



基于模型集成的在线学习投入 评测方法研究*

□ 李振华 张昭理 刘 海

【摘 要】

针对慕课等在线学习课程存在的完成率低、辍课率高等问题,不少研究者通过检测学习者的学习投入度来发现“问题”学生,对其进行干预以保证学习效果。本文以构建在线学习投入自动化评测模型为目标,通过构建集成评测模型,利用学习过程中产生的视频图片和鼠标流数据对学习者的投入水平进行评测。集成模型由3个子模型组成,其中两个子模型用于进行图片数据的处理,一个子模型用于进行鼠标流数据的处理,图片部分的评测采用VGG16卷积神经网络对源图片和相应的LGCP特征进行评测,鼠标流数据采用BP神经网络进行评测。最后,利用模型集成的方法对学习者的学习投入度进行综合评测,再将其结果与学习者填写的NSSE-China调查量表的结果进行相关性分析,结果显示两者的评测结果显著相关,表明该模型用于学习投入评测是可行且有效的。

【关键词】 在线学习;学习投入;自动化评价;视频图片;鼠标流数据;卷积神经网络;BP神经网络;模型集成

【中图分类号】 G434

【文献标识码】 A

【文章编号】 1009-458x(2020)10-0009-09

DOI:10.13541/j.cnki.chinade.2020.10.002

一、引言

自2012年“慕课元年”以来,大量在线学习平台井喷式推出,涌现出许多具有影响力的慕课平台,如国外的Coursera、EdX、Udacity和国内的中国大学MOOC、学堂在线等。慕课凭借其“大规模”“开放”“在线”等特点吸引了大量学习者(陈肖庚,等,2013),但在实践过程中研究者发现大部分在线学习平台存在着辍课率高、完成率低的问题。Ho等(Ho, et al., 2014)研究了MIT推出的17门慕课课程的学习统计数据,发现最初的8.4万名学习者中只有5.1%的人顺利拿到了证书。Jordan(2015)对多个慕课平台课程的学习情况进行研究后发现:慕课课程的平均完成率只有15%,最高也只能达到40%。蒋卓轩等(2015)分析了近8万人的中文慕课学习行为数据后也发现类似的现象:学习者参与课程的情况总体呈现下降趋势,且前期下降较为迅速,后期下降较为平缓。因此,提高在线学习的完成率、降低辍课率,提

升在线学习的学习效果,成为研究者普遍关心的问题。其实,辍课率高、课程完成率低的问题在传统的学校教育早已存在,并非在线教育所特有。早期的研究者认为这些问题的出现主要跟学习者的学习投入不高有关,不少研究发现学生学习结果与学习投入程度存在正向关系(Kuh, 2001, 2009)。因此,研究者们从学习者学习投入评测的角度出发,通过评测发现学习投入不高的学生并及时给予干预,以此保证学习效果。同时,对于学习者在线学习情况的精准评测有助于推动各类在线教学平台的迭代开发以及为研制学科工具提供参考数据(刘海,等,2018),也为打造个性化、智能化的教学环境提供技术支持(刘海,等,2018;张昭理,等,2018)。

二、学习投入评测相关研究

界定学习投入概念的结构是开展评测分析的重要前提和基础。学习投入被普遍认为是一个多维度的元概念,但学术界对于学习投入的维度构成并未达成一

* 基金项目:本文受国家重点研发计划课题“数据驱动的数字教育个性化服务支撑技术研究”(项目编号:2017YFB1401303)和西华师范大学英才科研基金“基于面部表情识别的在线学习情感投入检测”(项目编号:17YC496)资助。

致认识,从二维度到四维度均有支持者。但是Fredricks等人(Fredricks, Blumenfeld, & Paris, 2004)提出的“行为、认知、情感”三维度说获得了较多的支持和认可。

学习投入的评测研究是在明确了概念维度的基础上针对某一个或几个维度,运用特定的方法和技术手段来开展的。常用的评测方法包括:①自我报告法(Shernoff, Csikszentmihalyi, Shneider, & Shernoff, 2003);②经验抽样法(Fredricks & McColskey, 2012);③教师打分法(Wigfield, et al., 2008);④访谈法;⑤观察法(Greenwood, Horton, & Utley, 2002)。

学习投入的评测对于解决在线学习情境下学生辍课率高、完成率低的问题同样具有借鉴意义。不过,在在线学习过程中因受到师生时空分离、学生规模庞大、教师对学生的约束力减弱等因素影响,传统教育情境下的评测方法并不能很好地胜任。因此,在线学习情境下日志数据分析通常是评测学习者行为投入的常用方法。例如,Hamane等人(2014)、Sundar等人(2016)以及李爽等人(2016)的研究都通过分析在线学习系统中产生的日志数据来实现对学习投入的评测或成绩的预测。他们通常的做法是从日志数据中抽取或构建出反映学习投入不同维度的指标变量,利用回归分析、结构方程等方法构建评测模型。但是,日志数据主要针对学习行为进行记录,在表征情感和认知维度的投入情况方面存在局限,而且其记录的方式是基于事件触发,这使得所记录的投入状态数据是离散且不均匀的,影响了评测的准确性。而且在学习投入的各维度中认知、情感方面的投入是学习过程中的实质性投入,能促进高阶思维的发展并影响学习结果和体验(Nystrand & Gamoran, 1991)。因此,为了对情感和认知投入进行测量,不少研究者将脑电传感器、皮肤感应传感器、血压仪、心率仪等传感设备用于对学习者的投入状态和兴奋水平进行测量(Chaouachi, Chalfoun, Jraid, & Frasson, 2010; Goldberg, Sottolare, Brawner, & Holden, 2011)。基于传感器的评测方法可对学习过程中的生理指标进行全面的记录,能对学习中的情感和认知维度进行有效的表征,但是这些设备主要适用于实验室环境,易用性和经济性不高。Booth等人(2017)则利用摄像头对学习者的面部表情进行识别以判断学习投入。该方

法主要利用计算机视觉方法对学生的表情状态进行判断,因此对情感投入的评测具有较好的效果,加之使用方便、成本较低,易于普及。

从技术发展的视角来看,学习投入评测的发展过程大致呈现出从人工收集数据进行分析到借助信息系统自动收集数据进行分析,从粗粒度分析到细粒度分析,从定性分析到定量分析的发展轨迹。评测效率和精度的提升在很大程度上得益于信息技术的采用或技术本身的更新、升级。随着物联网、大数据和人工智能技术的兴起,各种新技术正尝试寻找与在线学习评测领域的最佳结合点,助推学习投入评测向自动化、智能化、精细化方向发展。

三、基于模型集成的学习投入评测

学习投入本身具有多维度的特点,涉及行为、情感和认知多个维度(Fredricks, et al., 2004)。现有的研究大多以单一来源的数据作为基础开展评测以证实对于学习投入评测的有效性。在这些研究成果的基础之上,我们希望能更进一步探索基于多种来源数据的学习投入评测,这不仅有助于得到更为精准的评测结果,而且是构建自动化评测系统的重要技术基础。因此,我们期望将多个评测结果进行融合来实现对多个维度投入的评测,提升评测的准确性。本研究针对在线学习过程中的情感和认知维度的评测开展研究,采用摄像头和鼠标作为主要的数据采集装置,对学习者的在线学习投入度进行评测。

(一) 实验设计的研究基础

由于人类的情感主要通过面部表情进行表达,采用计算机视觉技术对摄像头采集的人脸表情数据进行分析可以了解学习者在线学习过程中的情感投入状况。除此以外,也有研究发现学习者的头部和身体姿态与情感和认知投入之间存在确切的联系,并通过实验证实了头部或身体姿态用于情感和认知投入分析的可行性(Ballenghein & Baccino, 2019; D'Mello & Graesser, 2009)。

认知投入属于学习者的内隐状态,除了通过量表进行分析外,心理学领域通常对阅读过程中的眼动行为与认知加工之间的关系进行分析和研究。在接近一百多年的研究历史中,研究者提出了多个眼动控制模型,普遍认为阅读过程中的注视时间、凝视时间、眼



跳距离、回扫频率等眼动行为参数与认知加工之间存在着确切的基本对应关系(赵晶,等,2007)。Miller(2015)在此基础上通过分析阅读者在自主阅读情境下的眼动行为实现了在微观层面对学习投入进行测量。在以自主阅读为主要学习方式的在线学习中,眼动分析是测量学习投入较为有效的方法之一,但较高的设备成本限制了眼动分析方法的广泛使用。

鼠标作为学生与学习系统、学习资源进行交互的主要输入设备,具有较强的表征能力。鼠标点击流数据记录了学习者进出系统各学习模块和页面的时间、次数、先后顺序等信息,基于这些信息所形成的日志数据常用于对学习行为投入进行评测(李爽,等,2016;李爽,等,2017;李爽,等,2018)。此外,多项鼠标轨迹和眼动轨迹的相关性研究表明,在进行在线阅读或相关搜索任务的过程中,鼠标的移动轨迹和眼动轨迹之间存在着较强的相关关系和固定的“眼-鼠”跟随模式(Chen, Anderson, & Sohn, 2001; Huang, White, & Buscher, 2012; Liebling & Dumais, 2014),而且眼动轨迹通常引导着鼠标的移动(Liebling & Dumais, 2014)。Lagun和Agichtein(2011)在更大规模下开展实验后发现,在不使用眼动设备的情况下,鼠标移动数据的结果与实验室环境下的眼动跟踪结果是相关联的。可以看出,鼠标数据本身不仅仅表征了学习者在行为方面的投入,而且在在线阅读的情况下,通过鼠标轨迹数据可以对学习者的阅读速度、停留时长和回视浏览等情况有较为细致的了解。因此,鼠标移动数据也具备和眼动数据一样的表征学习者认知投入状态的能力。

本研究以摄像头和鼠标流数据作为来源数据分别构建评测模型,最后通过模型集成的方式将两个维度的评测模型进行融合,并将评测结果与学习者填写的量表数据进行对比来检测模型的实际评测效果。

(二) 实验流程

实验分为四个主要部分:数据收集、数据标注、模型训练和模型检验。

1. 数据收集

由于需要采集学习过程中的摄像头数据和鼠标数据,本研究主要针对学习者面对桌面电脑进行“一对一”自主学习的学习情景。所谓“一对一”是指一个学习者面对一台学习终端(台式电脑或笔记本电脑)进行学习的情况。其中,摄像头固定于显示屏上沿用

来采集学习者的正面表情,鼠标则记录学习者对学习页面的控制情况,实验中的学习页面主要由图片、文字、动态图和视频构成,以垂直滚动的方式进行浏览。整个实验示意图见图1。

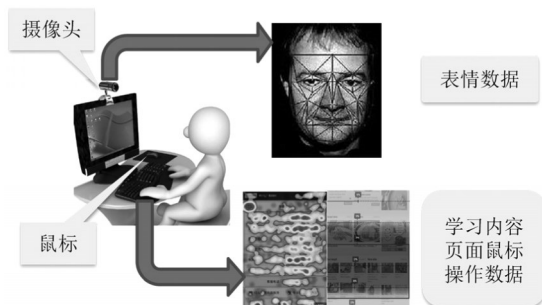


图1 学习者在线学习示意图

待相关数据收集完毕后,学习者在学习结束时还需要填写NSSE-China调查表以确定实际的投入状态水平,所谓NSSE-China是《全美学习投入调查量表》(National Survey of Student Engagement, NSSE)的中文版(罗燕,等,2009)。学习投入的评测量表除了NSSE-China外,还有《学生学习投入量表》(Utrecht Work Engagement Scale-Student, UWES-S),不过UWES-S主要从动机、精力和专注三个维度开展评测,对于学习过程中的交互、情感和认知维度的评测较为薄弱。相比之下,在对于行为、情感和认知维度的覆盖上NSSE-China问卷要更为全面。

2. 数据标注

采集到的视频流数据需要处理为离散的图片数据并进行投入水平的标注以方便进行评测模型的训练。标注人员对图片数据进行标注,为所有离散图片中学习者的投入状态水平进行打分,投入水平值为1、2、3,分别对应“不投入”“一般性投入”“非常投入”三种投入状态。

3. 模型训练

构造多模态数据评测模型,对图像数据和鼠标移动数据进行综合评测得到最终的学生学习投入水平。因为标记图片数据规模依然偏小,因此通过增加噪声、反转、旋转、尺度变换等操作扩充数据并构建训练数据集。在模型训练阶段,将数据集划分为训练集和测试集并对模型进行训练。训练过程中采用10折交叉验证的方式进行模型选择。

4. 模型检验

将评测模型得到的学生投入结果与学生填写的

NSSE-China 量表得到的投入结果进行对比以检验模型的准确性,如果两个结果存在较强的相关关系就说明面部表情数据和鼠标移动数据能够表征学习者在线学习投入状态,可以作为学习投入评测的依据。

(三) 实验组织

本研究以某高校2018年上学期网络公共选修课“知识管理方法和实践”的一个班学生作为研究对象,共47名同学,均来自各个学院研一研二年级。其中,男生19名,女生28名;最大的26岁,最小的23岁,整体平均年龄为 24.47 ± 0.87 岁;学生的院系来源组成,生命科学学院10.6%,计算机学院21.3%,经管学院17%,教育学院51.1%。主要的学习方式为学生使用电脑进行“一对一”在线自主学习,教师通过在线互动的方式开展辅导,在整个学习过程中学生可以使用系统中的博客、留言、Wiki等模块进行学习、交互等活动,同时也被要求尽量独立进行在线学习,不要出现多人同时学习交流的情况。在线学习管理系统记录学生的日志行为数据,摄像头记录学生的学习视频数据,鼠标则记录学习页面的滚动浏览情况和对应的时间戳。

(四) 图片数据的标注方法

由于在在线学习过程中学生表情较少且表情背后所代表的含义与一般情境下的表情含义有所不同,学习过程中主要以中立表情为主,其他表情则主要分布在无聊、困惑、满足等几种表情之中(D'Mello, 2013)。因此,通用的表情数据集并不适用于学习评测模型的构建,而应该使用专门的数据集进行模型训练。本研究中采用Whitehill等人(Whitehill, Serpell, Lin, Foster, & Movellan, 2014)提出的从视频数据中提取静态图像,然后进行人工标注的方法构建训练数据集。这些视频图片均是从视频数据中以固定时间间隔进行提取的。在固定时间间隔的设置上,我们主要从图像关键帧丢失和数据量大小两方面进行考虑:一方面,间隔时间如果设置过大会丢失含有关键表情的图像帧;另一方面,如果间隔时间设置过小则会导致提取的图片过多,进而增加图像标注的工作量。我们参照Whitehill等人(2014)提出的方法,并通过对学习者的学习视频进行抽样分析后发现以2秒为时间间隔将视频数据转换为图片序列在图像帧丢失和数据

量大小方面满足我们的要求。在获取了图像序列后,标注人员观察每张图片 and 与该图片时间戳相对应时刻的鼠标移动数据,然后给图片标注1到3的学习投入水平值。最终得到每位学习者不同学习时间学习投入水平的序列值,不同学习投入水平的图示见图2。

该方案的优点在于:①单张图片便于进行人工识别,处理上相对简单;②在统一的评判标准下,图片中学生的投入水平容易确定且发生模棱两可的情况较少。对此,我们基于Whitehill等人(2014)提出的标注方法拟定了一个指导性的学习投入度评判标准。标注人员根据标准分别从图片中学习者的头部姿态、视线位置、表情状态和鼠标轨迹对图片进行投入水平的评判。三种投入水平的参考标准为:

“非常投入”主要表现为学习者头部姿态是正位或者前倾,视线点位置落在屏幕区域内,表情以中性表情为主,并且主要呈现出聚精会神的状态,鼠标在整个学习页面中位置变化的幅度较小;

“一般性投入”主要表现为学习者头部基本处于正位或存在轻微偏斜,视线点位置在屏幕区域内,表情以自然表情为主,精力集中的程度较“非常投入”的状态要轻,鼠标位置变化的幅度较大;

“不投入”主要表现为头部姿态不正,无法看到正脸,视线不在屏幕上或呈现闭眼的状态,表情方面主要表现为漫不经心、无聊、沮丧等,鼠标移动的幅度变化一般处于静止和剧烈变化两个极端。

该标准对于标注员来说并不是硬性的,在实际标注过程中很多情况下依然需要标注员借助自身经验进行灵活判断。因此,图片标注的准确性受标注人员自身专业素质、图片可辨析程度等因素影响,不同标注人员针对同一图像的标注结果可能会有不一致的情况发生。我们通过Kendall系数来检验不同标注人员的标注结果一致性,以保证标注图像具有较高的一致

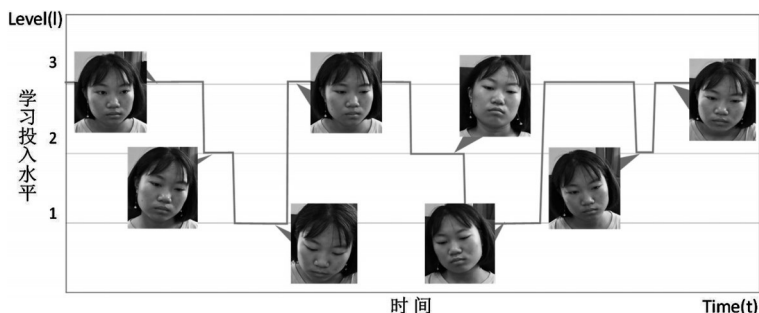


图2 某学习者不同学习投入水平的跳转示意图



性。所有表情图片的训练数据分为10个批次进行标注,每个批次图像均分给8个标注员完成,并确保标注数据满足一致性要求。最终,我们通过SPSS计算出所有标注员标注结果的Kendall系数,如表1所示。

表1 标注员标注结果的Kendall一致性检验

N	Kendall's W	Chi-Square	df	Asymp. Sig.
50	0.756	21.951	7	0.01

(五) 鼠标移动数据的获取

学习者与在线学习管理系统间的互动主要通过鼠标来完成,尤其是学习者在学习页面中的阅读情况,譬如注视时间、阅读速度、页面滚动或翻页频率等情况都可以借助鼠标的移动、停留、滚轮滚动等数据反映出来。Navalpakkam等人(2013)就发现用户在浏览网页时鼠标的移动轨迹和视线移动轨迹存在较大的相关性,而且可以用来预测用户在访问网页时的注意力。Huang等人(2012)总结出用户在浏览网页以及完成搜索任务时,鼠标与视线之间还存在较为固定的跟随模式。因此,通过检测鼠标在整个学习页面中的位置能够间接反映学习过程中学习者的视线和注意力的变化情况。但是和利用鼠标轨迹进行基于生物学测定的身份识别任务不同,身份识别的任务通常以100毫秒作为时间间隔(Rodden & Fu, 2007; 沈超,等,2010)详细记录鼠标轨迹以及移动过程中带有个人特征的动作细节,而本实验主要关注的是鼠标的位置变化,采样太过精细会增加与分析任务无关的冗余信息,因此实验中我们以200毫秒为间隔时间记录鼠标光标在页面中的绝对位置。该时间间隔下采集的轨迹数据在保留主要位置信息的同时也减弱了移动轨迹的起始和结束端附近区域鼠标点过于密集的情况。此外,在实验中与鼠标位置同时记录的还有时间戳信息。时间戳主要用于与视频图片的时间戳保持同步。在标注阶段,这些鼠标移动数据与图片数据都被标注相同的投入水平值。

(六) 数据集的划分

本研究以8:2的比例划分训练集和测试集,训练集的数据采用10折交叉验证(10-fold cross-validation)的方式进行模型的训练和参数的调优,测试集的数据用于测

试模型的泛化能力。在标记数据的过程中我们发现,学习者在每次学习过程中的投入行为并非均匀分布,表现出前段学习投入较高、中段较低、后段又较高的普遍情况。因此,为了确保训练集和测试集数据分布的一致性,我们根据课程时间段将数据分为前、中、后三部分,在三部分混合的基础上再进行训练集、测试集的划分以确保数据分布尽量均匀。数据集的划分方法如图3所示。

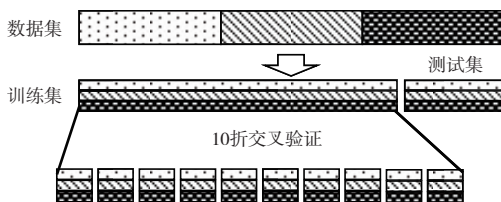


图3 数据集的划分方法

(七) 集成模型的构建

针对本研究的数据特点,我们构建了一个融合卷积神经网络和BP神经网络的集成评测模型,整个模型集成了三个子模型,其中两个卷积神经网络子模型用于对学习者的表情图片进行识别,一个BP神经网络子模型用于对学习者的鼠标移动数据进行识别,整个模型如图4所示。

卷积神经网络(CNN)多用于图像识别等视觉任务(Agrawal & Mittal, 2020),其“端到端”和“数据驱动”的特点相比传统视觉识别方法具有更好的识别效果。本研究中使用经典的VGG16模型进行图片的识别。为加强图片识别效果,其中一个子模型采用基于源图像的自适应加权局部格雷码模式(Local Gray Code Patterns, LGCP)的特征提取方法(吉训生,等,2017)来进行识别,该方法能够克

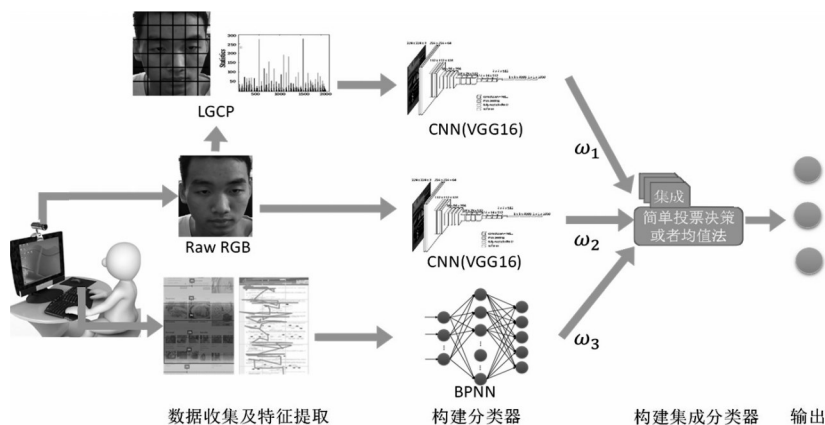


图4 集成评测模型

服LBP对非单调光线变化敏感的问题。针对鼠标移动数据的识别采用BP神经网络模型,其中输入层选取表情图片所属时间戳前20秒内的鼠标移动数据作为输入(总共100个鼠标坐标点),隐含层节点数为20,输出节点数为3。三个子模型的输出层节点数为3个,分别对应1到3的学习投入度水平值。每个模型通过softmax函数计算出每个输出节点取值的概率值大小,最后三个子模型通过投票法得到集成后的投入水平值。最终投入水平的计算公式为:

$$p(a_j|x, E_1, \dots, E_k) = \sum_{k=1}^K \omega_k p(a_j|x, E_k)$$

$$j = 1, 2, \dots, m; l = 1, 2, \dots, n.$$

(1)

式中 j 表示3个学习投入水平。 x_l 表示训练样本, E_k 表示第 k 个分类器, ω_k 是第 k 个分类器的权重值。不同分类器的权重值的计算采用投票的方式进行:

$$\omega_k = R_k / \sum_{k=1}^K R_k; k = 1, 2, \dots, K \quad (2)$$

式中 K 代表分类器总个数, R_k 代表第 k 个分类器的识别率。三个子模型的训练准确性曲线如图5所示。

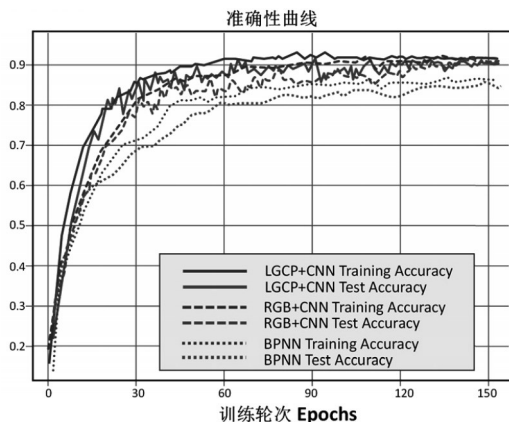


图5 三个子模型的准确性曲线

(八) 集成模型的检验

集成模型的检验是在三个子模型完成训练之后将三个模型利用投票的方法进行集成,并利用测试集数据对集成模型进行测试。最终,集成模型和子模型的对比结果如表2所示。

表2 三个子模型与集成模型对比情况

网络结构	训练集准确率	测试集准确率
LGCP+CNN(VGG16)	91.35%	89.40%
RGB+CNN(VGG16)	89.42%	86.73%
BPNN	79.56%	77.82%
集成模型	—	92.46%

可以看出,经过集成后模型对测试数据的准确率有一定的提升,这也说明通过将多个模型进行集成来提高评测准确率的方法是有效的。

最后,将学生学习数据输入集成模型计算出学生的投入水平,并计算这些结果与学生填写NSSE-China量表得到的投入水平结果之间的Spearman相关系数,进行相关分析。其中,量表的投入水平值的计算方法是:将所有反向计分题的分值进行反转,求得各个维度的得分均值,将各维度均值求和作为最终的学习投入水平值。最终结果如表3所示。

表3 集成评测模型与NSSE-China评测结果的对比

		评测模型	NSSE-China
集成评测模型	Spearman Correlation	1	0.788*
	Sig. (2-tailed)		0.023
	n	47	47
NSSE-China	Spearman Correlation	0.788*	1
	Sig. (2-tailed)	0.023	
	n	47	47

* 相关性在0.05的水平(双尾)上显著。

可以看出集成评价模型的最终评测结果与学生填写量表结果在 $P < 0.05$ 的水平上显著相关,这说明集成模型的评测结果与量表相近,即集成评测模型通过分析学习者的客观数据能够得到与量表评价相似的结果,也间接说明了学生的学习投入状态是会通过表情和阅读行为表现出来的,通过相关传感器设备对这些客观数据进行捕捉和分析能够实现学习投入度的自动化评测。

四、结果讨论

从本研究的结果可以明显看出:

1. 从单个评测模型的准确率来看,基于图像的评测模型效果优于基于鼠标移动数据的评测模型。我们认为主要原因在于表情数据本身对情感维度的表征比较准确,特定的情感状态一定对应着特定的表情,学习者不会在学习过程中露出与实际情感不相符的表情。鼠标数据在表征阅读行为时有一定的误差,虽然鼠标存在着追随视线的固定移动模式,但是在时间上存在着滞后性,而且存在一些突发的不可预知的鼠标动作,例如学习者左右晃动鼠标来确认光标位置的行为都会对分析造成影响。此外,鼠标停留的状态可能表示学习者在思考,也有可能表示学习者在开小差。针对



这个问题,一方面可以参考其他传感器数据来弥补仅依靠单传感器进行评测的不足,另一方面可以通过使用更合适的传感器来解决,如眼动仪等。

2. 基于多模态数据的集成模型比单模态数据模型有更好的评测准确度,不过准确度提升尚不够显著。对于集成模型能够提升评测效果较容易理解:由于现实中很难找到一种模态数据能够全面地表征学习投入,任何单一维度的数据只能从某一个或几个方面对学习投入进行表征。因此,对多维度数据进行综合分析能够得到更为准确的结果,增加其他模态的数据对于提升模型评测的准确性是有益的。对于准确度提升不够明显的问题,我们认为一方面是模型在融合策略的选择以及子模型训练、参数优化等方面存在不足,导致性能提升有限;另一方面,可能因为表情数据和鼠标移动数据在表征学习投入的维度上存在重叠,使图片数据和鼠标数据对某一学习投入维度或者子维度进行了重复评测。例如,本研究中我们认为学习过程中的情感主要通过表情体现出来,因此主要采用表情数据进行情感分析,但是一些基于具身认知理论的研究发现鼠标轨迹和情感之间同样存在着相关性(Yamauchi & Xiao, 2018)。在这种情况下,融合更多的模态数据以尽可能地全面覆盖学习投入的所有维度似乎是提升评测准确性的最佳解决办法。当然这也需要在后续的研究中进一步验证。

此外,本研究的局限在于学生样本数偏小,使得模型的通用性受到一定影响。另外,受实验组织的限制,学习内容方面无法估计不同科目和不同难度系数的学习内容对学习者的学习投入所造成的影响大小,这也使得收集到的与学习投入相关的表情和鼠标数据不一定能覆盖到所有可能的情况,进而会对模型的识别能力造成一定影响。不过,考虑到模型本身是通过数据驱动的,随着训练数据集质量的提高和数量规模的增大,集成模型的性能表现可能会得到进一步提升。因此,如果能在较大范围内开展试验,通过控制不同难度和不同学习内容对学习者的影响,以获得更为丰富和更具代表性的数据,对于提升模型的准确性和通用性是非常有益的。

五、小结与展望

本研究针对“一对一”在线学习情境下学习者的

在线学习投入自动评测问题开展研究,通过构建融合表情识别和鼠标移动数据分析的集成模型来进行评测,其中子模型分别采用了VGG16深度网络模型和BP神经网络模型,并利用自我标注的数据集进行训练。最终,集成模型的评测结果与学生填写的NSSE-China调查量表的结果显著相关。本研究结果说明在在线学习过程中学生所产生的各个维度的传感数据可被捕捉、量化,并可用于对学习投入的分析,也进一步说明了采用多模态数据进行学习投入评测的可行性。不过,由于实验数据来自较小的学生群体,影响了模型的通用性,需要在后续的研究中进一步改进和检验。本研究为未来的研究提供了几点思路:第一,在现有研究基础上能否再进一步集成多个模态的数据,且模型的表现随着集成模态数据的增多是否呈现简单的线性关系值得更进一步研究。第二,本研究的数据集中学生出现的表情类型不够丰富,而通用的表情数据集涵盖了人类较为典型的基本表情,这些表情数据能否用于评测模型的训练也需要进一步验证。第三,本研究的应用场景主要是针对使用桌面电脑或者笔记本的情况。对于使用手机和平板电脑的移动学习场景,如何实施有效的学习投入评测值得进一步探索。

【参考文献】

- 陈肖庚,王顶明. 2013. MOOC的发展历程与主要特征分析[J]. 现代教育技术,23(11):5-10.
- 吉训生,王荣飞. 2017. 自适应加权LGCP与快速稀疏表示的面部表情识别[J]. 计算机工程与应用,53(1):158-162.
- 蒋卓轩,张岩,李晓明. 2015. 基于MOOC数据的学习行为分析与预测[J]. 计算机研究与发展,52(3):614-628.
- 李爽,王增贤,喻忱,等. 2016. 在线学习行为投入分析框架与测量指标研究——基于LMS数据的学习分析[J]. 开放教育研究,22(2):77-88.
- 李爽,钟瑶,喻忱,等. 2017. 基于行为序列分析对在线学习参与模式的探索[J]. 中国电化教育(3):88-95.
- 李爽,李荣芹,喻忱. 2018. 基于LMS数据的远程学习者学习投入评测模型[J]. 开放教育研究,24(1):91-102.
- 刘海,陈莹莹,张昭理,等. 2018. 多学科资源工具研制与教学实践创新研究[J]. 电化教育研究,39(4):46-51.
- 刘海,李姣姣,张维,等. 2018. 面向在线教学平台的数据可视化方法及应用[J]. 中国远程教育(1):37-44.
- 罗燕,海蒂·罗斯,岑逾豪. 2009. 国际比较视野中的高等教育测量——NSSE-China工具的开发:文化适应与信度、效度报告[J]. 复旦教育论坛,7(5):12-18.
- 沈超,蔡忠阔,管晓宏,等. 2010. 基于鼠标行为特征的用户身份认证与监控[J]. 通信学报,31(7):68-75.

- 张昭理, 李阳, 刘海. 2018. 多屏多点触控教学系统支撑下教学创新与变革[J]. 电化教育研究, 39(3):82-89.
- 赵晶, 陈巍, 曹亮. 2007. 当代眼动研究对阅读心理学介入综述[J]. 浙江教育学院学报(1):25-30.
- Agrawal, A., & Mittal, N. (2020). Using CNN for facial expression recognition: A study of the effects of kernel size and number of filters on accuracy. *The Visual Computer*, 36(2), 405-412.
- Ballenghein, U., & Baccino, T. (2019). Referential processing during reading: Concurrent recordings of eye movements and head motion. *Cognitive Processing*, 20(3), 371-384.
- Booth, B. M., Ali, A. M., Narayanan, S. S., Bennett, I., & Farag, A. A. (2017). Toward active and unobtrusive engagement assessment of distance learners. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 470-476).
- Chaouachi, M., Chalfoun, P., Jraidi, I., & Frasson, C. (2010). Affect and Mental Engagement: Towards Adaptability for Intelligent. In *FLAIRS Conference*.
- Chen, M. C., Anderson, J. R., & Sohn, M. H. (2001). What can a mouse cursor tell us more? In J. A. Jacko, A. Sears, & J. Arnowitz (Eds.), *CHI 2001: Anyone, anywhere: CHI 2001 extended abstracts, conference on human factors in computing systems* (p. 281). New York, NY: Association for Computing Machinery.
- D' Mello, S. (2013). A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of educational psychology*, 105(4), 1082-1099.
- D' Mello, S., & Graesser, A. (2009). Automatic Detection Of Learner's Affect From Gross Body Language. *Applied Artificial Intelligence*, 23(2), 123-150.
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research*, 74(1), 59-109.
- Fredricks, J. A., & McColskey, W. (2012). The Measurement of Student Engagement: A Comparative Analysis of Various Methods and Student Self-report Instruments. In S. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 763-782). New York: Springer.
- Goldberg, B. S., Sottolare, R. A., Brawner, K. W., & Holden, H. K. (2011). Predicting Learner Engagement during Well-Defined and Ill-Defined Computer-Based Intercultural Interactions. In S. D' Mello, A. Graesser, B. Schuller, & J.-C. Martin (Eds.), *Affective Computing and Intelligent Interaction* (pp. 538-547). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Greenwood, C. R., Horton, B. T., & Utley, C. A. (2002). Academic engagement: Current perspectives in research and practice. *School Psychology Review*, 31(3), 328-349.
- Hamane, A. C. (2014). *Student engagement in an online course and its impact on student success* (Doctoral dissertation). Retrieved from Pepperdine University. (2014.3615910).
- Ho, A. D., Reich, J., Nesterko, S. O., Seaton, D. T., Mullaney, T., Waldo, J., et al. (2014). HarvardX and MITx: The first year of open online courses, fall 2012-summer 2013. *Social Science Research Network* (2014). <http://ssrn.com/abstract,2381263>.
- Huang, J., White, R. W., Buscher, G. (2012). User See, User Point: Gaze and Cursor Alignment in Web Search. *Proceedings of the 2012 ACM annual conference on human factors in computing systems* (pp. 1341-1350). New York, NY: ACM.
- Katy Jordan (2015, June 12). MOOC completion rates. Retrieved Oct. 23, 2019 from <http://www.katyjordan.com/MOOCproject.html>
- Kuh, G. D. (2001). The National Survey of Student Engagement: Conceptual framework and overview of psychometric properties. *Bloomington, IN: Indiana University Center for Postsecondary Research*, 126.
- Kuh, G. D. (2009). The national survey of student engagement: Conceptual and empirical foundations. *New Directions for Institutional Research*, 2009(141), 5-20.
- Lagun, D., Agichtein, E. (2011). ViewSer: enabling large-scale remote user studies of web search examination and interaction. In W.-Y. Ma (Ed.): *ACM Digital Library, Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval* (p. 365). New York, NY: ACM.
- Liebling, D. J., Dumais, S. T. (2014). Gaze and mouse coordination in everyday work. In A. J. Brush, A. Friday, J. Kientz, J. Scott, & J. Song (Eds.), *UbiComp' 14 adjunct: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing : September 13-17, 2014, Seattle, WA, USA* (pp. 1141-1150). New York: Association for Computing Machinery.
- Miller, B. W. (2015). Using Reading Times and Eye-Movements to Measure Cognitive Engagement. *Educational Psychologist*, 50(1), 31-42.
- Navalpakkam, V., Jentzsch, L., Sayres, R., Ravi, S., Ahmed, A., & Smola, A. (Eds.) (2013). *Measurement and Modeling of Eye-mouse Behavior in the Presence of Nonlinear Page Layouts*: Proceedings of the 22nd international conference on World Wide Web.
- Nystrand, M., Gamoran, A. (1991). Instructional discourse, student engagement, and literature achievement. *Research in the Teaching of English*, 261-290.
- Rodden, K., & Fu, X. (2007). Exploring how mouse movements relate to eye movements on web search results pages. *Proceedings of ACM SIGIR 2007*, 29-32.
- Shernoff, D. J., Csikszentmihalyi, M., Shneider, B., Shernoff, E. S. (2003). Student engagement in high school classrooms from the perspective of flow theory. *School Psychology Quarterly*, 18(2), 158-176.
- Sundar, P., Kumar, A. S. (2016). Disengagement detection in online learning using log file analysis. *International journal of computer technology and applications*, 9(27), 195-301.

(下转第 60 页)

- Prinsloo, P., & Slade, S. (2015). Student privacy self-management: implications for learning analytics. In: *LAK'15: Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 83-92). Retrieved from <https://dl.acm.org/doi/abs/10.1145/2723576.2723585>
- Reich, J., & Ruipérez-Valiente, J. A. (2019). Science Magazine. Retrieved from https://www.sciencemaginedigital.org/sciencemagazine/11_january_2019/MobilePagedArticle.action?articleId=1455537&ap=false#articleId1455537
- Schwartz, S. (2019). School Districts Are Banning Teachers From Using DonorsChoose. Education Week. Retrieved from, 2019.http://blogs.edweek.org/teachers/teaching_now/2019/03/donors_choose_district_ban.html
- Sokolova, K., & Perez, C. (2018). The digital ingredients of donation-based crowdfunding. A data-driven study of Leetchi projects and social campaigns. *Journal of Decision Systems*, 27:3, 146-186.
- United Nations. (2005). *Understanding Knowledge Societies: In twenty questions and answers with the Index of Knowledge Societies*. New York: United Nations. Retrieved from <https://publicadministration.un.org/publications/content/PDFs/E-Library%20Archives/2005%20Understanding%20Knowledge%20Societies.pdf>
- United Nations Educational, Scientific and Cultural Organization (UNESCO). (2019). *Recommendation concerning Open Educational Resources (OER)*. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000370936>
- University of Washington. (2020). Creating Accessible Videos. Retrieved from <https://www.washington.edu/accessibility/videos/#free>
- Weller, M., de los Arcos, B., Farrow, R., Pitt, R., & McAndrew, P. (2016). Identifying Categories of Open Educational Resource Users. In P. Blessinger & T. J. Bliss (eds.), *Open Education : International Perspectives in Higher Education* (pp. 73-91). New edition [online]. Cambridge: Open Book Publishers. Retrieved from <https://books.openedition.org/obp/3545?lang=en>
- Wenger-Trayner, E., & Wenger-Trayner, B. (2015). Introduction to communities of practice. Retrieved from <https://wenger-trayner.com/introduction-to-communities-of-practice/>
- Wiley, S. (2007). Noncommercial Isn't the Problem, ShareAlike Is. Iterating toward openness (weblog). Retrieved from <https://opencontent.org/blog/archives/347>
- Wiley, D. (2017). If We Talked About the Internet Like We Talk About OER: The Cost Trap and Inclusive Access. Iterating toward openness (weblog). Retrieved from <https://opencontent.org/blog/archives/5219>
- Woollacott, E. (2019). LinkedIn Data Scraping Ruled Legal. *Forbes*. Retrieved from <https://www.forbes.com/sites/emmawoollacott/2019/09/10/linkedin-data-scraping-ruled-legal/#4312d5a91b54>

收稿日期:2020-03-02

定稿日期:2020-04-29

作者简介:史蒂芬·道恩斯(Stephen Downes),加拿大国家研究委员会(National Research Council of Canada)高级研究员,慕课始创者之一。

译者简介:肖俊洪,汕头广播电视大学教授,Distance Education (Taylor & Francis) 期刊副主编。ID <https://orcid.org/0000-0002-5316-2957>

责任编辑 韩世梅

(上接第16页)

- Whitehill, J., Serpell, Z., Lin, Y.-C., Foster, A., & Movellan, J. R. (2014). The Faces of Engagement: Automatic Recognition of Student Engagement from Facial Expressions. *IEEE Transactions on Affective Computing*, 5(1), 86-98.
- Wigfield, A., Guthrie, J. T., Perencevich, K. C., Taboada, A., Klauda, S. L., McRae, A., et al. (2008). Role of reading engagement in mediating effects of reading comprehension instruction on reading outcomes. *Psychology in the Schools*, 45(5), 432-445.
- Yamauchi, T., & Xiao, K. (2018). Reading Emotion From Mouse Cursor Motions: Affective Computing Approach. *Cognitive Science*, 42(3), 771-819.

定稿日期:2020-05-26

作者简介:李振华,博士研究生,高级实验师;张昭理,博士,教授,博士生导师;刘海,博士,副教授,硕士生导师,本文通讯作者。华中师范大学国家数字化学习工程技术研究中心(430079)。

责任编辑 单玲

收稿日期:2019-10-28