

在线学习行为分析建模及挖掘

胡艺龄¹ 顾小清¹ 赵 春²

(1. 华东师范大学教育信息技术学系, 上海 200062; 2. 华东师范大学信息化办公室, 上海 200062)

[摘要] 在日趋精密数字技术条件下, 学习模式已通过互联网、社会化媒体实现数字化。海量的学习信息以数据的形式蕴含着学习者的隐性行为特征。文章从数据挖掘与领域应用、学习行为及行为分析、网络行为分析模型三个角度对在线学习行为可能的应用方向进行综述研究, 探讨学习者的在线学习行为的建模机制, 建立了数据、机制、结果三层次模型, 并从网络挖掘的角度对学习数据进行模式分类与解析。

[关键词] 学习分析; 在线学习; 行为分析; 数据挖掘; 系统建模

[中图分类号] G422 **[文献标识码]** A **[文章编号]** 1007-2179(2014)01-0102-09

一、引言

纵观21世纪教育研究的发展趋势, 探究学习者的深层次思维与行动成为教育者关注的焦点。学习是个体获得行为经验的过程(Dienes et al., 1995), 从教育心理学角度它可分为外显学习(Explicit Learning)与内隐学习(Implicit Learning)。传统学习行为理论研究由于技术与工具的局限性, 只探讨那些能直接测量的、可观察的行为, 而对内隐的学习行为潜能的研究, 一直缺乏一手的实践数据, 且尚未形成系统的理论分析体系。因此如何革新传统的学习行为分析模式, 将内隐分析与外显结合是教育者们一直思索的方向。

随着网络课程与学习技术系统的发展, 在线学习越来越多地发生在各种教育情境中, 这为深入研究学习者学习心理、行为带来了新契机。基于网络的学习能使学习凭自己的喜好进行自主、探究性的学习, 学习行为能自然地呈现, 网络也提供了记录学习者隐性学习行为的载体。这些学习行为常常是无意识、零散的, 行为之间潜在关联结构、底层规则不显著, 但比外显学习行为更能体现学习中微妙而复杂的逻辑关系, 能折射学生最真实的思维与学习情况, 是教师或其它评价机制不易捕获到的。隐藏在学习平台记录中的学生学习行为, 犹如丰富的矿藏,

如果采用适当的采掘方式进行挖掘, 这些反映学生学习行为(包括学习者的态度、动作、语言)的数据, 能够成为跟踪学习、评价学习以及改进学习的最佳数据来源。由商业领域孕育而生的数据分析技术——“商业智能”, 目前正在教育领域中悄然掀起并加剧进化, 学习分析(Learning Analysis)作为一种教育数据分析技术, 成为发展中不可或缺的核心力量(祝智庭等, 2012; 顾小清等, 2012)。它通过对数据背后所隐藏的信息加以发现和理解并有效进行利用(或干预、或预测)的研究, 从而追求最大的教育效益(Siemens et al., 2011)。为此, 最为核心问题是构建能满足学习跟踪、评价及改进需求的行为分析模型和数据挖掘算法。

本研究在广泛文献调研的基础上, 从学习分析需求出发, 构建了在线学习行为分析模型, 并设计了学习数据的挖掘模式, 最后结合案例探讨了学习分析研究的科学意义和应用价值。

二、现状分析

由于对学习者的在线学习行为分析涉及多个学术领域与技术, 如数据挖掘、计算机网络、行为认知学、人工智能、数理统计等, 因此本研究的前期调研聚焦以下三个方面: 数据挖掘与领域应用、学习行为及行为分析、网络行为分析模型。

[收稿日期] 2013-11-20

[修回日期] 2013-12-23

[基金项目] 本文系全国教科“十二五”规划教育部一般课题“以‘语义图示’实现可视化知识表征与建模的理论与实践研究”(BCA120024)部分成果; 教育部新世纪人才计划“基于个人数字终端的信息化创新应用研究: 资源、服务及应用实例”(NCET-11-0140)阶段成果。

[作者简介] 胡艺龄, 华东师范大学教育信息技术学系在读博士生(huyilingecnu@gmail.com); 顾小清, 华东师范大学教育信息技术学系主任, 教授; 赵春, 华东师范大学信息化办公室工程师。

(一) 数据挖掘与领域应用

从大量数据中抽取有用信息的相关研究、技术及工具,在计算机领域有多种提法,包括数据挖掘、知识发现、知识提取等(Agrawal & Shafer, 1996)。总体而言,研究人员更多地使用数据挖掘这一术语(Kurgan & Musilek, 2006)。

上世纪90年代曾兴起一股研究热潮,研究人员致力于寻找一种通用算法,以解决所有与从大量数据中搜寻有用信息相关的问题。在研究算法之外,许多特定的数据挖掘工具应运而生。譬如,Clementine(Khabaza & Shearer, 1995)、IBM Intelligent Miner(Tkach, 1998)、Weka(Witten & Frank, 2005)和DB Miner(Solé, 2010)等。在工具、方法论及模型得到快速发展后,数据挖掘在诸多领域得到了广泛且日益成熟的应用。国内外电信公司将数据挖掘引入核心业务管理系统,应用方向包括客户分析及客户关系管理(Cui, 2009);银行使用数据挖掘技术实施反洗钱监控(Cao & Do, 2012)、客户信用质量评估(Huang & Wu, 2011);证券交易监管部门使用数据挖掘技术实施欺诈交易监测(Golmohammadi & Zaiane, 2012);医院借助数据挖掘协助医生更好地开具处方和提供治疗参考(Wang et al., 2013);图书馆使用数据挖掘技术完成读者需求分析(Zhu & Zhang, 2011)、读者管理(Yu, 2011)、馆藏推荐(Huang et al., 2011)。

数据挖掘技术进入教育领域相对较晚,研究人员的兴趣主要集中在开发研究与分析教育情境下的海量数据的技术工具,如数据挖掘帮助提升高校学生的记忆力(Zhang et al., 2010)、数据挖掘在基础教育课程评估中的应用(Hung et al., 2012),而基于e-learning的数据挖掘研究则是近年来的研究前沿。譬如,在e-learning系统应用数据挖掘,协助学生完成个性化学习(Sun & Xie, 2009)、改善自适应学习的效用(Chellatamilan et al., 2011)、提升学生的学习绩效(Banu & Ramanan, 2012)、研究学生行为和认知风格(Jovanovic et al., 2012)。

(二) 学习行为及行为分析

对在线学习行为分析方面,国内的研究聚焦在数据指标与技术手段两个维度。在数据指标维度,徐红彩(2005)对在校大学生在线学习行为进行了调查与研究;曾祥跃(2008)对远程学习者在线学习

行为进行了调查与分析;靳丽(2008)在对在线学习者交互的行为特征进行分析的基础上构建了在线学习行为模型;杨金来(2008)提出了Who-Do-What的网络行为信息模型。在技术手段维度上,杨清莲(2005)等人结合Web日志挖掘和数据立方体技术对服务器端及客户端日志信息进行分析处理并加以评价;范洁(2005)采用数据挖掘技术设计实现了在线学习行为评估系统,利用这个系统中的日志对学习者进行形成性评价;李德江等(2010)设计与实现了分布式网上学习行为统计系统,介绍了如何从分散的客户端采集数据,再发送给服务器端统计的方法;廖竞等(2011)设计了一种基于数据流获取的在线学习行为数据采集与分析的方案,总结出在线学习者学习行为的规律。

在国际上,对在线学习行为分析的研究集中在三个方面:使用工具软件追踪和记录在线学习行为;关注学习者需求和在线学习环境;寻找在线学习行为和学习绩效的关系。其中,有代表性的研究主要包括:李等(Lee et al., 2002)研究了不同网络环境下学习者在线学习过程中与学习者个性特征相符的学习策略;杨和蔡(Yang & Tsai, 2009)采用问卷进行信息收集,对学习者在在线学习中的学习环境偏好和学习信念进行探讨;罗梅罗等(Romero et al., 2009)提出了在个性化学习系统中提高Web挖掘效率的高级架构;阿尔哈特比(Alkhatabi et al., 2011)针对目前在线学习系统中学习质量不佳的事实,使用Web数据挖掘技术建立学习质量评价的模型。洛(Lo, 2005)则通过学习者在线学习的路径、浏览顺序和习惯,确定该学习者可能具有的学习风格和学习偏好,为学习者推荐相应的学习资源。

(三) 网络行为分析模型

在国内,对网络行为进行分析的模型主要存在几个领域。在网络安全领域,王攀等(2008)针对Web用户行为分析面临的无规范、智能性差等问题,提出了基于动态行为轮廓库(DBP)的行为分析模型;李军等(2008)建立了新的基于流的统计分析模型,从网络异常行为预警的角度,实时监测和发现网络异常,检测和发现网络中的可疑代码;潘蕾等(2011)提出了通过分析用户使用行为获得用户行为模式,并判定其行为倾向、发现异常行为的主动模型。在商业领域,黄雄伟(2011)利用Web分析技术

对客户行为分析(customer behavior analytics, 简称CBA), 通过网络技术分析客户消费购买行为, 及时调整自身营销策略, 对客户成功实施交叉销售和关联销售; 肖英(2012)提出基于马尔科夫随机场(Markow Random Field)模型的行为意图分析模型。在在线学习领域, 彭文辉团队(2006)对网络学习行为进行了理论层面的划分, 提出了以多维度与多层次的角度进行模型的建构; 周岩(2009)针对大学生群体进行了网络学习行为的建模与分析, 并基于理性行为理论与技术接纳模型(TAM)构建出了包含九个潜在变量的因果关系模型。

在国外的网络安全领域, 孔特(Conte, 2008)等基于网络质量保证, 开发出可用性检查技术(WDP——网页设计视角建立可用性评估), 建立了基于Web应用程序的可用性评估模型; 盖齐(Géczy, 2008)等提出构建行为中心的用户配置文件模型, 该模型利用网络环境下用户交互进行行为分析; 多德罗(Dodero, 2011)等对如何扩展常规Web应用的关联数据以及利用数据集提供特定的Web服务进行了深入研究, 得到基于模型—视图—控制器的体系结构。在商业领域, 伊瓦塔和扎瓦德(Iwata & Sawada, 2013)基于价格信息的购买数据得到客户行为分析模型, 研究对每件物品的均值与方差实践估计。在在线学习领域, 贝克(Baker, 2008)等提出了用户认知模型, 建立了关于认知跟踪角度的新方法, 利用类似于建立上下关联的估算概率进行机器学习; 贝克与同事(2004, 2006)在关于行为模型研究中得出, 如果想得到较好的学习效果, 需向学习者提供开放式的学习环境, 让其自由地展现学习行为, 在此模型基础上他们提出自动训练技术, 对学生的练习行为进行规则获取, 使得智能教学辅导系统中的指导行为可以在不同课程之间通用; 麦克法德和道森(Macfayden & Dawson, 2010)则对学习管理系统中的数据进行追踪与分析, 得到与学习者期末成绩相关联的行为变量模型; 道森(2010)团队对高等教育机构如何高效利用数据捕获机制提高学习者的学习体验, 以及为实习生特定学习成果的完成提供信息情报进行了研究; 格雷勒和德拉克斯勒(Greller & Drachsler, 2012)从受益者、目标、数据、工具、外部限制、内部限定六个维度提出了设计模型; 埃莉莎(Elisa, 2012)则从资源使用的角度提出了过

程模型, 其核心是计算机、理论、机构与人员四种资源, 共同推动数据收集、信息加工、结果应用的循环发展。

(四) 研究现状总结

数据挖掘最近十年来在商用领域已经得到较为充分的研究及成功的应用, 其工具、方法论及模型已相对成熟。它通过对客户行为的分析提高客户关系管理质量、导引客户做出导向性的行为, 最终实现经济效益的提升; 在教育领域, 数据挖掘的研究和应用相对较少, 特别是在线学习行为分析方面的研究尚处于起步阶段。这从表一中可得到一定程度的印证。

表一 数据挖掘国内外研究现状

搜索引擎	关键字组合/题名	搜索结果/记录条数	
		~2008	2009~2013
WOK	data mining	10225	6425
	data mining+behavior	43	54
	data mining+education	18	28
	data mining+e-learning	6	13
CNKI	数据挖掘	8068	5465
	数据挖掘+行为	55	63
	数据挖掘+教育	70	67
	数据挖掘+学习+(系统或平台)	4	9

(时间截至 2013-02-19)

而已有的在线学习行为模型研究, 也仅针对学生在线学习平台的显性行为进行分析, 学生在线学习的习惯、情感、品质等隐性学习因子并未得到深入系统的研究; 且尚未有成熟的用于分析在线学习者行为的模型和工具, 还没有形成较为成型的研究与应用趋势。

三、建模思路

由综述可知, 已有学者对在线学习模式建模, 但均未体现对在线学习分析的过程描述。埃莉莎(Elisa, 2012)的四种技术资源模型虽然谈及资源循环过程, 但对数据处理分析的思路与结果对象未有诠释。因此, 本研究建模的总体目标为: 以学习分析需求为导向, 以理解和优化学习为目标, 自底向上对在线学习行为过程进行建模与分析; 遵循问题解决流程模式, 自底向上将模型分为数据、机制、结果三部分(见图1): 数据层呈现学习者可能产生的数据源与数据库; 机制层构建能够辅助理解与优化学习效果的技术与方法; 结果层体现利益相关者与数据之间的关联状态, 以及如何最终影响学习效果的关

系流程。模型虽然呈现自底向上的线性效果,但实为循环过程,数据受益者(学习者)又为数据产生者(即数据主体),从而使得该模型的运行机制周而复始,数据不断迭代更新。

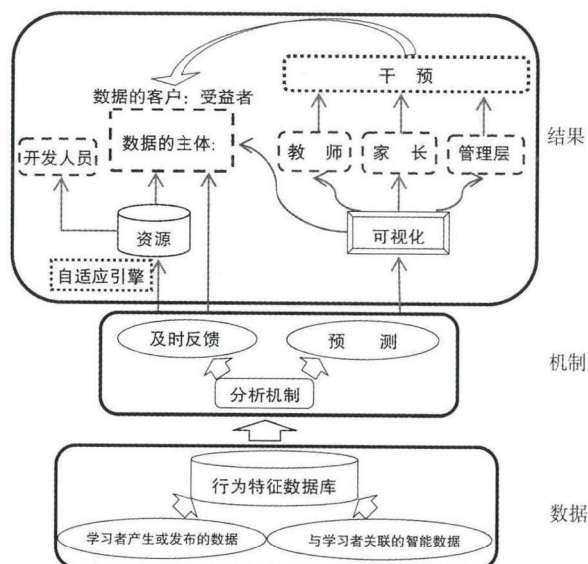


图1 在线学习行为分析模型

(一) 数据

模型的最底层由大量的学习者数据组成,来源分为学习者产生或发布的数据,如学习管理系统(LMS)、学生信息系统(SIS)、个人学习环境(PLE)、社会网络服务(SNS)的数据,和与学习者关联的智能数据,如课程数据、语义数据及其它可能关联的数据。在实现自底向上的行为数据特征提取、分类、存储与索引基础上,建立具有学习效益的行为特征数据库(如对学生的登录行为、资源浏览行为、信息交互行为等特征建立相关数据库)。如何对数据进行收集、清理是工作重点,清理时数据分类分两个维度,从存储类型上分为结构化、非结构化与半结构化数据,这是数据库存储的通用模式。结构化数据使用数据库二维逻辑表呈现,结构和数据长度是预先定义好的(如学习者的个人信息);非结构化数据则无固定结构与长度,允许重复字段,涵盖所有格式的文本、图像、音频、视频、超媒体、报表等(如学习管理系统里的文本和多媒体学习资源,社交网络中的资源信息),数据库中的模型结构呈树状和图状;半结构化数据介于上述两者之间,属结构变化大且与内容共同存储的形式,呈现一种自描述现象,HTML文档即为其中的代表(学习者在个人学习环境、学习管理系统中的描述性数据均为此类,如电子邮件、

评论)。

从可变性来看,可分为静态数据与动态数据。学生信息系统里的大部分学习者个人信息基本保持稳定,不会因时间变化而改变其属性,属静态数据(如姓名、性别等)。在线学习行为由结构性与层次性的活动组成,具有时间特性,产生的数据常直接反映事务过程,所以基本上为动态数据(如在线时间、访问量等)。在后期进行分析时,与行为特征关联的因素既包含了静态变量,又要采用大量的动态特征变量组。在构建研究模型中,对定义的在线学习行为因素进行筛选及分类,建立有效的特征子集,适当增加学习特征因素提取的层数,提高分析精度,并重点分析和测量行为认知因素的有用性、资源环境认同度等,在线学习行为类型具有多样性,在进行这些行为时学习者展现处理问题能力又具有复杂性分层,因此需要动态调整模型中的各参数,实现性能的不断优化。

(二) 机制

模型的第二层是分析机制模块,关注行为特征数据库中各因素间的关联与内涵的探究。此时待分析的数据已经过清理、标准化转换、整合、离散化、规约化的预处理,可按照应用目的设计相应分析机制,同时兼顾时间线将分析结果提供即时反馈和预测。

1) 即时反馈。用以描述和解释以当前时间点为终点的过程情况,包含对学习资源使用的反馈、对学习状态评价的反馈。在学习资源的反馈机制中可加入自适应引擎进行资源的智能化调整,以期更好地配合学习者的学习需求。学习者状态评价反馈涉及聚类分析、时间序列分析、规则推理、知识获取等技术的整合,实现客观、全面反映学习者学习记录的数据,以及同层次的比较数据,可为学习者当前的状态进行解释,提供诊断,便于及时进行学习改进。

2) 学习适应度。在学习者使用系统的过程中是否与学习环境相融合,以及在学习过程中的学习状态是否良好,可利用系统自适应模块进行分析。

3) 学习诊断与反馈。为了便于学习者及时了解自身学习效果,教育者及时进行教学干预,平台应用应包含即时性学习诊断与阶段性学习诊断。

预测用以对学习将来可能出现的状态进行预先推断。预测技术融合了可视化技术、回归分析、人

工智能、机器学习及人工神经网络,在建立因素与任务相关性映射的基础上,发掘学生的学习动机、归因和自我效能感,从认知、效果、矢量关联的角度进行应用;在应用过程中,对学习者的学习风格及偏好、学习计划、学习需要等诸多状态进行预测。

4)学习风格及偏好判断与反馈。对于学习主体,由于长期下意识行为而形成的学习习惯,可通过对学习者在使用的在线学习平台产生的访问路径、对象、频率等数据的采集和分析,预测出每位学习者的个人风格及偏好,有利于后续学习计划的制定,以及系统资源的推送。

5)学习计划制定。在了解学习者的风格及偏好基础上,分析出可能适宜学习者的学习计划,提供各类学习方案,便于学习者定制个性化学习。

6)学习需要。学习需要可以理解为学习者的学习动机,是学习者追求成功的一种心理倾向,包含内容兴趣和过程兴趣两部分。前者指学习者对学习目标及主题感兴趣;后者则反映学习者对学习策略的可预期,二者共同决定学习者对后续学习的参与程度。利用在线数据可预测学生的学习需要,通过系统智能化推荐相关知识点帮助其学习,或反馈给教师寻求帮助。

(三) 结果

模型的第三层即为分析结果的可视化呈现,最终对象包括学习者、教师、家长、管理层与开发人员。学习者作为数据的提供者又是数据的最大受益者,是整个模型的关键;可视化的分析结果呈现给教师、家长或课程领导者,便于他们通过干预机制对学习者的学习状态进行指导。对家长而言,分析结果提供了深入了解孩子学习状况的渠道,可以具体关注孩子学习过程的各个细节,弥补教师无法兼顾每位学生的缺点;对教师而言,分析结果便于教师动态、全面地掌握整体的学习情况(如什么时候学生准备

好进入下一阶段学习,哪些学生需要给予特别关注),对其有更清晰的认识从而开展适宜的教学活动及教学干预;对课程领导者而言,可以得到关于某门课程甚至学校的总体教学诊断情况,便于其对课程方案、教学方案及其它相关教育决策的制定;对学习本身而言,他(她)可以清晰地知晓自己当前的学习状态与未来可能的学习结果,从而提前采取措施,修正学习行为和计划。

资源分析结果可呈现给课程开发者、课程设计师,也可呈现给学习者。开发人员通过分析结果可以进一步改进系统设计以及资源结构调整,有助于在线课程的合理设计、在线学习平台的开发及教育资源的有效利用。学习者则获得适应性更高的学习资源,提高学习效率。在资源效果分析研究过程中,算法分析遵循以下流程:通过网站分析工具 linktag 标识,准确标识细分流量渠道,标识流量源;定制学习 KPI(Key Performance Indicator, 简称 KPI),监测学生浏览资源路径,改进用户访问体验,提升转化率;分析端到端的 ROI:从流量源-学习效益回报分析;确定优化方向,确定改进哪些资源或功能能有效提高学生的学习效率。

四、在线数据挖掘模式

借鉴商业领域的网络分析思路,将学习数据的挖掘工作分为三类(见图2),图中虚线部分表示是关联性不明确,待进一步分析的数据:

1) Web 内容挖掘:教育领域的 Web 内容挖掘体现在对网络文档中的内容及其描述进行有效信息的提取,包括内容的检索、过滤、去杂、异构集成、聚类、OLAP 等过程。从学习行为分析的角度,主要进行文本语义分析及多媒体挖掘两类。文本语义分析赋予网络文本以计算机可理解的语义逻辑,其中元结构本体的定义是关键,它是人类对特定领域信息

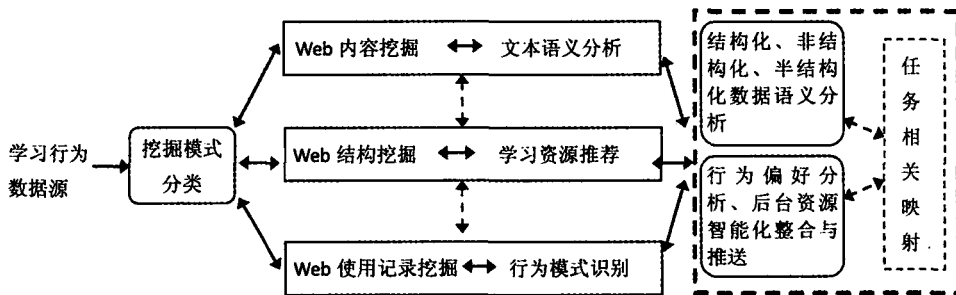


图2 在线学习数据挖掘模式

的通用解析,呈现对知识的结构化与形式化的陈述。利用 Web 语义技术从原本无结构的网络文本 HTML 或 XML 中抽取特征向量,建立可以表述和替代原始文本的抽象模型,使计算机能更好地实现自动化知识抽取、查询和搜索。随着多媒体技术的广泛使用,诸如图像和视频课程资源的兴起,多媒体挖掘也在在线学习行为分析中充当一定角色,它与文本挖掘的区别在于提取的特征向量不再是文本内容信息,而是与图像或视频相关联的类型值、文件 URL、关键值、颜色灰度值等信息,对这些特征向量建立 Web 仓库,重点进行关联发现,采用本体映射、结合规则发现,建立多维分析模式。

2) Web 结构挖掘:指从在线学习平台后台获取各文档页面之间的组织结构关系,并从网状链接的复杂多维关系中推导出有效知识的过程。应用于学习中的 Web 结构挖掘目前还没有引起人们较多的兴趣,但在整个 Web 信息空间中,页面超链接之间隐含的有用知识也是不可忽视的,比如结合最权威页面机制,在对学习页面进行分类和聚类后,进而找到最精华的学习资源,或根据学习者的访问结构为其推荐最合适的页面,构造辅助于教学的学习资源推荐功能。

3) Web 使用记录挖掘:通过挖掘后台服务器中保存的访问日志,可以获取学习者使用学习平台产生的各种访问及交互信息,从中抽取有用的数据,形成目标模式;分析得到用户的使用习惯、行为偏好、学习进度,有助于系统结构的调整与改善,并可以对学习者形成个性化的学习服务。这一挖掘分析在当前应用最广泛。已有研究表明,诸如“参加论坛中的讨论次数、线上练习测试次数、邮件发送次数”等数据均与学习者最终表现关联。使用记录的挖掘结合人工智能、信息论以及机器学习等领域的成熟技术,在数据处理时重点是如何进行有效数据甄别及事物识别,结合不同的技术分析访问模式,诸如预测技术,即通过数据集中的某一部分(预测变量)去推断另一部分的可能表现(被预测变量);聚类技术,即将一个整体数据集分成多个类别,每一类中都包含具有相似性能的一簇数据;关系挖掘技术(Relationship mining),即在一个数据集里发现变量之间的关联,并将其编码待用;时序模式挖掘(Sequential pattern mining),即捕捉一系列发生的

事件之间的关联;路径分析,即寻找用户频繁访问的路径模式。基于上述技术建模分析完成后,可以利用可视化技术和工具对结果进行呈现。

上述挖掘模式将在线学习行为可能产生的数据进行归类。随着应用目的变化,各种模式的研究和发展都在不断深化,如何将三种模式有效结合进而达到最优的分析结果是我们当前正在探究的课题,由于结构关系分析中需要涉及行为理论的表达与理解,从而影响任务相关性映射,因此如何进行自底向上的特征提取与关联分析,其中有许多机理有待进一步研究。

五、基于电子课本的数据挖掘个案设计

电子课本作为兼具数字阅读与数字学习双重特性(顾小清等,2012)的电子学习新形式,开始逐步应用于学前与中小学课堂教学中。本案例的设计拟依托电子课本对 K12 学生的知识点掌握情况进行数据挖掘与分析,最终实现对学习者的知识点掌握情况诊断与反馈。其前提是电子课本已对知识点进行了细粒度的资源结构化呈现,构建了可重用知识点的教学设计,提供了相关的练习资源(每道练习题配有答案提示,以辅助其完成答题),基于上述情境可以产生分析需要的各项细粒度学习数据(Fine-grained learning data)。遵循在线学习行为分析模型(见图1)的三层次结构,分析过程如下:

(一)数据

主要获取学习者在电子课本中产生、发布的数据,以及部分关联的智能数据。数据挖掘模式采用 Web 内容挖掘与 Web 记录挖掘相结合的形式体现任务的关联性。Web 内容挖掘获取的数据包括:学习者发布的关于知识点的评论文本、回答练习题的描述性答案、练习的错误率、错误的内容。Web 记录挖掘获取的数据包括:学习者练习的次数、每道练习题的提示次数和质量(即请求的深度)、错误答案的重复次数、回答每道题的反应时间。根据已有文献研究结果,学习者登录次数、页面停留时间及资源点击次数等常规性数据与学习行为无直接关联,故研究设计不考虑此类数据。

(二)机制

由于具有内容挖掘与记录挖掘两种模式,数据归一化前处理机制各有不同。内容挖掘采用文本内

容分析与知识获取技术对半结构化数据进行知识发现。记录挖掘则在对数据进行标准化转换、整合处理后采用规则推理、时间序列分析对各因素变量进行关联映射。归一化处理后,对整体数据进行聚类与回归分析,形成与最终结果正(负)相关的各因素权重,汇合形成掌握情况的相关报告。再结合人工智能与机器学习技术,实现个性化应用。

(三) 结果

本案例的设计最终希望实现对学习者的知识点掌握情况反馈。基于数据的收集与机制分析后可呈现两个角度的结果反馈:1)对学习者的分析结果直接呈现学习者的知识掌握水平报告及该过程的学习状态,实现对学习者的水平诊断,判断其是否可以进入下一阶段学习。如果达到标准则进行下一学习主题的推荐,如没有则建议巩固当前知识,加强练习;2)对教师,分析结果不仅反映每位学习者最终的知识掌握程度,并呈现具体的学习过程信息,如了解区分哪些学生没有尝试练习所以没能掌握知识,哪些是尝试了但仍然困惑的。这些信息能帮助教师针对不同类别采取不同教学策略;或是针对知识点的练习中大量学生寻求提示,并且重复练习的错误率较高,这些信息提示教师某知识点较难,学生普遍未能掌握,需要重复讲解。

六、结语

移动互联技术的飞速发展使得教育工作者和信息技术人员都坚信:在不远的未来,教育领域会掀起一波革命浪潮,线上教育将成为不可替代的学习模式。学习评价作为学习系统的反馈调节机制,在学习与教学过程中起重要作用,在基于测量、描述、判断和建构四种模式的基础上,如何利用网络行为数据辅助动态生成学习效果,是对在线教育最客观的分析。在线学习行为分析以其特殊的行为反馈机制,综合了自底向上的信息处理加工和自顶向下的行为关联分析的双向处理模式,成为在线教育蓬勃发展的伴随产物,是世界教育技术研究前沿所在,也对充分调动学习者的积极性,发掘学习者潜能,合理利用学习资源具有重大价值。

[参考文献]

[1] Agrawal, R., & Shafer, J. (1996). Parallel mining of associ-

ation rules [J]. Knowledge and Data Engineering, IEEE Transactions on, (8):962-969.

[2] Alkhattabi, M., Neagu, D., & Cullen, A. (2011). Assessing information quality of e-learning systems: A web mining approach [J]. Computers in Human Behavior, 27(2):862-873.

[3] Baker, R., Corbett, A., & Aleven, V. (2008). More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing [C]. In Proceedings of the 9th International Conference on Intelligent Tutoring Systems. Berlin, Heidelberg: Springer-Verlag:406-415.

[4] Baker, R., Corbett, A., Koedinger, K., & Roll, I. (2006). Generalizing detection of gaming the system across a tutoring curriculum [C]. In Proceedings of the 8th International Conference on Intelligent Tutoring Systems. Berlin, Heidelberg: Springer-Verlag:402-411.

[5] Baker, R., Corbett, A., Koedinger, K., & Wagner, A. (2004). Off-task behavior in the cognitive tutor classroom: When students "Game the System" [C]. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04). New York, NY: Association for Computing Machinery:383-390.

[6] Banu, R., & Ramanan, R. (2012). A study on data mining in e-learning - Empowering students education through social networks - A novel approach [C]. 2012 International Conference on Education and e-Learning Innovations, Sousse:1-6.

[7] Blikstein, P. (2011). Using learning analytics to assess students' behavior in open-ended programming tasks [C]. Proceedings of the First International Conference on Learning Analytics and Knowledge. New York, NY: Association for Computing Machinery:110-116.

[8] Cao, D., & Do, P. (2012). Applying data mining in money laundering detection for the vietnamese banking industry [M]. Intelligent Information and Database Systems:207-216.

[9] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide [M]. Technical report, CRISP-DM.

[10] Chellatamilan, T., Ravichandran, M., & Suresh R. M. (2011). Effect of mining educational data to improve adaptation of learning in e-Learning System [C]. Proceedings of the International Conference on Sustainable Energy and Intelligent Systems. Chennai:922-927.

[11] Conte, T., Vaz, T. V., Massolar, J., Emilia, M., & Guilherme, H. T. (2008). Process model elicitation and a reading technique for web usability inspections [C]. Web Information Systems Engineering - WISE 2008 Workshops: 5176, 36-47.

[12] Cristobal, R., Sebastian, V., & Amelia, Z. (2009). Applying Web usage mining for personalizing hyperlinks in web-based adaptive educational systems [J]. Computers & education, 53(3):828-840.

[13] Cui, Y. (2009). The research on the data mining technology application on telecom customer analysis [J]. Microcomputer Information, (15):155-156.

- [14] Dawson, S., Heathcote, L., & Poole, G. (2010). Harnessing ICT potential: The adoption and analysis of ICT systems for enhancing the student learning experience[J]. *International Journal of Educational Management*, 24 (2): 116-128.
- [15] Dienes, Z., Altmann, G. T. M., Kwan, L., & Goode, A. (1995). Unconscious knowledge in artificial grammars is applied strategically[J]. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21(5): 1322-1338.
- [16] Dodero, J. M., Ruiz-Rube I., Palomo-Duarte, M., & Juan, V. M. (2011). Open linked data model revelation and access for analytical web science[M]. *Metadata and Semantic Research*. Springer Berlin Heidelberg:105-116.
- [17] Elisa, T. (2011). Learning analytics: definitions, processes, potential[EB/OL]. [2012-10-10]. <http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf>.
- [18] 范洁(2005). 基于数据挖掘的在线学习行为评估系统设计与实现[D]. 国防科技大学.
- [19] 冯锐(2006). 网络学习支持系统中学生特征分析模型的构建及技术实现[J]. *中国电化教育*, (12):101-103.
- [20] Géczy, P., Izumi, N., Akagon, S., & Kôiti, H. (2008). Web behaviormetric user profiling concept[M]. *E-Commerce and Web Technologies*. Springer Berlin Heidelberg:134-143.
- [21] 顾小清, 傅伟, 齐贵超(2012). 连接阅读与学习: 电子课本的信息模型设计[J]. *华东师范大学学报(自然科学版)*, (2):81-90.
- [22] 顾小清, 张进良, 蔡慧英(2012). 学习分析: 正在浮现中的数据技术[J]. *远程教育杂志*, (1):18-25.
- [23] Golmohammadi, K., & Zaiane, O. R. (2012). Data mining applications for fraud detection in securities market[C]. *Intelligence and Security Informatics Conference, 2012 European*:107-114.
- [24] Huang, S. C., & Wu, C. F. (2011). Customer credit quality assessments using data mining methods for banking industries[J]. *African Journal of Business Management*, (15): 4438-4445.
- [25] Huang, S. M., Wang, L., & Wang, W. C. (2011). Adopting data mining techniques on the recommendations of library collections[C]. *Proceedings of the 2011 International Conference on Information & Knowledge Engineering*:46-52.
- [26] 黄雄伟(2011). 基于 Web 数据挖掘的客户行为分析研究及应用[D]. 武汉理工大学.
- [27] Hung, J. C., Hsu, Y. C., Rice, & Kerry(2012). Integrating data mining in program evaluation of K-12 Online Education[J]. *Educational Technology & Society*, 15 (3):27-41.
- [28] Iwata, T., & Sawada, H. (2013). Topic model for analyzing purchase data with price information[J]. *Data Mining and Knowledge Discovery*, 26, (3):559-573.
- [29] Jovanovic, M., Vukicevic, M., & Milovanovic, M. (2012). Using data mining on student behavior and cognitive style data for improving e-learning systems: a case study[J]. *International Journal of Computational Intelligence Systems*, 5(3):597-610.
- [30] Karin, A. H., & Helmut, H. (2003). Anytime, anywhere learning behavior using a web-based platform for a university lecture[C]. *Proceedings of the SSGRR 2003 Winter Conference*.
- [31] KdNuggets. Com(2002). Data mining methodology[EB/OL]. <http://www.kdnuggets.com/polls/2002/methodology.htm>.
- [32] KdNuggets. Com(2004). Data mining methodology[EB/OL]. http://www.kdnuggets.com/polls/2004/data_mining_methodology.htm.
- [33] KdNuggets. Com(2007a). Data mining activity in 2007 vs 2006[EB/OL]. http://www.kdnuggets.com/polls/2007/data_mining_2007_vs_2006.htm.
- [34] KdNuggets. Com(2007b). Data mining methodology[EB/OL]. http://www.kdnuggets.com/polls/2007/data_mining_methodology.htm.
- [35] Khabaza, T., & Shearer, C. (1995). Data mining with Clementine[C]. *Knowledge Discovery in Databases, [IEEE Colloquium on]*, 1-5.
- [36] Kurgan, L. A., & Musilek, P. (2006). A survey of knowledge discovery and data mining process models[J]. *Knowledge Engineering Review*, 21(1): 1-24.
- [37] 新丽(2008). 在线学习中学习者学习行为记录模块的设计开发及相关数据分析研究[D]. 内蒙古师范大学.
- [38] Lee, W. I., Shih, B. Y., & Tu, L. J. (2002). The application of KANO's model for improving web-based learning performance[C]. *32nd ASEE/IEEE Frontiers in Education Conference 2002*, (1): T3E-27-T3E-32.
- [39] Lo, J. J., & Shu, P. C. (2005). Identification of learning styles online by observing learners' browsing behaviour through a neural network[J]. *British journal of educational technology*, 36(1), 43-55.
- [40] 李德江, 张鹏, 田志英, 诸葛华, 茹鹏新(2010). 分布式学生网上学习行为统计系统设计与实现[J]. *新疆广播电视大学学报*, (2):5-9.
- [41] 李军, 曾文君, 李杨(2008). FB-NBAS: 一种基于流的网络行为分析模型[J]. *计算机工程*, 34(3):165-167.
- [42] 廖竞, 张晖(2011). 一种灵活的在线学习行为数据采集与分析系统[J]. *信息与电脑(理论版)*, (1):85-86.
- [43] Macfayden, L. P., & Dawson, S. (2010). Mining LMS data to develop an "Early Warning" system for educators: A proof of concept[J]. *Computers & Education*, 54 (2): 588-599.
- [44] 潘蕾, 朱红霞(2011). 网络访问行为分析模型的研究与设计[J]. *计算机与现代化*, (9): 140-143.
- [45] 彭文辉, 杨宗凯, 黄克斌(2006). 网络学习行为分析及其模型研究[J]. *中国电化教育*, (10):31-35.
- [46] Reber, A. S. (1993). Implicit learning and tacit knowledge: An essay on the cognitive unconscious[M]. New York: Oxford university press.
- [47] Siemens, G., Gasevic, D., Haythornthwaite, C., Buckingham, Shum S., Ferguson, R., Duval, E., et al. (2011). Open

Learning Analytics: an integrated & modularized platform[J]. Proposal to design, implement and evaluate an open platform to integrate heterogeneous learning analytics techniques. Society for Learning Analytics Research.

[48] Solé, M. (2010). DBMiner user manual[M]. Simon Fraser University, Intelligent Database Systems Laboratory.

[49] Sun, J. H., & Xie, Y. Q. (2009). A recommender system based on web data mining for personalized e-learning[C]. International Conference on Information Engineering and Computer Science:1-4.

[50] Tkach, D. S. (1998). Information mining with the IBM intelligent miner family[R]. An IBM Software Solutions White Paper:1-29.

[51] 王攀, 张顺颐, 陈雪娇(2008). 基于 DBP 的 Web 用户行为分析关键技术[J]. 电信快报, (8): 13-15.

[52] Wang, Y. F., Chang, M. Y., & Chiang R. D. (2013). Mining medical data: A case study of endometriosis[J]. Journal of medical systems, 37(2):1-7.

[53] Witten, I. H. & Frank, E. (2005). Data mining: Practical machine learning tools with java implementations[M]. Morgan Kaufmann.

[54] 肖英(2012). 基于 MRF 模型的网络交易行为分析研究[D]. 中南大学.

[55] 徐红彩(2005). 在校大学生网络学习行为的调查与研究[J]. 电化教育研究, (6): 61-63.

[56] Yang, F. Y., & Tsai, C. C. (2007). Investigating university student preferences and beliefs about learning in the web-based context

[J]. Computers & Education, 50(4):1284-1303.

[57] 杨金来, 张翼翔, 丁荣涛(2008). 基于网络学习平台的学习行为监控研究[J]. 计算机教育, (11): 65-68.

[58] 杨清莲, 周庆敏, 常志(2005). Web 挖掘技术及其在网络教学评价中的应用[J]. 南京工业大学学报(自然科学版), (27): 100-103.

[59] Yu, P. (2011). Data mining in library reader management[C]. Proceedings of the 2011 International Conference on Network Computing and Information Security, 2: 54-57.

[60] 曾祥跃, 袁松鹤(2008). 远程学习者网络学习行为的调查与分析[J]. 中国远程教育, (4): 47-51.

[61] Zhang, Y., Oussena, & Clark(2010). Use data mining to improve student retention in higher education: A case study[C]. Proceedings of the 12th International Conference on Enterprise Information Systems: 190-197.

[62] 周岩(2009). 基于 TRA 和 TAM 的大学生网络学习行为模型构建[J]. 中国电化教育, (11): 58-62.

[63] Zhu, T. T., & Zhang, L. L. (2011). Application of data mining in the analysis of needs of university library users[C]. 2011 6th International Conference on Computer Science & Education : 391-394.

[64] 祝智庭, 贺斌(2012). 智慧教育: 教育信息化的新境界[J]. 电化教育研究, (12): 5-13.

(编辑: 李学书)

A Continuing Process of Learning Analytics: System Modeling and Mining of Online Learning Behavior

HU Yiling¹, GU Xiaoqing¹ & ZHAO Chun²

(1. Department of Educational Information Technology, East China Normal University, Shanghai 200062, China; 2. Network & Information Center, East China Normal University, Shanghai 200062, China)

Abstract: With the fast development of sophisticated digital technology, learning mode has been digitized by the internet and social media. Massive data of learning behaviors have been created and are growing, showing various forms of learning behaviors, explicitly and implicitly, which carry the information critical to assess learning and promote instruction. In this paper, three probable directions of online learning behavior which including data mining and domain of application, learning behavior and analysis of behavior and analytical model of network behavior are reviewed. Then the behavior modeling mechanism of learners' online learning, establish three hierarchical model with data, mechanism and result are discussed. Finally, from the experience of web data mining, the pattern of learning data mining are categorized and analyzed. The research provides a framework and a good way to the analysis of online learning behaviors, which has important practical value to fully motivated learning. It widely explores learner's potential and makes good use of the learning resources.

Key words: learning analytics; online learning; analysis of behavior; data mining; system modeling