

基于“用户画像”的阅读疗法模式研究^{*}

——以抑郁症为例

□韩梅花 赵景秀

摘要 针对目前抑郁症阅读疗法施治对象的选取样本过于单一,私密性差,不能及时发现潜在患者,及时治疗的现状,提出了在大数据背景下基于“用户画像”的抑郁症阅读疗法新模式。该模式首先根据“伯恩斯抑郁症清单(BDC)”的内容来构造抑郁情绪的种子词,然后基于机器学习的方法,综合提炼用户网上行为和抑郁情绪的主观表露,构建用户抑郁情感词典。根据抑郁情感词典分析用户微博文本,计算其抑郁情感指数,得到“用户画像”,进而推送相应的阅读治疗资源。基于“用户画像”的阅读疗法模式大大突破了受众范围,便于准确把握诊断、治疗时机,使患者在无意识状态下接受阅读治疗,减轻了患者经济和精神的双重压力,具有较高的社会价值和重要的现实意义。

关键词 阅读疗法 大数据 抑郁症 “用户画像” 情绪分析

分类号 G252.17

DOI 10.16603/j.issn1002-1027.2017.06.013

1 引言

阅读疗法通常分为两类,第一类是发展阅读疗法,即阅读保健。只要我们承认人无完人,每个人性格上都有缺陷,那么阅读疗法就适用于所有人。读一读跟自己性格缺陷互补的书,对性格的发展、优化和完善会有帮助,性格好了,凡事理性面对、心平气和,乐观常在,忧戚不生,就会百病不侵。第二类是临床阅读疗法,即阅读治疗。就是当真患病了,医生在治病的过程中,配合药物和手术治疗,推荐一些书,辅助治疗和康复,那就叫临床阅读疗法^[1]。也就是说阅读疗法具有极为广泛的适应性,能够起到有病辅助治疗,无病愉悦身心的功效。

在大数据时代,如果能够利用无所不在的网络环境,根据用户在网上披露的情绪和行为,发现其不良情绪倾向,针对早期病症实施阅读疗法,对维护人们的心理健康,防止精神疾病的发生将大有裨益。

2 基于“用户画像”的抑郁症阅读疗法模式研究

2.1 基于“用户画像”的抑郁症阅读疗法模式

“用户画像”(User Profile),即用户信息标签化。是在大数据背景下,通过收集与分析用户的社会属性、生活习惯、情绪倾向、消费行为等信息之后,建立的精准的用户标签^[2]。

基于“用户画像”的抑郁症阅读疗法模式就是通过互联网获取用户在网络上的情绪表达,综合提炼用户网上行为和情绪极性的主观表露,利用用户无意识状态下有关认知、情绪和行为的关键词进行特征画像,计算其抑郁程度,然后针对症状推送阅读资料的阅读疗法新模式。相比传统的阅读疗法活动,通过互联网获取用户信息,进行“用户画像”和用于治疗的阅读资料智能推送,能够更加精准、快速、广泛地开展阅读治疗。

2.2 研究基于“用户画像”抑郁症阅读疗法模式的必要性和可行性

2.2.1 基于“用户画像”抑郁症阅读疗法模式的社会价值

根据世界卫生组织统计,预计到2020年,抑郁症可能成为仅次于心脏病的人类第二大疾患。而自

^{*} 2016年度校级哲学社会科学研究规划项目“新媒体环境下齐鲁传统文化经典藏读模式研究”(编号:XSK201615)的研究成果之一。

通讯作者:韩梅花,ORCID: 0000-0001-7066-381X, meihuahan69@163.com。

残、自杀行为是其最可怕的症状表现。据数据显示,当今我国抑郁症患者达到了3000万人,仅有10%的人接受过正规治疗^[3]。有九成抑郁症患者没有就诊,原因一是本人并不知道自已患上了抑郁症,再就是担心被周围人歧视,拒绝就医,还有就是迫于经济方面的压力难以就医。基于“用户画像”的抑郁症阅读疗法模式,通过识别分析用户隐匿于大数据中的负性情绪,进行“用户画像”,根据“用户画像”诊断用户的抑郁倾向,并针对抑郁症状,及时推送阅读材料,在潜移默化中纠正用户的认知,疏解用户的不良情绪,达到疗愈的目的。整个诊断治疗过程处于人机交互状态,自然性、私密性好,不会使患者产生畏惧和排斥心理。基于“用户画像”的阅读疗法的优势还在于,实现了受众范围的突破,能够使网络平台上的用户都有得到诊断、治疗的可能。特别是,创新了患者发现机制,能够在用户抑郁症状的初步表露中,及时捕捉到相关信息,通过“用户画像”把握诊断和治疗时机,而且诊断和治疗是在用户不自觉状态下进行的,能够有效疏解患者的抑郁情绪,减轻患者治疗的经济压力,患者通常会比较容易接受,治疗效果也会因此提高。

2.2.2 基于“用户画像”的抑郁症阅读疗法模式的可行性

理论上的可行性:每个人都有自己相对固定的认知图式。所谓图式,是人脑中已有的知识经验的网络。社会知觉的基础是被认知事物本身的属性,但认知者的主观因素也会对社会知觉的过程和结果产生重要影响^[4]。抑郁症患者经历较多的负性生活事件,加之敏感、过于内向等人格特点,使他们较易形成负性认知图式,在对外界信息进行编码时,也更多采用负性图式^[5]。抑郁症个体或抑郁症易感人群长期存在的这种负性图式,使他们习惯于以一种消极的、歪曲的方式选择性加工、编码、解释输入的信息,而缺乏对事件具体细节的认知,致使抑郁个体出现对信息的负性解释、选择性注意负性事件、特异性选择负性提取线索等情况。这种偏重负性信息的指向性会在对自身、他人、事物和社会的认知中无意识地流露。随着网络的普及,越来越多的用户在社交网络上分享自己的观点或体验,有意或无意地展露个人社会行为和内心体验。通过捕捉其在社交网络中关于认知、情绪及行为的描述中经常出现的负性词汇,就可以进行“用户画像”,计算其抑郁程度。比

如,在2016年2月份自杀的上海某高校教师的新浪微博中,短期内出现了“焦虑”“罪”“错”“辞别”“扭曲”“孤寂”“忧伤”“冲突”“亡灵”“沉滞”“惧怕”“压抑”“责罚”“平庸”“邪恶”“恐惧”“太累”“死”“黑暗”“笼罩”“没办法”“怕”“凄凉”“孤独”“自责”“失望”“痛苦”“惊惧”等大量的抑郁情感词汇,表明他的抑郁情绪积累浓重,具有显著的抑郁倾向,通过“用户画像”技术,完全能够在线分析出他的这种情绪变化,通过及时向他推送有针对性的阅读资料,持续进行干预治疗,疏解其内心的抑郁情绪,避免其自杀悲剧的发生是有可能的。

技术上的可行性:大数据技术和机器学习技术的发展,使得利用大数据分析技术,综合提取用户的网上行为以及情感极性词的无意识流露,对其进行画像,并根据画像情况进行抑郁症发生发展的预测和诊断,具有了技术上的现实可行性。同样利用现有的信息推送技术,由相关专业人员组成研究团队对阅读治疗抑郁症的文献信息、临床验方进行分析研究,找出与用户相关性最大的治疗方案,有针对性地推送给用户也是切实可行的。在这个过程中,将认知计算机科技、阅读治疗研究成果与循证治疗、大数据结合起来,能够在高效诊疗、用户认知关怀和用户健康管理等方面提供独特的价值。

伦理、法规的许可性:网络用户在微博上发表的内容是面向大众的公开表述,不是在私人空间中的私密信息,分析这些公开信息并不侵犯其隐私;根据分析结果,向表露出抑郁情绪的用户推送疏解其不良情绪的阅读治疗资料,是一种善意的行为,也不会构成对用户隐私的侵犯。各种合法的数据平台,也都有对用户隐私保护的相关承诺,例如新浪微博平台“微博服务使用协议”的第1.7款:“为提高用户的微博服务使用感受和满意度,用户同意无偿授权微博平台以任何形式(商业或非商业)将基于用户的操作行为对用户数据进行调查研究和分析,从而进一步优化微博服务。”^[6]第4.8款:“用户同意微博平台在提供微博服务过程中以各种方式投放各种商业性广告或其他任何类型的商业信息(包括但不限于在微博平台网站的任何页面上投放广告),并且,用户同意接受微博平台通过电子邮件、私信或其他方式向用户发送商品促销或其他相关商业信息。”^[7]这些条款都明确了平台使用用户数据的相关约定。而第6款,则专门阐述其隐私保护政策。总之,基于大

数据技术对用户情绪表露进行分析并根据分析结果向其推送有益于身心健康的阅读治疗资料,在伦理、法规层面也是可行的。

3 基于“用户画像”的抑郁症阅读疗法模式的构建

基于“用户画像”的抑郁症阅读疗法主要涉及以下5个环节:

(1)构建抑郁情感词典:建立一个描述抑郁情绪的词汇集合,每个词汇都有一个体现抑郁情绪强度的权重。首先根据“伯恩斯抑郁症清单(BDC)”的内容来构造抑郁情绪的种子词,通过 TF-IDF 方法或者聚类方法,扩充种子词,形成能刻画抑郁情绪的抑郁情感词典。

(2)获取用户情感词词袋:读取用户博文,使用 Jieba 分词工具,对用户博文进行中文分词,将用户博文切分成词汇,形成词袋。

(3)提取文本特征:遍历词袋,如果情感词典中有词袋中的词汇,则记录该词及其权重,得到文本的抑郁情绪特征向量。

(4)“用户画像”:对(3)得到的用户博文的抑郁情绪特征向量的各个分量加权求和,求得其抑郁情感指数。该指数如果大于某一设定的阈值,则用户具有抑郁情绪,完成“用户画像”。

(5)推送治疗:根据(4)中计算出的用户抑郁指数的大小,选择具有治疗作用的阅读资料,向用户推送。

3.1 构建抑郁情感词典

情感词典是对文本进行情感分析的基础。人们在文本表达中所选择的词汇,不仅能反映出其性别、年龄等人口学特征,还能折射其动机、人格和社会地位等心理或社会特质,每个词汇都体现出表达者的情感定向,不可避免地带有表达者的主观倾向,体现出表达者当时的情绪心境,是研究文本情感和主观感受的重要线索。目前已经开发出了一些汉语情感词典,比如《台湾大学情感词典》《中文网络情绪词典》和《中文基础情感词词典》等。借助于这些已有的情感词典,国内研究者已经成功地针对网络在线文本进行了多项情感分析研究,如对网络评论的语义倾向性分析研究^[8],对宾馆评论的情感分析研究^[9],对新闻媒体褒贬值的计算并据此预测股市的波动研究^[10],对网上产品评价信息的分析研究等^[11]。但是由于所构建的这些情感词库多是正负

性的简单二分,缺乏对复杂多样的人类情绪现象的深入洞察,尤其没有专门用于分析人类抑郁情绪的词典。因此,构建一个基于网络博文的抑郁情感词典是必不可少的,该情感词典应结合情绪结构理论或情绪心理学的研究成果,通过一段时间内累积的情绪能推断其主观感受体现出的抑郁情绪状况,快速识别出患者文本中所表达的情绪。

基于“用户画像”的抑郁症阅读疗法是通过分析网络用户在网上的行为和网络博文中的情感倾向,建立用户的抑郁情感指数,根据抑郁情感指数形成“用户画像”,也即分析用户的网络情绪。如何建立正确的“用户画像”,构建合理的抑郁情感词典是其中的最重要步骤。

抑郁情感词典是带有抑郁情感色彩的词或词组的集合,这些词可以是形容词,也可以是副词、名词或者动词,是进行抑郁情感分析的基础。抑郁情感词典的建设可以通过先确立抑郁种子词,再使用人工手动添加和机器学习获得新抑郁情感词结合的方法,对种子词进行扩充,最终形成具有相当规模的抑郁情感词典。

3.1.1 确立抑郁情感词典种子词

首先确立体现抑郁情绪的种子词。由于我国目前没有这方面的明确标准,可参考美国新一代心理治疗专家、宾夕法尼亚大学的伯恩斯(David D·Burns)博士设计的抑郁症自我诊断表“伯恩斯抑郁症清单(BDC)^[12]”的内容来构造抑郁情绪的种子词。这份清单共有15项,具体内容如下:

- (1)悲伤:你是否一直感到伤心或悲哀?
- (2)泄气:你是否感到前景渺茫?
- (3)缺乏自尊:你是否觉得自己没有价值或自以为是失败者?
- (4)自卑:你是否觉得力不从心或自叹比不上别人?
- (5)内疚:你是否对任何事都自责?
- (6)犹豫:你是否在做决定时犹豫不决?
- (7)焦躁不安:这段时间你是否一直处于愤怒和不满状态?
- (8)对生活丧失兴趣:你对事业、家庭、爱好或朋友是否丧失了兴趣?
- (9)丧失动机:你是否感到一蹶不振做事情毫无动力?
- (10)自我印象可怜:你是否以为自己已衰老或

失去魅力?

(11)食欲变化:你是否感到食欲不振或情不自禁地暴饮暴食?

(12)睡眠变化:你是否患有失眠症或整天感到体力不支、昏昏欲睡?

(13)丧失性欲:你是否丧失了对性的兴趣?

(14)臆想症:你是否经常担心自己的健康?

(15)自杀冲动:你是否认为生存没有价值,或生不如死?

围绕这 15 个方面的内容,同时考虑在国内大多数抑郁症患者感受到的主要不是情绪方面的症状,而是头痛、头晕、腹胀、心悸、身体疼痛等躯体症状,初步构建抑郁情绪的种子词,如:“酸痛”“头疼”“没有食欲”“头昏脑胀”“胸闷”“心慌”“真累”“没劲”“无聊”“死”“废物”“无能”“烦心”“郁闷”“痛苦”“内疚”“嫌弃”“崩溃”“无聊”“焦虑”等。

3.1.2 基于机器学习构建抑郁情感词典

人工建立情绪词库费时费力,所含的词汇较少,而且存在评分者偏差。随着计算机技术的发展,利用某种资源,通过机器学习的方法挖掘文本中的情绪特征,自动或半自动化地建立词库已成为主流方法。通过对已知抑郁症病人的文字表达资料(日志、博客等)使用 TF-IDF(词频-逆向文档频率)方法进行统计分析,得出资料中重要的高频词,然后分析这些高频词和种子词的同义或反义关系,根据词语的语义相似度计算词语的抑郁情感权值,获得新的抑郁情感词,扩充种子情感词,形成抑郁情感词典。

TF-IDF 是一种统计方法,用以评估一字词对于一个文件集或一个语料库中的其中一份文件的重要程度。其中词频(Term Frequency, TF)指的是某一个给定的词语 t_i 在文件 d_i 中出现的频率,也叫单文本词频。第 i 个词汇在第 j 个文件中的词频 $tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$ 。其中, $n_{i,j}$ 为词语 t_i 在文件 d_i 中的出现次数, $n_{k,j}$ 为文件 d_i 中第 k 个词出现的次数, $\sum_k n_{k,j}$ 则为文件 d_i 中所有词的出现次数之和。

逆文档频率(Inverse Document Frequency, IDF) $idf_i = \log \frac{|D|}{|\{j:t_i \in d_j\}|+1}$,

其中, $|D|$ 为文档总数, $|\{j:t_i \in d_j\}|$ 为包含词语 t_i 的文档数目。如果该词语不在语料库中,就会导致被除数为零,因此一般情况下使用 $|\{j:t_i \in d_j\}|$

+1。亦即逆文档频率是由总文档数除以包含该词语的文档数,再将得到的商取对数得到的一个量值,它是一个词语普遍重要性的度量,是表征这个词汇的重要性的权值。

$tf-idf_{i,j} = tf_{i,j} \times idf_{i,j}$,某一特定文档内的高词语频率,以及该词语在整个文档集合中的低文档频率,可以产生出高权重的 TF-IDF。因此,TF-IDF 倾向于过滤掉常见的词语,保留重要的词语。这种方法的主要思想是如果某个词或短语在一篇文章中出现的频率 TF 高,并且在其他文章中很少出现,则认为此词或者短语具有很好的类别区分能力,适合用来分类,也就可以作为情感词。

计算出文档中每个词的 TF-IDF 值之后,按照大小进行排序,选取其中值最高的几个作为候选词。将得到的候选词与基础情感词典逐个扫描,判断是否为已有情感词,如果属于已有情感词则结束,否则将候选词加入到词典中。

聚类分析是一种自动、快速实现文本信息挖掘的有效方式。通过对抑郁情感词聚类,可以实现抑郁情感词的自动识别,自动获取语料中的抑郁情感向量作为其情感特征。在训练文本集中,采用如下的半监督聚类算法获得用种子词表示的词类:

(1)分词:对训练文本集中的文本进行分词,去掉停用词及连词、代词、冠词等,得到分词集 $U = \{wd_1, wd_2, \dots, wd_n\}$ 。

(2)指定抑郁情绪种子词集 $V = \{ewd_1, ewd_2, \dots, ewd_n\}$,其中 $ewd_i (i=1, \dots, n)$ 是体现抑郁情绪较重的情感词。

(3)聚类:对种子词集 V 中的每一个种子词 $ewd_i (i=1, \dots, n)$,分别计算它与分词集 U 中每一个词 $ewd_i (i=1, \dots, n)$ 的相关指数 $Cor(ewd_i, wd_j)$,如果 $Cor(ewd_i, wd_j)$ 大于某给定的阈值,则将词 wd_j 归入种子词 ewd_i 表示的词类,否则继续计算种子词 ewd_i 与 U 中下一个词的相关指数。直到 V 中所有种子词都计算完毕,得到由种子词表示的词类。对相关指数高的,给予较高的权值,低的给予较低的权值。词的相关度由互信息(Mutual Information)度量^[13]:

$$Cor(ewd_i, wd_j) = \frac{n(\alpha\lambda - \mu\beta)^2}{(\alpha + \mu)(\beta + \lambda)(\alpha + \beta)(\mu + \lambda)}$$

其中, n 表示训练语料中文本总数, α 表示 ewd_i 出现且 wd_j 也出现的文本频数, β 表示 ewd_i 出现但 wd_j

不出现的文本频数, μ 表示 wd_j 出现但 ewd_i 不出现的文本频数, λ 是 ewd_i 与 wd_j 均不出现的文本数。

(4)词类权重计算:以词类所含信息量作为该词类权重,并用种子词权重表示该词类权重。种子词 ewd_i 的权重按式(2)计算:

$$\rho(ewd_i) = \sum p(wd_{ij}) \log p(wd_{ij})$$

其中 $\rho(ewd_i)$ 表示种子词 ewd_i 的权重, $p(wd_{ij})$ 表示种子词 ewd_i 的第 j 个元素的概率分布。 $\rho(ewd_i)$ 值越大,表明它们之间的相关度越大。设定权重阈值,选取权重大于阈值的词作为候选词,将得到的候选词与基础情感词典逐个扫描,判断是否为已有情感词,如果属于已有情感词则结束,否则将候选词加入到词典中。

3.2 根据词典分析用户抑郁情绪——“用户画像”

随着Web2.0的发展,越来越多的用户乐于在互联网上分享自己的观点或体验,使得社交网络成为人们生活中不可或缺的一部分。人们在社交网络上的种种行为特征,体现了其在现实生活中的行为、思想、运动轨迹等。在社交网络产生的海量数据中,感情失意、生活不顺、事业不顺等用户情绪往往有着丰富的表达。比如中国最大的社交网络平台——微博平台中的女性用户,她们热衷于在微博平台上分享个人生活,宣泄生活中的各种情绪表达。这些情绪表达有的是文字,有的是表情符号。表情符号可以作为重要的情绪线索,来提高文本情绪分类问题的精度^[14-18]。为了研究分析人们的抑郁情绪表达,我们可以通过网络爬虫获取某一用户的微博文本,利用Jieba分词工具,将文本分词,去除停用词和“@、V、#、http://”等与情感无关的字符,得到词袋。遍历词袋,如果情感词典中有词袋中的词汇,则记录该词及其权重,形成文本抑郁情绪特征向量。最后对该特征向量的各分量作加权求和运算,获得该用户的抑郁情绪指数,形成该用户的“用户画像”。

3.3 根据“用户画像”自动推送阅读治疗的资源

中医认为,利用“情志相胜”原理选取适当的图书阅读,能够达到疗愈功效。在引导读者情绪方面,几乎所有的书刊都具有一定的作用,而心理咨询、人生哲理、小说等类书刊的作用较大。人的心理系统通过阅读的方式保持适度的开放,可以保持和外界的信息交流,让内心的能量流动循环起来,减少迟滞感造成的抑郁情绪。社交媒体同样也能够帮助用户对负面情绪进行调节。研究发现社交媒体的一些功

能,例如用户间沟通、信息搜索以及娱乐活动,都能对用户的情绪产生积极影响^[19],并且用户从其朋友处获得的积极反馈,能够协助他们构筑自尊心以及社交能力^[20]。当我们在微博平台分析出具有抑郁情感的用户,完成“用户画像”后,可以根据其抑郁情感指数的高低,向其推送具有治疗功能的文字资料,用户通过阅读这些内容,在心理上能够产生:共鸣—认同—净化—领悟这么一系列的递进心理过程,使自身的不良情绪得以宣泄,心理郁结得以疏通,认知水平得以提高,从而达到一定的治疗效果。

4 结论

本文提出的大数据背景下基于“用户画像”的阅读疗法,一方面在实践上为阅读疗法开创了一个与时俱进的新方向。通过提取用户无意识状态下有关认知、情绪和行为的关键词进行特征画像,能够客观、准确地反应用户的抑郁症状,进而针对症状制定相应的阅读治疗方案。“用户画像”还提前了发现患者的时间,能够在患者初步表现出抑郁症状时,就把握住治疗时机,契合了中医上的治未病原理,会收到有病治病、无病防病的效果。另一方面,基于大数据的“用户画像”模式阅读疗法实践,必然会进一步丰富阅读疗法的新理论,使阅读疗法研究随着时代的进步而获得更深入地发展。

参考文献

- 1 王波.图书馆阅读推广的定义、类型、方法——在“图书馆阅读推广理论与实践”专题研讨会上的演讲[J].上海高校图书情报工作研究,2017,27(1):6-19.
- 2 余孟杰.产品研发中“用户画像”的数据建模——从具象到抽象[J].设计艺术研究,2014(6):62-64.
- 3 抑郁症:“精神感冒”如何破解?.[2017-05-23].<http://news.0898.net/health/n/2014/0523/c231201-21273659.html>
- 4 百度百科.[2017-05-07].<http://baike.baidu.com/item/%E5%9B%BE%E5%BC%8F>.
- 5 刘庆英,冯正直.抑郁症的整合情绪记忆模型述评[J].心理科学进展,2009,17(5):938-943.
- 6 微博服务使用协议.[2017-06-06].<http://weibo.com/signup/v5/protocol>.
- 7 同6.
- 8 丁建立,慈祥,黄剑雄.网络评论倾向性分析[J].计算机应用,2010,30(11):2937-2940.
- 9 杨鼎,阮爱民.一种基于情感词典和朴素贝叶斯的中文文本情感分类方法[J].计算机应用研究,2010,27(10):3737-3743.
- 10 王超,李楠,李欣丽,梁循.倾向性分析用于金融市场波动率的研

- 究[J]. 中文信息学报, 2009, 23(1): 95 - 99.
- 11 杨超, 冯时, 王大玲, 杨楠, 于戈. 基于情感词典扩展技术的网络舆情倾向性分析[J]. 小型微型计算机系统, 2010, 31(4): 691 - 695.
 - 12 [美]大卫·伯恩斯. 伯恩斯新情绪疗法[M]. 李安龙 哈尔滨: 北方文艺出版社 2007: 1-4.
 - 13 Yang ZR, M. Zwolinski. Mutual Information Theory for Adaptive Mixture Models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(4): 396 - 403.
 - 14 Hu X, Tang J, Gao H, et al. Unsupervised sentiment analysis with emotional signals[C]. //Proceedings of the 22nd international conference on World Wide Web. Brazil: ACM, 2013: 607 - 618.
 - 15 Liu K L, Li W J, Guo M. Emoticon Smoothed Language Models for Twitter Sentiment Analysis [C]. //AAAI. Canada: AAAI Press, 2012.
 - 16 Liu S, Li F, Li F, et al. Adaptive co-training SVM for sentiment classification on tweets[C]. //Proceedings of the 22nd ACM international conference on Information & Knowledge Management. USA. ACM, 2013: 2079 - 2088.
 - 17 Ptaszynski M, Rzepka R, Araki K, et al. Research on emotions: review of the field and proposal of research framework [J]. Proceedings of 17th Association for Natural Language Processing, 2011: 1159 - 1162.
 - 18 Purver M, Battersby S. Experimenting with distant supervision for emotion classification[C]. //Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics. France: ACM, 2012: 482 - 491.
 - 19 Apaolaza V, He J, Hartmann P. The effect of gratifications derived from use of the social networking site Qzone on Chinese adolescents' positive mood[J]. Computer in Human Behavior, 2014, 41: 203 - 211.
 - 20 Valkenburg P M, Peter J, Schouten A P. Friend networking sites and their relationship to adolescents' well-being and social self-esteem[J]. CyberPsychology & Behavior, 2006, 9(5): 584 - 590.

作者单位: 曲阜师范大学日照校区图书馆, 日照, 276826

曲阜师范大学信息科学与工程学院, 日照, 276826

收稿日期: 2017年6月6日

Research on Bibliotherapy Model Based on User Profile

——Take Depression as an Example

Han Meihua Zhao Jingxiu

Abstract: In this paper, a new mode of bibliotherapy for depression based on user profile is put forward in view of the current situation of depression bibliotherapy with weakness such as samples selected too simple, poor in privacy, and unable to detect and treat potential patients in a timely manner. This mode first constructed the seed words of depressive mood according to “The Burns Depression Checklist (BDC)”, and then based on the machine learning method, comprehensively refined the subjective revealing of user's online behavior and depressive mood to construct the user's depressed emotional dictionary. According to the depressed emotional dictionary to analyze user's blog, calculate the depression emotional index, get the user profile, and then push the corresponding bibliotherapy resources. The bibliotherapy mode based on the user profile exceeded the audience scope, accurately grasp the timing of diagnosis and treatment, so that patients could accept bibliotherapy in an unconscious state, so as to reduce the patient's double pressure of economic and spirit.

Keywords: Bibliotherapy; Big Data; Depression; User Profile; Emotional Analysis