DOI:10.13811/j.cnki.eer.2016.09.006

基于学习分析的在线学习测评建模与应用

——学习者综合评价参考模型研究

郑勤华[1]， 陈耀华[1]， 孙洪涛[2]， 陈 丽[1]

(1.北京师范大学 远程教育研究中心,北京 100875;2.中央民族大学 现代教育技术部,北京 100081)

[摘 要] 学习分析技术为提升学习质量提供了新的思路。而从学习分析在学习评价、诊断、预测、干预的实际应用来看,评价模型的建立是重要的基础。本研究以学生综合评价为目标,通过理论演绎和专家访谈构建了以投入度、完成度、调控度、联通度和主动性为核心的五维度综合评价参考理论模型;并通过学习行为数据聚合特征变量,构建了相应的计算模型。在此基础上,进行了相应的工具设计和研发,并通过实践应用验证了学生综合评价参考模型的科学性和可用性。本研究为学生的综合测评提供了理论上的参考,为学习分析技术的实践应用提供了方法和思路上的借鉴。

[关键词] 学习分析技术; 学生综合评价参考模型; 学习评价

[中图分类号] G434 [文献标志码] A

[作者简介] 郑勤华(1978—),男,江西上饶人。副教授,博士,主要从事远程教育管理与在线学习分析研究。E-mail: zhengqinhua@bnu.edu.cn。

一、引 言

教育发展的目标是通过改革提升教育服务能力,能够为所有有学习意愿的人提供教育服务;同时提升教育服务水平,能够为每一个学习者提供个性化的教育服务,即构建面向全体受众的大规模个性化开放教育体系。要实现这种大规模个性化教育体系,我们认为解决的思路有两个方向,一是发挥互联网的社会交互特性,鼓励学习者建立自己的学习空间,使用自己想用的互联网服务建立连接、分享内容、贡献内容、合作学习或者扩展自己的个人网络和专业网络。交互尤其是社会交互是这种教与学的核心,学习者通过交互建构自己的数字身份,贡献自己的智慧,通过持续不断的创新推动课程的进化与发展,并不断发展自己的概念网络、技术网络和社会网络,注重问题的解决和创新[1]。此时实现的是在线教育在社会性交互方面的潜力,主要是利用技术促进人与人之间的通信,改善师生之间、生生之间的交互[2]。二是充分利用大规模学习者在平台中的学习和交互行为产生的海量数据,这些数据的应用和挖掘将成为平台价值的更大发源地。对于机构来说,最有价值的是通过数据挖掘为学习者提供有针对性的教育服务,而这类教育服务并不是学习者“自有价值”的天然获得,而是在此之上为提高个体学习效率和绩效而进行的增值服务[3]。按照美国教育部教育技术办公室 (U.S. Department of Education, Office of Educational Technology)2012 年简报中的提法来进行概括, 这些增值服务主要包括用户知识模拟、用户行为分析、用户经验分析;用户分类、分组;知识域模拟如学习课题分类排序、知识元素与相应的教学原则分析;趋势分析;自适应和个性化学习。其中,对学习者最有价值的就是自适应和个性化学习,平台能够根据已记录的学习者在线学习活动的行为数据、交互内容数据以及相应的绩效表征文本,对学习者的行为和经验进行建模和分类,从而提供最直接、最有效的学习诊断和预测,进而给学习者提升学习效率和效果提出有价值的建议[4]。

基金项目:北京师范大学自主科研基金项目“学习者在线学习状态分析与可视化工具研发”(项目编号:SKZZB2015013)

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

学习分析技术强调基于数据的学习描述、诊断、预测和干预。第一个关键节点就是对于学生学习状态的综合性描述，这种描述不是类似于传统商业智能的特征提取，而是应该基于教育教学理论，从评估和发展两个方面构建在线学习者的综合评价模型。本研究以在线学习者的综合评价为主题，尝试通过完整的学习分析技术，构建基于学习行为数据的综合评价参考模型，从而为学习者的个性化学习服务提供智能化决策支持。之所以选择评价作为切入点，是因为学习分析技术所期待的个性化学习服务，来源都在于有一个科学的评价模型，只有建立了评价基准，才能够做出相应的预警和预测服务，从而为有效的干预提供支持。

与所有的传统教育评价一样，在线教育中学习者的综合测评也一直存在着科学性和可行性之间的矛盾。从科学性的角度来说，如何全面、正确地评价学生的学习绩效和个体发展，目前主要寄希望于终结性评价和形成性评价结合的思路进行考量。形成性考核是为激发、提高学生的学习所设计的所有活动，并且要为学生提供他们学习过程中进步、发展的标志。终结性评价是用于对学生所取得的成就作出评估的记录或报告。[5]在我国的教学评价中，考核类型的划分往往是根据时间段来进行，形成性考核是指平时成绩，终结性考核则是指期末成绩。虽然我们形成了这种综合评价，但终结性考核主要考察的是认知绩效，而形成性考核的基本形式有：平时作业、阶段性学习测验、课程实践教学、专题讨论、合作学习、学习记录等[6]，主要的考察点仍然是学生学习的认知结果。如今，对人才培养的定位越来越指向综合性，尤其是在线教育领域，我们一直提出的是应用型人才的培养目标。这又涉及一个可行性的问题，因为如果没有合适的评价维度和指标体系，以及相应的数据采集方法，这种理想的综合性评价就会因缺失可操作的具体方法和工具而难以为继。

因此，如何在评价方面切实体现人才培养目标，就需要对目标本身进行综合建模，从更充分的维度综合形成科学、可操作的评价目标体系。基于此，本研究重点回答以下两个问题，一是如何构建全面衡量学习质量的综合评价模型；二是在学习分析技术使用中，如何根据平台记录的有效教与学行为数据聚合到综合评价模型，从而为可操作的自动化参考性评价提供基础。

二、学生综合评价参考模型的理论研究

学生综合评价参考模型的提出主要依赖于理论演绎和专家访谈。《教育评价辞典》中将教育评价定义为："根据一定的教育价值观或教育目标，运用可行的科学手段，通过系统地搜集信息资料和分析整理，对教育活动、教育过程和教育结果进行价值判断，从而使评价对象不断自我完善和为教育决策提供依据的过程"[7]。学生评价作为教育评价中的核心内容之一，可以将基于学习分析的学生评价界定为：根据一定的教学目标，运用可行的科学手段，通过对学生学习产生数据的分析和整理，对学生的学习活动、学习过程和学习结果进行价值判断，从而使学生不断自我完善的过程。

从教育评价和学生评价的界定可以看出，进行学生评价，需要以教学目标作为价值判断依据和准绳。对于教学目标，布鲁姆将其分为认知领域、情感领域和动作技能领域等三大领域；加涅将教学目标按照学习结果分为言语信息、智力技能、认知策略、动作技能和态度等五个大类。由于在线学习基本不涉及动作技能的目标习得，因此，我们主要在加涅的目标分类基础上，结合在线教育自身的特点以及在线学习者（主要是成人学习者）的特点，构建了如下的五维度理论模型。

从智力技能方面，我们在构建评价模型的时候，引入了完成度这一维度来反映这类教学目标的具体完成情况。所谓的完成度，指的是学习者按照教师的教学设计，所完成的相应的学习活动的情况。需要特别说明的是，我们的前提是教师的在线课程教学设计是完整而合理的，这样完成度才能够较好地反映学生的在线学习的结果情况。

从态度和情感方面，我们认为不同的学习动机和学习需求是在线学习者的重要属性，是评价学生在学习准备的态度方面的重要指标。为此，我们引入主动性这一评价维度，综合判断学生学习的态度准备，从而保证评价的全面性。

由于在线学习者的受教育背景、学习需求和兴趣不同，其学习动机、学习风格、认知策略等相应地也会有相当大的差异，并且由于在线学习大多数情况下是由学生自主进行，因此对学生的自主学习能力也有较高的要求。这些能力、方法上的差异对学生的在线学习效果有着重要的影响，因此需要进行全面的评价，以帮助教师针对不同类型的学习者采取有针对性的教学策略和教学干预，由此我们构建了调控度的维度。

在在线学习中，学生学习的投入水平也是我们关注的重点，如果说完成度主要考虑的是学生学习的质

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

量评价,那么投入度考虑的就是学生学习的过程数量指标。在线教育相较传统课堂教学有其特殊性,由于教与学的时空分离,教师、管理者仅仅通过作业、单元测验、考试等一些传统的评价手段很难实时掌握学生的学习过程和学习状态,很难及时作出针对性的调控和干预。学习分析基于学生的学习行为数据进行评价,在对在线学习的过程监控方面有其天然的优势,构建学生评价模型时需要发挥这方面的长处,弥补在线教育过程性评价方面的短板,由此我们引入了投入度作为表征学生在线学习过程的评价维度。

在信息资源呈几何级数增长的今天,学生的在线学习目的不再仅仅是从固化的课程资源中获取知识或技能,更重要的是培养在海量信息资源中发现、获取乃至创造知识的能力。联通主义学习理论认为,网络时代的学习者需要具备从更大范围发现和组织信息资源的能力,具备辨析信息资源内容和发现信息间关联性的能力,具备建立社会化连接以交流共享信息的能力以及通过创造内容与他人联通的能力。[8]在联通主义学习中连接的建立和网络的形成都依赖于交互的开展,整个网络以交互为核心。学习者参与到与实践者、其他学习者、教师和导师的交互之中,这种社会交互对学习者了解课程内容并在空间中进行定向至关重要[9]。从在线学习的发展趋势和对学习者的学习素养要求的角度出发,我们构建了联通度这一维度来评价网络时代学习者所需要具备的能力素养。

初步模型构建后,研究者邀请了在线教育领域、学习分析领域和教学评价领域的七位专家进行了半结构化访谈,获得了相应认可后,我们最终以在线教育人才培养目标为核心,结合教育评价理论和远程教育相关理论,提炼构建了学生综合评价参考模型(如图 1 所示),称为 S-SERI (Student-Systematically Evaluation Reference Indicator)模型。

S-SERI 模型由维度和指标构成,其中维度满足对学生的某方面评价需求,具有一定概括度和抽象层次,可能包含多个指标;指标则是对维度的具体分解。例如衡量学生在线学习投入的“投入度”维度,包括“行为投入”、“认知投入”、“情感投入”等一级指标,在“行为投入”指标下又包括“活跃度”、“持续度”等二级指标。维度和指标的确立主要基于以下两个原则。(1)典型性原则。评价指标应选取最能反映学生在该评价目标上发展水平的典型特征,并不需要将所有反映评价目标的指标一个不漏地堆积起来,评价模型最终应用于在线教育教学实践,过多的指标可能反而让教师和学生难以解读、无所适从。(2)可操作性原则。维度一级级分解成的最终指标应该是具体的,可以观察、描述和测量的。构成 S-SERI 模型的五个维度具体如下。

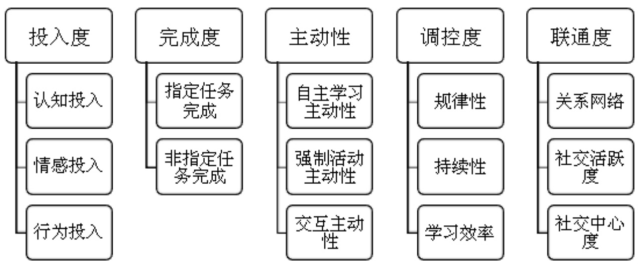

图 1 学生综合评价参考模型

1. 投入度

从活跃性、持续性等方面对学生在线学习的投入程度进行评价。对投入度的评价除了行为投入以外,还包括更深层次的认知以及情感投入。投入度是对学生学习过程进行表征的重要维度。

2. 主动性

包括完成自主学习任务的主动性、参与教师指定教学活动的主动性以及进行交互的主动性等,主动性在一定程度上表征了学习者的学习动机水平与变化情况。

3. 调控度

从学习的规律性、持续性、效率等方面对学生调控自己学习过程的水平进行评价,调控度是对学生认知策略、自主学习能力进行表征和评价的重要维度。

4. 完成度

以课程的教学目标为标准对学生实际完成情况进行评价,对于不同类型的课程,教学目标的侧重点不同,完成度的指标以及权重设置也会进行相应的调整。

5. 联通度

对学生建立社会化认知网络的能力进行评价,包括建立连接的能力、维护连接的能力等。联通度的评价核心是交互,包括学生与资源的交互、学生与教师以及与其他学习同伴的交互。

进一步,我们根据构建的理论模型,开展基于学习数据的综合建模工作。由于维度建模来源于理论演绎和专家访谈,下一步的重心是如何将构建的维度模型进一步细化,并提炼出相应的数学结构,从而为量化评估提供可操作的依据。

三、数据模型的构建

学生评价的理论模型构建是由上而下,而算法模型的构建是由下而上,即先确定各维度下最底层指标的计算方法,再将这些指标进行聚合,最终得到维度算法模型,模型构建的具体流程如图 2 所示。

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

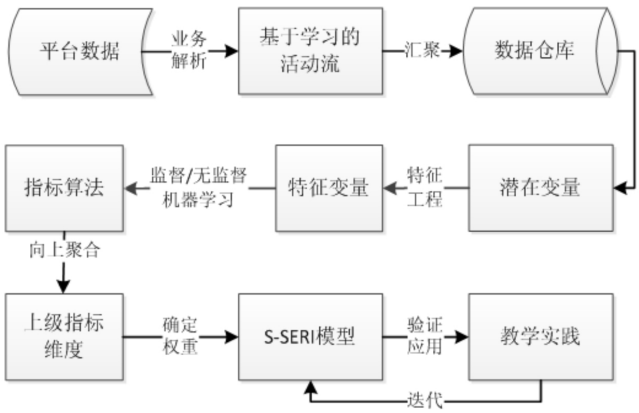

图 2 学生综合评价参考模型数据分析应用流程

1. 原始数据汇聚与预处理

学生的原始学习数据往往是多源异构的，分散存储在内容管理系统、学习管理系统、教务管理系统等各个平台中，将这些数据汇聚到一起，尽可能对学生学习时所发生的过程进行还原是科学和全面评价的基础。这就要求对各个平台的数据有一个全面完整的认识，理顺数据中蕴含的业务逻辑，提取出在线教育的各类业务工作中的关键主体、要素和流程。如针对课程教学业务，其主体包括教师、学生和课程；要素包括资源、交互、作业和测试等；流程包括资源学习、交流讨论等。对于业务所涉及的流程，则是相互交织的一系列复杂过程，包含多种动态和静态信息。流程往往体现了要素的状态变化和行为过程。在此基础上，将在线教育的核心教学与管理业务分解，形成覆盖学生学习过程的完整数据描述。

参考 xAPI 的理念[10]，基于业务解析提取出的关键学习主体、要素和流程，我们构建了基于学习的活动流。活动流由必要属性和可选属性构成：必要属性包括活动主体、动作以及活动对象三类，可选属性包括活动时间、活动结果等若干类。通过这些属性，可以对学生的学习活动过程进行描述，例如小张（活动主体）在某个时间（活动时间）观看了（动作）关于某个知识点的视频（活动对象）。表 1 列出了一些可能的必要属性，不同教学设计、教学情境中这些属性需要作出相应调整。

构建基于学习的活动流，方便不同数据源按照统一的规范进行整合，提高数据汇聚的效率。按照这种规范描述的学习过程，最终会以具有语义结构的数据进行存储，也便于机器理解和学习。同时，由于学习分析大部分基于行为数据进行，对数据预先进行基于（行为）活动的描述和封装，有助于后续研究的进行。

在解析业务方面，学习活动数据化的过程中会遇到数据质量的问题。当前在线教育的数据质量整体水平还比较低，缺失值比例较高，可能会出现异常值、不确定的定义和编码、无效或错误的数据以及数据不一致等情况。在这个阶段要及时地发现数据质量存在的问题，对数据进行清洗，尽可能地提高数据质量，为后续的分析奠定基础。

表 1　　活动流中一些必要属性的描述

活动主体	动　　作	活动对象
教师 学生 小组	浏览、观看、编辑、评论、提问、回答、笔记、分享、标注、创建、收藏、取消收藏、得分、下载、发表、回复、搜索、删除、参与、尝试、暂停、继续、跳过、完成、退出、通过、失败	文件、问题、协作、视频、页面、讨论、作业、消息、笔记、学习计划、测试、帖子、档案、模块

2. 潜在变量生成

潜在变量是指从原始数据库中获得的、在意义上与模型可能相关的变量。这些变量与评价需求相关联，且数据质量较好。在潜在变量生成阶段中，我们在之前数据汇聚的基础上，提取出能够描绘学习活动核心要素和流程的关键数据，这些数据是后续模型构建的基础。前一步中汇聚的数据集合是对教学活动和学生学习经历的完整描述，潜在变量是在此基础上进行的提炼，是一系列能够表征教学与学习情况的有效标准数据项。

潜在变量生成阶段，仅通过对于机构教学要素和教学流程的分析得出有效的数据项。应尽量避免复杂计算。如必须进行复杂计算，则只有通用性的、能够对机构的教学与管理起到重要监测作用的变量才可以纳入到潜在变量中。这些数据项可以是基本的属性量，如课程的资源数等，也可以是行为计数，如每日学生资源访问量，还可以是简单计算的结果，如作业提交时间间隔，即教师发布作业与学生提交作业之间的时间差值；在少数情况下，可以是需要较为复杂计算的结构，如教学交互的有效性等。

3. 特征变量选取

特征变量是从潜在变量中选出的，经过统计分析后证明具有足够好的数据质量，可以进一步进入算法构建的变量。特征变量选取又称作特征工程（Feature Engineering），根据具体的评价指标，用特定领域知识或机器学习的方法来选择、删除或者组合潜在变量得到特征，作为指标的计算变量。之所以需要进行特征变量选取，是因为多余的变量会对模型训练产生干扰，有可能造成模型的过拟合。另外变量过多时，在数据体量越来越大的情况下，会大大降低模型的训练和计算效率。

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

特征变量与评价指标有着较大的不同,它是评价指标构建的基础,可以通过不同类型的计算,成为监测教学与管理状态的重要指标。这些特征变量有可能是显而易见的,如学生在线观看教学视频的次数、时长等,也有些可能需要经过一定的计算变换,如学生观看教学视频后进行练习的概率、正确率等。

当需要计算的指标较为简单、潜在变量较少的时候,可以让专家基于特定的领域知识来人工选取特征变量。而指标算法复杂,涉及较多变量的情况下,人工的方法存在一定的局限性,可能会影响到指标算法的准确性,所以我们综合采用如下几种方式来选取特征变量。

(1)计算变量与对应指标的相关性,选择与指标有相关性的变量作为特征变量;

(2)将潜在变量组合后再来选择特征变量,如对学生的人口学数据(性别、年龄、前置学历等)进行组合,来获得较大的特征集,组合特征能够同时兼顾全局模型和个性化模型;

(3)在数据构造较为复杂的时候可以采用深度学习的方法来自动识别特征变量,深度学习又被称作无监督特征学习(Unsupervised Feature Learning),具有自动学习特征的能力,可以从大规模的未标注数据集中体现学习数据的本质特征。

表 2　　特征值的计算与维度计算示例

投入度	v1 学习频次		v2 学习时长		v3……	
	值	得分	值	得分	值	得分
主动性	v1 答疑提问频次		v2 论坛发帖频次		v3……	
	值	得分	值	得分	值	得分
调控度	v1 作业提交延迟时间		v2 单次视频学习时长		v3……	
	值	得分	值	得分	值	得分
完成度	v1 视频完成率		v2 作业完成率		v3……	
	值	得分	值	得分	值	得分
联通度	v1 社交活跃度		v2 社交参与度		v3……	
	值	得分	值	得分	值	得分

维度值和特征变量的关系见表 2,由于篇幅关系,这里并没有列出所有的特征变量。所有的特征值都有两个数值,一个是原始特征值,如学习频次,某个学生以周为单位的学习频次是 15 次,那么记录的特征值就是 15,但由于综合评价需要将特征值聚合成维度值,所以就需要将原始特征值转换为百分制。具体的百分制的得出,一般通过聚类、Z 分数、逻辑回归等方法将特征值加以转换。如学习频次 15,转换为百分制则为 90。

4. 模型构建

对于指标算法构建,主要分两种类型,一种类型是可以通过问卷、量表或者其他测量方法获得指标因变量的值,这种情况下采用有监督的机器学习,例如人工神经网络或者一些分类算法,通过调整参数进行模型训练,选择准确度较高的算法即可。

另一种类型指标构成复杂,且对应因变量的值获取相对较为困难,如何采用无监督的机器学习方法对这些指标进行分析是模型构建的关键和难点问题。要完成这种分析,需要先将专家知识通过某种方式传授给机器,再由机器完成专家所不能够完成的整合分析。如对于学生的活跃度这个指标的分析,可以找到在线时长、登录次数等特征数据,由专家对不同活跃水平的学生贴标签,然后再对各类学生进行聚类等各种分析。对分析的结果进行初步的模型计算。接下来,根据计算结果,确定学生活跃度的分析算法。

两种类型的指标算法确定以后,我们采用专家模糊层次分析和机器学习相结合的方法确定各个低一级指标的权重,最后一步步向上聚合成更高级别的指标、维度乃至 S-SERI 模型。

5. 模型验证迭代

模型的信度和效度是其重要的衡量指标。除了一般性的数据挖掘指标之外,需要将其同业务数据进行对比,并根据平台当前的数据情况确定其置信区间。模型构建需要多轮迭代,需要尝试各种特征变量的组合与各种不同类型的模型计算方式的结果,通过比较计算结果最终确定模型算法。

四、基于数据模型的评估实践

基于国内某高校网络教育学院的数据,我们对上述的 S-SERI 模型进行了应用和验证,该学院使用了第三方开发运维的专门学习管理系统,由于数据可获得性和数据质量等问题,对网络学习系统相关页面另外进行了埋点处理,最终形成了相应的数据仓库和数据集,如前文图 2 所示。根据特征值计算,最终聚合出相应的维度值和 S-SERI 值,构建应用系统进行可视化呈现,如图 3、图 4、图 5 所示。

其中,图 3 展示了学院实时的 S-SERI 总体均值,表征的是学院学生综合评价的总体情况;图 4 展示了某位学习者的实时 S-SERI 值,表征的是学习者个体的学业综合评价情况;图 5 展示了两个月内学院学生的 S-SERI 总体均值,表征的是学习者总体变化的趋

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

势。基于上述数据分析，进一步开展了基于 S-SERI 值的管理和教学干预建议参考。如对于个体学习者来说，工具将根据其五个维度的得分情况，分析分值高低的原因，从而从底层数据采集点的教育意义出发，提出适应性的干预建议。以图 4 所示的学习者为例，对他的综合评分所处的相对位置进行了展示，对学习主动性的维度得分偏低提出了有针对性的提醒，当然，我们进一步根据主动性维度具体的特征值计算规则，可以发现是他在完成教师布置的必做学习活动中，往往是在提交的最后节点之前，甚至之后才完成相应的学习活动，我们对其进一步的学习改进建议则是合理地安排学习时间、更加主动地参与学习活动等。

图 3　学院总体 S-SERI 计算

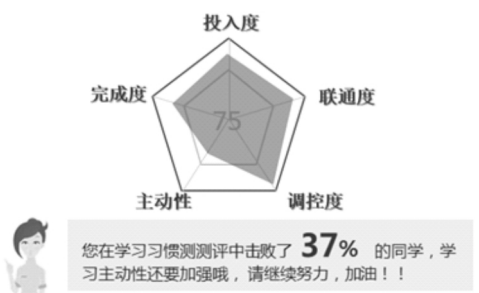

图 4　学习者个人 S-SERI 计算

图 5　学院总体 S-SERI 趋势分析

当然，系统设计和应用的最终目的是判断我们基于学习数据开展的学生综合评价参考模型的科学性和有用性。科学性这一点在前文的理论推演和专家访谈中予以了部分保证，还需要在在线教育机构的实践应用中予以不断的反馈和迭代，方能进行系统性的优化。从可行性的角度来说，本研究的实践表明，利用学习分析技术，形成系统化的学习者测评模型是能够部分代替传统的评价设计的。之所以是部分代替，原因之一是学习分析技术严重依赖于数据的可获得性，从理想的设计角度，研究者希望获得一切学生相关的数据，从而更加深入和全面地解读在线学习，但从实践角度出发，我们的原型和机构进行了多轮碰撞，最终形成了当前的最优解决方案，尤其是我国的网络教育机构大多并不是用统一的系统进行招生、教务、学习、社交、管理等业务，这给数据汇聚带来了极大的困难。另一个重要原因也在于，我们的在线教育本身的数据质量的问题，在此不再赘述。尽管有上述困难，本研究的实践应用很大程度上证明了基于学习分析的 S-SERI 模型的科学性和有效性，能够为在线教育的教与学服务改进提供支持。

五、讨论与建议

本研究是以学习分析技术对在线学习者的综合评价展开了建模和研发应用。研究从投入度、完成度、调控度、联通度和主动性等五个维度构建了学生综合评价参考模型，在此基础上，通过在线学习平台汇聚的学习数据，进行了从行为数据到特征数据，再到维度计算，最终形成综合评价参考值的相关研究。本研究通过设计的理论模型和分析计算方法，综合设计开发了相应工具，在网络教育机构中展开实际应用，证明了本研究构建的 S-SERI 模型的科学性和可行性。总结本轮研究，我们认为，在基于学习分析技术进行学习者综合评价的研究中，需要重点考虑以下几点。

首先是评价的目的。本研究构建的评价模型是针对学习者的全面综合评价，并不一定适合所有的课程类型，而且研究中并没有考虑实践技能等方面的习得（最主要的原因是在线学习数据中这方面的特征指标难以获得）。在具体的课程教学模式中，有的课程可能采取的是真正意义的自主学习，以学习者自我监控、自定步调、自定学习方式等为主体，对学习中交互的发生并没有活动要求，甚至目标中并不强调，此时联通度这一维度的价值并不大；而有的课程以师生和生生之间的教学交互为核心，在讨论中演进整个学习过程，并不涉及大量的形成性自主测试，此时的完成度维度则对该类课程价值不大。我们目前设计的是一个较为普适的框架，更多针对的是以视频浏览、在线测试、形成性作业、论坛交互、答疑反馈等为主要学习活动的教学模式，各维度的计算权重也是以此作为依据由专家模糊层次分析确定。为了考虑不同的评价目的，我们除了制定通用的 S-SERI 计算模型外，也在工具的设计中提供了教师自定权重的相关调整机制，从而保证工具本身的通用性。需要着重强调的是，评价

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

的目的决定了我们的模型设计,并对最后的应用提供直接参考,这也是为什么本研究中将提出的模型命名为“参考”模型。

其次是特征数据的聚合问题。学习分析的数据源头是平台记录的各类行为数据本身,基于原始行为数据的分析在综合测评中直接应用并无可能,需要将其按照实际应用聚合成相应的特征变量。但是特征变量的选择是一个自上而下和自下而上的统一进程。前者要求特征变量的选择必须有维度的解读意义,即能够表征出该维度的某个方面;后者要求特征变量的选择和行为数据能够匹配,能够用最底层的行为数据聚合而成,而非简单的理论逻辑推演得出,两者共同对特征工程提出了大量的研发和数据清洗要求。在本研究中,为解读学生综合评价参考模型,依据底层数据的可获得性,共聚合形成了 32 个特征变量,其中一半左右的变量通过简单的统计分析就能聚合而成,如周均学习时长等;另一半的特征变量需要通过基于行为数据的负责运算而得,如基于社会网络分析的社交中心度和基于数据包络分析的学习效率等。特征工程是本次研究中承上启下的关键环节, 从本研究的情况来看,它是决定综合测评能否实践应用的核心,是工具研发的数据科学基础。

最后,本研究构建的理论模型和工具,在实践应用中,曾经在方法论层面有所犹豫。一般进行此类的评价研究,希望能够通过有验证的数据对照环节进行比较从而论证评价模型的科学性。比如,基于学习行为数据构造综合测评模型对学习者进行测量形成相应的数据集,将该数据集和按照传统评价方式(如考试)的成绩进行验证,通过两套数据的相关性或拟合度来验证评估模型的成功与否。在本研究中,我们最终放弃了该研究范式,主要原因在于,我们构建的是综合测评模型,而传统的成绩度量主要是考察认知结果,在在线学习领域,并没有特别合适的评价数据作为比较验证对象;二来本研究的出发点在于,学生的学习状态客观真实地体现在了日常学习的点点滴滴行为中,这种基于行为的综合测评能够更加全面而客观地体现学习者的综合绩效,而这也是未来评价的理念指向。因此,我们最终选择了设计性研究的范式,通过多轮的迭代来验证模型的科学性和有效性,在每一轮的迭代中, 我们将搜集教师和学生的反馈信息,从而对模型本身和工具应用作出修正。

总之,学习分析技术为教育教学的改进提供了新的思路和方法,研究者对此投入了极大的热情。在线教育领域由于学习行为的记录相对便捷和全面,因此成为此类研究的突破口。大量的学习分析应用都需要以评价作为基准展开,本研究以学生综合测评作为对象,构建了相应评价参考模型,并进行了实际应用。研究成果并不完美,但我们认为汇报我们的研究范式和研究思路同样具有重要的参考意义。我们将在已有研究的基础上,不断迭代完善学生综合测评模型,并据此进一步为教学和学习个性化干预提供指导意见,从而实现提升学习质量的最终目的。

[参考文献]

[1] [3] 郑勤华.学习理论与远程教育成本—效益实现路径研究[J].开放教育研究,2014,(10):29~37.

[2] Moore, M., Kearsley, G.. Distance Education: A Systems View [M]. Belmont: Wadsworth Publishing Company,1996.

[4] U.S. Department of Education Office of Educational Technology. Enhancing Teaching and Learning through Educational Data Mining and Learning Analytics[EB/OL]. [2016-07-21]. http://tech.ed.gov/learning-analytics/.

[5] Morgan, C. , O'Reilly, M.. Assessing Open and Distance Learners [M]. London: Kogan Page, 1999.

[6] 陈庚,谢浩,郑勤华. 网络教育教学管理[M]. 北京:中央广播电视大学出版社, 2012.

[7] 陶西平.教育评价辞典[M].北京:北京师范大学出版社,1998.

[8] 王志军,陈丽. 联通主义学习理论及其最新进展[J]. 开放教育研究,2014,(5):11~28.

[9] Siemens, G.. Orientation: Sensemaking and Wayfinding in Complex Distributed Online Information Environments[D].Aberdeen:The University of Aberdeen,2011.

[10] Experience API -ADL NET [EB/OL].[2016-07-21].https://www.adlnet.gov/adl-research/performance-tracking -analysis/ experience-api/.

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

The Construction and Application of An Online Learning Evaluation System Based on Learning Analytics Technology: A Study of Students Systematic Evaluation Reference Indicator

ZHENG Qin-hua, CHEN Yao-hua, SUN Hong-tao, CHEN Li

[Abstract] Learning analytics technology provides a new idea to improve learning quality. Looking at the practical applications of learning analytics in learning assessment, diagnosis, prediction and intervention, we can find that the construction of assessment model is the most important. Aiming at systematic evaluation of students´ learning, this study constructed a theoretical model through theoretical deduction and expert interviews. The model contained five core dimensions of reference indicator: engagement, completeness, regulation, connectivity and initiative. A corresponding computational model was developed through characteristic variable aggregation using learning behavior data. Based on the theoretical and computational models, an evaluation tool was designed and developed. The scientificity and usability of the students systematic evaluation reference indicator were verified in practices. This study informed the theoretical research on systematic evaluation of students and the application of learning analytics technology in practices.

[Keywords] Learning Analytics Technology; Student Systematic Evaluation Reference Indicator (S-SERI); Learning Evaluation

---

（上接第 32 页）

between education and technology as well as difficulty to achieve real integration in practices. It will be conducive to solve these issues if we study educational technology by examining the essence of educational technology innovation process by coupling "Education & Technology". Previous integration theory considered educational technology as the external resources or tools of classroom instruction, and re-organized them based on the external role of teachers. Coupling theory studies the internal interaction process of the main body of "Education & Technology" from a dynamic and generative view. This theory highlights the essence of the process of generating the educational technology that meets the teachers´ and students´ needs. In the generation of new educational technology, there exists not only two-way and overall coupling mechanism of "Educational & Technology", but also the interaction of the key factors such as external environment, genes, theoretical guidance, intrinsic motivation. Therefore, new educational technology can be generated in a comprehensive way. In order for the innovated educational technology to meet the needs of the entire educational system and the classroom teaching practices, it is necessary for the research experts to design guidance at the top-level as well as for the teachers and students to couple the educational technology that reflects the experts´ ideas with the classroom reality. As a consequence, "close-fitting" and new educational technology can be generated continuously.

[Keywords] Education & Technology; Coupling; Creation; New Educational Technology; Essence of the Process

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net