



基于患者画像的抑郁症阅读材料推荐模型研究^{*}

□刘忠宝 赵文娟

摘要 阅读疗法是治疗抑郁症的有效方法,然而阅读疗法往往需要具有图书馆和心理学背景的专业人员为患者提供阅读材料,这项工作非常繁杂而且费时费力。鉴于此,以微博“树洞”存在的抑郁症患者数据为研究对象,在建立患者画像和阅读材料画像的基础上,通过比较两类画像的匹配度,为患者提供适合的阅读材料。本研究对于抑郁症的治疗起到了一定的推动作用,也对阅读疗法的深入研究和落地实践具有一定的参考价值。

关键词 抑郁症 患者画像 阅读疗法 阅读材料 推荐模型

分类号 G256

DOI 10.16603/j.issn1002-1027.2021.01.016

1 引言

时至今日,抑郁症已经成为一种常见的心理疾病,是目前世界第四大疾病,2020 年成为仅次于心血管病的世界第二大疾病,到 2030 年抑郁症患者人数预计将超过所有心血管患者的总和,成为世界第一大疾病^[1]。抑郁症主要症状是心情低落、思维迟缓、意志衰退并伴随着其他心理和生理症状,严重影响患者的情绪和行为,并且经常反复发作。世界卫生组织发布的最新数据表明,心理问题导致 12% 的疾病,其中 46% 的疾病与抑郁症直接相关^[1]。抑郁症患者自杀的死亡风险比普通人群高 20 倍以上,预计到 2030 年抑郁症将成为致残和自杀的第一诱因^[2]。目前,人们对抑郁症的认识不足,对自身是否处于抑郁症患病状态没有意识,部分人对心理疾病存在偏见,不愿意主动寻求帮助。此外,抑郁症是比较敏感的心理疾病,主动就诊的抑郁症患者可能会再次遭受心理伤害,因此,超过 70% 的抑郁症患者在早期阶段不会寻求医疗帮助,这导致他们的病情进一步恶化^[3]。

作为一种典型的心理疗法,阅读疗法现已被证明是缓解抑郁症患者抑郁风险的一种有效方法。中医自古以来就注重“情志调摄治未病”和“情志相胜”的治疗理念和方法,《吕氏春秋》记载了名医文挚采用心理疗法来治愈齐国君主齐湣王忧郁症的案

例^[4]。与传统治疗方法相比,阅读疗法有效地扩大了受众范围,减轻了患者在经济上和精神上的双重压力。然而,阅读疗法往往需要具有心理学和图书馆背景的专业人员为患者提供阅读材料,这项工作非常繁杂而且费时费力。推荐模型的出现为解决上述问题提供了可能。因此,本研究以微博“树洞”存在的抑郁症患者数据为研究对象,在建立患者画像和阅读材料画像的基础上,通过比较两类画像的匹配度,来为患者推荐适合的阅读材料,进而缓解患者的抑郁倾向。

2 研究进展

阅读疗法是指以文献为媒介,将阅读作为养生、保健以及辅助治疗疾病的手段,使自己或指导他人通过对文献内容的学习、讨论和领悟,养护或恢复身心健康的一种方法^[5]。目前,阅读疗法引起国内外学者的广泛关注,并取得了不少研究成果。国外研究主要侧重于心理学、医学等领域。麦卡利斯(McCulliss C)等针对阅读疗法的模式和内容给出建议,并分析了阅读疗法对改善心理健康的作用^[6];布鲁斯特(Brewster E)对英国阅读疗法的书方模式进行了梳理^[7]。在国内,王波针对阅读疗法的基本原理、概念和最新进展进行研究^[8],并探讨了中国历史上的阅读疗法医案,包括先秦到汉代的阅读疗法思想

^{*} 国家社会科学基金一般项目“大数据环境下面向图书馆资源的跨媒体知识服务研究”(编号:19BTQ012)的研究成果之一。

通讯作者:刘忠宝,ORCID:0000-0002-0038-2462,邮箱:zblu@blcu.edu.cn。



和医案^[9],魏晋南北朝的阅读疗法思想和医案^[10],以及明清时期的阅读疗法思想和医案^[11],进一步整理了中国古代和国外典型的阅读疗法书目,并对阅读疗法的书目类型进行划分并给出实例^[12];宫梅玲等主要研究阅读疗法的书方以及实践模式,包括对失恋大学生的抑郁情感书方分析及配伍等^[13],并探讨了阅读疗法的运作模式^[14];林佳瑜对高校具有育儿焦虑情绪的女职工实施阅读疗法,为阅读疗法实践提供了可借鉴的方案^[15];潘悦利用交互式阅读疗法为学龄前儿童提供有针对性的情绪疏导^[16];赵晓华等对移动互联网时代基于自媒体的高校图书馆开展阅读疗法的可行性和实现方式进行了探讨^[17];张怀涛等从学科融合角度,探讨了阅读疗法与相关学科之间的联系,明确了促进阅读疗法发展的主干学科和辅助学科,进一步丰富了阅读疗法的理论体系与方法体系^[18]。

近年来,一些学者逐渐意识到书目对于阅读疗法成效的重要性,并开展了一系列深入研究。吴笑兰全面梳理了近年来我国高校图书馆阅读疗法书目编制研究情况,并对其中的书目选择和文献揭示问题进行了深入研究^[19]。王景文等利用文献计量方法,在梳理我国阅读疗法书目研究进展的基础上,指出存在的两大问题:一是我国阅读疗法书目研究缺乏突破性成果;二是我国阅读疗法书目实践严重不足。这为后续阅读疗法书目研究指明了方向^[20]。宫梅玲等通过研读心理学家推荐且学生反馈良好的阅读材料,给出了适用于治疗童年创伤性抑郁症的书目分类:第一类是童年创伤性抑郁症作家的写实文学,第二类是有童年创伤的心理学家撰写的疗伤文献,第三类是精神病专家和心理学家撰写的治疗抑郁症的文献^[21]。徐海军等以宫梅玲研制童年创伤引发的抑郁障碍配伍书方为例,重点探讨了阅读疗法师选择阅读疗法理论书、中医理论书、抗抑郁的书的方法^[22]。田花蔓等在分析大学生心理健康问题的基础上,从适应环境挫折、学习挫折、情感挫折、自我认知挫折、人际交往挫折等方面对高校阅读疗法书目选择标准进行了研究^[23]。解荣等以现有阅读疗法研究成果为研究对象,利用例证法揭示书目特征,探讨阅读疗法书目的选择方法,建立特征化的阅读疗法书目体系^[24]。此外,他们还从读者视角探讨了阅读疗法书目的评价指标。该研究立足于阅读疗法实践,利用扎根理论方法对访谈数据的深入分

析,得到阅读疗法书目评价指标,并引入阅读疗法心理学原理和吸引力因素理论揭示评价指标的内涵^[25]。

随着大数据时代的到来以及人工智能技术的兴起,一些学者针对阅读疗法视域下的患者画像与阅读材料推荐方法展开研究。患者画像指基于抑郁者患者真实数据形成的描述患者属性及其行为的标签集合,其中包含了患者的需求、行为、兴趣、心理、性格等。阅读资料推荐方法根据患者画像描述的患者抑郁程度和兴趣偏好,智能推送有助于治疗的阅读材料。韩梅花等探讨了大数据环境下基于“用户画像”的阅读疗法新模式^[26]。该模式首先依据“伯恩斯抑郁症清单”得到抑郁情绪种子词;接着利用 TF-IDF 算法或聚类算法扩充种子词,进而得到抑郁情感词典;然后,利用抑郁情感词典,通过计算微博用户的抑郁情感指数得到“用户画像”;最后,根据用户抑郁情感指数高低,推送具有一定疗效的阅读材料。刘丹等对基于用户画像的高校图书馆阅读疗法模式展开研究^[27]。该模式首先通过扩充 HowNet 词典得到情绪词典;接着,基于情绪词典分析大学生的微博文本,得到大学生的“情绪画像”;然后,根据得到的负面情绪倾向,发现潜在的患者;最后,利用现有的信息推送技术,向患者自动推送阅读材料。辛云星对阅读疗法在青少年减压方面的有效性问题进行了研究并搭建了“心悦读”青少年阅读推荐平台^[28]。该平台每周为用户推荐一定数量的阅读材料。该平台分别用用户配置文件和阅读材料配置文件表示用户画像和阅读材料画像,利用基于内容过滤的方法进行推荐。此外,该平台还能够根据用户的压力状态和阅读行为,及时更新用户配置文件,以便使推荐结果更符合用户需求。

上述研究重点关注阅读疗法的理论、方法和实践,而对于如何利用阅读疗法为患者提供有针对性的阅读疗法服务方面的研究尚不成熟,特别是这些研究对患者的抑郁倾向和偏好分析不够深入,这在一定程度上限制了阅读疗法的实际效果。本研究旨在引入用户画像和智能推荐技术和方法,在阅读疗法实施过程中突出患者抑郁倾向和偏好分析,通过建立患者画像和阅读材料画像,智能地为患者推荐适合的阅读资料,以期为患者提供更为精准的阅读疗法服务。



3 患者画像

建立患者画像的目标是通过对患者数据的分析,为患者打上特征标签,以获得患者的阅读偏好和行为习惯。患者画像建模的基本流程包括数据采集层、数据处理层、标签分析层。数据采集层借助 API 接口和网络爬虫程序从微博“树洞”抓取患者的静态数据、动态数据以及微博文本数据;数据处理层对收集到的海量数据进行必要的去噪、转换和分词等预处理,并利用机器学习算法进行深层次的数据挖掘和关联分析;标签分析层为每位患者打上特征标签,并采用聚类算法将患者划分成若干个具有相似特征的患者群体。患者画像建模如图 1 所示。

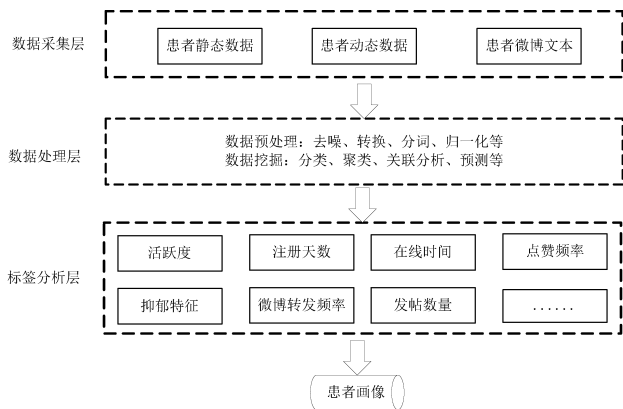


图 1 患者画像建模

3.1 数据采集层

目前,抑郁症患者常常到微博“树洞”倾诉迷茫和痛苦。“树洞”来源于《皇帝长着驴耳朵》一书,理发师最先发现皇帝长了驴耳朵,他担心自己不小心将该秘密泄露出去,于是就跑到山上将秘密告诉了树洞。后来,“树洞”就被引申为袒露心声、隐藏秘密的地方。微博“树洞”结合了微博和“树洞”的优势,具有匿名发布、言论自由、内容丰富、发布迅速等特点,成为目前广大网民匿名倾诉的主要平台。本文依据 3.2 节构建的情感词典,利用文本挖掘技术对微博“树洞”发布的微博信息进行分析,进而识别出哪些博主是抑郁症患者。本文借助 API 接口和网络爬虫程序从微博“树洞”获取患者的静态数据、动态数据以及微博文本数据,其中患者的静态数据和动态数据分别指:

(1) 患者静态数据

患者静态数据描述患者的基本属性,主要包括发帖数、关注者数、粉丝数、用户性别、用户年龄、注册天数和区域代码。注册天数是指患者自注册微博之日起的总天数,区域代码是指患者所处地理位置,

微博提供了区域代码列表,例如 11 表示北京市。

(2) 患者动态数据

患者动态数据描述患者的交互属性,主要包括转发频率、评论频率、点赞频率、发帖频率和微博发布时间。其中,转发频率、评论频率、点赞频率、发帖频率表明患者与其他患者互动的行为特征,通过数量/注册天数进行归一化处理。由于患者相比正常人晚上更容易失眠,因此在患者动态数据中加入了微博发布时间。

3.2 数据处理层

数据处理层的任务是通过对患者数据分析,以期得到患者特征表示。其基本流程是:首先,构建抑郁情感词典,该词典由基础情感词典、表情符号词典以及自杀基础词典组成;接着,利用 Jieba 工具对微博文本数据进行分词;然后,利用 Word2vec 算法将分词后的微博文本数据表征为向量形式,基于抑郁情感词典,对分词后的微博文本数据进行遍历与匹配,得到患者的抑郁情感词,利用 TF-IDF 算法对抑郁情感词进行加权求和运算,得到患者抑郁指数,根据抑郁指数大小,可以将抑郁症患者分为轻度、中度、重度三类;对于抑郁成因,一方面可以参考患者在微博上的注册信息,其中部分信息可能涉及抑郁成因,另一方面可以利用文本分析工具抽取微博文本特征,利用典型相关分析算法将微博文本特征与抑郁成因集合(如遗传因素、内分泌因素等)进行关联分析,进而得到患者的抑郁成因;最后,融合患者的静态数据、动态数据、抑郁指数、抑郁成因得到患者特征表示{发帖数,关注者数,粉丝数,用户性别,用户年龄,注册天数,区域代码,转发频率,评论频率,点赞频率,发帖频率,微博发布时间,抑郁指数,抑郁成因}。在上述流程中,抑郁情感词典构建至关重要,下面给出该词典构建的具体过程。

抑郁症情感词典构建基于基础情感词典。目前,基础情感词典主要包括“知网情感分析用词语集”^[29]和“大连理工大学中文情感词汇本体库”^[30]。“知网情感词典”分为中文和英文两部分,其中,中文词典包括正负面评价词典和正负面情感词典。由于抑郁症是一种由负面情绪引发的疾病,因此选用中文负面情感词典。该词典包括 1254 个负面情感词。“大连理工大学中文词汇本体库”从词语词性类别、情感类别、情感强度及词汇极性等不同角度描述中文词汇或者短语。鉴于抑郁症与心理学的密切关



联,故选用哀、惧、恶和惊等四类共计 3457 个负面情感词。上述两类负面情感词存在部分交叉,经去重处理后,共计得到 3916 个负面情感词,将其作为基础情感词典。

基础情感词典是从现有的公开词典中提取并建立的,该词典对负面情绪的覆盖面大,但抑郁症的针对性不强,因此,有必要根据抑郁症的特点对基础情感词典进行扩展。目前常见的词典扩展方法是通过计算知识库和语料库的相似度,将相似度较高的词语扩展到词典中。词典扩展包括情感种子词选取、词向量表示和相似度计算等步骤。首先,根据临床医学广泛使用的汉密尔顿抑郁测量表和伯恩斯抑郁自查表,并结合心理学给出的抑郁症症状,构造抑郁情感种子词,如“妄想”“焦虑”“惊恐”“叹气”“痛苦”“压抑”“烦闷”等;接着,将微博文本数据作为语料库,并利用 Word2vec 算法得到词向量表示;然后,利用相似度计算方法计算与种子词相似的词;最后,选取与每个种子词相似的前 n 个词扩充到基础情感词典。由于语料库规模较大,基础词典扩展可以采用迭代的方式,即将上一轮的扩展词作为下一轮的种子词,实现词典的增量式扩展。

表情符号是微博文本数据的重要组成部分,这些符号直观地表明了患者的情绪。抑郁症患者的情感表达相对是消极的,因此,选取微博文本数据中的消极表情符号建立表情符号词典。表 1 给出了微博文本数据中的消极表情符号。

表 1 微博文本数据中的消极表情符号

表情符号	描述	表情符号	描述
	委屈		抓狂
	可怜		泪
	失望		衰
	生病		伤心

抑郁症患者属于自杀高风险人群,而且倾向于在微博中表达自杀意图和自杀方式。目前还没有关于自杀意图和自杀行为的公开词典,因此,采用人工筛选方式将 16 个代表性词语作为自杀基础词典(见表 2)。

表 2 自杀基础词典

安眠药	跳楼	自焚	触电
跳崖	砒霜	沉湖	咬舌
上吊	割动脉	卧轨	撞车
绝食	割腕	投井	农药

3.3 标签分析层

标签分析层通过对患者数据的分析,提取关键信息并发现信息之间的潜在联系,为患者打上特征标签。单一患者和患者群体的标签集合可分别视为单一患者画像和患者群体画像。与单一患者画像相比,患者群体画像从多个维度表征患者特征的相似性,如患者群体的性别、年龄、发帖数等基本属性的相似性、患者群体之间转发频率、点赞频率、发帖频率等互动属性的相似性等。此外,根据患者动态数据的变化情况,及时调整单一患者和患者群体的标签集合,进而完成患者画像和患者群体画像的更新。

在患者画像建模过程中用到一系列机器学习算法。引入相似度计算方法比较患者与特征标签之间的相似性,并为学生打上特征标签。利用聚类算法对患者进行聚类分析,得到患者群体。聚类的目的是将具有相近兴趣和需求的患者聚合在一起。聚类的意义主要体现在两方面:一方面,对于某患者,推荐模型不仅能根据该患者画像推荐阅读材料,而且还能推荐其所在患者群体的阅读材料,有效地扩大了阅读材料的推荐范围;另一方面,对于一些新患者,其动态数据较少,无法建立有效的患者画像,但能根据该患者的静态数据初步判断其所属的患者群体,将其所在患者群体的阅读材料推荐给该患者,随着患者动态数据规模的增大,患者画像的刻画能力增强,推荐模型能够提供更具针对性的阅读疗法服务。在缺乏先验知识的情况下,K-means 聚类算法能够根据患者的相似性,将患者划分为不同的类簇,各类簇内部元素之间具有较大相似性,各类簇即为患者群体。该算法的具体流程是:

第一步:随机选取 K 个患者的特征向量作为聚类中心;

第二步:遍历全体患者,并计算每个患者的特征向量与聚类中心之间的距离,当距离小于事先给定阈值时,则认定该患者与位于聚类中心的患者具有较大相似性,将两者归并至同一类簇;

第三步:根据上述聚类结果,重新计算 K 个类簇的中心,更新聚类中心;

第四步:不断迭代,重复第二步和第三步,直到 K 个聚类中心不再变化或此次聚类结果与上次聚类结果趋同;

第五步:得到最终聚类结果。

聚类中心 K 的选取至关重要,会直接影响聚类



结果。聚类中心的选取目前主要有经验选择、随机选择、最小最大原则等方法,其中最小最大原则根据聚类对象的相似性来选择聚类中心,对经验知识的依赖较小,也可克服随机选择的盲目性。鉴于心理学已经明确给出抑郁症的类别划分,因此,聚类中心选取可结合最大最小原则和心理学给出的抑郁症类别划分进行选取。

4 基于患者画像的推荐模型

在建立单一患者画像和患者群体画像的基础上,对阅读材料进行画像建模,通过比较两类患者画像与阅读材料画像的匹配度,分别为单一患者和患者群体推荐适合的阅读材料。

4.1 阅读材料画像

阅读材料选取依赖于合理的分类标准。本研究采取的阅读材料分类标准是:第一类是医学专业的阅读材料,即精神病学家和心理学家撰写的治疗抑郁症的阅读材料;第二类是患有抑郁症的作家撰写的纪实类阅读材料;第三类是患有抑郁症的精神病学家和心理学家撰写的治疗抑郁症的阅读材料。根据上述分类标准,给出阅读材料的主要来源:一是中国图书馆学会阅读推广委员会《面向大学生的常见心理困扰对症书目》给出的 96 种阅读材料;二是重度抑郁症患者李兰妮在《旷野无人——一个抑郁症患者的精神档案》一书中提到的 30 余种阅读材料;三是阳光工程心理互助网提到的 80 余种抑郁症治疗阅读材料。三类阅读材料共计 200 余种。为了扩大阅读材料搜集范围,定期地(如一周)在当当网、京东图书等网站上,利用“抑郁”“抑郁症”“心理健康”“忧郁”“童年创伤”等检索词搜索关于抑郁症普及、治疗以及抑郁症患者撰写的写实书籍。由图书馆学、心理学、精神病学等相关领域专家组成的选书团队认真研读这些书籍,从其专业角度出发,根据书籍与抑郁症的匹配度,分别给出每本书籍的综合评价。此外,结合网友的书评以及对作者及其经历的分析,选出备选阅读材料。上述选书方式增加了阅读材料选择的可信度。

根据前期的实地调研和文献梳理,本文认为在基于患者画像的推荐模型中,阅读材料主要考虑作者、图书类型、出版时间、适用人群、热度、抑郁主题等方面的特征。对于作者,患者可能对曾经的患者所著的书籍较为感兴趣,如李兰妮、奥卡拉罕、斯瓦

多等;对于图书类型,患者可能偏爱某些特定类型的图书,如医学、心理学、励志等类型的图书;对于出版时间,患者可能对某一时段的阅读材料感兴趣;对于适用人群,抑郁成因不同的患者对阅读材料的选取也不尽相同;对于热度,患者对销量高、阅读次数或下载量多的阅读材料具有一定的兴趣;对于抑郁主题,为了精确地获得阅读材料特征,利用 LDA(Latent Dirichlet Allocation)主题模型对阅读材料的摘要、引言、正文等进行抑郁主题抽取,利用 TF-IDF 算法对抑郁主题词进行加权运算,取权值最高的抑郁主题词作为阅读材料的主题,该权值称为抑郁主题隶属度。综上所述,阅读材料画像可描述为{作者,图书类型,出版时间,适用人群,热度,抑郁主题隶属度}。

4.2 推荐模型

目前主流的推荐方法有基于内容的推荐方法、基于协同过滤的推荐方法、基于深度学习的推荐方法和混合推荐方法。这些方法由于缺乏患者对阅读材料的浏览、评分、评论等历史数据,故无法适用于基于患者画像的阅读资料推荐。因此,借鉴协同过滤思想,利用相似度计算方法分别计算单一患者和患者群体与阅读材料之间的匹配度,将匹配度较高的阅读资源分别推荐给单一患者和患者群体。两种推荐方式的流程基本相同,当患者群体内患者数量很少导致不足以形成类簇时,利用单一患者画像进行推荐。推荐模型如图 2 所示。

该推荐模型的基本流程是:

第一步:建立面向抑郁症患者的单一患者画像、患者群体画像以及面向阅读材料的阅读材料画像;

第二步:对各类画像的标签进行聚类并归纳为若干主题,利用主题描述患者特征和阅读材料特征;

第三步:借鉴协同过滤思想,通过 Pearson 相关系数计算患者和阅读材料之间的相似度;

第四步:患者和阅读材料之间的相似度即为两者的匹配度,按匹配度降序排列,将匹配度高的阅读材料推荐给患者。

4.3 阅读疗法实施路径

基于患者画像和阅读材料推荐模型,围绕阅读疗法服务模式、阅读材料提供路径、阅读材料提供方式、阅读材料使用方式等方面,探讨图书馆开展阅读疗法的实施路径。

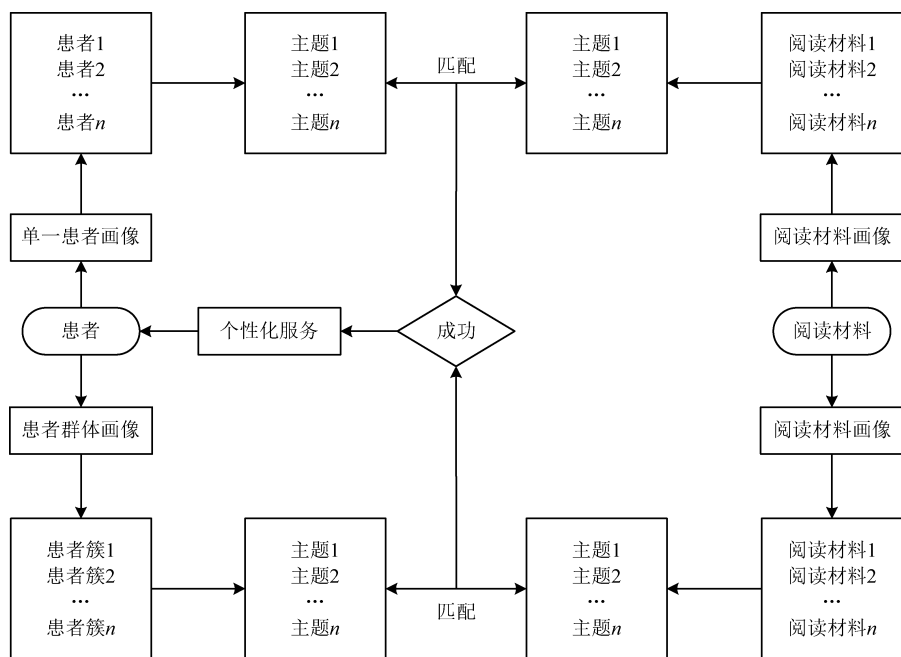


图2 基于患者画像的推荐模型

(1) 阅读疗法服务模式

融合线上线下优势创新阅读疗法服务模式。心理学的研究表明,环境会对人产生重要影响。图书馆应专门设立阅读疗法阅览室,并在其中布置画作、绿植以及舒适的桌椅,通过轻柔的音乐使抑郁症患者放松心情,营造适合阅读疗法的阅读环境。引入虚拟现实技术,让无法到图书馆的患者在线能够享受到同样的阅读疗法服务。虚拟现实是一种集图形学、仿真、传感、多媒体于一体的信息技术,虚拟的视觉感知和视觉互动,创造了一个虚拟的现实世界。阅读疗法具有增加患者参与度、构建虚拟现实环境的需求。因此,将阅读疗法与虚拟现实技术相结合是可行的。患者可以通过视觉、听觉和触觉等多种感官来阅读推荐模型推送的阅读材料。虚拟现实阅读治疗基于患者的目标驱动,治疗模式可以实现沉浸式的心理体验和深层次的虚拟互动体验。线上阅读方便快捷,但容易引起患者视觉疲劳、心情浮躁等问题。因此,图书馆可以根据本文提到的阅读材料来源,购置一些纸质书籍,为患者提供更多的阅读选择。

(2) 阅读材料提供路径

根据患者画像与阅读材料画像的匹配度,可以为患者提供定制化的阅读材料。图3给出了三条阅读材料提供路径。路径1是将患者画像与阅读材料进行粗粒度匹配,并通过服务调度引擎,向患者提供

所需的阅读材料;路径2是将患者画像与从阅读材料中抽取出的阅读单元进行细粒度匹配,阅读单元由阅读材料的章节组成,对阅读单元进行多维度融合,向患者提供细粒度的阅读材料;路径3是向第三方服务提供者开放患者画像接口,允许经授权的服务提供者对患者画像进行抓取,根据患者需求,为其提供相应的阅读材料。

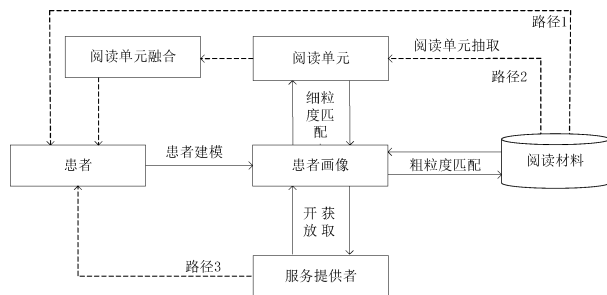


图3 阅读材料提供路径

(3) 阅读材料提供方式

阅读材料不仅可以通过图书馆的系统平台向患者的计算机终端进行推送,也可以利用新媒体技术和移动通信技术进行推送,如微博、微信、电子邮件、腾讯QQ等。此外,图书馆可以面向抑郁症患者群体,开通微信公众号,定时或不定时地推荐一些阅读材料,以适合不同抑郁程度的患者阅读。一般情况下,推荐模型推送的阅读材料均是文字信息,这种单调的信息呈现方式严重地影响了患者的阅读体验。



因此,有必要探索将阅读材料可视化、情境化、游戏化的新路径。利用文本挖掘技术,对阅读材料进行文本分析,构建阅读材料的知识图谱,借助可视化技术,直观地呈现阅读材料的知识架构和知识体系,方便患者掌握阅读材料的全貌,患者可以选择感兴趣的章节进行详细阅读,一定程度上提高了患者的阅读效率。引入语义分析技术,对阅读资料进行语义挖掘,生动描绘阅读材料的情境化语境,复现作者和主人公的生活场景,有助于增强患者对阅读材料的认同感。借鉴游戏化阅读思想,将阅读过程视为游戏过程,每完成一个阶段的阅读任务,给予患者一定的积分,当患者完成全部阅读任务时,其所获得的积分可以兑换相应的奖励。上述做法有助于提高患者阅读的参与度和完成度。

(4) 阅读材料使用方式

根据患者画像中的“抑郁指数”和“抑郁成因”可以得到患者的患病程度和患病原因。对于抑郁指数较高的患者,其抑郁程度较重,需要借助临床医生开具的处方药物进行治疗。对于抑郁指数较低的患者,可以寻求阅读疗法来缓解病症。阅读材料画像中的“适用人群”和患者画像中的“抑郁成因”明确了阅读材料的适用症状。在阅读疗法实践中,阅读顺序对于阅读疗法效果具有重要影响。因此,建议患者先阅读患有抑郁症的作家撰写的纪实类阅读材料,这些阅读材料具有一定的情感宣泄和审美教育功能,对于缓解苦恼、不安情绪具有一定作用,能够辅助治疗患者心理和精神上的某些轻微疾病。等到患者心情平静,享受到阅读带来的疗愈体验后,可以再阅读患有抑郁症的精神病学家和心理学家撰写的治疗抑郁症的阅读材料以及医学专业的阅读材料。在实施阅读疗法过程中,以下几个问题值得注意:一是在阅读初期,患者可能无法静心阅读,建议患者先听轻柔的音乐放松心情,或找知心朋友宣泄负面情绪,等情绪稳定后,再进行阅读;二是患者可以对自己喜欢的阅读材料反复阅读,并做好阅读疗法日记和读书感悟;三是如果条件允许,图书馆可以组织群体阅读活动,通过个体阅读和群体阅读相结合的方式,患者能够及时交流阅读心得,纾解心理压力,坚定治疗抑郁症的信心和决心。

4.4 应用案例

从微博“树洞”爬取 114 名抑郁症患者的静态数据、动态数据以及微博文本数据。以患者“走饭”为

例,给出单一患者画像。通过对患者数据的分析可知,患者性别为女,患者年龄为 28 岁,注册时间为 2009 年 3 月 17 日,粉丝数为 565611 人,发帖数为 1896 个,关注者数为 368 人,发帖频率为 1.49,评论频率为 0.21,转发频率为 0.11,点赞频率为 0.07,地理位置在江苏省,发帖时间为晚上 11 点 48 分。根据抑郁情感词典,得到该患者的抑郁情感词有:死、离开、哭、誓死、离去、孤立、苦、疼、创伤、害怕、纠结、痛苦、悸动、葬礼、残疾、受苦、焦虑、不适、悲伤、车撞、丧心病狂、安眠药、惊悚、惊吓、伤感、自卑、车祸、死尸。利用 TF-IDF 算法对抑郁情感词进行加权求和运算,得到该患者的抑郁指数为 0.535030722618103。抑郁成因集合为{1,2,3,4,5},其中 1 表示遗传因素,2 表示内分泌因素,3 表示压力因素,4 表示性格因素,5 表示日常琐事因素。利用典型相关分析算法得到该患者的抑郁原因为 5。该患者的注册时间为 2009 年 3 月 17 日,自杀时间为 2012 年 3 月 18 日,由此可知,该患者的注册时间为 1275 天。微博提供的江苏省代码为 32。性别男和女分别由 1 和 0 表示。融合患者的静态数据、动态数据和抑郁指数,可得“走饭”的患者画像为{1896, 368, 565611, 0, 28, 1275, 32, 0.11, 0.21, 0.07, 1.49, 23:48, 0.535030722618103, 5}。

将 114 名患者采用 K-means 聚类算法聚合成不同的类簇,得到图 5 所示的聚类结果。聚类效果由准确率 P(Precision)、召回率 R(Recall)和 F1 值(F-measure)来衡量,三者的计算公式如下:

$$P = \frac{TP}{TP + FP} \times 100\%$$

$$R = \frac{TP}{TP + FN} \times 100\%$$

$$F1 = \frac{2 \times P \times R}{P + R} \times 100\%$$

其中,TP、FP、FN 分别表示正确分类的患者数、错误分类的患者数以及无法分类的患者数。

在本实验中,TP=105,FP=4,FN=5,由此可知,P=96.33%,R=95.45%,F1=95.89%。上述实验结果表明 K-means 聚类算法适用于患者聚类场景。由图 4 可以看出,114 名患者聚为三个簇,即三类患者群体,分别由 A、B、C 表示。三类的中心分别为患者“卖火柴的女孩 01a”“寻_劫余”和“旅行的孤独风”。

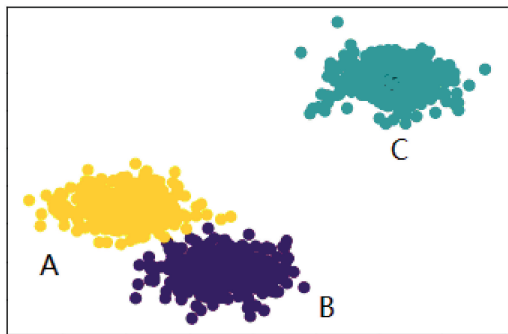


图4 患者画像聚类结果

以书籍《当下的力量》为例,给出阅读材料画像。该书的作者为埃克哈特·托利,图书类型为心灵读物,出版时间为2017年10月,该书的热度由当当网的销量136087册来表示,利用LDA主题模型对该书的摘要、引言、正文等进行主题抽取,得到宁静、狂喜、欢乐、成就、安全、力量、平和、摆脱、信念、信心、活力、青春、充实等13个主题词。利用TF-IDF算法对主题词进行加权运算,得到该书籍的主题是平和,其主题隶属度为0.671129312506041。适用人群集合基本上与抑郁成因集合相对应,可以表示为{1, 2, 3, 4, 5, 6},其中1表示因遗传因素引起的抑郁症人群,2表示因内分泌因素引起的抑郁症人群,3表示因压力因素引起的抑郁症人群,4表示因性格因素引起的抑郁症人群,5表示因日常琐事因素引起的抑郁症人群,6表示面向全体抑郁症人群。该书的阅读材料画像为{埃克哈特·托利, 心灵读物, 2017年10月, 6, 136087, 0.671129312506041},利用Word2vec算法将阅读材料画像中的非数值转化成数值,以便于后续匹配度计算。

将患者画像和阅读材料画像归纳映射为若干主题,利用Pearson相关系数计算患者与阅读材料之间的匹配度,将匹配度高的阅读材料推荐给患者。表3和表4分别给出了为患者“走饭”和患者群体A推荐的阅读材料列表。根据需要,将排名靠前的阅读材料推荐给患者和患者群体。

表3 为患者“走饭”推荐的阅读材料列表

阅读材料	匹配度
当下的力量	0.6188630487922147
思维改变生活	0.5480235894250735
走出抑郁	0.4984368045687428
永远成长的苹果树	0.4058891270569825

阅读材料	匹配度
其实没那么糟	0.3259097029842339
活下去的理由	0.2209450987415884
生命的重建	0.1509453690423887
.....	

表4 为患者群体A推荐的阅读材料列表

阅读材料	匹配度
构建生活	0.5953608745981209
当下的力量	0.5239975331480924
永远成长的苹果树	0.4628940388423981
新世界,灵魂的觉醒	0.4250992480634793
一颗找回自我的心	0.3983652300958123
发现未知的自己	0.2809644896314803
回到当下的旅程	0.1889607896320862
.....	

5 结语

抑郁症患者往往选择在微博分享他们的迷茫和痛苦。因此,本研究立足于阅读疗法,从微博“树洞”获取患者数据,建立患者画像和阅读材料画像,通过比较两者的匹配度,为患者提出适合的阅读材料。本研究不仅推动了阅读疗法的深入研究,而且丰富了阅读疗法的落地实践。后续研究将主要侧重于建立推荐结果的反馈系统,患者可以通过反馈系统提交推荐结果的反馈信息,该信息有助于进一步完善推荐模型的性能,也会对研判抑郁症的治疗效果具有一定的参考价值。

参考文献

- World Health Organization. Depression: a global crisis [EB/OL]. [2020-12-03]. http://www.who.int/mental_health/management/depression/wfmh_paper_depression_wmhd_2012.pdf.
- Friedrich M J. Depression is the leading cause of disability around the world[J]. JAMA, 2017, 317(15): 1517.
- Shen G Y, Jia J, Nie L Q, et al. Depression detection via harvesting social media: a multimodal dictionary learning solution[C]// Proceedings of the 20th International Joint Conference on Artificial Intelligence, San Francisco, USA, 2017: 3838-3844.
- 许维通. 吕氏春秋集释[M]. 北京: 中华书局, 2009: 245-246.
- 王波. 阅读疗法[M]. 北京: 海洋出版社, 2007: 12.
- McCulliss C. Bibliotherapy: historical and research perspectives [J]. Journal of Poetry Therapy, 2012, 25(1): 23-38.



- 7 Brewster E. An investigation of experiences of reading for mental health and well-being and their relation to models of bibliotherapy [D]. Sheffield: University of Sheffield, 2011.
- 8 王波.阅读疗法概念辨析[J].图书情报知识, 2005(1): 98-102.
- 9 王波.先秦到汉代的阅读疗法思想和医案[J].图书馆, 2004(6): 40-43.
- 10 王波,张慧君.魏晋南北朝的阅读疗法思想和医案[J].图书与情报, 2004(1): 42-47.
- 11 王波.明清时期的阅读疗法思想和医案[J].中国图书馆学报, 2004(4): 89-91.
- 12 王波.高校图书馆阅读推广中的新生教育书目[J].图书情报研究, 2015, 8(2): 3-15.
- 13 宫梅玲,于学美,冀宝苹,等.失恋引发的大学生抑郁障碍阅读疗法方分析及配伍[J].大学图书馆学报, 2016, 34(2): 80-87.
- 14 宫梅玲,楚存坤,张洪涛,等.阅读疗法立体化运作模式探究[J].大学图书馆学报, 2011, 29(5): 84-88.
- 15 林佳瑜.焦虑情绪的阅读疗法实证研究——以高校女职工育儿焦虑为例[J].图书情报工作, 2019, 63(12): 59-65.
- 16 潘悦.学龄前儿童情绪管理中的交互式阅读疗法实践[J].图书馆研究与工作, 2019(7): 47-52.
- 17 赵晓华,尹智龙,杜慰纯.基于自媒体的高校图书馆阅读疗法应用研究[J].图书馆, 2015(8): 76-79.
- 18 张怀涛,崔萌.学科融合环境下的阅读疗法[J].高校图书馆工作, 2020, 40(1): 70-76.
- 19 吴笑兰.我国高校图书馆“阅读疗法书目”建设[J].图书情报研究, 2010, 3(4): 29-35.
- 20 王景文,黄晓鹏,唐品,等.我国阅读疗法书目研究与实践进展评述——基于文献分析与书目建设的视角[J].图书馆杂志, 2015(4): 46-50.
- 21 宫梅玲,徐海军,雷菊霞,等.童年创伤引发的抑郁障碍阅读疗法方分析及配伍[J].大学图书馆学报, 2017(3): 36-45.
- 22 徐海军,郭洪霞.阅读疗法师如何读书和选书——以宫梅玲研制童年创伤引发的抑郁障碍配伍书方为例[J].图书馆论坛, 2018(11): 126-129.
- 23 田花蔓.高校图书馆阅读疗法书目研究[J].浙江交通职业技术学院学报, 2019, 20(2): 73-77.
- 24 解荣,周金元,杨桦.阅读疗法书目特征多样性分析研究[J].图书馆理论与实践, 2019(4): 1-5.
- 25 解荣,周金元,杨桦.读者视角下的阅读疗法书目评价研究及内涵揭示[J].图书情报研究, 2019(4): 52-57.
- 26 韩梅花,赵景秀.基于“用户画像”的阅读疗法模式研究:以抑郁症为例[J].大学图书馆学报, 2017, 35(6): 105-110.
- 27 刘丹,张兴刚,任淑敏.基于用户画像的高校图书馆阅读疗法模式[J].中华医学图书情报杂志, 2018, 27(7): 68-71.
- 28 辛云星.面向电子阅读疗法的青少年阅读推荐和导读方法研究[D].北京:清华大学, 2018.
- 29 董振东.知网情感分析用词语集[EB/OL]. [2020-12-03]. http://www.keenage.com/html/c_bulletin_2007.htm.
- 30 大连理工大学信息检索研究室.大连理工大学中文情感词库本体库[EB/OL]. [2020-12-03]. <https://www.iteye.com/resource/zhaohzeng-8248355>.

作者单位:北京语言大学语言智能研究院,北京,100083

收稿日期:2020年9月9日

修回日期:2020年9月29日

(责任编辑:关志英)

Research on the Reading Material Recommendation Model Based on the User Profile of Patients with Depression

Liu Zhongbao Zhao Wenjuan

Abstract: As an effective way to treat depression, bibliotherapy often requires professional staff with both library and psychological background to provide reading materials for patients, which is very complicated and time-consuming. In view of this, this paper takes the microblog information published by the patients with depression in the “tree hole” as the research subject, and constructs user profile of both the patients and the learning materials. The similarity of the two user profiles is compared to provide suitable reading materials for patients. The paper plays an important role in promoting the treatment of depression, and also provides a certain reference values for the research and practice of bibliotherapy.

Keywords: Depression; Patient Profile; Bibliotherapy; Reading Material; Recommendation Model