武汉大学中南医院护理部, 湖北 武汉 430071; 3. 武汉大学中南医院脊柱与骨肿瘤科, 湖北 武汉 430071

**[作者简介]**柯洁 (1986-), 女, 本科, 助理研究员, 研究方向为移动医疗服务。

**[通讯作者]**王璇 (1977-), 女, 本科, 主管护师, 护士长, 研究方向为护理管理与互联网医院服务。E-mail: whc1203@163.com

overall evaluation opinions of users is clarified by calculating the weight values of each dimension. Finally, the LSTM-CNN model was used to perform fine-grained sentiment analysis on user comments in each dimension. **Results** The main dimensions of mobile medical Apps that users pay attention to are: professionalism, reliability, interactivity, ease of use and characteristics. Users' overall satisfaction with mobile medical Apps is not high, and there are few

positive comments in terms of reliability, interactivity, and features. There are also obvious different advantages and disadvantages in different dimensions of Apps in various sub-fields of mobile medical care, which need to be optimized and improved according to their own characteristics and development goals. **Conclusion** The integrated multi-strategy deep learning method has good applicability, stability and generalizability in the dimension mining and sentiment classification of mobile medical App online reviews, and can provide important method support for the text analysis of App online reviews and user satisfaction evaluation.

**[Key words]** Mobile medical App; Deep learning; Online comment mining; Sentiment analysis

随着互联网医疗行业的快速发展,各类移动医疗 App 不断涌现。根据移动应用市场知名咨询公司 data.ai 的统计数据显示,截至 2022 年 2 月,中国移动医疗 App 数量已近 2 000 个,涉及在线问诊、网上就医、医药服务、健康管理等多个细分领域<sup>[1]</sup>。受新型冠状病毒肺炎(以下简称“新冠肺炎”)疫情影响,人们对医疗 App 的使用需求持续增长,进一步推动了传统医疗服务向移动端延伸。移动医疗 App 建立了医院信息系统、移动网络平台和用户就医需求之间的多元连接,使用户可以随时随地访问医疗资源并获取相应的医疗健康服务,是未来互联网医疗行业发展的主要模式之一<sup>[2]</sup>。但随着移动医疗 App 数量的爆发式增长,也引发了同质化竞争加剧、服务质量参差不齐等一系列问题。如何及时准确地了解移动医疗 App 的用户使用体验和评价意见,进而推动移动医疗的创新与发展,已成为国内外学者和互联网医疗行业共同关注的焦点。

App 在线评论是洞察用户体验和满意度的重要数据来源,表达了用户使用 App 后的真实感受,蕴含着重要研究价值。一方面有助于 App 开发商、运营商等主体基于用户反馈信息进行 App 优化与创新,提升用户体验和使用意愿;另一方面可以帮助潜在用户全面比较同类医疗 App 之间的功能、内容等差异,从而做出更加理性和正确的决策。而在线评论中往往包含用户对 App 不同属性维度的评价,并带有强烈的感情色彩。通过对在线评论进行维度挖掘和情感分析,可以准确、全面地掌握用户对移动医疗 App 的关注重点和满意程度,进而为优化用户体验指明方向。

目前,常用的评论挖掘与情感分析方法包括情

感词典、机器学习算法和深度学习算法。其中,情感词典中情感词的判别与选择取决于先验知识与实验设计,在完备性和准确性方面还存在不足;而传统机器学习算法的精确度在很大程度上取决于标注训练集的质量,对人类经验和领域知识的依赖性很高。随着深度学习的快速发展,基于深度学习的文本情感分析方法逐渐得到广泛应用,并在自适应特征学习和模型泛化能力上体现出明显优势。因此,本文将综合应用多种深度学习算法,对移动医疗 App 在线评论进行维度挖掘与情感分析。

## 1 相关研究方法

### 1.1 评论维度抽取方法

评论维度抽取是指通过挖掘评论中的特征词提取出用户对产品或服务的具体评价维度,是对评论文本进行细粒度观点挖掘和情感识别的基础<sup>[3]</sup>。常用的评论维度抽取方法主要包括模式匹配法、词频统计法和监督学习法。

模式匹配法是通过词典中的每个词与被处理文本之间逐一进行匹配,或者根据文本特点与定制规则进行匹配<sup>[4]</sup>。有学者基于中国知网(CNKI)提供的词典,采用语义相似度计算方法实现观点维度词的抽取<sup>[5]</sup>;有学者专门构建了一个汽车领域的本体词典,用于从相关评论中抽取客户对不同品牌汽车性能和配置指标的评价<sup>[6]</sup>。模式匹配法使用简单,准确率高,但一般需要领域专家针对不同领域建立相应的词典和规则库,投入的时间和人力成本很高,并且需要定期更新维护。

基于词频统计的方法是将词频作为衡量词语对文本重要性的测量指标,主要分为词频统计法和加权统计法<sup>[7]</sup>。有学者使用 TF-IDF 算法和位置加权算法在新闻和博客评论文本中提取评论维度特

征<sup>[8]</sup>；有学者提出了一种基于 C-value 度量统计的方法，用于对用户评论进行维度抽取<sup>[9]</sup>。词频统计法的优势在于可以快速提取主题词及分析特征词词频与主题类别之间的关系，但对评论中不常出现的特征词分析效果较差。

监督学习法是将评论维度抽取作为自然语言处理领域中常见的序列标注问题，主要包括两大类。一是基于语料库的无监督抽取，典型算法有隐式语义分析（latent semantic analysis, LSA）、隐含狄利克雷分布（latent dirichlet allocation, LDA）等，但这些算法没有考虑评论文本中各个词之间的位置关系，对细粒度的特征词抽取效果并不理想；二是基于单文本的抽取，典型算法包括基于词图网络的快速关键词提取算法（rapid automatic keyword extraction, RAKE）、主题排序算法（topic rank, TR）、位置排序算法（position rank, PR）等<sup>[10-12]</sup>，主要通过计算候选词向量与主题的语义相似度排序并提取出特征词。但常用于计算词向量的 Word2vec 等模型均为静态模型，在语义环境变化较大的情况下容易产生误差。近年来，有研究者尝试通过将预训练深度神经网络模型作为语言模型，针对具体任务使用微调的方式进行自然语言处理，取得了很好的效果。BERT 模型就是预训练模型之一，该模型利用其独特的 Transformer 结构对文本进行双向学习和处理，利用自注意力机制（self-attention）学习词间的关系，从而使词向量产生更加丰富的内涵，使原有的静态向量具有“动态”语言意义，甚至能在一定程度上让同一个词表达不同含义<sup>[13]</sup>。

## 1.2 文本情感分析方法

情感分析也称为“情感倾向分析”或“意见挖掘”，是目前自然语言处理领域的研究热点，其目的是从文本中识别出用户积极、中立或消极的情感态度<sup>[14]</sup>。情感分析方法主要包括情感词典方法、传统机器学习情感分析方法和深度学习方法。

情感词典方法主要依赖词典的构建，即先基于情感词典获取文档中情感词的情感值，再通过加权计算确定文档的整体情感倾向，因此需要针对特定领域建立相应的情感词典，提高分类准确率<sup>[15]</sup>。

对于小样本数据而言，该方法能够取得较好的分类结果且易于理解，但随着在线评论数据量的快速增长，情感词典需要不断更新扩展才能提高结果的准确率<sup>[16]</sup>。

传统机器学习情感分析方法是通过训练情感分类器来预测文本的情感倾向<sup>[17]</sup>。与情感词典方法相比，机器学习的扩展性和可重复性更好，能取得更加准确的分类结果，但其准确度很大程度上依赖于标注训练集的质量，受人工成本、专家经验等因素的影响较大。

近年来，深度学习方法在情感分析领域得到广泛应用。经典的深度学习模型包括卷积神经网络（convolutional neural networks, CNN）、循环神经网络（recurrent neural network, RNN）、长短期记忆神经网络（long short-term memory, LSTM）、双向长短期记忆神经网络（Bi-directional long short-term memory, BiLSTM）和注意力机制等<sup>[18]</sup>。深度学习方法比情感词典方法和传统机器学习方法有更强的表达能力和模型泛化能力。与此同时，在单一深度学习模型基础上，研究者还进一步对集成多种神经网络方法的文本情感分析展开了探索。有学者结合 CNN 和 LSTM 两种模型的优点，利用 CNN 模型提取较高层次的单词或词组，用向量进行数据建模，并将其输入 LSTM 模型中得到更加复杂的向量表示，这样既能获取文本的局部特征，又能得到句子的时态语义<sup>[19]</sup>；有学者应用 CNN 和 RNN 结合的方法进行在线评论情感分类，也取得了良好的分类效果<sup>[20]</sup>。

在已有研究成果的基础上，本文针对移动医疗 App 在线评论数据量大、口语特征明显、一词多义等特点，采用集成多策略的深度学习方法，通过词频-逆向文件频率算法（term frequency-inverse document frequency, TF-IDF）、BERT 模型与混合聚类的协同过滤推荐算法（Canopy+K-means）相结合的方式挖掘移动医疗 App 评论维度，明确用户重点关注的 App 属性维度；应用 LSTM-CNN 模型进行维度情感分析，从不同层面评估用户对移动医疗 App 的满意度，并通过实证研究，对所提出的评论维度挖掘和情感分析模型的准确性和稳定性

进行有效验证,为移动医疗 App 在线评论挖掘和用户满意度研究提供有益参考。

## 2 研究设计

### 2.1 研究框架

本文采用一种集成多策略的深度学习方法对移动医疗 App 在线评论进行维度挖掘和情感分析。维度挖掘的步骤:首先对采集到的移动医疗 App 在线评论数据进行清洗、分词、词性标注和去停用词等预处理,然后使用 TF-IDF 算法对评论进行抽词标引,提取特征词,再应用预训练的 BERT 模型生成特征词的词向量,接着利用 Canopy+K-means 算法进行特征词分析,投射到高维空间进行聚类,

由此得到移动医疗 App 用户在线评论中的不同维度。情感分析的步骤:进一步通过分句和依存句法分析 2 种文本分析方法从评论样本中得到包含具体维度特征词的短句,作为 LSTM-CNN 情感分类模型的输入语料,并根据各维度在所有评论中被提及的概率和该维度的情感倾向与整体评论情感倾向的一致性计算各维度权重。同时,利用训练集与测试集评论语料对 LSTM-CNN 情感分类模型进行训练,并基于训练好的模型进行评论情感倾向分类,进而计算用户对移动医疗 App 具体属性维度的积极情感倾向率,在此基础上评价用户满意度。具体研究框架如图 1 所示。

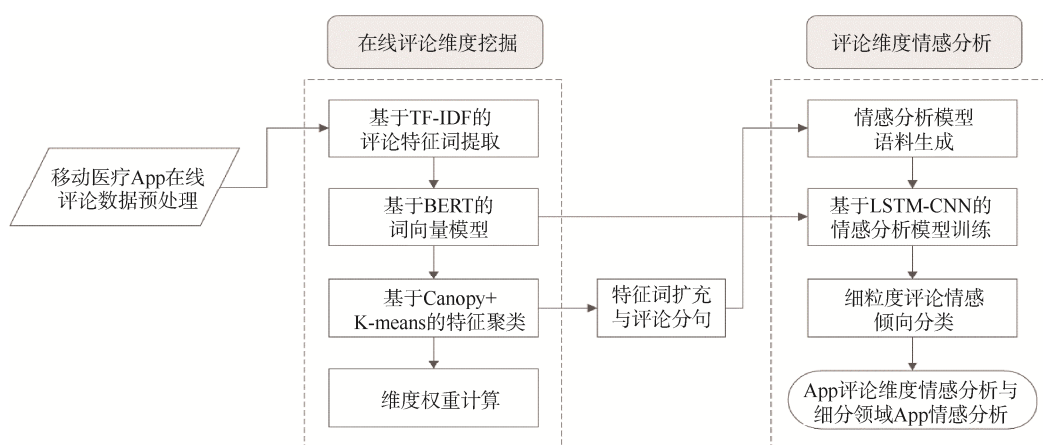


图 1 移动医疗 App 在线评论维度挖掘与情感分析研究框架

### 2.2 移动医疗 App 在线评论维度特征的抽取与聚类

移动医疗 App 在线评论具有明显的口语化和网络用语化特点,而且存在很多低价值甚至无效的信息。因此,首先需要对采集的评论数据进行清洗、文本分词、词性标注和去停用词等预处理。其次,为了兼顾评论中特征词抽取的时效性和准确性,本文使用 TF-IDF 算法进行特征词抽取,并选取 TF-IDF 值较高的名词表征移动医疗 App 的属性特征,进行向量数据建模,作为后续特征聚类的输入向量语料。

在此基础上,将候选特征词输入经过预训练的 BERT 模型,得到相应词向量,接着进行“二阶段”聚类:第一阶段使用 Canopy 聚类算法对在线评论中的属性特征和向量语料进行“粗”聚类,得到初始分类 K 值;第二阶段采用第一阶段输出的 K 值

进行 K 均值聚类(K-means),从而获得更加精准的分类结果,由此得到在线评论中用户所关注的移动医疗 App 的主要维度<sup>[21]</sup>。与传统 K-means 聚类算法相比,Canopy 聚类算法的优势在于无须事先指定 K 值,能够显著提高聚类效率,具有更高的实际应用价值。

### 2.3 在线评论维度权重计算

由于移动医疗 App 各维度的评价意见对用户满意度会产生不同程度的影响,因此需要进一步通过权重计算明确不同维度的重要性。具体而言,维度  $a_x$  的权重主要与两方面因素有关。一是维度  $a_x$  在所有评论中被提及的频次;二是用户对维度  $a_x$  的评论情感倾向与用户整体评论情感倾向的一致性<sup>[22]</sup>。因此,可以通过公式(1)和公式(2)进行计算。

$$w_x = \frac{\sum_{i=1}^R a_x \in r_i}{\sum_{x=1}^A \sum_{i=1}^R a_x \in r_i} \times \frac{1}{R} \sum_{i=1}^R eq(s_i, S_i) \quad \text{公式 (1)}$$

$$eq(s_i, S_i) = \begin{cases} 1, & \text{if } s_i = S_i \\ 0, & \text{if } s_i \neq S_i \end{cases} \quad \text{公式 (2)}$$

其中  $w_x$  表示维度  $a_x$  的权重值,  $A$  表示在线评论中所提取的维度个数,  $s_i$  表示涉及维度  $a_x$  的评论情感倾向,  $S_i$  表示整体评论情感倾向。计算得到的维度  $a_x$  的权重值越高, 说明其受用户的关注度越高, 对用户整体评论情感倾向的决定性作

用越强。

#### 2.4 基于 LSTM-CNN 的在线评论情感分析

本文构建的 LSTM-CNN 情感分析模型结构如图 2 所示。该模型主要由 LSTM 层、输入层、卷积层、池化层和全连接层构成。LSTM 模型是一种改进的基于 RNN 模型的网络结构, 能够有效保存长序列的历史信息; 而 CNN 模型在整个分析过程中能够比较精准地定位核心语义信息, 从而为后续数学模型判定得到文本的关键语义。因此, 本文综合应用 LSTM 和 CNN 模型来提高移动医疗 App 在线评论情感分类的准确率。

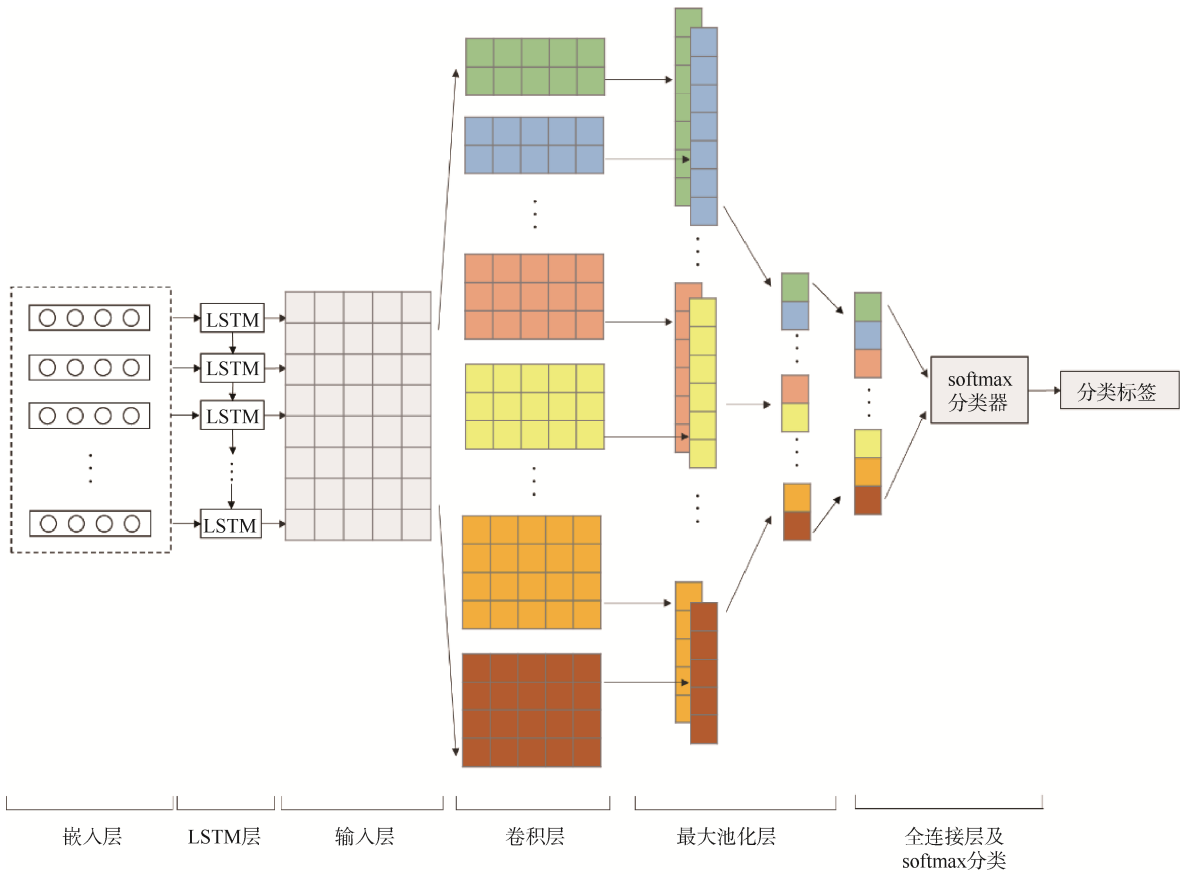


图 2 LSTM-CNN 模型结构

首先利用提前训练好的 BERT 模型将评论语料映射成向量矩阵, 每 1 个词为 1 个向量, 将其按时序输入 LSTM 模型, 从中提取有用的语义特征, 从而使卷积层能提取到更加丰富的局部信息, 提高模型分类的准确率。接着, 将 LSTM 模型输出的向量矩阵作为 CNN 模型的输入, 在卷积层对 LSTM

每一时刻的输出数据进行卷积操作, 提取出最重要的语义特征, 再由池化层对卷积层输出的数据进行最大池化。其中, 池化操作是设定一个 Max 值作为中间变量, 取每个过滤器中的 Top-K 个最大值作为该过滤器的语义信息。最后, 通过全连接层将前面提取到的局部特征进行整合与归一化处理, 并基

于柔性最大回归 (softmax regression) 得到最终分类结果。

### 3 实验结果和分析

本文以 App Store 中国应用市场中医疗类免费 App 下载排行榜为依据, 在“七麦数据”平台上逐一查询排名前 50 位的移动医疗 App 的评论数据。

“七麦数据”是权威的移动应用数据分析平台, 记录了 App Store、Google Play 和中国 8 个主要的安卓手机应用市场上 App 的各类榜单、评分、评论等多维数据<sup>[23]</sup>。根据查询结果, 本文选取了 2020 年 1 月 1 日至 2021 年 12 月 31 日期间评论量超过 500 条的 15 个医疗类 App 作为实证研究对象。根据这些 App 的详细介绍, 将其进一步细分为互联网医

院 App (4 个)、医药服务 App (2 个)、健康管理 App (4 个) 和综合服务 App (5 个) 四大类。首先, 应用 Python 爬虫技术采集这些 App 的评论数据, 包括 App 评分、评论数量、评论内容、评论人、评论时间等, 同时对用户 ID 进行了隐私保护处理。采集格式如表 1 所示。然后, 对这些评论数据进行清洗, 得到 23 771 条有效评论。接着从 15 个样本 App 的在线评论中各随机提取 500 条评论数据作为情感分析语料, 同时将没有被提取的评论数据分成两部分, 其中 80% 的语料作为 LSTM-CNN 模型深度学习算法的训练集, 20% 的语料作为测试集。之后, 使用 jieba 工具对所有评论数据进行分词和词性标注, 并进行停用词过滤, 作为初始化语料。

表 1 移动医疗 App 在线评论数据采集样例

评论时间	用户 ID	评论星级	评论标题	具体评论内容
2021-10-19 14:03:21	杭**民	5	很有帮助的平台	特别是对于现在这种面对疫情的情况, 我们可以通过这个平台学习到很多医学知识, 也能得到专业的帮助, 很赞!
2021-09-28 18:11:57	Tay**cd	4	很好用	非常实用的一款医疗软件, 上面的医师都是知名医院的专业医师, 平时有什么医疗健康方面的问题都可以在上面找到解决方案, 推荐推荐!
2020-07-10 22:53:29	哈**!	3	建议改进	上传图片的功能太坑了吧, 只能传 9 张, 我自己拼接一下还压缩了, 传上去看都看不清楚, 请问核磁结果动不动就好几张片子, 不管怎么拍都不可能 9 张内拍完吧, 不能给医生看检查结果还怎么问诊? 强烈建议改进!
2021-11-19 18:50:42	曾**镜	2	关注为什么每次都要弹出来?	大部分内容都是可以的, 也涨了见识, 但是每次看文章, 只要是没关注的医生都会弹出关注的页面, 弹出来就算了, 还不会自行退出, 我想关注我不会自己关注吗? 每篇文章都弹出来, 难道我每个医生都要关注吗?
2020-09-22 16:47:07	桃**岁	1	从以前用到现在了	以前还好, 现在花钱了指派医生太慢了, 医生回复也慢, 而且感觉没有以前专业。打算卸载换其他的软件, 无语!

#### 3.1 评论维度提取结果

采用 TF-IDF 算法计算得到移动医疗 App 在线评论对应的 TF-IDF 值, 将这些值进行排序, 选择排序靠前的特征词作为候选特征词, 然后进一步剔除词频较低和与医疗类 App 属性特征无关的词语, 最终得到 437 个特征词。采用经过预训练的 BERT 模型得到每个词的词向量, 并进行聚类分析。为测试基于 BERT 的 Canopy+K-Means 聚类算法的效果, 本文尝试提取 BERT 倒数第 2 层、倒数第 3 层的特征, 并用 10% 的语料进行简单的 FINE-TUNE 分类微调, 对提取调整后的 BERT 特征进行评估, 结果如表 2 所示。

表 2 基于 BERT 模型不同聚类方法的效果比较

聚类方法	聚类效果
BERT-CLS	0.37
BERT-MEAN	0.50
BERT 倒数第 2 层 CLS	0.45
BERT 倒数第 2 层 MEAN	0.51
BERT-FINE-TUNE-MEAN	0.90
BERT-FINE-TUNE 倒数第 2 层 MEAN	0.87

表 2 中聚类效果得分越高表示模型聚类效果越好。可以看到, 将 BERT 模型微调后得到的句向量进行求平均后得出的特征可以较好地生成聚类结果, 因此本文采用 BERT-FINE-TUNE-MEAN 方法



进行特征词聚类。最终,通过 Canopy+K-Means 聚类分析方法确定 K 值为 5。聚类后提取出专业性、易用性、可靠性、特色性、交互性 5 个评论维度。各维度下的具体特征词如图 3 所示。为了使提取到

的维度尽可能全面覆盖评论中涉及的 App 属性特性,本文利用通过维基百科中文语料训练的 BERT 词向量对每个维度下的特征词进行了扩充,得到移动医疗 App 评论特征词集。

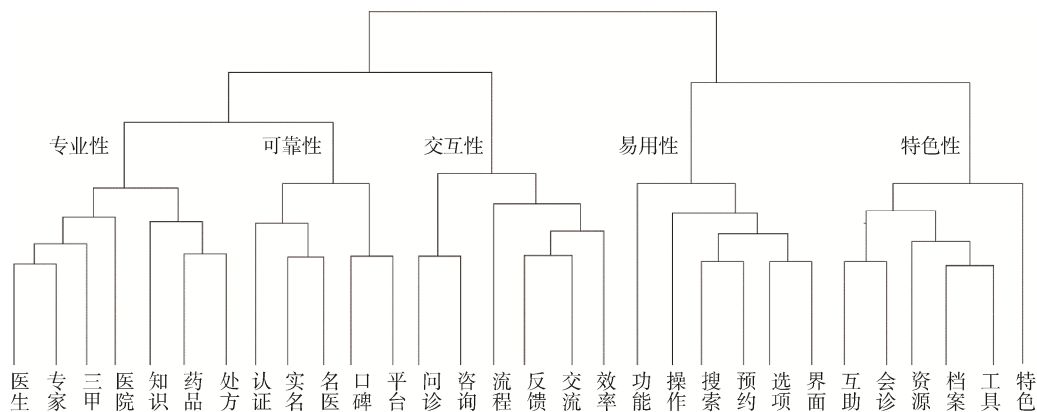


图 3 移动医疗 App 在线评论维度提取结果

### 3.2 评论维度权重计算结果

根据 App 在线评论维度提取结果,随机选择在线评论子集 R,对 R 中每条评论所包含的维度特征词及其情感倾向进行人工标注。根据评论星级设定评论的整体情感倾向,5 星和 4 星代表积极情感,3 星代表中性情感,1 星和 2 星代表消极情感。其中,积极情感评论所占比例为 58.3%,中性情感评论所占比例为 12.6%,消极情感评论所占比例为 29.1%。根据前述公式(1)和公式(2),计算得到移动医疗 App 5 个属性维度的权重值,从高到低依次为:专业性(0.367)、可靠性(0.296)、交互性(0.153)、易用性(0.125)、特色性(0.059),如图 4 所示。其中,专业性和可靠性是用户提到频次最多且与评论整体情感一致性最高的 2 个维度,所以权重值最高。可见,移动医疗 App 提供的信息内容与服务是否专业和是否可靠是用户最为关注的方面。而“医生的回复”“用户间的社交联系”“平台的反馈”等内容也经常被用户提及,这些内容反映了医患之间、用户之间和用户与平台之间的多元交互性,也是医疗 App 用户关注的重点,因此权重值也较高。此外,还有部分用户对 App 的界面是否简洁、功能是否易于使用、信息发布是否及时等易用性特征做出了评价。分析结果显示,用

户评论里较少提及移动医疗 App 的特色性。通过对样本 App 进行观察发现,大部分 App 的功能和内容同质化现象严重,缺乏核心特色,导致用户对医疗 App 属性特色的敏感度较低,也使该维度相关评论在整体评论中不具有主导作用,所以权重值最低。

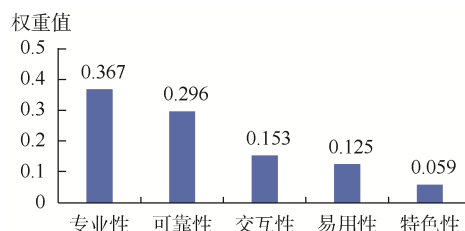


图 4 移动医疗 App 的在线评论维度权重值

### 3.3 评论维度情感分析结果

本文从 15 个样本 App 中各抽取 500 条在线评论,然后进行特征分句,并过滤掉其中不包含上述 5 个维度特征词的短句,将过滤后的评论内容作为情感分析依据。同时,对包含不止 1 个属性特征的短句进行依存句法分析,即通过语言单位内成分之间的依存关系揭示句子的句法结构,如“医生非常专业但回复速度太慢”,就需要分别提取包含专业性和交互性维度特征词的两个短句进行情感分析,由此得到进行移动医疗 App 在线评论情感分

析的语料。

采用 LSTM-CNN 模型作为情感分析模型。为了验证模型的有效性 with 精确性, 本文同时使用决策树 (Decision Tree)、LSTM、CNN 等 4 种模型进行情感分类效果对比 (表 3)。

表 3 不同模型的情感分类结果比较

模型	准确率/%	召回率/%	F1 值/%
Decision Tree	78.62	78.34	78.43
LSTM	90.25	85.73	86.18
CNN	93.06	90.17	90.92
LSTM-CNN	95.35	94.28	94.62

情感分类结果显示, LSTM-CNN 模型在准确率、召回率和 F1 值上均取得了比单一深度学习模型

更佳的效果, 证明它能够保证移动医疗 App 在线评论情感分类的准确性与可靠性。在此基础上, 本文利用 LSTM-CNN 模型对经过特征分句处理的用户评论语料进行情感分析, 分别得到 15 个样本 App 在 5 个维度上的用户评论积极情感倾向、中性情感倾向和消极情感倾向分析结果。其中, 样本 App 的用户评论积极情感倾向分析结果有显著差异, 而中性评论与消极评论由于数量相对较少, 情感分析结果并未呈现出较大差异。根据已有研究提出的 App 用户满意度评价标准, 在线评论的积极情感倾向率 (积极情感倾向的评论在所有评论情感倾向中所占的比率) 可作为衡量用户体验和满意度的重要依据<sup>[21,24]</sup>, 因此本文通过计算 5 个维度上每个样本 App 的用户评论积极情感倾向率来衡量用户满意度 (表 4)。

表 4 移动医疗样本 App 各维度的用户评论积极情感倾向率

序号	APP	细分领域	专业性	可靠性	交互性	易用性	特色性
1	好大夫在线	互联网医院	86.23	84.68	73.07	72.32	54.25
2	慧医	互联网医院	75.34	76.62	63.85	76.08	50.76
3	微医	互联网医院	68.65	70.05	65.48	70.22	88.27
4	微脉	互联网医院	76.77	74.18	68.36	73.53	70.18
5	叮当快药	医药服务	81.36	82.40	76.31	82.65	70.28
6	1 药网	医药服务	79.53	80.27	77.62	79.33	68.78
7	健康云	健康管理	68.25	67.82	56.08	48.23	82.21
8	小豆苗	健康管理	67.61	72.37	74.25	68.50	65.86
9	京东健康	健康管理	80.85	81.30	80.02	82.51	73.73
10	爱康	健康管理	65.93	68.70	67.22	60.17	68.39
11	小荷健康	综合服务	77.10	75.87	79.31	81.78	67.30
12	医鹿	综合服务	80.72	85.15	81.48	83.78	60.53
13	丁香医生	综合服务	87.76	85.34	78.26	75.61	50.35
14	春雨医生	综合服务	83.29	82.07	76.15	73.33	66.76
15	健康 160	综合服务	75.32	70.15	78.77	78.24	67.78

样本 App 在线评论的情感分析结果显示, 在专业性和可靠性维度上, “丁香医生” “好大夫在线” “春雨医生” “叮当快药” 等移动医疗 App 的积极情感倾向率都超过 80%, 说明用户对它们认可度较高, 而这些 App 在我国移动医疗服务起步阶段就已上线, 经过多年运营与发展, 已拥有庞大的用户基础和极高的知名度, 并与国内众多三甲医院、医疗服务机构、药品生产企业建立了广泛的合作关系, 积累了丰富的在线医疗资源, 因此在专业性和可靠性上获得了用户高度认可。“医鹿” 和 “京东健康” 两个 App 在专业性和可靠性维度上

的积极情感倾向率也位居前列, 它们分别由阿里巴巴和京东两个互联网领军企业开发。由此可见, 开发者的规模、实力、信誉等因素会对用户观点产生重要影响。在交互性和易用性维度上, 由阿里巴巴、京东、字节跳动等大型互联网企业开发的 “医鹿” (阿里巴巴)、“京东健康” (京东)、“小荷健康” (字节跳动) 的用户评论积极情感倾向率明显高于其他 App, 主要原因可能是这些互联网企业具有丰富的 App 设计与开发经验, 更加注重产品的交互设计细节与可用性测试, 因此能带给用户更佳的使用体验; 而其他 App 则更多关注内容生态和



服务体系的建设,忽视了交互的便利性与功能的易用性。在特色性维度上,各 App 在线评论积极情感倾向率均较低,基本都在 70% 以下,仅有“微医”(88.27%)和“健康云”(82.21%)的用户反馈较好。其中,“微医”除了预约挂号、在线问诊等基本医疗服务外,还推出了“病友帮”“服务包”等特色服务,而“健康云”则推出了亲人健康管理、慢病管理、医疗资源智能匹配等特色化功能,在同类 App 中表现出明显的特色优势。

不同类别的移动医疗 App 用户在线评论积极情感倾向率显示,医药服务类 App 和综合服务类 App 在专业性、可靠性、交互性和易用性 4 个维度上的用户满意度均较高,而健康管理类 App 在这 4 个维度上的消极评价则较多(图 5),说明目前我国移动医疗应用市场中,医药服务类 App 和综合服务类 App 的整体发展情况较好,用户接受程度较高;而健康管理类 App 的医疗资源质量和服务水平则参差不齐,有待进一步提升。此外,在特色性维度上,各类 App 的用户在线评论积极情感倾向率均较低,说明目前移动医疗 App 的同质化现象严重,缺乏符合用户个性化需求的特色功能与服务。虽然各大医疗 App 开发商已经开始重视特色化和精准性医疗健康服务的建设,但在医疗资讯、在线问诊、在线购药等核心功能方面的创新性还有待进一步加强。

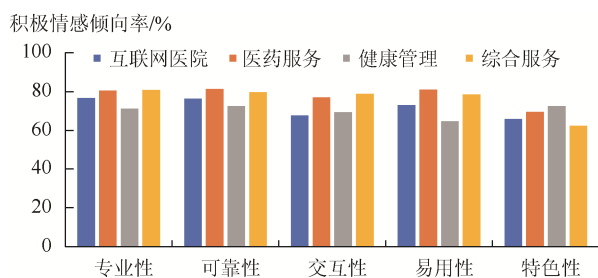


图 5 移动医疗细分领域 App 在线评论积极情感倾向率分布

### 3.4 结果分析

本文综合应用 TF-IDF 算法、BERT 模型和 Canopy+K-means 聚类分析方法,以当前我国移动医疗应用市场中用户广泛使用的 App 为实证对象,基于真实的大规模评论数据,提取得到受移动医疗

App 用户关注的专业性、可靠性、交互性、易用性、特色性 5 个主要维度作为衡量用户满意度的关键指标,进一步细化了基于传统问卷调查和专家经验提出的评价框架。已有研究大多侧重于从服务质量、信息质量、交互质量 3 个维度衡量 App 用户满意度,未能充分体现用户对移动医疗 App 专业性、可靠性的特殊需求,也难以突出激烈市场竞争下 App 的特色性对用户满意度的重要影响<sup>[25-27]</sup>。因此,本文的研究结果能更全面、精准地帮助 App 开发商和运营商把握用户对移动医疗服务的现实需求。与此同时,在评论维度挖掘基础上,本文还进一步计算了各维度的权重值,比较了不同维度对用户满意度的影响,明确了移动医疗 App 未来发展的重点方向。

通过基于 LSTM-CNN 模型的移动医疗 App 在线评论情感分析,本文通过计算得到各 App 在不同维度上的用户评论积极情感倾向率,并由此详细比较了不同 App 的优势与劣势,深入揭示了影响用户满意度的具体因素,有助于 App 开发商和运营商在“健康中国”战略引导下,以用户需求为中心,更有针对性地进行内容、功能、界面的优化与完善。同时,通过医疗健康各细分领域 App 间的横向比较,进一步明确了不同类型移动医疗 App 的核心优势和发展方向,有助于推动我国移动医疗 App 的创新与发展,带动我国传统医疗行业的数字化转型与智能化升级。

## 4 结语

随着移动医疗服务的快速发展和新冠肺炎疫情的常态化防控,移动医疗 App 正逐渐成为人们有效使用医疗资源和服务的便捷方式。通过对移动医疗 App 在线评论进行维度挖掘和情感分析,可以深入了解用户在使用移动医疗服务时的关注重点和体验效果,对 App 开发商和运营商进行产品迭代创新和用户体验优化具有重要价值。针对已有研究方法的局限性,本文提出了一种集成多策略的深度学习方法对移动医疗 App 在线评论进行维度挖掘和情感分析,综合应用 TF-IDF 算法、BERT 模型和 Canopy+K-means 聚类分析方法提取 App 在线评论的主要维度,在此基础上进一步利用

LSTM-CNN 模型进行在线评论细粒度情感分析,从而全面、准确地衡量了用户对不同移动医疗 App 的满意度。实验结果显示,本文提出的集成多策略的深度学习方法在移动医疗 App 在线评论维度挖掘和情感分类上具有很好的适用性、稳定性和可推广性,可为 App 评论文本精准分析和用户满意度评价提供重要理论依据和技术支持。

但本文也存在一定的局限性。考虑到深度学习方法对在线评论挖掘和情感分类的数据量要求,本文仅选择了目前用户数量较多且评论数据丰富的代表性移动医疗 App 进行实证研究,样本规模不足,未能充分反映我国移动医疗应用市场的整体情况。在今后的研究中,随着用户评论数据的不断累积,我们将进一步扩大样本规模,并针对更多医疗细分领域的 App 展开可持续研究。同时,进一步优化基于深度学习的细粒度评论情感分类模型,为推动我国移动医疗 App 的创新发展提供更有效的理论和实践指导。

#### 【参考文献】

- [1] data. aiIntelligence. 应用排行[EB/OL]. [2022-03-10]. <https://www.data.ai/intelligence/top-apps/store-rank/ios>.
- [2] Pires I M, Marques G, Garcia N M, *et al*. A research on the classification and applicability of the mobile health applications[J]. *Journal of Personalized Medicine*, 2020, 10(1): 11.
- [3] Yang X Y L, Xu S, Wu H, *et al*. Sentiment analysis of Weibo comment texts based on extended vocabulary and convolutional neural network[J]. *Procedia Computer Science*. 2019, 147: 361-368.
- [4] 王思丽, 刘巍, 杨恒, 等. 基于自然语言处理和机器学习的实体关系抽取方法研究[J]. *图书馆学研究*, 2021(18): 39-48.
- [5] 朱嫣岚, 闵锦, 周雅倩, 等. 基于 HowNet 的词汇语义倾向计算[J]. *中文信息学报*, 2006, 20(1): 14-20.
- [6] Zhang Q H, Wang J, Fan K. Research on passenger car recommendation based on comments mining of internet[C]// 2017 12th IEEE Conference on Industrial Electronics and Applications. IEEE, 2017: 122-127.
- [7] 侯汉清, 章成志, 郑红. Web 概念挖掘中标引源加权方案初探[J]. *情报学报*, 2005, 24(1): 87-92.
- [8] Ku L W, Liang Y T, Chen H H. Opinion extraction, summarization and tracking in news and blog corpora[J] AAAI Spring Symposium- Technical Report, 2006, SS-06-03: 100-107.
- [9] Zhu J B, Wang H Z, Tsou B K, *et al*. Multi-aspect opinion polling from textual reviews[C]//Proceedings of the 18th ACM Conference on Information and Knowledge Management. ACM, 2009: 1799-1802.
- [10] Rose S, Engel D, Cramer N, *et al*. Automatic keyword extraction from individual documents[M]. *Text Mining: Applications and Theory*, UK: John Wiley & Sons, Ltd, 2010: 1-20.
- [11] Bougouin A, Boudin F, Daille B. TopicRank: graph-based topic ranking for keyphrase extraction[C]//6th International Joint Conference on Natural Language Processing, IJCNLP 2013 - Proceedings of the Main Conference, 2013: 543-551.
- [12] Florescu C, Caragea C. PositionRank: an unsupervised approach to keyphrase extraction from scholarly documents[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver. Association for Computational Linguistics, 2017: 1105-1115.
- [13] 祖成. 基于 BERT 的中文文本向量化表示[J]. *科技与创新*, 2021(21): 107-108.
- [14] 陈龙, 管子玉, 何金红, 等. 情感分类研究进展[J]. *计算机研究与发展*, 2017, 54(6): 1150-1170.
- [15] Joshi A, Bhattacharyya P, Ahire S. Sentiment resources: lexicons and datasets[M]// A Practical Guide to Sentiment Analysis. Cham: Springer International Publishing, 2017: 85-106.
- [16] Deng S Y, Sinha A P, Zhao H M. Adapting sentiment lexicons to domain-specific social media texts[J]. *Decision Support Systems*, 2017, 94(3): 65-76.
- [17] Neethu M S, Rajasree R. Sentiment analysis in twitter using machine learning techniques[C]//Fourth International Conference on Computing. IEEE, 2014: 1-5.
- [18] 钟佳娃, 刘巍, 王思丽, 等. 文本情感分析方法及应用综述[J]. *数据分析与知识发现*, 2021, 5(6): 1-13.
- [19] Zhang Y, Yuan H, Wang J, *et al*. YNU-HPCC at EmoInt-2017: Using a CNN-LSTM Model for Sentiment Intensity Prediction[C]// Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2017.
- [20] Tang D Y, Qin B, Liu T. Document modeling with gated recurrent neural network for sentiment classification[EB/OL]. [2022-01-21]. <https://aclanthology.org/D15-1167.pdf>.
- [21] 赵杨, 李齐齐, 陈雨涵, 等. 基于在线评论情感分析的海淘 App 用户满意度研究[J]. *数据分析与知识发现*, 2018, 2(11): 19-27.
- [22] 赵志滨, 刘欢, 姚兰, 等. 中文产品评论的维度挖掘及情感分析技术研究[J]. *计算机科学与探索*, 2018, 12(3): 341-349.
- [23] 七麦数据[EB/OL]. [2022-01-21]. <https://www.qimai.cn/about/index>.
- [24] Song B, Lee C, Park Y. Identifying critical factors for customer satisfaction in mobile application service: a semantic text mining and bayesian network approach[J]. *International Proceedings of Economics Development & Research*, 2011, 123(8): 33-37.
- [25] Akter S, D'Ambra J, Ray P. Development and validation of an instrument to measure user perceived service quality of mHealth[J]. *Information & Management*, 2013, 50(4): 181-195.
- [26] Oppong E, Hinson Robert E, Adeola O, *et al*. The effect of mobile health service quality on user satisfaction and continual usage[J]. *Total Quality Management and Business Excellence*, 2018, 32(1): 1-22.
- [27] 张阁, 于琦, 贺培风. 我国移动医疗 App 核心功能分析及发展建议[J]. *中华医学图书情报杂志*, 2017, 26(4): 21-25, 33.

[收稿日期: 2022-04-08]

[本文编辑: 黄思敏]