

学习分析：从源起到实践与研究

吴青 罗儒国

(武汉大学教育科学学院,湖北武汉 430072)

[摘要] 学习分析是“大数据”在教育领域的应用,引发了教育技术发展的第三次浪潮,并获得学术界的广泛关注。本文梳理了学习分析的形成过程,然后从利益相关者、研究目标、研究对象、技术方法四个维度,回顾了近五年来国内外学者在学习分析方面的研究成果,并提出未来发展趋势和可能遇到的挑战,便于相关人员制定教育决策、优化教育管理过程以及完善学习过程。研究表明,学习分析研究主题主要涵盖学习者知识建模、学习情绪建模、学习行为特征抽取、学习活动跟踪、学习者建模、学位获取分析、教学资源 and 教学策略优化、自适应学习系统和个性化学习、在线学习影响因素分析九个方面;分析数据主要来源于集中式学习环境、分布式学习环境以及身体活动数据;常用分析方法包括统计分析、信息可视化、数据挖掘、社会网络分析、话语分析和网站分析。目前,学习分析研究遇到的挑战包括教育数据预处理难度大、数据访问权限不明确、学习分析适用性有限。虽然学习分析尚处于发展初期,但由于能够为教育系统各级决策提供科学参考,已经成为教育信息化的重要内容之一。

[关键词] 学习分析;教育数据挖掘;社会网络分析;信息可视化

[中图分类号] G434 **[文献标识码]** A **[文章编号]** 1007-2179(2015)01-0071-09

互联网的迅猛发展推动了大数据研究,越来越多的教育研究者意识到数据分析在改善学习体验方面的潜能。因此,学习分析(Learning Analytics,简称LA)逐渐从分析领域独立出来,吸纳数据挖掘、社会网络分析、统计分析等多种分析方法,形成一个独立的新兴领域。从2010年该领域正式形成至今,学习分析被应用于教育系统各个层面。例如,微观层面关注个体学习活动,为学习过程提供实时帮助。中观层面充分利用教育管理数据,实现教育资源配置优化。宏观层面则综合考量中观与微观层面,有利于教育行政决策的制定和完善。然而,鉴于学习分析是个相对年轻的跨学科研究领域,各国学者对其理解和界定众说纷纭,本文从利益相关者、研究目标、研究对象、技术方法四个维度,厘清学习分析的源起、适用对象、目标范围和技术实现等,从而促进学习分析在我国的应用和发展。

一、源起

在线学习和大数据出现之前,教育机构和个人已经开展了教育数据的分析和评估。例如,英国开放大学在1969年开始记录和分析远程学习者每学年的课程成绩(McIntosh, 1979)。廷托(Tinto, 1997)经过持续20年的数据收集和整理,于1997年发表了关于学习者保持率影响因素的研究成果。随着21世纪初第二代“读/写”网络的兴起,越来越多可用于分析的海量数据集逐步显现。例如,1994年英国只有7%的高等教育机构使用虚拟学习环境,2001年在线学习覆盖率达40%,2003年超过85%。为了更好地了解学生及其学习环境,数据挖掘技术在教育领域的应用范围逐渐扩大。为探索上述数据蕴含的教育特征和规律,“教育数据挖掘”(Educational Data Mining,简称EDM)作为数据挖掘的子

[收稿日期] 2014-10-13 **[修回日期]** 2014-12-25 **[DOI编码]** 10.13966/j.cnki.kfjyjj.2015.01.008

[基金项目] 湖北省教育科学“十二五”规划2013年度课题“基于网络学习行为的学习风格挖掘及其教学策略研究”(2013B004);全国教育科学“十二五”规划2013年度教育部青年课题“基于数据挖掘的网络学习行为分析及其教学策略研究”(ECA130375)。

[作者简介] 吴青,博士,武汉大学教育科学研究院讲师,研究方向:学习分析、计算机支持的协作学习(58554930@qq.com);罗儒国,博士,武汉大学教育科学研究院副教授,研究方向:课程与教学基本理论、教师发展研究。

集,于2005年被首次提出。2006年,罗梅罗(Romero)等首次公开发表《在线学习中的数据挖掘》(Data mining in e-learning)一书,引起学术界的广泛关注。2008年,第一届教育数据挖掘国际学术会议召开。

然而,教育数据挖掘专注于技术层面,侧重教育数据模型和模式的抽取,过多强调结果自动化反馈(Rebecca, 2012a)。随着以学习为中心观点的崛起以及政治、经济因素的推进(Rebecca, 2012b),教育研究者更加专注于研究有利于改善学习的干预措施。2011年,美国德州大学阿灵顿分校(University of Texas at Arlington)的西蒙斯(Siemens)、加拿大阿萨斯卡大学(Athabasca University)的德拉甘(Dragan)和悉尼科技大学(University of Technology Sydney)的岑(Shum)等国际知名学者成立了一个跨学科网络组织——学习分析研究协会(The Society for Learning Analytics Research, 简称 SoLAR)。该组织旨在通过运用学习分析,提高学术研究标准、促进开放教育资源发展、提高政策制定者和决策者的分析意识、推进利益相关者的协作沟通和讨论。这标志着学习分析逐渐从教育数据挖掘领域独立出来,融合包括学习科学、统计学、计算机科学、信息科学、社会学等其他学科的技术方法,形成一个独立的新兴领域。迄今为止,该组织已经成功举办了四届学习分析与知识国际会议(The International Conference on Learning Analytics & Knowledge),为学习分析在学术界的推广及在教育界的实践起了极大作用。与此同时,SoLAR于2013年和2014年分别在斯坦福大学和哈佛大学举办学习分析暑期班,搭建知识分子和社会之间的桥梁,目的在于加快学习分析的学科成熟度。此外,该组织创办《学习分析杂志》(The Journal of Learning Analytics)作为官方出版物,促进学习分析研究成果的出版和传播。

美国新媒体联盟(New Media Consortium)发布的地平线系列报告从2011年开始将学习分析作为未来发展的重要主题。2011年该报告指出学习分析将在4-5年内普及,到2014年认为近1年将得到推广。2011年的报告较为狭隘地认为学习分析的受益者是教师和管理人员,而到了2013年将受益者扩大到教育系统各个层面,包括管理者、政策制定者和立法者、教育工作者和研究人员、学生。2011年的报告仅强调数据挖掘在该领域的应用,2012年突

出谷歌分析软件在在线学习系统的使用,2013年引入网络跟踪工具,2014年则重点强调需借鉴商业智能的成功应用,并结合大规模开放在线课程作为推广背景。此外,2014年位于美国波士顿的皮尔森数据分析和自适应学习中心的资深科学家拉尔森(Larusson)较早出版了相关著作《学习分析》。

目前,国内尚未形成一个专门的学习分析学术研究组织,但该技术已经引起国内学者的广泛关注。例如,北京师范大学黄荣怀和华东师范大学祝智庭将学习分析作为构建智慧学习环境的支撑技术之一,用于理解和优化学习过程和学习情境。华东师范大学顾小清将学习分析应用于“下一代互联网教育创新支持系统”项目,实现对学习过程的监测与评价。国家开放大学魏顺平选取该校学习管理系统中存储的海量学习数据,分析成人学习者在线学习现状及影响因素。清华大学张羽将学习分析应用到MOOCs中,促进教育质量的提升。由此可见,随着大数据在教育领域的渗透,学习分析已经成为国内外教育工作者关注的热点。

二、研究方法

本研究以“学习分析”作为主题词,选择近五年发表在CSSCI来源期刊的国内相关学术文献;以“learning analytics”为关键字,从学习分析与知识国际会议论文、美国高校教育信息化协会和英国开放大学发布的系列报告、SSCI国际权威学术期刊、Google Scholar检索国际相关文献。国内外文献总计269篇,其年度分布如图1所示。自2010年学习分析概念被提出后,相关学术论文数量呈上升趋势,其中2012年到2013年的文献数量保持相对峰值,体现了学术界对该领域较高的关注度。由于2014年发表的部分文献还未上传到相关数据库,2014年的数量相对较少。

三、利益相关者

大部分研究者趋向基于教育系统中参与角色的不同,将学习分析利益相关者划分为学生、教师和研究人员、教育管理者、政策制定者和立法者(Greller et al., 2012; Chatti, 2012; 李青等, 2012; 李艳燕等, 2012)。而岑(Shum, 2012)根据学习分析数据访

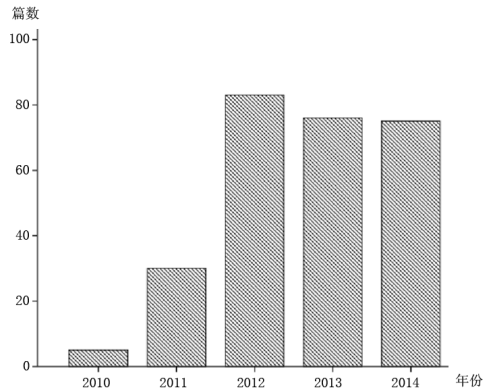


图1 学习分析文献的年度分布

问的级别和环境差异,将教育系统划分为微观、中观和宏观三个层面。微观层面主要针对个体学习者的过程数据,目的在于识别风险学习者并提供干预,同时为学习者提供观察自己学习习惯的机会,并给出改善建议。微观层面的利益相关者主要包括学生、教师和研究人员。例如,美国亚利桑那州立大学为全面了解学生个人信息而开展的“Student 360”项目(Crow,2012)。中观层面重点关注制度层面的教育数据,便于更好地理解变量影响差异,优化教育资源分配,提高组织工作效率。中观层面的利益相关者主要包括教育管理者,如利用学习分析提高学生成绩合格率和毕业率(邓文俊等,2014),制定财务计划(EDUCAUSE,2012)等。宏观层面的分析旨在开展跨机构分析,合并中微观层面数据,比较不同运营模式、学术模式或教学方法等,完善教育行政决策。宏观层面的利益相关者主要包括政策制定者和立法者。例如,“美国毕业生计划”指出2020年需要额外500万高等教育毕业生(Smith et al.,2012),为教育政策制定提供方向。上述两种观点只是在粒度划分上存在差异,本质高度一致,即学习分析服务对象包括教育系统自下而上的各个层面。基于利益相关者聚类学习分析文献的年度分布如图2所示。绝大多数文献关注微观层面,占81.4%;宏观层面的文献最少,仅占5.6%。由此可见,学生、教师和研究人員是学习分析研究主体,主要关注如何利用学习分析的结果增强教学实践有效性。而宏观和中观研究随着年度增长,文献数量呈逐年上升趋势。这说明教育管理者、政策制定者和立法者开始意识到学习分析在教育数据管理中的重要性和实用价值,主要通过研究报告的形式发布相关研究成果。

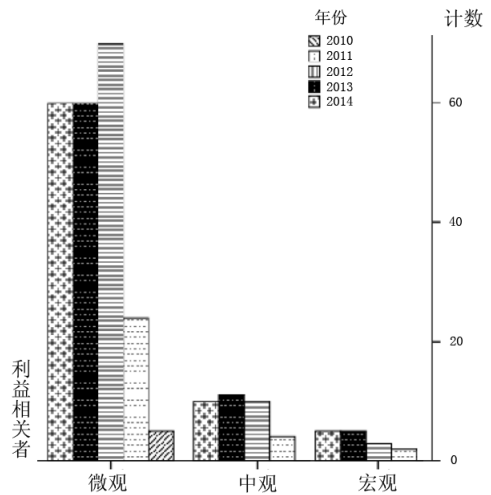


图2 按利益相关者聚类的学习分析文献年度分布

四、研究目标

佛罗里达州立大学信息技术领导学院院长、执行董事梅(May,2011)较早提出学习分析的研究目标包括描述性和预测性两大类。德国亚琛大学的沙提(Chatte et al.,2011)在此基础上,基于具体分析过程将研究目标细分。然而沙提的分类存在一定的内容交叉,例如,“预测和干预”同时需要对学习者的知识结构或学习行为开展“监测和分析”。因此,宾科夫斯基(Bienkowski et al.,2012)从学习分析需实现的功能模块角度提出新观点,其分类涵盖了教育数据挖掘和学习分析两大领域,如“领域知识建模”隶属教育数据挖掘领域,并非学习分析的关注重点。然而,该分类体系粒度划分不均衡。例如,学习者情绪、学习行为范式、学习者当前状态等都是通过观测外显学习行为表达的,导致过半数文献均可归结于“学习者行为建模”子类,粒度划分过于粗犷。此外,“趋势分析”“学习者体验建模”等分类太过模糊,导致部分文献可同时隶属多个类别。李艳燕等(2012)的观点也存在类似情况。上述四位学者的观点及观点之间的对应关系如表一所示。

本文借鉴上述学者观点,对269篇文献重新梳理,归纳出学习分析的研究目标主要包括:

(一)学习者知识建模

为了描述学习者知识和技能的掌握情况,研究者从课程、知识单元和知识点等多个层面,抽取在线学习系统中学习者的交互数据,构建学习者知识模型。该模型主要应用于自适应学习系统和智能教学

表一 学习分析研究目标

May, T. A	M. A. Chatti	Bienkowski, M.	李艳燕等
描述性	监测和分析	学习者知识建模 学习者行为建模	教学交互分析 学习者之间网络分析 学习者行为与情感分析
	辅导和指导	学习要素分析和 教学原理分析	教学资源分析
	个性化和推荐 智能适应	学习者建模 自适应性和个性化	学习者特征分析
		领域知识建模	
预测性	预测和干预	趋势分析	
	评估与反馈	学习者体验建模	
	反思		

系统等,体现了学习者过程性知识和高阶思维能力,便于系统在恰当时间采用恰当的学习方式推送恰当的学习内容。例如,弗格森(Ferguson et al., 2013)基于线索词匹配和k-近邻聚类技术,利用语篇特征和局部特征相结合的方式,构建学习者知识模型。

(二)学习情绪建模

为检测学习者无聊、沮丧、兴奋等不同情感状态对学习积极性和学习进展的影响,研究者利用心脏速率监视器、视觉跟踪器等多种可穿戴技术,收集和分析学习者心跳速率、微笑次数、专注时间等身体活动数据,掌握学习情绪对学习成绩的影响。例如,瓦特拉普(Vatrapu et al., 2013)通过技能测量、微笑等九个指标描述学习者学习状态。视觉追踪研究表明,较高的情绪容易促进学习者学习积极性。

(三)学习行为特征抽取

为研究不同学习行为范式和学习成绩之间的关系,研究者收集在线学习系统的网页点击次数、点击顺序、停留时间等信息,抽取学习行为范式或形成的网络结构特征,并研究其与成绩之间的联系。例如,加拿大萨斯卡通大学(University of Saskatoon)布鲁克斯(Brooks et al., 2014)获取在线学习系统中学习者交互数据,采用非监督机器学习技术,抽取五种学习者行为范式,包括活跃型、早期型、及时型、最少活动型和延期型,指出活跃型对提高学习成绩并无显著性影响。吴怵等(2014)对学习者图示化学习过程开展序列分析,发现具有“概念建构—假设提出—推理论证”学习行为模式的学习者比较容易取得好成绩。

(四)学习活动跟踪

为了展示学习者当前学习状态,研究者收集在线作业完成情况、教学视频学习时长、在线测评得分、论坛参与等信息,以可视化形式呈现知识建构过程和个人在小组学习中的贡献情况,有助于学习者调整学习计划和学习进展。例如,有学者针对可汗学院在线学习平台大量学习活动数据,基于分类器将全体学习者的学习状态进行等级划分,然后利用个体可视化工具查看个体相对于班级平均学习状态所处的位置(Ruipérez-Valiente et al., 2014)。

(五)学习者建模

为聚类学习特征相似的学习者,有研究者通过分析个人基本信息及相关数据,例如人口统计、学习风格和学习偏好、学习目标、学习背景等数据,构建描述个人学习特性的学习者模型,并运用该模型将学习者分组,提供有针对性的个性化学习环境,从而提高学习效率。例如,希腊开放大学针对在线论坛中学习者的讨论内容,利用文本挖掘和社会网络分析技术探索学习者的参与模式,并使用统计软件R和数据挖掘工具Weka将学习者按照特征分类(Lotsari, 2014)。北京师范大学武法提等(2014)分析电子书包中电子学档系统记录的数据,从学习内容、学习活动、学习方式和学习评价四个方面构建学习者个性化模型。

(六)学位获取分析

为探究学习群体信息与学位获取之间的关系,即学习者保持率或毕业率,研究者通过采集入学信息、生源信息、完成情况、学位信息等数据,从课程、学校和政府三个层面探索其中蕴含的序列模式或规律。例如,美国纽约州立大学谢伊等选取弗吉尼亚州和华盛顿州的入学信息和最终学位获取数据,发现参与远程教育课程的学习者比接受传统课程的学习者更易获得学位(Shea & Bidjerano, 2014)。

(七)教学资源 and 教学策略优化

为帮助教师完善在线课程,研究者通过收集在线学习系统中的学习行为和成绩,评估课程设计效果,找出能够有效促进学习的教学实践类型。例如,列文(Leeuwen, 2014)提出基于人工智能框架,从有效性、可用性和效率三方面评估教学大纲质量。北京师范大学沈欣忆等(2014)通过分析MOOCs学习的不足,提出十二种网络课程教学策略,以提高在线学习参与度。

(八) 自适应学习系统和个性化学习

为实现网络环境下学习行为的自动化反馈,研究者综合运用学习者知识模型等,获取学习偏好、学习效果等信息,结合预定义的教学策略和学习路径,为学习者提供个性化学习建议,从而调整和改善学习体验。例如,德国杜伊斯堡-埃森大学(University of Duisburg-Essen)哈金(Hecking, 2014)在学习管理系统中通过增加操作日志服务、通知代理、分析服务和人工制品检索服务四大服务组件,实现基于上下文的学习内容个性化推荐。深圳大学曹晓明等(2014)提出使用智能 Agent 技术构建“一对一”的泛在自主学习系统。

(九) 在线学习影响因素分析

由于各级教育机构问责制的加强,教育机构需要解释在线学习过程和学习效果,部分学者以调查问卷的形式分析在线学习影响因素。例如,加拿大西蒙特雷泽大学阿里等(Ali et al., 2013)指出在线学习工具的易用性和实用性将影响在线学习效果。中山大学舒忠梅(2014)指出多元能力的培养氛围、学习资源丰富程度、教师授课水平及相关系统支撑度是在线学习满意度的关键影响因素。

按研究目标聚类的学习分析文献年度分布如图3所示。其中,以教学资源和教学策略优化、学习活动跟踪为研究主题的文章分别占23.7%和23.4%,成为学习分析的主要研究内容。这与大多数研究者基于微观层面开展学习分析相吻合。2012年以来,学习行为特征抽取、学习情绪建模、在线学习影响因素分析逐渐成为热门主题。而自适应学习系统和个性化学习过多强调学习软件的自动反馈机制,更适于教育数据挖掘,因此近三年来相关文献数量呈下降趋势。随着学习分析应用的推广,其研究主题日益趋向多角度、多样化,例如学习者动机建模、大数据访问中正常访问行为的变化情况均成为新的研究内容。可见,在各类学习数据日益丰富的时代背景下,学习分析对教育信息化发展将起到更为突出的推动作用。

五、研究对象

研究对象的选取将直接影响学习分析的有效性和预测准确率。贝克(Baker, 2011)指出数据的层次性是学习分析数据的重要特征,包括按键层次、答

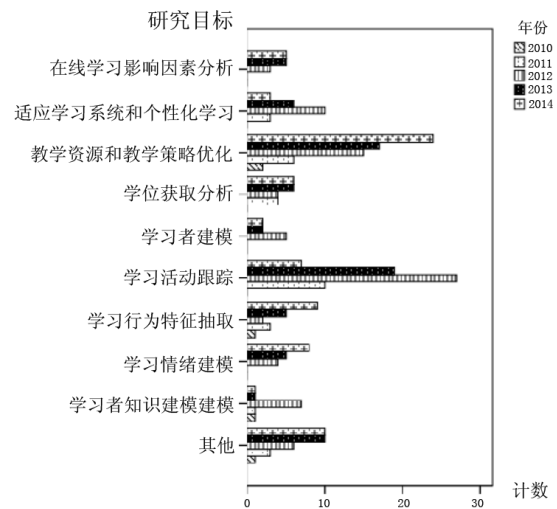


图3 按研究目标聚类的学习分析文献年度分布

案层次、会话层次、学生层次、课堂层次、教师层次和办学层次,每一层次都嵌套在上一层次中。布朗(Brown, 2012)基于数据的内容表征差异,将学习分析研究对象划分为个性特征指标(dispositional indicators)和行为表现指标(activity and performance indicators)。个性特征指标一般是可量化的事实性数据,如年龄、性别、种族、平均分、学习经验等;行为表现指标主要反映在线学习环境中的学习行为数字轨迹,如登陆学习管理系统的次数、在学习网站的时间、发帖次数、测验分数等。顾小清等(2012)认为包括学习者数据和学习过程相关数据两大类。前者指在移动终端、社会性交互软件和学习管理系统中记录的数据;后者包括与学习过程相关的课程、学期考试信息。

本文基于数据来源环境差异,认为学习分析研究对象主要包括来自集中式学习环境、分布式学习环境和可穿戴传感器的数据,文献年度分布如图4所示。其中,集中式学习环境的数据分析占主导地位,达41.6%。该数据主要来自学生信息系统、学习管理系统、网络课程等正式学习环境。例如, Moodle 中积累的大量学习资料访问和上传行为,及写作、考试等学习活动日志。分布式学习环境的数据分析从2012年开始猛增,逐渐成为新的热点,占27.5%。该数据包括来自各社交网络软件、个人学习环境等非正式学习环境的数据。这些数据往往表现为不同存储格式,分布于多个媒体和网站,例如,利用电子邮件、短信和社交网络开展的学习讨论和交互行为。随着可穿戴传感器的推广,移动数据、生

物特征数据和情绪数据等身体活动数据开始引起研究者的关注(San Pedro et al., 2013)。学习者学习活动过程中的实时物理交互活动数据或将成为扩展学习分析渠道和深化学习分析内涵的重要途径,这也为研究者近年开展学习情绪建模提供了数据基础,例如问卷调查数据、年度数据报表等其他来源数据一直是学习分析的研究对象。

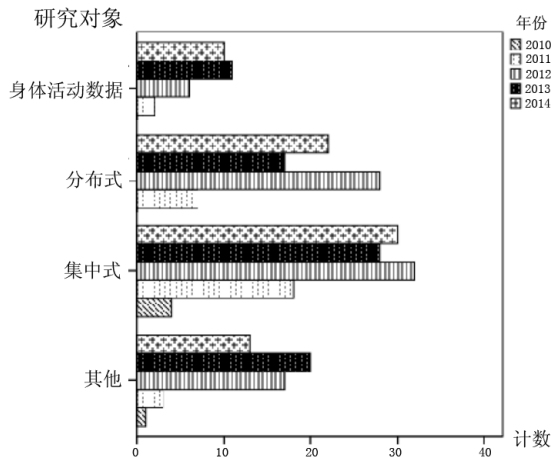


图 4 按研究对象聚类的学习分析文献年度分布

六、技术实现

学习分析作为一个新兴研究领域,学术界倾向将其作为各种教育数据分析和检测方法的综合,并不存在专用的配套分析技术和方法。例如,伊莱亚斯(Elias, 2011)指出学习分析的技术方法是商业智能、网络分析、数据挖掘、学术分析、社会网络分析、话语分析法和内容分析法等在教育领域的继承和渗透。然而,其观点存在一定偏颇,例如商业智能是一个应用领域而并非技术方法,可以运用数据挖掘、社会网络分析等方法实现其商业目的。祝智庭等(2013)指出,除上述方法外,还包括运筹学、行动分析等技术手段。本文通过文献梳理,总结使用较多的分析方法,分为统计分析、信息可视化、数据挖掘、社会网络分析、话语分析和网站分析六大类,(见表二)。

基于分析方法聚类的学习分析文献年度分布如图 5 所示。其中,统计分析和数据挖掘是被广泛使用的研究方法,分别占 37.9% 和 19.3%。而话语分析使用频率最低,仅为 3.7%。2011 年至 2013 年间,教育数据挖掘的推广以及商业智能的成熟应用,使得研究者开始将数据挖掘技术大量应用到学习分

表二 分析方法

分析方法	常用技术	常用目的	案例
统计分析	均值比较、方差分析、相关分析、奇偶分析、因子分析等	现有学习管理系统或开放学习环境大部分内置了教育数据的统计分析功能,以表格或图形等多种形式呈现学习者与学习系统之间的交互数据,帮助教师和学习者积极开展教学反思。	佛罗里达州立大学通过统计维基课程中学习讨论的交互频次、交互范围、交互类型等信息构建相应的因果图,分析如何通过改变学习行为改善学习输出(Jeong et al., 2012)。
信息可视化	概念图、散点图、雷达图、交互图、帕累托图、控制图等	信息可视化将教育相关数据转换成图形,为用户呈现不同的性能指标,旨在借助图形化手段,清晰有效地传达与沟通信息内涵,帮助用户理解和分析教育数据。	比利时天主鲁汶大学开发面向目标的学习活动跟踪仪表盘,用于实时跟踪记录并可视化呈现学生在文字处理软件、编程环境和网络浏览器中学习任务的执行情况(Santos, 2012)。
数据挖掘	基于分类器或回归技术的模型预测、基于聚类的结构发现、基于关联规则挖掘、序列模式挖掘的关联挖掘等	数据挖掘并不需要像统计分析那样预设模型,通过对教育环境中各类数据的探索,可以了解学习者的个人特征、学习进程和学习环境等。	模型预测通常用于预测学习者成绩或高校选择等情况(胡艺龄等, 2014)。结构发现通常用于对学校、学习者或学习行为进行分组(Shanabrook et al., 2010)。关联挖掘通常用于抽取学习者请求帮助的行为模式,研究何种行为模式有利于学习提高(Kinnebrew et al., 2012)。
社会网络分析	中心度计算、隶属于群分析、位置和角色分析等	社会网络分析以独立行动者作为节点,行动者之间的互动或关系作为边,测量学习网络的结构、特征等属性。	澳大利亚昆士兰大学研发的 SNAPP 利用学习者在学习管理系统中发帖互动行为构建社会网络图,识别核心和边缘参与者,抽取学习讨论交互模式,分析学习者自包含小团体或派系(Bakharia et al., 2011)。
话语分析	结构分析、文化分析、认知分析、批判性话语分析等	话语分析分解学习者讨论言论,并运用合适的认知编码框架进行标注,然后开展多角度研究,便于回答学习者如何建构知识等问题。	澳大利亚麦考瑞大学使用话语分析研究在线协作学习过程中以学习者为中心的教学设计的效果(Bower, 2010)。
网站分析	流量引入分析和优化、受访页面分析、人口页分析、退出页分析、访客分析、转化分析等	网站分析通过分析在线系统的访问数据,发现该网站被访问的规律和特征,将其与网络教学资源 and 教学策略的适用性评估相结合,从而发现在线学习系统可能存在的问题和机遇。	美国德克萨斯大学运用 SimilarWeb 分析科学仿真网站 PhET 近 13 个月以来的访问量,以预测用户偏好、用户层次和网站未来发展趋势(Zhang, 2014)。

析领域。而 2013 年至 2014 年间,Google Web Analysis 的兴起,使得网站分析方法逐渐成为在线学习效果分析的利器。由此可见,统计分析和数据挖掘依然占据分析方法的主导地位,信息可视化、网站分析和社会网络分析等多样化的分析方法也逐渐被学者所采用。

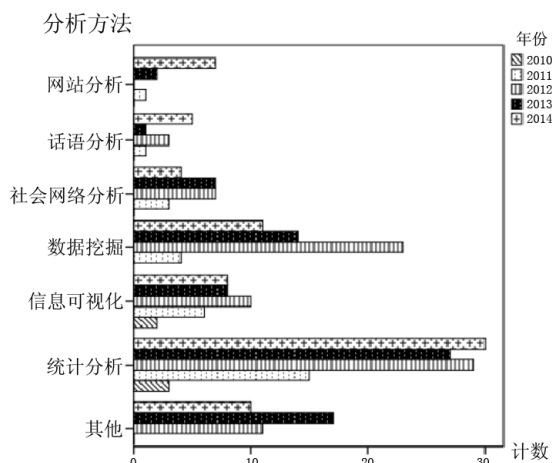


图5 按分析方法聚类的学习分析文献年度分布

七、结论和挑战

MOOCs的推广、教育大数据的出现以及相关教育政策的重视,促使学习分析这一新兴研究领域蓬勃发展。越来越多教育机构和研究人员认识到学习分析在制定教育决策、调整教学活动和改善学习环境方面的潜能。本文回顾近五年来的研究成果和相关文献,梳理出利益相关者(who)为实现特定的研究目标(why),从网络环境中捕获哪些数据(what),并应用何种技术方法(how)。研究表明,学生、教师和研究人員所关注的微观层面是目前学习分析的主要研究团体。来自学生信息系统、在线学习系统的集中式学习环境数据依然是学习分析的主要分析对象,然而分布式学习环境中的数据逐渐成为新的分析热点。研究者关注的焦点从传统的学习活动跟踪、教学资源 and 教学策略优化逐渐转移到学习情绪建模、在线学习影响因素分析等更为多样化的研究主题。统计分析和数据挖掘是主导分析方法,但信息可视化、网站分析和社会网络分析逐渐被研究者接受。

学习分析研究成果为促进我国教育信息化发展带来如下启示:一方面,学习分析视角应多立足于宏观和中观层面。目前我国研究者大多基于微观视角,探索在线学习过程中隐含的学习规律,从而显得研究主题相对狭隘。研究成果大多基于集中式学习环境中存储的学习数据,围绕教学资源 and 教学策略优化、学习行为特征抽取等展开。《教育信息化十年发展规划(2011-2020年)》多次提及利用教育信息化促进教育决策的制定,我国研究者应充分利用

教育信息化产生的海量数据,实现教育政策的系统化、教育制度的规范化和教育环境的智能化。另一方面,研究者应多关注基础教育。现有成果的分析样本大多取自高等教育阶段的混合学习或远程教学。研究团队大多集中在北京师范大学、华东师范大学等高校。然而,我国“三通两平台”的推广以及电子书包等的应用,为学习分析提供了大量基础教育数据。研究者应结合基础教育领域的具体目标,帮助一线教师把学习分析融入于实际教学中,从而提高教育质量。

学习分析虽然具有开阔的研究前景,但目前仍处于发展初期,相关领域还不够成熟。学习分析研究仍面临如下挑战:

1)教育数据预处理问题。教育数据呈增量惊人、分布式存储、来源多样、异质性、碎片化等特点,对研究人员提出了诸多挑战。一方面,如何从学习者与学习内容、学习者之间、学习者和软件系统之间的交互数据中选取合适的开展分析是学习分析难点之一。另一方面,在数据分析过程中,至少70%到85%的工作量花费在数据清理、格式化和数据对齐方面,影响学习分析的使用和推广。

2)数据访问权限。利益相关者应该拥有何种访问权限,才能避免学习分析手段的滥用,从而构建需求驱动的学习型社会,成为学界关注的另一个重点。而解决该问题的主要思路是让利益相关者明确数据所有权和管理权。

3)学习分析适用性。由于学习是个创造性过程,需要新的观念、思路和方法,而学习分析主要解释和识别已经存在的客观规律。又由于学校并非是一成不变的机械化知识加工场,因此学习分析的适用范围具有一定局限性。虽然学习分析可以为研究者提供教学数据的定量分析结果,但过分依赖其释意是不合理的,因此有必要结合学习者的自主性特征综合分析。

[参考文献]

- [1] Ali, L., Asadi, M., Gašević, D., & Hatala, M. (2013) Factors influencing beliefs for adoption of a learning analytics tool: An empirical study[J]. *Computers & Education*, 62: 130-148.
- [2] Baker, R. S. J. D. (2011). Data mining for education[A]. B. McGaw, P. Peterson, & E. Baker. *International Encyclopedia of Ed-*

ucation[C]. Oxford, UK: Elsevier.

[3] Bakharia, A., & Dawson, S. (2011). SNAPP: A bird's-eye view of temporal participant interaction[A]. Proceedings of the 1st international conference on learning analytics and knowledge. ACM[C]. 168-173.

[4] Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics[DB/OL]. Washington, DC: U. S. Department of Education. Retrieved from <http://www.ed.gov/edblogs/technology/files/2012/03/edm-labrief.pdf>.

[5] Bower, M., & Hedberg, J. G. (2010). A quantitative multimodal discourse analysis of teaching and learning in a web-conferencing environment - The efficacy of student-centred learning designs[J]. Computers & Education, 54(2):462-478.

[6] Brooks, C., Erickson, G., Greer, J., & Gutwin, C. (2014). Modelling and quantifying the behaviours of students in lecture capture environments[J]. Computers & Education, 75:282-292.

[7] Brown, M. (2012) Learning analytics: Moving from concept to practice[DB/OL]. <http://www.educause.edu/library/resources/learning-analytics-moving-concept-practice>.

[8] 曹晓明,朱勇(2014). 学习分析视角下的个性化学习平台研究[J]. 开放教育研究, (5):67-74.

[9] Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thus, H. (2012). A reference model for learning analytics[J]. International Journal of Technology Enhanced Learning, 4(5): 318-331.

[10] Crow, M. M. (2012). No more excuses[DB/OL]. <http://www.educause.edu/ero/article/no-more-excuses-michael-m-crow-analytics>.

[11] 邓文俊,曹阳波,刘英群,程建钢(2014). 基于教务管理系统的数据分析工具筐研究与设计[J]. 中国电化教育, (5):138-143.

[12] EDUCAUSE(2012). 7 things you should know about analytics[EB/OL]. <http://www.educause.edu/ir/library/pdf/ELI7059.pdf>.

[13] Elias, T. (2011). Learning analytics: Definitions, processes and potential [EB/OL]. <http://learninganalytics.net>.

[14] Ferguson, R., Wei, Z., He, Y., & Buckingham Shum, S. (2013). An evaluation of learning analytics to identify exploratory dialogue in online discussions[A]. Proceedings of the Third International Conference on Learning Analytics and Knowledge, ACM[C]. 85-93.

[15] Greller, W., & Drachler, H. (2012). Translating learning into numbers: A generic framework for learning analytics[J]. Educational Technology & Society, 15(3):42-57.

[16] 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.

[17] Hecking, T., Manske, S., Bollen, L., & Hoppe H. U. (2014). A flexible and extendable learning analytics infrastructure[M]. Advances in Web-Based Learning - ICWL 2014. Springer International Publishing: 123-132.

[18] 胡艺龄,顾小清,赵春(2014). 在线学习行为分析建模及

挖掘[J]. 开放教育研究, 20(2):102-110.

[19] Jeong, A., & Lee, W. J. (2012). Developing causal understanding with causal Maps: The impact of total links, temporal flow, and lateral position of outcome nodes[J]. Educational Technology Research and Development, 60(2): 325-340.

[20] Kinnebrew, J., & Biswas, G. (2012). Identifying learning behaviors by contextualizing differential sequence mining with action features and performance evolution[A]. Proceedings of the 5th International Conference on Educational Data Mining[C]. 57-64.

[21] Leeuwen, A. V., Janssen, J., Erkens, G., & Brekelmans, M. (2014). Supporting teachers in guiding collaborating students: Effects of learning analytics in CSCL[J]. Computers & Education, 79: 28-39.

[22] 李青,王涛(2012). 学习分析技术研究与应用现状述评[J]. 中国电化教育, (8):129-133.

[23] 李艳燕,马韶茜,黄荣怀(2012). 学习分析技术:服务学习过程设计和优化[J]. 开放教育研究, 18(5):18-24.

[24] Lotsari, E., Verykios, V. S., Panagiotakopoulos, C., & Kalles, D. (2014). A learning analytics methodology for student profiling [M]. Artificial Intelligence: Methods and Applications. Springer International Publishing:300-312.

[25] May, T. A. (2011). Analytics, university 3.0, and the future of information technology. EDUCAUSE Review Online [DB/OL]. <http://net.educause.edu/ir/library/pdf/ERM1159.pdf>.

[26] McIntosh, N. E. (1979). Barriers to implementing research in higher education[J]. Studies in Higher Education, 1(4):77-86.

[27] Rebecca, F. (2012a). The state of learning analytics in 2012: A review and future challenges [DB/OL]. Technical Report KMI-12-01. Milton Keynes: Knowledge Media Institute, The Open University. <http://kmi.open.ac.uk/publications/techreport/kmi-12-01>.

[28] Rebecca, F. (2012b). Learning analytics: Drivers, developments and challenges[J]. International Journal of Technology Enhanced Learning, 4(5/6):304-317.

[29] Ruipérez-Valiente, J. A., Muñoz-Merino, P. J., Leony, D., & Kloos, D. (2014). ALAS-KA: A learning analytics extension for better understanding the learning process in the Khan Academy platform [J]. Computers in Human Behavior,

[30] San Pedro, M., Baker, R., Bowner, A., & Heffernan, N. (2013). Predicting college enrollment from student interaction with an intelligent tutoring system in middle school[A]. Proceedings of the 6th International Conference on Educational Data Mining: 177-184.

[31] Santos, J., Govaerts, S., Verbert, K., & Duval, E. (2012). Goal oriented visualizations of activity tracking: A case study with engineering students[M]. Buckingham Shum, Gasevic and Ferguson:143-152.

[32] Shanabrook, D. H., Cooper, D. G., Woolf, B. P., & Arroyo, I. (2010). Identifying high-level student behavior using sequence-based motif discovery[A]. Proceedings of the 3rd International Confer-

ence on Educational Data Mining[C]. 191-200.

[33] Shea, P., & Bidjerano, T. (2014). Does online learning impede degree completion? A national study of community college students [J]. *Computers & Education*, 75: 103-111.

[34] 沈欣忆, 李爽, 丹尼尔·希基, 李营 (2014). 如何提升 MOOCs 的学生参与度与学习效果——来自 BOOC 的经验[J]. *开放教育研究*, 20(3): 63-70.

[35] 舒忠梅, 徐晓东 (2014). 学习分析视域下的大学生满意度教育数据挖掘及分析[J]. *电化教育研究*, (5), 39-44.

[36] Shum, S. B. (2012). Learning analytics. UNESCO policy brief[DB/OL]. Retrieved from <http://iite.unesco.org/pics/publications/en/files/3214711.pdf>.

[37] Smith, V. C., Lange, A., & Huston, D. R. (2012) Predictive modeling to forecast student outcomes and drive effective interventions in online community college courses[J]. *Journal of Asynchronous Learning Networks*, 16(3): 51-61.

[38] Tinto, V. (1997). *Colleges as communities: Taking research*

on student persistence seriously[J]. *The Review of Higher Education*, 2(21): 167-177.

[39] Vatrapu, R., Reimann, P., Bull, S., et al. (2013). An eye-tracking study of notational, informational, and emotional aspects of learning analytics representations[A]. *Proceedings of the Third International Conference on Learning Analytics and Knowledge*[C]. 125-134.

[40] 吴忭, 顾小清 (2014). 图示化学习工具促进知能发展的学习行为模式分析[J]. *开放教育研究*, (5): 59-66.

[41] 武法提, 牟智佳 (2014). 电子书包中基于大数据的学生个性化分析模型构建与实现路径[J]. *中国电化教育*, (3): 63-69.

[42] Zhang, M. (2014). Who are interested in online science simulations? Tracking a trend of digital divide in Internet use[J]. *Computers & Education*, 76: 205-214.

[43] 祝智庭, 沈德梅 (2013). 学习分析学: 智慧教育的科学力量[J]. *电化教育研究*, (5): 5-19.

(编辑: 魏志慧)

Learning Analytics: From the Origin to Practice

WU Qing & LUO Ruguo

(College of Educational Science, Wuhan University, Wuhan 430072, China)

Abstract: *Learning analytics is a big data application in the field of education, triggering the development of the third wave in educational technology, and receiving wide attention from international and domestic academics. The article reviews the origin of learning analytics, and then from four dimensions of whom, why, what, how, recalling the past five years relevant research results, and puts forward future development challenges that may be encountered. Specifically, based on macro, meso and micro perspective, this research facilitates related personnel to make education decisions, optimize the educational management process and improve the learning process.*

Research topics cover nine aspects including: learners' knowledge modeling, learners' emotion modeling, learning behavior feature extraction, learning activities tracking, the learner modeling, obtaining degree analytics, the teaching resources and the teaching strategy optimization, adaptive learning system and personalized learning, and online learning influence factors analysis. The data mainly comes from the centralized learning environment, distributed learning environment, and physical activity data. The common analysis methods include statistical analysis, information visualization, data mining, social network analysis, discourse analysis, and web analytics. The challenges include: the difficulty of data preprocessing, the uncertainty of data accessing rights, and the limited application scope. Although the study is still in the initial stages of development, it has become an important part of educational informatization as it provides scientific reference for decision-making at all levels of the education system.

Key words: *learning analytics; educational data mining; social network analysis; information visualization*