

社交网络用户认知域特征预测研究综述

郑敬华¹ 郭世泽² 高 梁² 钟晓峰¹

(电子工程学院 合肥 230037)¹ (北方电子设备研究所 北京 100083)²

摘 要 网络空间认知域安全是网络空间安全的重中之重,而认知域特征预测是研究网络空间认知域安全的基础。首先,明确了社交网络用户个体认知域特征预测在网络空间认知域安全中的地位;接着,从预测流程、特征选取和模型建立3个方面综述了国内外在社交网络用户认知域特征预测方面的研究,并针对国内典型社交网络用户样本特征,指出了研究中存在的问题,进而提出一些可能的研究思路和方法;最后,总结了该领域当前面临的挑战与存在的不足,以及有待重点研究的相关问题。

关键词 社交网络,认知域特征,预测,网络空间安全

中图分类号 TP391 文献标识码 A DOI 10.11896/j.issn.1002-137X.2018.03.003

Survey on Cognitive Domain Feature Prediction of Social Network Users

ZHENG Jing-hua¹ GUO Shi-ze² GAO Liang² ZHONG Xiao-feng¹

(Electronic Engineering Institute, Hefei 230037, China)¹

(Institute of North Electronic Equipment, Beijing 100083, China)²

Abstract Cognitive domain security is of primary importance in cyber space security, and cognitive domain characteristic prediction is the basis of researching on cognitive domain security. This paper showed a kind of cyber space cognitive domain safety model. Then, this paper is clear that cognitive domain characteristic prediction of social network users has a vital role in cyber space cognitive domain security. It summarized home and abroad researches about the prediction of social network users from three aspects: prediction process, characteristics choosing and model building. Aiming at characteristics of domestic typical social network users samples, this paper pointed out the problems existing in researches, and showed some possible thoughts and methods. At last, this paper summed up current challenges and shortages in this field including some related problems to be solved as stress.

Keywords Social network, Cognitive domain feature, Prediction, Cyberspace security

网络空间横跨物理域、信息域和认知域。物理域由实体网络组成,支撑着整个网络空间的运行;信息域包含各种网络业务,是网络空间运行的驱动力;认知域是人类认知活动所涉及的范围和领域,反映网络对人的情感、意志、知识和信念的影响。传统的网络空间安全主要涉及物理域安全和信息域安全,但决定网络空间甚至国家安全的最主要因素仍然是“人”。随着社交网络的普及以及移动终端的智能化,人们对网络的依赖越来越强,与此同时,网络对人的情感、意志、知识和信念的控制和影响也越发凸显。为了确保网络空间的安全,实现网络空间认知域的安全已成为重中之重。

网络空间认知域的主体是用户。为了保护用户的认知域安全,应预测用户的认知域特征,分析其存在的认知域缺陷并针对性地实施认知域安全防护措施。其中,用户认知域特征预测最为重要,是进行网络空间认知域安全分析的基础。

根据不同的分类标准,可将认知域特征分为多种类型。

本文中网络用户的认知域特征预测主要是针对社交网络用户,包括预测社交网络用户的身份属性、人格特征以及心理健康等。身份属性包括用户的性别、年龄、种族、宗教、政治倾向等;人格是描述一个人在不同时间、不同情境中能够保持一致的行为方式的一种倾向,一般包括专注性、情感的表达能力、安全感、对新事物新环境的适应能力以及客观性等;心理健康是人的健康不可分割的重要方面,是一种持续的心理情况,描述个体在各种环境中的一种高效且满意的心理状态,包括情绪健康、意志健全及行为协调等多个方面的内容。

网络空间认知域涉及的是人的认识领域,反映的是人的思维活动,人的思维活动又可以通过人的心理来体现。而人的心理由内因和外因两种因素决定,内因可以通过人格因素来区分和衡量,不同的人格决定了人的不同行为;外因可以通过人的身份属性来区分和衡量,身份属性是人的心理活动最核心的内容,个人对自己身份的认定能够直接改变其价值体系。

到稿日期:2016-12-17 返修日期:2017-05-19 本文受省部级重大项目(AWS13J003)资助。

郑敬华(1976—),女,博士,讲师,主要研究方向为机器学习、信息安全,E-mail:zhengjh1001@163.com(通信作者);郭世泽(1969—),男,高级工程师,主要研究方向为网络空间安全;高 梁(1982—),男,助理研究员,主要研究方向为网络空间安全;钟晓峰(1981—),男,讲师,主要研究方向为信息安全、机器学习。

因此,依据社交网络数据预测用户的人格和身份属性,进而分析其心理健康状况,能够推断用户向某种认知缺陷方面发展的可能性,即其存在的认知域缺陷。基于认知域缺陷,可以有针对性地对其进行认知干预和引导,从而达到保护网络空间认知域安全的目的。

本文第1—3节主要针对社交网络用户的身份属性、人格以及心理健康等认知域特征的预测,分别从预测流程、特征选取和模型建立这3个方面综述了国内外的研究进展、研究思路和方法;同时,针对新浪微博用户数据的特点,分析目前研究中可能存在的问题和不足,并提出一些新的研究思路和方法。最后总结了该领域存在的不足与面临的挑战,以及有待重点研究的相关问题。

1 社交网络用户身份属性预测

Kosinski等^[1]的研究表明,社交网络中的信息可以用来预测用户的性别、年龄及性取向等身份属性。目前,随着社交网络在人们生活中的广泛应用,对用户身份属性的判别也由以前的基于人脸、指纹等生理特征的检测拓展到通过社交网络数据进行判别,这样不仅可以进行更实时、精确的预测,还大大降低了设备使用率。

1.1 预测流程

目前,在线社交网络用户的身份属性预测主要通过3类数据进行,分别是动态关系数据、静态状态数据和操作行为数据。

在通过社交网络用户动态关系数据预测用户身份属性方面,Kosinski等^[2]通过社交网络用户的好友连接关系来推测用户的身份属性,认为相同属性的用户更有可能成为朋友或者形成一个社区团体,最后他们利用社团发现方法对用户身份属性进行了识别。He等^[3]通过贝叶斯网络创建社交网络用户因果关系模型,并证明利用社交网络用户的朋友连接关系能够推断用户隐藏的身份信息。Lindamood等^[4]通过用户的节点信息以及节点之间的连接信息推测社交网络用户的政治倾向,并通过创建贝叶斯分类器进行预测建模。Wu等^[5]通过不同的社交关系来推测用户的身份属性,证明同学关系可被用来推测用户的年龄,好友关系可被用来推测用户的性别等。Zheleva等^[6]通过社交网络的好友关系和组关系,将社交网络描述为图,然后通过创建敏感信息预测模型对Facebook用户的性别进行预测,最终的准确率约为73.4%。

在通过社交网络静态关系数据(如微博信息、粉丝数、收藏数、点赞数、网站浏览日志等)来预测用户的身份属性方面,Kosinski等^[1]通过实验证明,根据社交网络用户的网站浏览记录可以预测用户的年龄、性别、职业、教育水平、政治倾向、种族观点以及性取向等身份属性。Bin等^[7]通过Facebook用户的Likes数,利用ODP分类算法来推断用户的人口统计特征,包括性别、年龄、宗教及政治观点。

在通过操作行为数据预测用户的身份属性方面,西安交通大学的蔡忠闵团队^[8]率先进行了探索性研究。他们首先记录用户鼠标操作、击键、触摸等产生的操作行为数据,然后利用这些数据来提取行为特征,进而建立用户的身份属性预测模型,对用户的性别、年龄、种族等身份属性进行检测和判别。

现有研究结果表明,根据动态关系数据预测用户身份属性时,预测准确率相对较高,但数据结构复杂,计算量大;根据静态状态数据预测用户身份属性时,建模过程相对简单,但一般需要大量的标签数据,且预测准确率不高;根据操作行为数据预测用户身份属性时,准确率较高,但只适合于对少数几类身份属性进行预测,如性别、年龄、左右手习惯等。

对已有的研究过程进行归纳总结,得到如图1所示的社交网络用户身份属性预测的基本流程:1)获取用户社交网络数据,可通过网站授权API接口或通过网络爬虫获取;2)对数据进行统计分析,并根据数据分布确定可用于预测的数据;3)对数据属性进行分类和预处理,包括基于先验知识的类别划分以及必要的数据格式转换、数据归一化、缺失数据处理等预处理;4)提取数据特征,借助相应的机器学习算法构建身份属性预测模型,一般采用监督算法、半监督算法进行建模;5)验证模型并进行模型修正,通常采用预测相关性指标、预测误差指标和分类准确性指标进行验证。

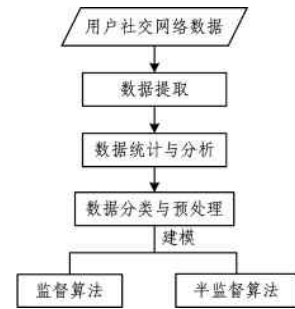


图1 基于社交网络的用户身份属性预测流程

Fig.1 Prediction process of identity attribute for social network

1.2 特征提取

特征提取是建立预测模型的关键步骤,目的是选取大多数用户都存在的并且不同类型用户表现不同的数据特征^[9],同时尽量寻找能够代表不同类型社交网站的特征。

关于特征提取,目前尚没有系统的论述,各研究者从不同的角度对其进行描述。Kosinski等^[1]通过搜集用户的Facebook Likes数,建立用户和Likes数矩阵,然后通过奇异值分解达到降维和特征提取的目的。丁宇新等^[10]通过计算用户特征的信息增益比率,分析特征的分类能力,并选择分类能力强的特征。信息增益比率的定义为:

$$GainRation(S,A) = \frac{Gain(S,A)}{SpilitInformation(S,A)} \quad (1)$$

其中:

$$Gain(S,A) = entropy(S,A) - \sum_{v \in values(A)} \frac{|S_v|}{|S|} entropy(S_v) \quad (2)$$

$$SpiliInformation(S,A) = \sum_{v \in values(A)} \frac{|S_v|}{|S|} lb \frac{|S_v|}{|S|} \quad (3)$$

其中, A 表示特征集合, S 表示特征取值的所有样本集, S_v 表示特征 A 取值为 v 的所有样本的样本集。

由于通过社交网络数据对用户进行身份属性预测的研究还处于初始阶段,因此目前的研究结果中,选取的特征主要涉及两类:动态关系和静态状态。其中,动态关系具体表现为用

户好友关系,静态状态具体表现为 Facebook Likes 数。而对于众多的社交网络特征,如语言特征、人格特征,还没有涉及,因此在特征提取方面还有很大的研究空间。

1.3 模型建立

目前主要利用机器学习中的分类算法来实现用户身份属

性的预测,主要有朴素贝叶斯分类器^[4]、决策树分类器^[8]、逻辑回归^[1]、线性回归^[1]、对数线性模型^[11]、神经网络、随机森林以及由多种方法联合实现的分类器^[12]等。表 1 列出了这些算法所使用的平台数据和数据类型,并给出了具体的预测属性和预测结果。

表 1 身份属性预测研究中使用的算法

Table 1 Algorithms used in identity attribute prediction

文献	平台	算法	数据类型	身份属性	预测结果
[1]	Facebook	线性,逻辑回归	Likes 数	性别,年龄	AUC:性别 0.87,年龄 0.7
[8]	封闭模拟平台	决策树,多种分类器组合	鼠标的单击、双击、偏移	性别,年龄,种族,左右手	误检率:8.74%,漏检率:7.69%
[7]	Facebook	贝叶斯分类	Likes 数	性别,年龄,宗教,政治观点	AUC:年龄:77%,性别:84%
[10]	人人网	基于图的半监督分类算法	好友关系,组关系	兴趣,毕业院校	准确率:85.6%
[11]	Twitter	Log-linear models	tweet	性别,年龄,情绪等	预测概率提高 0.04~0.14
[13]	新浪微博	基于位置的方法,多数投票方法	地理位置,社交关系	地址,学校,工作单位等	误差<3.2%

此外,有少数研究者试图通过半监督学习算法对社交网络用户进行身份预测,如丁宇新等^[10]采用基于图的半监督学习方法,利用在线社交网络用户的个人信息、交友信息和组群关系进行用户属性的预测。最近,也有研究者通过微博的位置信息、社交关系等信息推测微博用户的身份属性,如胡开先等^[13]提出一种基于位置和社交关系的社交网络用户身份识别方法,并通过两种方法的融合、互补的方式推测用户的地址信息、学校、工作单位和兴趣等身份特征,最后采用新浪微博数据验证了方法的有效性。

1.4 研究小结

目前,通过社交网络预测用户身份属性的研究,仍然集中在建立社交网络某一特征数据与某一身份属性之间的关系上,并且采用的是单任务有监督的机器学习算法^[1,7]。然而,现实中获取网络用户身份属性标签数据存在一定的困难,需要耗费大量的人力、物力和财力,并且无法保证大部分网络用户填写信息的真实性。

针对标签数据不足的情况,可以采用半监督学习方法。半监督学习用大量的无标注数据来改进监督学习,训练更好的分类器或回归函数,设法获取具有泛化能力和良好性能的学习器。丁宇新等^[10]采用基于图的半监督学习方法预测用户的兴趣以及毕业院校等信息,但其仅仅是通过用户之间的关系进行预测,因此半监督学习方法在基于社交网络预测用户的身份属性方面还有很大的研究空间。

此外,社交网络用户的身份属性众多,包括性别、年龄、种族、政治倾向等,并且很多任务之间具有相关关系,因此可考虑结合多任务学习和半监督学习方法搭建一种新的学习框架,以便更好地构建身份属性特征分类器。

2 社交网络用户的人格特征预测

目前,大五人格模型(Big-Five Model)是心理学界公认的最可靠、最主流的人格模型,因此对社交网络用户进行人格预测时,大部分都采用了大五人格模型^[14]。

大五人格模型把人格划分为 5 种类型,分别是开放性(Openness)、外向性(Extraversion)、神经质(Neuroticism)、责任感(Conscientiousness)和宜人性(Agreeableness)。开放性描述个体的认知风格,高分个体具有想象、审美、情感丰富、创造、智慧等特性;外向性表示人际互动的数量和密度、对刺激

的需要以及获得愉悦的能力,高分个体表现出热情、果断、活跃、冒险、乐观等;神经质也叫情绪稳定性,反映个体的情感调节过程,高分个体倾向于焦虑、敌对、压抑、自我意识、冲动、脆弱等;责任感指个体控制、管理和调节自身冲动的方式,高分个体具有公正、条理、尽职、成就、自律、谨慎、克制等特性;宜人性描述的是个体对他人的态度,包括亲近以及敌对,高分个体表现出信任、利他、直率、谦虚等特性。为便于计算机进行计算,大五人格模型一般表示成(O,E,N,C,A) 5 个维度。

2.1 预测流程

目前,已经有许多学者证明利用社交网络数据进行用户人格预测是可行的。其中,Schrammel 等^[15]通过发放调查问卷的形式得出,不同人格的人在社交网络上的表现和使用规律是不同的;Golbeck 等^[16]搜集 Twitter 用户信息,利用机器学习回归算法中的高斯过程回归和 ZeroR 回归来预测 Twitter 用户的人格,结果表明两种方法基本能获得相同的结果,在开放性和宜人性上的预测结果一致,表明了社交网络中的行为数据是可以预测用户人格信息的;Shuotian 等^[17]通过实验证明利用微博在线行为数据能够很好地预测用户的人格;Back 等^[18]研究证明不同人格的网络行为是不同的,相反,网络行为能够影响人的人格;Yoram 等^[19]通过实验验证了利用 Facebook 的一些静态数据可以预测用户的人格;Jennifer^[20]利用基于词语计量的文本分析工具(Linguistic Inquiry and Word Count,LIWC)对 4 类社交媒体中近 9000 个用户的文本数据进行实验,验证了文本数据与用户大五人格之间存在着一定的关系。

综上所述,国外学者主要是利用 Facebook 和 Twitter 等社交网络对用户人格进行预测^[21],而国内学者主要是利用新浪微博、人人网等社交平台对用户进行人格预测,不过无论使用哪种社交网络,或对哪种人格进行预测,研究思路基本一致。通过归纳总结,得到如图 2 所示的基础预测流程:1)获取网络用户人格标签数据,通过让参与者填写人格问卷的形式确定不同维度的人格特征得分;2)获取参与者的网络行为数据,可以通过网站授权 API 接口,或者网络爬虫获取;3)提取与人格相关的数据特征,通过特征提取、特征选择来确定用于创建预测模型的数据;4)创建人格预测模型,主要通过机器学习算法实现,大部分采用传统的有监督分类或回归算法;5)验证模型的正确性并修正模型,通常采用预测相关性指标、预测

误差指标和分类准确性指标进行验证。

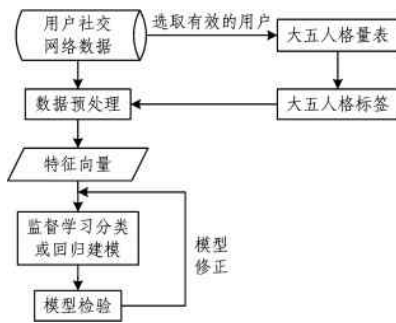


图2 基于社交网络的用户人格特征预测流程

Fig. 2 Prediction process of Big-Five personality traits for social network

2.2 特征选取

目前,基于社交网络的用户人格特征主要从3类特征中提取,即静态特征、动态特征、语言特征以及语义特征,如表2所列。

表2 社交网络用户人格预测的数据类型

Table 2 Data types used in personality traits prediction for social network

数据分类	内容	文献
静态特征	性别,年龄,朋友数,粉丝数,图像信息等	[7,18,21-22,38]
动态特征	点赞,发消息,收藏,转发,评价等	[1,22]
语言特征	文本信息	[15,20,23-25,27]
语义特征	情感,观点等	[28-29,33]

静态特征包括性别、年龄、朋友数、粉丝数、图像信息等。Quercia等^[22]通过分析Twitter的3种特征进行用户大五人格的预测,这3类特征分别是粉丝数、关注数以及用户出现在其他人阅读列表中的次数。白朔天等^[17]通过提取新浪微博的29种特征数据进行用户人格预测,这些特征主要包括微博状态数、自我描述长度、是否默认头像、关注数、粉丝数、互粉数、安全信息等。Bachrach等^[19]通过提取Facebook用户的朋友数、组数、likes数、照片数以及更新的状态数,对Facebook用户的人格特征进行预测。Ferwerda等^[23]通过采集113名Instagram用户的22398幅照片信息,并提取图片数据特征,如色调、亮度、饱和度等,最终验证图片的数据特征与人格特征之间是存在关联的,如神经质人格与图片的亮度相关联,宜人性人格与图片中黑暗与光明区域的多少相关联。

动态特征主要包括用户点赞、发消息、转发等一系列行为数据。李琳等^[24]通过统计方法提取了新浪微博用户的静态特征,此外还考虑了点赞、发消息等动态数据特征,并通过计算时间序列统计量来刻画动态行为特征,以表现微博用户人格随时间变化的情况。结果表明,观察时间越长,预测模型的准确率越高,并且每种人格维度都有不同的最优观察周期。

语言学中的很多研究证明,不同人格的人使用语言时的风格是不同的,也就是说通过用户社交网络发表的一些文本数据(如微博内容、回复、自我描述等)也能够预测用户的人格特征。文献^[17,25-27]等都在试图通过分析用户的语言特征来预测用户人格;Schwartz等^[26]研究了N-Grams短语及关键词这两种语言特征与人格之间的关系,并在随后的研究中发现差分语言分析(Differential Language Analysis,DLA)方

法在分析文本中的开放词汇方面具有很多优势。

语义特征能够分析出用户的情感、观点、意见以及人格魅力等信息^[28]。这种方法基于与人格特质相关的语料库,因此语料库的不同影响了人格的预测结果。有研究者在针对某一语料库进行人格预测时,预测结果的准确率可以达到83%,但是当扩大语料库时,其准确率迅速下降到55%^[29]。针对这一问题,Francisco等^[28]通过使用一种大规模的语料库,采用回归及排序算法对各种文本特征提取进行比较,从分类准确率、与基准回归算法相比提高的百分比以及排序算法的误差3个角度进行验证,结果表明预测结果有了很大提高。

之前的研究者大多通过单一社交网络(如Facebook)或者单一属性(如likes数)预测用户人格,并且主要是探寻个别数据特征与不同人格特质之间的关系,对于特征的选取并没有固定的方法或理论,一般通过大脑风暴方式^[30]或统计方式得到。近年来,有研究者试图通过融合不同社交网络数据或者不同类型的数据预测用户人格。Marcin等^[31]采集Twitter和Instagram的文本、图像以及用户的元数据进行人格预测,表明融合的结果使得人格预测的错误率有所下降。随后,Kseniya等^[32]也从多个社交网络分析用户的人格,他们采集Twitter,Foursquare和Instagram3类社交网络的数据进行用户人格预测,并将其与单一数据源预测人格进行比较,结果显示在某些情况下,所提方法的预测性能可以提高17%。Wei等^[33]提出一种异构信息集成框架,将社交网络用户的语义、头像以及表情作为预测人格的数据特征,相比单一特征的人格预测,其预测性能有所提高。

2.3 模型建立

在模型建立方面,目前大多采用有监督学习算法创建预测模型,具体算法及预测结果如表3所列。其基本思路是把每个维度划分成高分、中分、低分3类或更细化的5类,然后利用分类算法训练分类器,通过分类器对用户的人格特征进行预测。现有的研究一般是将5个维度当成5个任务来分别进行训练、建模,预测的准确率能达到70%左右,典型的结果有Ortigos等^[34]采用朴素贝叶斯和C4.5算法对5类人格分别进行预测建模,其中三分类预测精度达到约70%,五分类精度达到约63%。

表3 人格特征预测研究中使用的算法

Table 3 Algorithms used in personality traits prediction

文献	平台	算法	预测结果
[16]	Twitter	高斯过程,ZeroR	MAE约为0.11
[35]	Facebook	朴素贝叶斯,C4.5	正确率为63%
[21]	Facebook	高斯过程,M5	MAE约为0.11
[30]	人人网,新浪微博	增量回归,多任务回归	精度>85%
[27]	Facebook	线性回归,RepTree,决策表	可预测出74.5%的用户
[34]	新浪微博	半监督线性回归	RAE约为75%

最近有许多学者尝试采用几种不同的机器学习算法来建立预测模型,包括多任务回归学习^[35]及半监督学习^[36-37]等。研究结果显示,其预测准确率有所提高。

2.4 研究小结

当前,基于社交网络进行用户人格特征预测虽然取得了一定的研究成果,但仍然存在着很大的局限性:大多数的人格

预测研究仅限于对某一类微博特征与某种人格特质之间的相关性分析^[1,16],并且采取的都是单任务学习方法,即将5种人格特征的预测作为毫不相干的任务单独执行,而实际上这5种特征之间存在着一定的相关性,多个任务之间潜在的关系没有得到充分利用。因此,可以考虑利用多任务学习方法来弥补这种缺陷。白朔天等^[35]率先使用多任务回归模型进行人格预测建模,但只是通过寻找关联系数矩阵来并行学习多个任务,并没有充分利用5种特征中的共享信息;同时,使用弗罗贝尼乌斯范数进行建模计算虽然防止了模型的过拟合现象,但是并没有考虑到数据的稀疏性及模型的鲁棒性。因此,如何有效利用5种人格特征间的相关信息来提高预测准确率,是有待进一步研究的课题。

3 社交网络用户的心理健康预测

心理学指出个体周边环境中能够包含一些可以预测其心理健康状况的信息^[39-41],而社交网络是现实世界的延伸,是个体的周边环境,也是行为总体的一部分,因此可以考虑借助社交网络来预测个体的心理健康状况。

国内外不少学者已经就网络行为和心理健康之间的关系进行了研究。Campbell等^[42]研究发现网瘾与互联网使用时间和频率有关;Wei等^[43]发现网络游戏与抑郁明显相关;张梦茜等^[44]发现网瘾与焦虑也存在着一定的关系;管理等^[41]试图通过微博语言特征与个体自杀性行为之间的关系识别出自杀可能性高的个体。最近也有研究表明,通过社交网络用户发布的语言能够识别出具有某种心理健康问题的个体,如

自杀倾向、伤害他人倾向和精神分裂等^[45-46]。Munmum等^[47]的研究表明,语言特征可以用来预测用户的抑郁症和是否有自杀念头。Mrinal等^[48]第一次通过社交网络研究推特效应,即自杀模仿效应。Danielle等^[49]从数据挖掘的角度研究了Twitter用户抑郁症分类的数据特征选择问题。

基于社交网络数据预测用户心理健康状态的主要文献及结果如表4所列。其中,Randall等^[21,25]利用Twitter用户的静态信息和文本信息,通过构建逻辑回归、多层感知器、随机森林和支持向量机等4种分类模型来预测具备精神变态倾向的用户;朱廷劭等^[30]通过分析用户网络行为数据,基于PHI(Psychological Health Inventory)七维度心理健康量表,利用决策树创建心理健康状况预测模型,其平均预测正确率约为70%;白朔天^[12]等采用多任务回归学习预测社交网络用户的抑郁和焦虑两种心理健康状况,最终证明心理健康问题可以通过网络行为反映出来;George等^[45]使用聚类算法,针对Reddit社交数据的语言信息,确定与心理健康相关的语言特征,通过分析不同心理健康问题的subreddits论坛内容(包括发布的帖子和评论内容),分析出16种覆盖不同心理健康问题的语言特征;Margaret等^[46]通过Twitter API采集了174位自认为存在精神分裂症的Twitter用户的3200个帖子,首先通过LIWC,LDA(LatentDirichlet Allocation),Brown Cluster,Character Language Models和Perplexity等5种自然语言处理方法挖掘与精神分裂症相关的语言特征,然后采用支持向量机SVM和最大熵MaxEnt两种机器学习算法进行分类,其最好分类准确率分别为82.3%和81.9%。

表4 心理健康状态预测研究中使用的算法
Table 4 Algorithms used in mental health prediction

文献	平台	算法	数据类型	健康状况	预测结果
[25]	Twitter	LR,MLP,RF100,SVM	微博信息,文本信息	精神变态	AUC为0.736
[30]	Internet	决策树	调查问卷	PHI心理健康状态	准确率为60%
[12]	新浪微博	多任务回归	调查问卷	SCL90	最大MAE为0.61
[45]	Reddit	聚类	语言特征,如句子结构,词语的用法等	精神分裂,抑郁等	16种心理健康问题
[46]	Twitter	SVM,MaxEnt	语言特征	精神分裂	准确率为82.3%

从目前的研究来看,利用社交网络进行用户心理健康预测主要依据的是社交网络本身的数据,研究的对象主要是社交网络的语言特征,研究的重点在于数据特征的选取,研究的方法主要是机器学习中的有监督分类算法。大部分研究仅停留在具有某种心理健康问题的个体在社交网络中的语言特征分析上。

基于社交网络预测用户心理健康的最大障碍在于数据的获取。一般来说,具备某种心理健康问题(比如焦虑、抑郁等)的用户,很可能不会再使用社交网络,因此研究社交网络用户具有某种心理健康问题的倾向性也许更为关键;同时,应该将对用户心理健康有重要影响的用户身份和人格两类数据作为识别某种心理健康个体的参考。总之,关于社交网络用户心理健康预测的研究刚刚起步,还有很多问题值得探索。

结束语 本文在综述社交网络用户认知域特征预测研究进展的基础上,指出了现有研究存在的问题和不足,并给出了一些新的研究思路,可以看出该领域还有很大的研究空间。就目前的发展来看,亟需在以下两个方面做更加深入的研究。

1)更系统地提取表征用户网络行为的特征。社交网络行为特征的提取是创建预测模型的关键,只有恰当地选择特征才能提高预测的准确度。目前,信息的表现形式越来越多样,常见的有视频、图像、语音、表情符号等,如何系统地多样化信息形式中提取表征用户网络行为的特征,将是认知域特征预测面临的第一个难点问题。

2)寻找更加适用于该领域的机器学习模型。目前,认知域特征的预测一般采用传统的机器学习算法进行建模,基本没有考虑认知域各特征之间的内在联系,而这种潜在的联系往往会对预测结果产生较大的影响。因此,在后续研究中需要寻找一种更加适合的机器学习模型,并利用这些关联信息,以提高预测的准确率和效率。

虽然目前认知域特征预测领域的研究还不成熟,并且面临着各种困难和障碍,但是从社会发展的长期趋势来看,无论从安全角度还是发展角度,关键的核心仍然是人,因此心理学与计算机科学以及社会科学之间的融合将成为未来的必然发展趋势。

参考文献

- [1] KOSINSKI M, STILL WELL D, GRAEPEL T. Private traits and attributes are predictable from digital records of human behavior[J]. *Proceedings of the National Academy of Sciences*, 2013, 110(15): 5802-5805.
- [2] MISLOVE A, VISWANATH B, GUMMADI K P. You are Know: Inferring User Profiles in Online Social Networks[C]// 34th ACM International Conference on Web Search and Data Mining. New York, USA, 2010: 251-260.
- [3] HE J, CHU W, LIU Z. Inferring privacy information from social networks[C]// International Conference on Intelligence and Security Information. 2006: 154-165.
- [4] LINDAMOOD J, HEATHERLY R, KANTARCIOGLU M. Inferring private information using social network data[C]// Proceeding of International Conference on World Wide Web. 2009: 1145-1146.
- [5] XU W, ZHOU X. Inferring privacy information via social relations[C]// Proceeding of IEEE 24th International conference on Data Engineering Workshop. 2008: 525-530.
- [6] ZHELEVA E, GETOOR L. To Join or not to Join: the Illusion of Privacy in Social Networks with Mixed Public and Private User Profiles[C]// Proceedings of the 19th International Conference on World Wide Web. ACM, 2009: 531-540.
- [7] BIN B, MILAD S. Inferring demographic of Search Users: Social Data Meets Search Queries[C]// International World Wide Web Conference Committee. ACM, 2013: 13-17.
- [8] SHEN C, CAI Z, GUAN X, et al. User Authentication Through Mouse Dynamics[J]. *IEEE Transaction on Information Forensics and Security*, 2013, 8(1): 16-30.
- [9] ZHANG L, CHEN Z X, YANG B. Personality Analysis and Prediction of Social Network Users[J]. *Chinese Journal of Computers*, 2014, 37(8): 1877-1894. (in Chinese)
张磊, 陈贞翔, 杨波. 社交网络用户的人格分析与预测[J]. *计算机学报*, 2014, 37(8): 1877-1894.
- [10] DING Y X, XIAO X, WU M J, et al. Predicting Users' Profiles in Social Network based on Semisupervised Learning[J]. *Journal on Communications*, 2014, 35(8): 15-22. (in Chinese)
丁宇新, 肖晓, 吴美晶, 等. 基于半监督学习的社交网络用户属性预测[J]. *通信学报*, 2014, 35(8): 15-22.
- [11] SVITLANA V, YORAM B, MICHAEL A, et al. Inferring Latent User Properties from Texts Published in Social Media[C]// Twenty-ninth Aaii Conference on Artificial Intelligence. 2015: 4296-4297.
- [12] BAI S T, HAO B B, LI A, et al. Depression and anxiety prediction on microblogs[J]. *Journal of University of Chinese Academy of Sciences*, 2014, 31(6): 814-820.
- [13] HU K X, LIANG Y, XU H B, et al. A Method for Social Network User Identity Feature Recognition[J]. *Journal of Computer Research and Development*, 2016, 53(11): 2630-2644. (in Chinese)
- [14] GOLDBERG L R, JOHNSON J A, EBER H W. The international personality item pool and future of public-domain personality measures[J]. *Journal of Research in Personality*, 2006, 40(1): 84-96.
- [15] SCHRAMMEL J, KOFFEL C, TSCHELIGI M. Personality traits, usage patterns and information disclosure in online communities[C]// Proceedings of the 23rd British HCI Group Annual Conference on People and Computer. 2009: 169-174.
- [16] GOLBECK J, ROBLES C, EDMONDSON M. Predicting personality from twitter [C] // Proceedings of the 2011 IEEE 3rd International Conference on Privacy, Security, Risk and Trust (Passat). 2011: 149-156.
- [17] BAI S, HAO B, LI A, et al. Predicting Big Five Personality Traits of Microblog Users[C]// International Conferences on Web Intelligence (WI) and Intelligent Agent Technology (IAT). 2013: 501-508.
- [18] BACK M D, STOPFER J M, VAZIRE S. Facebook profiles reflect actual personality, not self-idealization [J]. *Psychological Science*, 2010, 21(3): 372-374.
- [19] BACHRACH Y, KOSINSKI M, GRAEPEL T, et al. Personality and patterns of Facebook usage[C]// Proceedings of the 3rd Annual ACM Web Science Conference. 2012: 24-32.
- [20] JENNIFER G. Predicting Personality from Social Media Text [J]. *AIS Transactions on Replication Research*, 2016(2): 1-10.
- [21] GOLBECK J, ROBLES C, TUENER K. Predicting personality with social media[C]// Proceedings of the 2011 Annual Conference Extended Abstracts on Human Factors in Computing Systems. 2011: 253-262.
- [22] QUERCIA D, KOSINSKI M, STILLWELL D, et al. Our Twitter profiles, our selves: Predicting personality with Twitter[C]// Proceedings of the 2011 IEEE 3rd International Conference on Privacy, Security, Risk and Trust. Boston, USA, 2011: 180-185.
- [23] FERWDA B, SCHEDL M, TKALCIC M. Predicting Personality Traits with Instagram Pictures[C]// Proceedings of the 3rd workshop on Emotions and Personality in Personalized Systems. 2015: 7-10.
- [24] LIN L, ANG L, BIBO H, et al. Predicting Active Users' Personality Based on Micro-Blogging Behaviors[J]. *Plos One*, 2014, 9(1): e84997.
- [25] WALD R, KHOSHGOFTAAR T M, NAPOLITANO A. Using twitter content to predict psychopathy[C]// Proceedings of the 2012 11th International Conference on Machine Learning and Applications. 2012: 394-401.
- [26] SCHWARTZ H A, EICHSTAEDT J C, DZIURZYNSKI L. Toward personality insights from language exploration in social media[C]// Proceedings of the AAAI Spring Symposium Series. 2013: 72-79.
- [27] WALD R, KHOSHGOFTAAR T, SUMNER C. Machine prediction of personality from Facebook profiles[C]// Proceedings

- of the 2012 IEEE 13th International Conference on Information Reuse and Intergration (IRD), 2012:109-115.
- [28] FRANCISCO I, ALASTAIR J G, SCOTT N, et al. Large Scale Personality Classification of Bloggers[J]. International Conference on Affective Computing & Intelligent Interaction, 2011, 6975:568-577.
- [29] NOWSON S, OBERLANDER J. Identifying more bloggers: Towards large scale personality classification of personal[C]// International Conference on Weblogs and Social Media, 2007:1-7.
- [30] TINGSHAO Z, YUE N, ANG L. Using Decision Tree to Predict Mental Health Status based on Web Behavior[C]// Proceedings of the 2011 3rd Symposium on web Socitey, Port Elizabeth, South Africa, 2011:27-31.
- [31] MARCIN S, BRUCE F, MARKO T, et al. Fusing Social Media Cues: Personality Prediction from Twitter and Instagram[C]// The 25th International World Wide Web Conference, Montréal, Canada, 2016:1-2.
- [32] KSENIYA B, ALEKSANDR F, ANDREY F, et al. Towards User Personality Profiling from Multiple Social Networks [C] // AAAI, 2017.
- [33] WEI H H, ZHANG F Z, NICHOLAS J Y. Beyond the words: Predicting User Personality from Heterogeneous Information [C]// 10th ACM International Confernece on Web Search and Data Mining(WSDM 2017). Cmbridge, United Kingdom, 2017: 305-314.
- [34] ORTIGOSA A, CARRO R, QUIROGA J. Predicting user personality by mining social interactions in Facebook[J]. Journal of Computer and System Sciences, 2013, 80:57-71.
- [35] BAI S T, YUAN S, CHENG L, et al. Application of Multi-Task Regression in Social Media Mining[J]. Journal of Harbin Institute of Technology, 2014, 46(9):100-110. (in Chinese)
白朔天, 袁莎, 程立, 等. 多任务回归在社交媒体挖掘中的应用[J]. 哈尔滨工业大学学报, 2014, 46(9):100-110.
- [36] NIE D, GUAN Z, HAO B, et al. Predicting Personality On Social Media with Semi-supervised Learning[C]// International Joint Conferences Web Intelligence(WI) and Intelligent Agent Technologies(IAT), 2014:158-165.
- [37] CORREA T, HINSLEY A W, DE ZUNIGA H G. Who interacts on the Web?: The intersection of users' personality and social media use[J]. The Intersection of Users' Personality and Social Media Use Computers in Human Behavior, 2010, 26(2): 247-253.
- [38] LEQI L, DANIEL P P, ZAHRA R S, et al. Analyzing Personality through Social Media Profile Picture Choice[C]// Proceedings of the Tenth International AAAI Conference on Web and Social Media(ICWSM), 2016.
- [39] BESSIERE K, KIESLER S, KRAUT R, et al. Effects of Internet use and Social Resources on Changes in Depression[J]. Information, Community & Society, 2008, 11(1):47-70.
- [40] GAO C, ZHANG Y, MA X F. Study on Social Psychological factors and Countermeasures of College Students' online Behavior [C]// Chinese Medical Association the 10th National Conference on Behavioral Medicine, 2008:132-133. (in Chinese)
高昶, 张妍, 马潇斐. 当代大学生上网行为的社会心理因素及对策探讨[C]// 中华医学会第十次全国行为医学学术会议, 2008: 132-133.
- [41] GUAN L, HAO B B, CHENG Q J, et al. Behavioral and Linguistic Characteristics of Microblog Users with Various Suicide Ideation Level: an explanatory study[J]. Chinese Journal of Public Health, 2015, 31(3):349-352. (in Chinese)
管理, 郝碧波, 程绮瑾, 等. 不同自杀可能性微博用户行为和语言特征差异解释性研究[J]. 中国公共卫生, 2015, 31(3):349-352.
- [42] CAMPBELL A J, CUMMING S R, HUGHES I. Internet use by the Social Fearful: Addiction or Therapy? [J]. Cyber Psychology & Behavior, 2006, 9(1):69-81.
- [43] WEI P, LIU M. Online gaming dependency: a preliminary study in China[J]. Cyberpsychology Behavior and Social Networking, 2010, 13(3):329-333.
- [44] ZHANG M H, ZHAO X Y, SUN Y M. Study on Internet Addiction Disorder among College Students and Its Influence on Interactive Anxiety[J]. Chinese Journal of Social Medicine, 2013, 30(1):39-41. (in Chinese)
张梦茜, 赵笑颜, 孙易蔓. 大学生网络成瘾现状及其对交往焦虑的影响[J]. 中国社会医学杂志, 2013, 30(1):39-41.
- [45] GEORGE G, ANIKA O, TIM J H, et al. The Language of Mental Health Problems in Social Media[C]// Proceedings of the 3rd Workshop on Computational Linguistics and Clinical Psychology, San Diego, California, 2016:63-73.
- [46] MARGARET M, KRISTY H, GLEN C. Quantifying the Language of Schizophrenia in Social Media[C]// Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology, Denver, Colorado, 2015:11-20.
- [47] MUNMUM D C, EMRE K, MARK D, et al. Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media [C]// Proceedings of 2016 Special Interest Group on Computer-Human Interaction(SIGCHI), CA, USA, 2016:2098-2110.
- [48] MRINAL K, MARK D, GLEN C, et al. Detecting Changes in Suicide Content Manifested in Social Media Following Celebrity Suicides[C]// Cyprus, 26th ACM Conference on Hypertext and Social Media, 2015:85-94.
- [49] DANIELLE M, CRAIG B, MIKE C. Feature Studies to Inform the Classification of Depressive Symptoms from Twitter Data for Population Health [OL]. <https://arxiv.org/pdf/1701.08229.pdf>.