

基于在线学习行为数据的人格特质识别研究

赵宏¹ 刘颖² 李爽¹ 徐鹏飞¹ 郑勤华¹

(1. 北京师范大学 教育学部, 北京 100875; 2. 北京大学附属小学, 北京 100875)

[摘要] 获取学习者个性特征是实现以学生为中心的精准化、个性化教学的重要前提,而学习行为是分析学习者个性特征的重要依据。本研究以参加奥鹏公共研修平台在线学习者研究对象,基于人格特质生成学习行为偏好假设,探索利用机器学习分类算法实现在线学习行为的人格特质识别;同时基于文献构建人格特质类型与在线学习行为之间的映射关系,采用 Rapid Miner 数据挖掘工具探索决策树、朴素贝叶斯和支持向量机三种算法对五种人格特质的识别效果。结果发现:决策树算法对人格特质类型的识别准确率高于其他两种算法,对大五人格特质的综合识别效果最好;不同人格特质识别灵敏度不同,尽责性人格特质类型的识别灵敏度最高,神经质人格特质最低。

[关键词] 人格特质;在线学习行为;学习行为偏好;分类算法

[中图分类号] G434 **[文献标识码]** A **[文章编号]** 1007-2179(2019)05-0110-11

一、前言

获取学习者个性特征是实现以学生为中心的精准化、个性化和智能化教育的重要前提。如何通过大数据分析和人工智能技术在无意识、非侵入、非配合情况下自动识别和掌握学习者的个性特征,对其进行综合分析、判断和干预是基于大数据开展个性化教学、科学评价、精细化管理和智能化决策等迫切需要解决的问题。

人格特质作为不同时间与不同情况相对一致的影响行为的品质或特性,是学习者外在行为的内在动机。研究表明,人们无论在现实生活还是虚拟世界都会不可避免且“非故意地”留下反映人格特质的行为痕迹。随着社交媒体的普及和数据量爆炸式增长,人格特质已被广泛应用于解释用户的社交网

络行为。同样,人格特质已被证明对学习者在在线学习行为起重要的影响作用(Anderson & Simpson, 2004)。在线学习者的人格特质不同,其学习过程差异极大(Kanuka & Nocente, 2003),如高度开放的在线学习者重视在线学习的功能价值,而高神经质在线学习者喜欢在线学习的自由(Watjatrakul, 2016);内向性人格特质的学习者愿意在社会情境中与教师网络互动,而外向性人格特质的学习者更愿意在学习情境中与教师网络互动等(Anderson et al., 2004; 李红波等, 2004)。

由此可见,人格特质作为决定行为的关键因素,会对学习者的在线学习意愿、在线学习行为偏好等产生举足轻重的影响。因此,随着网络数据的日益丰富,学习者人格特质研究已成为在线学习领域关注的热点。

[收稿日期] 2019-06-11

[修回日期] 2019-08-10

[DOI 编码] 10.13966/j.cnki.kfjyyj.2019.05.011

[作者简介] 赵宏,博士,副教授,硕士生导师,北京师范大学教育学部,研究方向:在线学习与认知发展(zhaohong@bnu.edu.cn);刘颖,硕士,北京大学附属小学,研究方向:在线学习者特征;李爽,博士,副教授,硕士生导师,北京师范大学教育学部,研究方向:学习支持;徐鹏飞,博士,硕士生导师,北京师范大学人工智能学院,研究方向:教育大数据和人工智能;郑勤华,博士,教授,北京师范大学教育学部,研究方向:教育大数据。

二、研究现状

人格是内在心理物理系统的动力组织,它决定人体对环境的适应(Allport, 1937)。特质是人格的基本结构单位,一个特质就是一个人格维度,代表机能上人格个体差异的基本类别。因此,现代心理学的人格特质指个体所具有的神经特性,具有支配个体行为的能力,使个人在变化的环境中给予一致的反应(黄定华,2009)。也有研究者指出,人格特质是个体在各种情境下表现出的影响行为的品质或特性。它既保障个体一般化、稳定而持久的行为倾向,又可以通过个体认知情绪的调控、情绪应对的策略而间接影响行为决策(朱从庆等,2016)。不同学者对人格特质解释的差异显示,人格特质指组成人格的因素引发个体行为和主动引导人的行为,并使个体面对不同种类的刺激作出相同反应的心理结构。

传统上,研究者一般采用自陈量表采集人格特质数据,即要求受试者按照实际情形,对自己的人格特质进行自我评价。由于人格特质测量量表主要依赖于被试的主观感受和自我陈述,因此目前大多数人格特质测量量表都是标准化测验,但被试回答过程中会出现倾向于用更符合社会价值和更有利于自我表征的方式表达自己,即故意对问卷题目做出或夸大或缩小的歪曲反应,从而影响测量结果的客观性(胡雯,王有智,2014)。信息技术的飞速发展和互联网技术的全面普及,研究者可以实现在线学习行为数据自动记录和采集。因此,随着大数据广泛应用,基于大数据视角的人格特质识别成为人格计算领域重要的研究方向和研究热点。相对于心理测验被动地测量人格特质类型来说,人格特质识别是基于用户的在线语言学特征、社交网络行为特征和社交网络结构特征等在线行为数据对其人格特质类型进行自动识别和判断。这一方面克服了传统人格特质测量方法的主观性和静态性,另一方面也避免了由被试的个人因素带来的测量偏差,为学习者个性特征获取提供了新的方法和思路。

国外开展用户人格特质识别研究起步较早,成果丰富。国外早期研究以心理学为主,主要是基于文本内容的语言学特征判断人格特质。梅尔等(Mehl et al., 2006)认为,基于咒骂词和负向情感词等可以判断用户的宜人性人格。巴克等(Back et

al., 2008)证明邮件地址的语言学特征能准确判断个体的神经质、开放性、宜人性、尽责性和自恋等人格。邱等人(Qiu et al., 2012)证明基于推特的语言学特征可以判断用户的宜人性和神经质人格。随着计算机科学发展,研究者开始尝试基于机器学习的人格识别研究。奥伯兰德等(Oberlander et al., 2006)使用朴素贝叶斯方法成功识别博客用户的尽责性、外向性和宜人性等人格特质。研究者不断探索基于机器学习的人格识别方法,致力于研究什么样的特征以及学习模型能有效提高人格特质识别的准确度。分类和回归是目前两种典型的利用机器学习识别用户人格特质的方法。分类法是将人格特质识别看成是二分类或三分类,回归法则估计用户人格特质的强度。

相比国外研究而言,国内关于人格特质识别研究起步较晚,成果相对较少,主要采用分类法,即基于微博用户的文本特征和网络社交结构特征,采用机器学习的分类算法识别人格特质,比较不同分类算法的准确性。例如,白朔天等(Bai et al., 2013)选取444名新浪微博参与者,从中挖掘29种用户行为特征,采用单任务线性回归、增量回归和多任务回归算法验证识别效果。杨洁(2015)通过提取新浪微博用户发布的微博文本和社交网络关系特征,采用支持向量机算法建立人格特质识别模型,准确率达70%~80%。刘多星(2017)基于LIWC(一种基于词语计量的文本分析工具)词类别使用和大五人格各维度的映射关系,构建人格特质打分模型并对微博用户的人格特质进行打分,以各人格特质维度的平均值为阈值对用户的人格特质分类,结果发现划分的训练样本数为5时,分类器的性能最佳,准确率达86%~95%。王佳佳(2017)基于非参数层次贝叶斯主题模型NHBT构建社交网络用户大五人格特质和主题偏好之间的关系,提出PT-LDA主题模型识别未知用户的人格特质。娜迪热和胡俊(2018)使用线性回归、高斯回归和支持向量机算法训练预测模型,基于新浪微博社交用户的行为特征和社交关系网络特征识别用户的大五人格特质,结果表明:三种算法中支持向量机的预测效果最好,且预测模型对开放性人格特质的预测效果最好,所选用的社交关系网络特征对预测模型贡献显著。

综上,人格特质识别研究近年取得了明显进展,

但多集中于脸书、推特和微博等在线社交网络,对在线学习的研究不多。人格特质作为决定个体行为的根本因素,对学习者在在线学习行为的影响举足轻重。识别其人格特质能帮助教育机构从根本上把握学习者在线学习行为的原因,从而为其提供全方位个性化诊断和针对性的学习支持服务。另外,当前人格识别研究主要以基于文本的语言学特质分析为主,辅于社交网络结构特征,行为特征数据偏少。因此,本研究以在线交互行为特征为对象,开展在线学习者人格特质识别研究,同时检验机器学习不同分类算法的准确度,识别人格特质的最优算法模型。

三、研究方法及过程

(一) 研究对象

本研究以参加奥鹏教师网络研修平台“2017年湖北省中小学教师信息技术应用能力提升工程项目”的学员为研究对象,收集该平台在线学习者的人格特质数据和在线学习行为数据。学习形式包括观看视频、论坛互动、提交作业等,课程学习时间为每人1200分钟,学生在线参与度高,在线行为数据丰富。

(二) 研究方法及工具

1. 大五人格测量量表

本研究使用经张新建等人修订的科斯塔和麦克雷(Costa & McCrae)2004年更新出版的简化版大五人格量表(NEO-FFI)。该量表广泛应用于教育领域的人格特质测量,有非常高的信度和效度,且被证明有跨语言和跨文化的一致性与稳定性(McCrae & Costa, 1989; McCrae & Costa, 2004; McCrae, 2005; John et al., 2008)。量表包括60个条目,主要测量神经质、外向性、开放性、宜人性和责任性五个人格特质维度,每个维度含12条目,每个条目采用李克特5点量表(即“非常不符合”“大部分不符合”“不确定”“大部分符合”“非常符合”),计分为1、2、3、4、5。部分题项采用反向记分。

2. 机器学习分类算法

本研究主要探讨基于在线行为数据的人格特质类型识别,即人格特质的分类,不考虑学习者人格特质的强度。因此本研究采用决策树、朴素贝叶斯和支持向量机三种分类算法,探讨其对五种人格类型的识别效果。

(三) 数据收集

1. 大五人格量表数据收集

本研究采用在线问卷方式收集数据,共回收问卷230份,有效问卷219份(有效率95%)。调查对象中,男生116名,女生103名,年龄在25~48岁间,基于调查问卷数据,本研究采用克隆巴赫 α 系数分析大五人格测量量表的信效度(见表一)。

表一 大五人格测量量表信度

人格特质维度	项目数	Cronbach's α
神经质	12	0.849
外向性	12	0.794
开放性	12	0.863
宜人性	12	0.845
尽责性	12	0.831
总体	60	0.840

由表一可知,本研究样本大五人格量表总体Cronbach's α 值是0.840,五个维度分量表Cronbach's α 值介于0.79~0.86之间,表明量表有较好的内部一致性。

2. 在线行为数据收集

本研究在线学习行为的人格特质识别建立在人格特质测量的基础上,学习者行为数据和人格数据与其对应。鉴于此,本研究在奥鹏教师网络研修平台读取数据,并收集有效问卷219名调查对象的在线学习行为数据,具体行为指标后面详细论述。

(四) 研究过程

1. 人格特质与在线学习行为关系构建

建立人格特质与在线学习行为的映射关系是实现基于在线学习行为数据人格特质识别的基础和依据。已有研究表明,在线学习行为存在层次的高低,且不同层次行为对在线学习的促进作用不同(Gunawardena et al., 1997; Berge, 1999; Gilbert & Moore, 1998; 陈丽, 2004; Ally, 2004)。本研究根据远程学习教学交互层次塔理论,将在线学习行为层次分为操作交互、信息交互和概念交互。其中,信息交互层作为可被观察的层次,是促进学生产生概念交互,进而实现学习目标的重要环节。因此,研究者倾向于从信息交互层次讨论在线学习者的学习行为(孙月亚, 2015; 冯晓英等, 2016; 李爽等, 2016; 李超等, 2017)。

同样,研究者也从信息交互行为角度探讨不同

人格特质学习者在线学习行为偏好。如外向性学习者乐观向上,待人热情,好交际,其典型在线学习行为是与同伴交互和教师交互(李阳,2016;衷克定,2016;陶楚楚,2018;张琪等,2019);神经质人格特质学习者敏感,焦虑水平高,与人交往容易产生不安和压抑情绪,其在线学习行为偏好是与学习资源交互,不喜欢与同伴交互(张晓丽,2008;王陆等,2011);开放性学习者思维活跃,乐于思考,寻求变化,富有创造性,其典型在线学习行为是与学习资源深度交互,且通过与同伴交互和教师交互碰撞思维火花(Anderson et al., 2004);宜人性学习者热心合作,顾全大局,对人友好,试图为大家营造共同学习的氛围,其典型在线学习行为是与同伴交互和教师交互

(李阳,2016);尽责性学习者自律性强,对个人要求高,学习态度积极负责,其学习主动性、计划性和持久性、管控制等表现良好,其典型在线学习行为是与学习资源交互、与教师交互(张艳红等,2013;Fong & Mar, 2015)。本研究对不同人格特质映射的在线学习行为的相关研究进行梳理,构建对应不同人格特质维度的在线学习行为测量指标(表二)。

基于表二的行为指标,本研究收集有效问卷219名,经过数据清洗,剔除无用和重复数据,在保持数据原貌和不影响建模结果的前提下,选取最常用的离差标准化,对原始数据进行特征转换和归一化,最终形成在线学习行为特征数据。

2. 在线学习行为指标验证

表二 不同人格特质的在线学习者典型学习行为

人格特质类型	典型在线学习行为表现	测量指标
外向型	与同伴在线交流	发帖和回帖频次
	资源分享	上传资源频次
	互动评价	评价作业或作品频次
	与教师在线交流	浏览和回复教师贴的频次
	参与答疑	参与答疑频次
神经质	浏览平台	浏览在线学习平台频次、时长
	观看视频	观看视频的频次、时长
	浏览教学资源	浏览导学信息、课件和扩展资源的频次
	与同伴在线交流	发帖和回帖频次
	互动评价	评价作业或作品频次
	作业考核	完成作业频次、参加考试频次及成绩
开放性	浏览教学资源	浏览导学信息、课件和扩展资源的频次
	与同伴在线交流	发帖和回帖频次
	互动评价	评价作业或作品频次
	与教师在线交流	浏览和回复教师贴的频次
	参与答疑	参与答疑频次
宜人性	与同伴在线交流	发帖和回帖频次
	资源分享	上传资源频次
	互动评价	评价作业或作品频次
	作业考核	完成作业频次、参加考试频次及成绩
	与教师在线交流	浏览和回复教师贴的频次
	参与答疑	参与答疑频次
尽责性	浏览平台	浏览在线学习平台频次、时长
	观看视频	观看视频的频次、时长
	浏览教学资源	浏览导学信息、课件和扩展资源的频次
	作业考核	完成作业频次、参加考试频次及成绩
	与教师在线交流	浏览和回复教师贴的频次
	参与答疑	参与答疑频次

为确认不同在线学习行为表征人格特质类型的敏感性和准确性,本研究使用斯皮尔曼等级相关法对人格特质和在线学习行为测量指标进行分析(见表三)。根据分析结果,删除不相关的测量指标,调整后的人格特质对应的测量指标分别为:外向性 10 条,神经质 15 条,宜人性 9 条,尽责性 6 条,开放性 10 条。

3. 基于在线学习行为数据的人格特质识别

基于调整后的在线学习行为指标,本研究采用 Rapid Miner 工具对同一种人格特质类型分别采用决策树、朴素贝叶斯和支持向量机算法进行识别:

1) 预先对输入的人格特质数据进行标签处理,作为训练数据,将各人格特质类型按分数分为高低两组,其中各维度分数高于平均分阈值的为高分组,代表高倾向性,统一标记为“1”;各维度分数低于平均分阈值的为低分组,代表低倾向性,统一标记为“0”(刘多星,2017)。

2) 将标签化处理的人格特质和在线学习行为数据的 70% 作为训练数据,用来找出最优分类系统(模型);30% 作为验证数据,用于判断找出的分类系统(模型)是否足够有效。

3) 采用决策树、朴素贝叶斯和支持向量机三种算法分别识别五种人格特质(见表四)。本研究采用精准率、召回率、准确率和综合评价指标(F-Measure)评估三个分类算法的识别效果,其中 Rapid Miner 可自动输出精准值、召回率和准确率,但综合评价指标需手工计算,计算公式如下:

$$F = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

由表四可知,决策树算法识别五种人格特质的效果最好,识别准确率和综合评价指标值最高。决策树算法识别五种人格特质的准确率达 80% ~ 90%,其中,除神经质人格的识别准确率为 80% 外,其余四种人格特质的识别准确率都超过 90%。此外,决策树算法在外向性、开放性、宜人性、尽责性四种人格特质识别的综合评价指标值也均大于 90%,神经质人格特质的综合评价指标值高于 75%。

为了更好地分析不同分类算法识别不同人格特质的效果,本研究比较了不同算法识别不同人格特质的准确率(见图 1)。

横向比较发现,除决策树算法外,朴素贝叶斯和

支持向量机算法对不同人格特质类型的识别准确率差异较大。朴素贝叶斯算法识别尽责性特质效果最好(识别准确率为 93.33%),而识别外向性人格特质效果最差(识别准确率为 66.67%),识别其他三种人格特质的效果相近;支持向量机识别神经质效果最差(识别准确率为 57.33%),识别其他四种类型的效果相差不大。

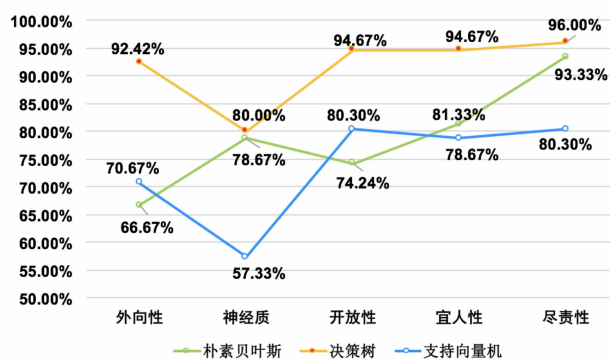


图 1 不同分类算法的识别准确率

纵向比较而言,三类算法识别尽责性人格特质的效果最好。朴素贝叶斯和决策树算法分类器识别尽责性人格特质准确率在 90% 以上,支持向量机的识别准确率也超过了 80%。同时,三类算法识别神经质人格特质效果最差,尤其是支持向量机算法识别神经质人格特质的准确率只有 57.33%,决策树和朴素贝叶斯识别准确率分别为 80% 和 78.67%。

4. 人格特质识别结果验证

为进一步验证识别效果,本研究以参与奥鹏教师网络研修平台“2017 年南平市中小学教师继续教育全员远程培训项目”(简称南平教师培训项目)学员为对象,重新收集人格特质数据和在线学习行为数据,验证三种分类算法识别人格特质的准确程度。

第二轮人格数据仍然采用在线问卷采集,共回收文卷 102 份,有效问卷 88 份。通过 SPSS Modeler 23.0 分析人格量表数据的信度表明:量表总体和各维度 Cronbach 系数均大于 0.7,说明量表前后一致性较好,有较高的信度。同时从奥鹏教师网络研修平台数据库中读取并收集 88 名有效问卷调查对象的在线学习行为数据,用于分析其人格特质。将清理、转换后的人格特质数据和在线行为数据输入 Rapid Miner 中,用三种算法分别识别五种人格特质(见表五)。

表三 斯皮尔曼相关分析

人格维度	学习行为	在线学习行为指标	相关系数	双尾
外向性	与同伴论坛交流	发布主题贴	0.494 **	0.000
		回复同伴主题贴	0.463 **	0.000
	资源分享	上传共享资源	0.629 **	0.000
		上传学习笔记	0.363 **	0.000
	互动评价	评论在线作品	0.439 **	0.000
		评论个人作业	0.335 **	0.000
		直接交流的人数	0.768 **	0.000
	与教师论坛交流	浏览教师主题贴	0.246 **	0.000
		回复教师主题贴	0.225 **	0.000
	参与教师答疑	参与线上答疑	0.322 **	0.002
观看答疑视频		0.047	0.376	
神经质	浏览平台	浏览学习平台时长	0.670 **	0.000
		浏览学习平台频次	0.577 **	0.000
	观看视频	观看教学视频时长	0.535 **	0.000
		观看教学视频频次	0.496 **	0.000
	浏览教学资源	浏览导学信息	0.407 **	0.000
		浏览课件	0.477 **	0.000
		浏览拓展材料	0.184	0.440
	与同伴论坛交流	发布主题贴	0.512 **	0.000
		回复同伴主题贴	0.472 **	0.000
	互动评价	评论在线作品	0.389 **	0.000
		评论个人作业	0.504 **	0.000
		直接交流的人数	0.411 **	0.000
	作业考核	完成作业	0.475 **	0.000
		参加期末考试	0.408 **	0.000
		作业成绩	0.482 **	0.000
总成绩		0.325 **	0.000	
开放性	浏览教学资源	浏览拓展材料	0.668 **	0.000
	与同伴论坛交流	发布主题贴	0.719 **	0.000
		回复同伴主题贴	0.587 **	0.000
	互动评价	评论在线作品	0.777 **	0.000
		评论个人作业	0.646 **	0.000
		直接交流的人数	0.695 **	0.000
	与教师论坛交流	浏览教师主题贴	0.459 **	0.000
		回复教师主题贴	0.367 **	0.002
	参与教师答疑	参与线上答疑	0.580 **	0.000
		观看答疑视频	0.455 **	0.000
宜人性	与同伴论坛交流	发布主题贴	0.448	0.256
		回复同伴主题贴	0.614 **	0.000
	资源分享	上传共享资源	0.540 **	0.000
		上传学习笔记	0.568 **	0.000
	互动评价	评论在线作品	0.713 **	0.000
		评论个人作业	0.501 **	0.000
	直接交流的人数	0.791 **	0.000	

人格维度	学习行为	在线学习行为指标	相关系数	双尾
宜人性	与教师论坛交流	浏览教师主题贴	0.434 **	0.000
		回复教师主题贴	0.219	0.400
	作业考核	完成作业	0.401 **	0.002
		参加期末考试	0.423 **	0.000
		作业推优比率	0.334	0.453
		作业成绩	0.232	0.327
		总成绩	0.189	0.249
尽责性	浏览平台	浏览学习平台时长	0.400 **	0.000
		浏览学习平台次数	0.379 **	0.000
	观看视频	观看教学视频时长	0.511 **	0.000
		观看教学视频频次	0.491 **	0.000
	浏览教学资源	浏览导学信息	0.472 **	0.002
		浏览课件	0.335 **	0.006
		浏览拓展材料	0.134	0.230
	与教师论坛交流	浏览教师主题贴	0.135	0.147
		回复教师主题贴	0.061	0.540
	作业考核	完成作业	0.483 **	0.000
		参加期末考试	0.596 **	0.000
		作业推优比率	0.278	0.163
		作业成绩	0.669 **	0.000
		总成绩	0.633 **	0.000
	参与教师答疑	参与线上答疑	0.379 **	0.000
		观看答疑视频	0.467 **	0.000

注: * p < 0.05, ** p < 0.01

表四 三种分类算法识别五种人格特质的结果

人格特质	分类算法	精准率 (%)	召回率 (%)	综合评价指标 (%)	准确率 (%)
外向性	朴素贝叶斯	96.30	52.00	80.00	66.67
	决策树	95.74	93.75	94.73	92.42
	支持向量机	70.67	100.00	82.81	70.67
神经质	朴素贝叶斯	63.89	88.46	74.1	78.67
	决策树	68.57	85.71	76.19	80.00
	支持向量机	57.33	100.00	72.88	57.33
开放性	朴素贝叶斯	92.86	73.58	82.10	74.24
	决策树	100.00	92.98	96.36	94.67
	支持向量机	80.30	100.00	89.07	80.30
宜人性	朴素贝叶斯	86.05	82.22	84.09	81.33
	决策树	91.49	100.00	95.56	94.67
	支持向量机	69.23	87.10	77.14	78.67
尽责性	朴素贝叶斯	94.44	96.23	95.33	93.33
	决策树	100.00	94.23	97.03	96.00
	支持向量机	80.30	100.00	89.07	80.30

表五 验证阶段三种分类算法的识别结果

人格特质	分类算法	精准率(%)	召回率(%)	综合评价指标(%)	准确率(%)
外向性	朴素贝叶斯	93.75	68.18	78.95	75.00
	决策树	95.45	87.50	91.30	84.00
	支持向量机	84.00	100.00	91.30	84.00
神经质	朴素贝叶斯	77.78	84.00	80.77	68.75
	决策树	74.67	100.00	85.50	74.67
	支持向量机	65.33	100.00	79.03	65.33
开放性	朴素贝叶斯	78.26	90.00	83.72	72.00
	决策树	100.00	90.57	95.05	92.42
	支持向量机	63.64	93.33	75.68	64.00
宜人性	朴素贝叶斯	86.05	82.22	84.09	81.33
	决策树	91.49	100.00	95.56	94.67
	支持向量机	60.00	100.00	75.00	60.00
尽责性	朴素贝叶斯	94.87	92.50	93.67	92.42
	决策树	100.00	92.00	95.83	94.67
	支持向量机	70.67	100.00	82.81	70.67

表五显示,第二轮识别结果与第一轮基本一致。决策树分类算法识别五种人格特质类型的整体效果最好,准确率和综合评价指标值都高于朴素贝叶斯和决策树分类算法。本研究进一步比较分析了三种分类算法的识别效果,结果见图2。

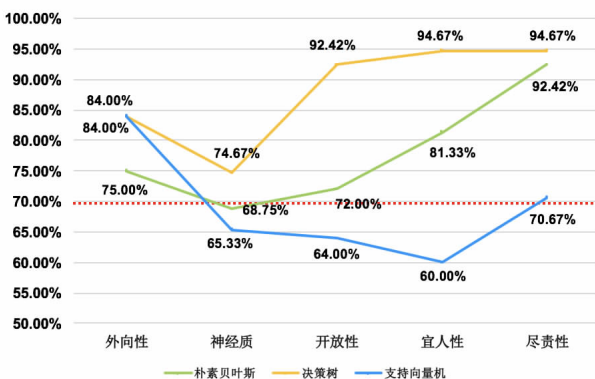


图2 三种算法识别五种人格类型的准确率

综合而言,三种分类算法对尽责性人格特质识别准确率较高,尤其是决策树和朴素贝叶斯算法识别准确率均为94%;其次是对外向性人格特质的识别准确率也相对较高,高于75%;同时三种算法识

别神经质人格特质的准确率普遍较低,这与第一轮识别结果基本一致。

四、结果与讨论

本研究利用机器学习分类算法探索基于在线行为数据的人格特质自动识别,为学习者个性特征获取和自动识别探索了新方法和思路。同时,人格特质类型与在线行为之间映射关系的建立为教育机构预测和判断学习者的兴趣及行为偏好,实现以学习者为中心的精准化、个性化和智能化的学生评价和学习服务提供了依据。

(一)人格特质与学习行为之间存在映射关系

本研究在文献调研基础上建立了人格特质类型和学习行为之间的映射关系,并通过斯皮尔曼等级相关法对其进行验证,研究结果与以往相关研究一致,再次证明不同人格特质学习者有不同的学习行为偏好。例如,神经质学习者的人格特质与浏览平台和教学资源、观看视频以及参加作业考核呈显著正相关;但与同伴交流和互评呈显著负相关。这一结果充分显示神经质学习者喜欢与学习资源交互,不喜欢与同伴交互的特点(张晓丽,2008;王陆,

2011)。同样,尽责性学习者与浏览视频及课件、完成作业、参与答疑等直接影响完成学业的行为显著相关,但与教师交流等行为不相关。这说明尽责性学习者自律性强,对个人要求较高,对学习有主动性和计划性等特点(李阳,2016)。同时,本研究结果证明不同在线学习行为表征不同人格特质类型的敏感性不同,因此基于在线学习行为数据识别人格特质是可行的(Ghorbani & Montazer, 2015)。本研究在一定程度上解决了个性化推荐中的数据稀疏问题,有助于提高教育机构个性化支持服务的准确度。

(二)不同分类算法识别准确率不同,决策树算法的识别准确率最高

研究发现决策树算法识别五种人格特质的效果最好,验证阶段也再次证明了这一结论。这与以往研究结果一致(Bai et al., 2013)。总体来说,不管是比较人格特质的识别准确率,还是比较综合评价的指标值,决策树的识别效果都明显高于朴素贝叶斯和支持向量机算法。决策树算法能很好地识别在线学习者的人格特质类型,准确率达80%~90%。其中,除神经质人格的识别准确率为80%外,其余四种人格特质的识别准确率超过90%。此外,召回率及综合评价指标在外向性、开放性、宜人性、尽责性的值都大于90%,神经质的值也超过70%。这一结果与分类算法及在线学习行为数据有关。本研究采用了不同层次的在线学习行为数据类型,而不同类型行为数据量不同,例如学习者浏览教学资源行为多,而论坛中的发帖和回帖行为少;还有部分学习者行为数据缺失。这些噪声干扰都会影响分类算法的结果。决策树算法不仅能运用信息增益率排除所选特征值繁多的缺陷,而且可以按照不同的剪枝标准进行排除以保证决策树的平衡,因此决策树算法对噪声数据的干扰有较好的鲁棒性(robust),识别人格特质计算量相对较小且效率高,识别准确率较高(蔡萌萌等,2019)。相比之下,朴素贝叶斯算法过分依赖严格的属性值独立性假设前提(魏茂胜,2017),而支持向量机对缺失数据非常敏感(杨剑锋等,2019)。因此,后两种分类算法识别的准确率明显偏低。本研究也验证了决策树算法不仅识别准确率高,具有很好的扩展性和适用性,能应用于不同在线学习环境中不同群体的人格特质识别。

(三)不同人格特质识别的灵敏度不同,尽责性

识别灵敏度最高

三种分类算法识别尽责性人格特质的灵敏度最高,这一结果与奥伯伦德尔(Oberlander)和诺森(Nowson)的研究结果一致。在本研究中,朴素贝叶斯和决策树算法分类器识别尽责性人格特质类型准确率都在90%以上,支持向量机的识别准确率也超过80%,这与不同人格特质的特征有关。高尽责性学习者自我管理能力强,不管是线上还是线下,都能按照课程大纲和上课进度,做好学习计划和安排,按时完成作业,积极参与学习活动,取得较好的学习成绩(王艳芝,2004;孙圣涛等,2006;黄真浩等,2014;王月明,2015)。由于高尽责性学习者的现实生活和网络学习行为表现基本一致,故其识别灵敏度较高。

外向性、开放性和宜人性格特质学习者,其典型的在线学习行为表现接近,都表现出倾向于与人交流,乐于分享自己的学习心得和经验,喜欢和同伴在论坛中讨论,积极参加在线学习活动等(Ghorbani et al., 2015;李阳,2016;Correa et al., 2010)。因此,这三种人格特质的识别一定程度上容易混淆,识别的准确率与尽责性准确率相对偏低。

同时,本研究发现神经质识别灵敏度最低,这一点和王佳佳及娜迪热等人的研究结果一致(王佳佳,2017;娜迪热等,2018)。原因可能在于:网络环境的虚拟性和匿名性为个体隐藏自己的人格特质,并产生与现实生活截然不同的行为表现提供了良好的场所(刘颖等,2018)。对神经质学习者来说,因其不善交际而在现实生活中少与老师和同学交流,但在在线学习环境中,神经质学习者往往表现出与现实完全不同的人格特质(Ahn et al., 2012;Wang et al., 2016)。因此,神经质学习者“伪装性”相对较高,造成识别准确率较低。

人格特质作为决定人行为的根本因素,对学习者的学习意愿和行为产生重要影响。实现在线学习者人格特质的自动识别对准确掌握学习者在线学习行为偏好和潜在需求有重要意义。本研究结果证明了基于在线学习行为数据的人格特质识别是可行的,实现了通过数据挖掘和人工智能技术,在无需学习者配合的情况下,自动认知和理解学习者的学习行为,获取学习者的个性特征,为教育机构探索智能化、个性化教学提供了依据。但人格特质结构复杂,

其反映在言语、行为、情感和态度等方面,在线学习行为数据只是其中的一个子集。随着计算机技术和学习分析技术的发展,学习者网络学习活动和行为可以得到全过程的记录、分析和监控,并由此形成以学习者为中心的学习活动、社交网络、情感状态等多维信息的立体化数据(Tommassel, A., et. al., 2016),因此如何将学习者的学业情绪、学习态度等特征纳入在线学习数据体系,实现多模态学习数据的精准量化,探讨基于多模态在线学习数据的学习者个性特征自动识别是后续研究的重点。

[参考文献]

- [1] Ahn, S. J., Fox, J. & Bailenson, J. N. (2012). Leadership in science and technology: A reference handbook[M]. SAGE Publications; 695-702.
- [2] Allport, G. W. (1937). Personality: A psychological interpretation[J]. American Journal of Sociology, 45(1):48-50.
- [3] Ally, M. (2004). Foundations of educational theory for online learning[J]. Theory and Practice of Online Learning, (2): 15-44.
- [4] Anderson, B., & Simpson, M. (2004). Learning and affective support online in small group and class contexts[J]. The International Review of Research in Open and Distributed Learning, 5(3):15-20.
- [5] Back, M. D., Schmukle, S. C., & Egloff, B. (2008). How extraverted is honey. bunny77@hotmail.de? Inferring personality traits from email addresses[J]. Journal of Research in Personality, 42: 1116-1122.
- [6] Bai, S., Hao, B., & Li, A. (2013). Predicting big five personality traits of microblog users[C]//Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)-Volume 01. IEEE Computer Society, 501-508.
- [7] Berge, Z. L. (1999). Interaction in post-secondary web-based learning[J]. Educational Technology-Saddle Brook, (39): 5-11.
- [8] 蔡萌萌,张巍巍,王泓霖(2019). 大数据时代的数据挖掘综述[J]. 价值工程, (38):155-157.
- [9] 陈丽(2004). 远程学习的教学交互模型和教学交互层次塔[J]. 中国远程教育, (5):24-28.
- [10] Correa, A. W., Hinsley, H. G., & De, Z. (2010). Who interacts on the Web?: The intersection of users' personality and social media use[J]. Computers in Human Behavior, 26(2): 247-253.
- [11] 冯晓英,郑勤华,陈鹏宇(2016). 学习分析视角下在线求知水平的评价模型研究[J]. 远程教育杂志, (4):39-45.
- [12] Fong, K., & Mar, R. A. (2015). What does my avatar say about me? Inferring personality from avatars[J]. Personality and Social Psychology Bulletin, 41(2): 237-249.
- [13] Ghorbani, F., & Montazer, G. A. (2015). E-learners' personality identifying using their network behaviors[J]. Computers in Human Behavior, 51: 42-52.
- [14] Gilbert, L., & Moore, D. R. (1998). Building interactivity into web courses: Tools for social and instructional interaction[J]. Educational Technology, 38(3): 29-35.
- [15] Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1997). Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing[J]. Journal of Educational Computing Research, 17(4): 397-431.
- [16] 黄定华(2009). 论人格特质[J]. 湖南城市学院学报,30: 10-13.
- [17] 黄真浩,白新文,林琳(2014). 尽责性和神经质影响拖延行为的中介机制[J]. 中国临床心理学杂志, (1):140-144.
- [18] John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy[J]. Handbook of personality: Theory and Research, 3(2): 114-158.
- [19] Kanuka, H., & Nocente, N. (2003). Exploring the effects of personality type on perceived satisfaction with web-based learning in continuing professional development [J]. Distance Education, 24(2): 227-244.
- [20] 李超,程昱,石磊(2017). 基于层次分析法的在线学习行为指标体系构建[J]. 河北广播电视大学学报,22:18-21.
- [21] 李红波,胡建兵(2004). 网络学习的评价模式[J]. 电化教育研究, (5): 47-50.
- [22] 李爽,王增贤,喻忱,宗阳(2016). 在线学习行为投入分析框架与测量指标:基于LMS数据的学习分析[J]. 开放教育研究, (2):77-88.
- [23] 李阳(2016). 大学生网络学习行为与人格特征的相关性研究[D]. 西安:西北大学:13-14.
- [24] 刘多星(2017). 基于微博文本情感计算的用户个性分析方法的研究[D]. 石家庄:河北科技大学:42-43.
- [25] 刘颖,赵宏(2018). 网络环境下虚拟人格研究进展与热点分析[J]. 开放学习研究, (4):20-26.
- [26] McCrae, R. R., & Costa, P. T. (1989). More reasons to adopt the five-factor model. [J]. American Psychologist, (44):451-452.
- [27] McCrae, R. R. (2005). 78 members of the personality profiles universal features of personality traits from the observer's perspective [J]. Journal of Personality and Social Psychology, (88):547-561.
- [28] McCrae, R. R., & Costa, P. T. (2004). A contemplated revision of the NEO Five-Factor Inventory [J]. Personality and Individual Differences, 36:587-596
- [29] Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life[J]. Journal of Personality and Social Psychology, 90(5), 862-877.
- [30] 娜迪热,胡俊(2018). 基于用户社交网络数据的人格倾向性分析及预测模型的建立[J]. 电脑知识与技术, (7):6-11.
- [31] Oberlander, J., & Nowson, S. (2006). Whose thumb is it an-

yway? Classifying author personality from weblog text [C]//Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions, 627-634.

[32] Qiu, L., Lin, H., Ramsay, J. E., & Yang, F. (2012). You are what you tweet: Personality expression and perception on Twitter [J]. *Journal of Research in Personality*, 46(6): 710-718.

[33] 孙圣涛,姚誉羚(2006).大学生时间管理倾向与应对方式、人格特质的相关研究[J].*中国临床心理学杂志*, (2):186-195.

[34] 孙月亚(2015).开放大学远程学习者在在线学习行为的特征分析[J].*中国电化教育*, (8):64-71.

[35] 陶楚楚(2018).基于人格特征的网络学习交互行为研究[D].锦州:渤海大学:29.

[36] 王佳佳(2017).社交网络用户人格特质挖掘及其在个性化推荐领域的应用研究[D].合肥:合肥工业大学:11.

[37] 王陆,方平,刘维民(2001).关于学生个性因素与远程教育相关性的探讨[J].*电化教育研究*, (6):28-32.

[38] Wang, Y. C., Hinsberger, H., & Kraut, R. E. (2016). Does saying this make me look good? How posters and outsiders evaluate facebook updates [C]//Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, 125-129.

[39] 王艳芝(2004).人格特质与大学生英语学习成绩关系研究[D].长春:吉林大学:33-38.

[40] 王月明(2015).研究生学业拖延相关问题.[D].杭州:浙

江理工大学:92.

[41] Watjatrakul, B. (2016). Online learning adoption: Effects of neuroticism, openness to experience, and perceived values [J]. *Interactive Technology and Smart Education*, 13(3): 229-243.

[42] 魏茂胜(2017).数据挖掘中的分类算法综述[J].*网络安全技术与应用*, (6):65-66.

[43] 杨剑锋,乔佩蕊,李永梅,王宁(2019).机器学习分类问题及算法研究综述[J].*统计与决策*, (6):36-40.

[44] 杨洁(2015).基于用户情感和网络关系分析的人格预测模型[D].上海:东华大学:11-37.

[45] 张晓丽(2008).大学生网络学习行为与人格特征之间的关系研究[J].*西北医学教育*, (16):847-852.

[46] 张艳红,佐斌(2013).网络师生互动的影响因素研究[C].北京:华人心理学家学术研讨会:109.

[47] 朱从庆,董宝林(2016).人格特质与大学生锻炼动机:情绪调节方式的中介效应[J].*武汉体育学院学报*, (50):94-100.

[48] 张琪,王红梅,庄鲁,赖松(2019).学习分析视角下的个性化预测研究[J].*中国远程教育*, (4):38-45, 92-93.

[49] 衷克定(2016).人格类型理论视角下的在线学习活动设计验证[J].*现代远程教育研究*, (2):81-89.

(编辑:赵晓丽)

Online Learning Behavior based Personality Recognition

ZHAO Hong¹, LIU Ying², LI Shuang¹, XU Pengfei¹ & ZHENG Qinhu¹

(1. Faculty of Education, Beijing Normal University, Beijing 100875, China;
2. Peking University Elementary School, Beijing 100875, China)

Abstract: Learners' personality is critical for student-centered and personalized teaching, and their learning behavior is essential for analyzing learners' personality. Based on the hypothesis that personality generates learning behavior preference, this study explored personality recognition based on online learning behaviors, using machine learning classification algorithm. Firstly, Big Five Personality questionnaire was administrated to online learners from Open Edutainment. Secondly, the mapping relation between personality traits and online learning behaviors was constructed based on literature and verified by Spearman correlation analysis. Finally, Rapid Miner was used to explore the identification effects of Decision Tree, Naive Bayesian, and Support Vector Machine (SVM) algorithms on five personality traits. The study found that the recognition accuracy and effect of the Decision Tree algorithm on each personality type is higher than the other two algorithms. Moreover, the sensitivity of different personality recognition is found to be different: the conscientiousness is the highest, and the neuroticism is the lowest.

Keywords: personality trait; online learning behavior; learning behavior preference; classification algorithms.