

学习分析工具比较研究

孟玲玲 顾小清 李 泽

(华东师范大学 教育信息技术学系, 上海 200062)

[摘要] 近年来,随着智慧学习环境的飞速发展,海量、丰富、多样、异构的学习数据急剧积累,如何有效地利用这些数据引起了学界的广泛关注。学习分析技术应运而生,并成为研究热点。学习分析工具在学习分析过程中具有举足轻重的作用,好的研究工具可以使研究过程事半功倍。本文从多个角度对学习分析工具进行了分类,并从使用环境、数据支持格式、是否可视化等维度详细比较,讨论了24种学习分析工具的特点,旨在为学习分析提供相应的技术支持。

[关键词] 学习分析;工具;分类;比较

[中图分类号] G434

[文献标识码] A

[文章编号] 1007-2179(2014)04-0066-10

当前,随着智慧学习环境的飞速发展,海量、丰富、多样、异构的学习数据急剧积累,如何有效地利用这些数据,使之转变成有价值的信息、知识,从而监控学习过程、改进学习实践、优化学习效果,辅助教育决策,并提供更好的教学、学习服务,引起了学界的广泛关注。学习分析技术应运而生。学习分析技术是“测量、收集、分析和报告有关学生及其学习环境的数据,用以理解和优化学习及其产生的环境的技术”(Knewton, 2011)。近年来,学习分析已经成为教育领域的研究热点。美国新媒体联盟(NMC)与美国高校教育信息化协会(EDUCAUSE)合作的“地平线项目”在2010年和2011年年度报告中均预测学习分析技术将在未来四到五年内成为主流(The New Media Consortium, 2011;李青, 2012)。

学习分析工具目前已有很多,且仍在快速发展之中。例如, Nvivo、Atlas.ti 可以对文本内容、多媒体内容等进行标注、编码,还可以进行简短评论; Gephi、JUNG、Guess 可以用来进行社会网络分析; WMatrix、CATPAC 可以进行学习内容分析; SPSS 可以进行数据统计分析等(魏顺平, 2013)。当前,已有一些研究者对学习分析工具进行初步研究,但总体来说这些研究较为零散,基本聚焦在社会网络分

析工具的比较上(孙洪涛, 2013; Huisman et al., 2005;王陆, 2009; Hansen et al., 2009; Smith et al., 2009; Kirschner, 2012)。缺乏对学习分析工具的总体分析研究。本文综述了当前学习分析领域中可用的24种工具,从多个角度对学习分析工具进行分类,并从使用环境、数据支持格式、是否可视化等十个维度详细比较,讨论了24种学习分析工具的特点,旨在为学习分析提供相应的技术支持。

一、学习分析工具分类

学习分析工具依据不同的规则,可进行不同的分类,如根据专门程度,可分为通用工具和专用工具;根据分析结果的描述方法,可分为可视化的和非可视化的;根据工具的可扩展性,可分为提供开发接口的和不提供开发接口的。本文根据各学习工具所侧重的分析对象与类型,将其分为:学习网络分析工具、学习内容分析工具、学习能力分析工具、学习行为分析工具及其他综合分析工具。

(一)学习网络分析工具

社交网络近年发展迅速,成为平台中学习交互的重要工具,从而产生了大量数据。学习网络分析工具就是以这些数据为分析对象,重点分析两类

[收稿日期] 2014-06-30

[修回日期] 2014-07-04

[作者简介] 孟玲玲,博士,华东师范大学教育信息技术学系副教授,硕士生导师,主要研究方向为学习分析、知识工程、计算机教育应用;顾小清,博士,华东师范大学教育信息技术学系教授,博士生导师,主要研究方向为教学训练系统设计、数字化学习环境及用户行为(xqgu@ses.ecnu.edu.cn);李泽,华东师范大学教育信息技术学系在读硕士研究生。

问题:

第一,以学习个体为研究对象,关注某个体在学习平台中的活跃程度如何?与哪些学习同伴交互信息?从哪些同伴处得到了哪些启示?对哪些内容存在认知困难?哪些因素影响了学习者的知识建构?等等。

第二,以学习网络作为研究对象,分析网络中各个体之间的关系、角色、网络形成的过程与特点、网络中学习信息的分布以及学习进展等,还可以分析学习者如何在网络学习中建立并维持关系从而为自己的学习提供支持(Haythornthwaite et al., 2010)。

(二) 学习内容分析工具

它主要分析两类问题:第一,以学习过程中教师-学生、学生-学生交流所产生的内容作为研究对象,如面对面的对话、网络课程与会议中产生的文本、网上同步、异步交流等。关注学习对话,分析学习交流中话语的文本含义,关注学生知识建构过程,使研究者对学习发生的过程有更清晰的认识(Ferguson, 2011),既可以作定性分析,也可以作定量分析(O'Halloran, 2011)。第二,以学生学习内容本身为研究对象,如文本内容分析、多媒体内容分析等,通过对相似内容的分析、标注,发现特征相似的文本内容,从而提供更符合用户需求的学习内容,实现个性化推荐(Little et al., 2011; Clow et al., 2011; Jovanovic et al., 2008)。

(三) 学习能力分析工具

它以学习者的学习能力、学习水平为分析对象,通过量表、博客或考试等形式对学生的学习能力、学习水平等维度(如探究意识、创造力、心理弹性等)进行测试,评估学习进展,并将结果以可视化形式(如蛛网、漩涡图、折线图)反馈给学习者及教师。学习者根据图示了解自身的学习状况并反思,教师依据图示对学习者的学习提供针对性的建议和策略,帮助他提高学习能力。

(四) 学习行为分析工具

它以学习者与系统的人机交互数据为研究对象。学习者登录系统的时间、访问时间、完成作业等都被系统自动捕获并记录。研究者通过对这些学习过程数据的分析,以及监测学习者访问系统的行为,了解学习者的学习轨迹、学习特征,挖掘学习者的行为模式,从而了解学习过程的发生机制,分析学习者

的学习需求、学习进展等,预测学习者未来可能的行为,从而优化教学,为学习者提供个性化的服务(DeWever, et al., 2006; Chen, 2010; 顾小清, 2010; The New Media Consortium, 2011)。

(五) 其他综合性分析工具

综合性分析工具通常功能强大,本身具有一定的适用性。它一般从数据挖掘、统计分析、联机分析处理等领域借鉴而来,可以对数据进行预处理、分类、回归、聚合、描述性统计、因子分析、聚类分析等,还可产生报表、可视化显示结果,能与其它数据处理软件进行数据迁移,并支持数据决策。

二、学习网络分析工具比较

常见的学习网络分析工具有:SNAPP、Gephi、NetMiner、UCINET、Pajek、GUESS、JUNG、NodeXL、Cohere。九种工具的比较见表一。

(一) SNAPP

SNAPP 可从学习管理系统的论坛中提取数据。例如,根据参与讨论者、参与讨论人数、发帖数、回帖数等,进行自我中心社会网络分析、小团体分析、中心性分析等社会网络分析,并实时以社会网络关系图呈现可视化结果。此外,它关注学习者发帖互动的频率和反应时间,以网络图的形式展现学习者的讨论动态,方便教师了解所有学习者参与讨论的情况,及时对不活跃不积极学习者或小群体进行干预,针对问题调整教学策略。它可与 Blackboard、WebCT、Moodle、Desire2Learn 等在线学习平台实现无缝对接。

(二) Gephi

Gephi 定位于复杂网络分析,主要用于分析各种网络和复杂系统,可分析多达 5 万个节点、100 万条边界的复杂网络。它提供 10 余种不同的布局算法,可实现实时动态分析、时段动态分析、无标度网络分析、分层图示等交互可视化与数据探测。它还可用作探索性数据分析、链接分析、社交网络分析、生物网络分析、新媒体分析等,具有较强的多媒体展示功能,提供 API 接口,便于功能扩展。

(三) NetMiner

NetMiner 可分析多达 100 万个节点 1 亿条边界的复杂网络,可进行可视化、交互式探索数据分析,发掘数据隐藏的潜在联系与结构,具有影响力、结构

表一 学习网络分析工具比较

编号	工具名称	发起者	使用环境	专门程度	使用对象	分析对象	数据支持格式	输出格式	可视化	商业软件/开源	优势与不足
1	SNAPP	澳大利亚 Shane Dawson 团队	支持 Java 脚本运行的/IE/Firefox/Safari 浏览器	专用	Blackboard /WebCT/Moodle/ Desire 2Learn 学习平台	LMS 的论坛中学习者的社会网络关系	直接从论坛提取数据, 通常是文本数据。	JPG, PNG, VNA, GEFX	是	免费软件	不需安装, 可从 LMS 直接导入论坛数据, 使用方便; 需支持 JavaScript 浏览器。
2	Gephi	法国贡比涅技术大学学生研发	Windows/Mac/Linux 操作系统下 Java 环境	通用	具备数据分析和图论基础的专业研究人员	社交网络分析、无标度网络分析、动态网络分析、生物网络调控分析及网络流量图分析等。	GEXF, GDF, GML, GraphML, NET, DOT, CSV, DL, TPL, VNA, Spreadsheet	GEXF, GDF, SVG, PNG	是	开源软件, 不收费	分析内容多样, 探索分析功能完备, 可分析大型复杂网络, 具有可扩展功能接口。有 Firefox 插件, 可在线提取数据分析。
3	Net-Miner	韩国	Windows	通用	具备社会网络分析和数据挖掘基础或编程能力研究人员	社会网络分析、数据挖掘	Excel, TEXT, CSV, PY, NMF, Pajek, DL, StOCNET, GML	Excel DL, CSV, TEXT, Pajek, NIF, StOCNET, BMP, JPG, SVG	是	商业软件, 免费试用 28 天	可 3D 网络地图形式展示结果, 支持 Python 编程分析。
4	UCI-NET	原是加州大学分校的校园数据通讯系统	Windows/Linux/Mac	通用	具有一定网络分析基础的社会网络分析初学者	节点数较少的社会网络分析	ASCII, Excel, DL, KrackPlot, Pajek, Ucinet3.0, txt, Negopy, VNA	DL, KrackPlot, Ucinet3.0 原始数据文件	是	商业软件, 免费试用 90 天	适合新手的社交性网络分析工具。
5	Pajek	斯洛文尼亚的 Vladimir Batagelj 和 Andrej Mrvar	Windows/Linux/Mac	通用	具有网络分析知识基础的研究人员	节点较多的大型复杂网络	NET, MAT, TIM, DAT, GED, MGR, JS, MAC, MOL, CLU, VEC, PER, CLS, HIE, VID, PAJ, LOG, MCR	SVG, X3D, EPS, JPEG, BMP, VRML 等制表符分割的文本文件	是	商业软件, 非商业用途可免费使用	可处理超大型复杂网络关系, 配有 R 软件接口可扩展统计功能。
6	GUESS	密歇根大学 Eytan A-dar 开发	Java 环境	通用	具有 Python 编程能力及相关专业知识的网络分析研究人员	社会网络、知识网络、电脑网络、生物网络等分析。	GD, GraphML, Pajek 文件	PDF, GIF, JPG, PNG, PS, EPS, SVG, JAVA, CGM, EMF	是	开源软件, 不收费	软件专业化程度、使用门槛较高, 适合高级分析人员。
7	JUNG	加州大学三位计算机专业博士生开发	Java 环境	通用	具有图论、社会关系基础及 Java 编程能力的专业研究者	社会网络分析、Web 流量分析、图论的理论推导及关系类型数据的挖掘等。	CSV, 也可用户自行编码调用其他格式	CSV, 图片格式及一般的 K 部图, 超图, 用户自定义格式。	是	开源软件, 不收费	可对数据实现动态分析呈现, 使用门槛较高, 需掌握算法及 Java 编程技能。
8	No-deXL	Marc Smith 在微软工作时由其团队开发	以模板形式嵌套在 Excel 中使用	通用	具备基本社会网络分析知识的研究者	社会网络关系	Email, CSV, TXT, XLS, No-deXL, DL, NET, GraphML 及矩阵格式文件, 还可从 Twitter 等社交网站上直接提取关系数据。	CSV, TXT, NET, Excel, GraphML, DL	是	开源软件, 不收费	使用方法简单, 不需编程, 可直接从社交网站中提取关系, 但这些社交网站国内均不可用。
9	Cohere	英国开放大学项目	可运行 Java 程序的浏览器	通用	任何注册用户	用户与用户之间的分享联系	TXT, URL, XML, RDF	社会网络关系图、信息发布排名榜、个人信息发布条等图片格式。	是	免费软件, 不收费	有火狐浏览器插件; 但目前用户数很少, 处于测试阶段。

洞等关系和邻近结构分析、子图布局、中心性分析、派系分析、核分析、社团发现等基本的复杂网络分析功能。此外,它还包含为数众多的基于过程的分析程序,如聚类分析、多维量表、矩阵分解、对应分析、结构对分析等,支持一些标准的统计:描述性统计、ANOVA、相关和回归。NetMiner 4 还提供基于 Python 的脚本编程环境,允许用户通过编辑代码实现复杂分析,也允许没有编程基础的用户,通过各类 GUI 工具实现复杂分析。社会网络关系可用立体 3D 可视化地图表示,用户可随意改变观察视角,更方便观察。在展示、调整社会网络布局时,用户可录屏记录,方便回放和结果展示。

(四) UCINET

UCINET 是目前被使用最多的网络分析软件,最大可处理 32 767 个结点的网络数据,适用于子群分析、中心性分析、个体网络分析和角色分析等,还包含许多基于过程的分析,如聚类分析、多维标度、二模标度(奇异值分解、因子分析和对应分析)、角色和地位分析(结构、角色和正则对等性)和拟合中心-边缘模型以及中位数、标准偏差、回归分析、方差分析、自相关、QAP 矩阵相关、回归分析、t 检验。此外,该工具有强大的矩阵分析程序,如矩阵代数 and 多元统计等(Huismanm 2005)。

(五) Pajek

Pajek 可分析多达 1 亿个节点的大型复杂网络关系,支持对有向、无向、混合网络,多关系网络,二值网络关系的分析,可同时处理多个网络。此外,Pajek 还可处理动态网络,即随时间动态改变的时间事件网络。动态网络分析数据文件可以包含指示行动者在某一观察时刻的网络位置的时间标志,因而可以生成一系列交叉网络,可以对这些网络进行分析并考察网络演化。它具有强大的画图功能,可将分析结果可视化,还可嵌套宏命令进行统计分析,配有 R 软件接口,能将 Pajek 的数据探索功能与 R 强大的统计功能结合起来。PajekXXL 是 Pajek 的加强版,在更高配置环境下可更快速处理网络关系(胡长爱 2010)。

(六) GUESS

GUESS 是基于 Gython 语言的数据探索及可视化分析工具,可直观地对图形结构进行操作,能将输入的文本与可视化对象进行关联分析。该工具适用

于社会网络、知识网络、电脑网络、生物网络等分析,支持动态和时间敏感数据,允许操纵图片状态,允许将图形和网络无限放大,提供可视化前端,支持静态图像和动态的视频输出(闫冠韞 2013)。

(七) JUNG

JUNG 是基于 Java 图形处理及可视化 API 和库,为图形或网络的建模、分析、数据可视化提供了通用、可扩展语言,具有可扩展性,含有多种算法、过滤机制等。用户可通过 Java 自主编程实现所需功能,动态分析读取数据,支持各种实体及其关系的表示,如定向图、无向图、多模态图和超图等。它提供注释图、实体和元数据关系的机制,方便针对复杂数据集的分析工具创建,包括图论、数据挖掘、社会网络分析等一系列算法,如聚类、分解、优化、随机图形生成、统计分析以及网络距离计算、流量和重要的测量等。该工具可建模数据、分析及数据的可视化呈现,通常用于与社会网络分析、信息可视化、知识发现及数据挖掘相关的数学及算法图形处理。

(八) NodeXL

NodeXL 是微软研究院产品,建立在 Excel 基础上的交互式网络可视化与分析工具。该工具利用 Excel 作为数据展示和分析平台,能将存储在 Excel 中的边、顶点数据以网络图的形式呈现并分析,支持 Excel 2007、Excel 2010 和 Excel 2013 版本,可定制图像外观、无损缩放、移动图像,动态过滤顶点和边,提供多种布局方式,查找群和相关边,支持直接从多种社交网站或邮件中导入网络关系进行分析。

(九) Cohere

Cohere^①是一款在线分享工具,可将网上交流的学习内容进行结构化,也可用于分析学习者相互间的分享联系。用户可创建自己的想法、问题、见解、假设、预测、软件、理论等讯息,可附加提供相关网址、自定义标签。该工具会根据内容的相似程度,自动连接相匹配的用户,帮助用户找到与自己有共同想法的人,并与其建立联系。该在线工具能将同一主题下的用户及相互关系以社会网络关系图可视化展示并进行分析,还能统计出每个用户与他人“连接”的次数,并按次数由多到少排名(Buckingham, 2014)。

三、学习内容分析工具比较

常见的学习内容分析工具有:WMatrix、CAT-

PAC、LIWC 等。七种工具的比较见表二。

表二 学习内容分析工具比较

编号	工具名称	发起者	使用环境	专门程度	使用对象	分析对象	数据支持格式	输出格式	可视化	商业软件/开源	优势与不足
1	Wmatrix	英国工程和自然科学委员会的项目	Chrome/Firefox/IE 浏览器	通用	文本分析人员	英文文本	纯文本格式 (ASCII 编码)	文本格式 (ASCII 编码)	否	商业软件, 免费试用 1 个月	仅支持少于 10 万单词的英文文本的分析, 不能脱机使用。
2	CAT-PAC	布法罗大学的社会学家 Joseph K. Woelfel	Windows (32 位)	通用	分析大量的商业、政府、科研人员	文本	文本格式 (ASCII 编码)	TEXT 文件, DEN 文件, LBL 文件, WIN 文件, CRD 文件	否	商业软件, 收费	可对多种语言进行语义及扩展分析, 功能多样; 但自身不具备可视化功能, 需结合 ThoughtView 软件实现数据可视化。
3	LIWC	心理学家 James, Roger 与 Martha 开发	Windows/Mac	通用	文本分析研究人员如心理学家、问卷调查者等	文本格式的书信、问卷、访谈、文献等	文本文件 (ASCII 编码, Word)	制表符分隔的文本文件 (SPSS, Excel 等)	否	商业软件, 收费	可对多语言进行统计分析, 内嵌词典完备; 但在 Mac 上运行的版本可统计短语。
4	Nvivo	Tom Richards 开发, 后由 QSR 公司完善	Windows/Mac	通用	质性研究人员	论坛内容、问卷、访谈音频、课堂观察视频等	Word, RIF, TXT, PDF/MP3, M4A, WMA, WAV/MPG, MPEG, MPE, MP4, AVI, WMV, MOV, QT, 3GP, MIS, GIF, MZIS, BMP, JPG, JPEG, PNG, TIF, TIFF 制表符分割的文本和 Excel 格式的数据集, SQL, Access 的数据库文件, 可直接导入 Twitter, Facebook, 社交网站的讨论数据, 可导入 SurveyMonkey 调查问卷数据。	Word RIF, TXT, PDF, MP3, M4A, WMA, WAV, MPG, MPE, MP4, AVI, WMV, MOV, QT, 3GP, MIS, MZIS, JPG, HIM, HIML, Excel 和制表符分割的数据集	是	商业软件, 收费, 有 30 天免费试用期	支持多种数据格式与研究方法, 支持协作研究。可导入 EverNote、OneNote 和 EndNote 笔记数据作为源数据。可对导入的 Word 等文本文件进行编辑, 导出为 PDF 文件时中文和日文不能有效识别, 需先导出保存为 Word 再转换为 PDF。导出图片只能保存为 JPG 格式。
5	Atlas.ti	柏林科技大学心理系开发	Windows/Android/IOS	通用	质性研究人员	论坛内容、问卷、访谈音频、课堂观察视频	Word RIF, TXT, PDF, AVI, WAV, MOV, QT, AU, SND, MPG, M4A, MV, MP3, MP4, BMP, JPEG, TIFF, Excel, SQL 数据库文件。	RTF, TXT, PDF, BMP, EMF, XML, HTML, Excel, SPSS, WORD, CSV	是	商业软件, 收费	可直接将谷歌地图图片数据导入工具进行分析。在 iPad 与 Android 手机上已有 app; 只能手动或半自动编码, 不能自动编码。
6	Wordle	Jonathan Feinberg	支持 Java 脚本运行的浏览器	通用	仅需简易统计分析的用户	在线上上传的文本文件	TXT, URL	PDF, SVG	是	免费软件	可实现简易词频统计功能并制作美观的词云。适合文本量较小的统计。不支持中文及日文统计。
7	LOCO - Analyt	澳大利亚	浏览器	专用	教师、学生	学生学习轨迹, 内容	自动记录用户访问数据	图表格格式	是	开源软件, 不收费	对学生讨论用词的标注只支持中文。

(一) WMatrix

WMatrix 是一款在线文本分析工具, 可对文本

内容进行定量研究。它通过浏览器上传分析文本、语料库, 对文本进行自动标注, 统计词频并按字母或

频数排序, 通过查看相应单词的频率, 了解单词出现次数是否异常, 从而对文本进行语义分析、频次分析, 形成频数云。它还可以比较不同目标文本中分类词汇的频次差异, 最终形成文字云显示统计结果。字体越大表示单词出现的频次越多, 不常用的单词以斜体字呈现。Wmatrix 提供 11 种统计方法用于计算词汇之间的关联强度, 以估测目标文本的词汇搭配习惯(O'Halloran, 2011)。

(二) CATPAC

CATPAC 可识别多种语言, 能读取任何文本, 不需预编码和语言假设。它通过词频统计、自学习式语义分析等概述文本想表达的关键信息, 提供较为精准又客观的文本分析, 以类似人神经中枢处理信息的方式挖掘信息间联系, 可进行自动聚类、编码、多维度标记。软件自身不具备可视化功能, 需通过 ThoughtView 等软件读取数据, 形成概念图。

(三) LIWC

LIWC 支持包括中英文在内的多种语言, 内嵌近 80 本词典供文本分析使用。它可根据用户选定的词典统计目标文本中各维度使用的词汇频次, 如表达积极消极感情维度、因果关系维度。用户可在给定词典中选择需要分析的维度, 也可自定义词典。LIWC 每秒能分析数百个标准的 ASCII 文本或 Word 文本。

(四) Nvivo

Nvivo 是一款质性分析工具, 可方便地收集、整理和分析访谈、小组讨论、问卷调查、音频、视频等内容, 还可协助处理社交媒体和网页内容, 分析无特定结构或半结构化数据, 如会话访谈音频、课堂视频、问卷调查、课堂观察、论坛互动等。它支持团队协作编辑分析项目, 不仅可以导入多种类型数据, 还可以导入 EverNote 或 OneNote 笔记作为源数据, 能对各种源数据创建节点并编码分析, 对编码内容添加备忘、链接等。对分析结果的处理, Nvivo 能建立可视模型、图表, 并将结果网络视图化, 也能将结果导入 Excel、SPSS 等软件进行统计分析, 支持团队协作, 多用户同时处理同一项目等。

(五) Atlas. ti

Atlas. ti 功能与 Nvivo 类似, 擅长对文本和多媒体数据进行专业分析, 支持多种研究方法, 支持多种半结构化、非结构数据分析, 如文本、动态网页链接、

图片、音频、视频以及内嵌对象(如 Excel 表格、PowerPoint 幻灯片)等。它可直接访问数据库系统, 并能嵌入谷歌地图功能, 可直接将谷歌地图图片数据导入工具进行分析, 还可同时对多种材料进行编码分析, 通过网络编辑器显示编码、引文、备注和其它实例之间的复杂关系, 并将结果可视化。目前, 它已有可在平板、安卓手机上运行的 APP。

(六) Wordle

Wordle 是一款在线文字云制作工具。该工具自动统计源文本中词汇出现频次, 并将频次统计结果可视化。出现次数越多, 该文字越大, 且用户可调整字体、布局和配色方案。此外, 用户可以输入文本内容, 还可以输入网页链接, 由 Wordle 自动到相应的链接上, 捕获文本内容。Wordle 网站高级应用中也可实现对源文本中词汇、短语的统计, 用户可将生成的“文字云”以 PDF 或 SVG 的格式保存下来, 也可以截屏将文字云保存。

(七) LOCO - Analyst

LOCO - Analyst 是运用语义 web 技术、基于 LOCO (learning object context ontology) 理念创设的可嵌入到学习平台中的分析工具, 也是一款基于网络学习环境的上下文感知学习分析工具, 旨在为教师提供反馈信息, 帮助教师调整课程内容和结构。该工具可以记录学生的学习轨迹、平台资源使用、学习活动、参与课程、在线学习社区中学生之间互动等, 自动标注学生讨论内容方便教师查看, 剔除无意义内容。教师还可以查看群体和个体讨论中运用最频繁的词汇。部分数据能以图表、网络图的形式可视化呈现(Jovanovic J. et al. 2007)。

四、学习能力分析工具比较

常见学习能力分析工具有: ELLIment、Enquiry Blogger、Socrato。三种工具的比较见表三。

(一) ELLIment

ELLIment 根据“有效终身学习量表”(Effective Life Long Learning Inventory) 将学习者的学习能力分为七维度——改变与学习、探究意识、意义构建、创造力、学习关系、战略意识和心理弹性。学习者在完成量表问卷后, 问卷结果会以蛛网的形式反馈给学习者及教师。学习者根据蛛网图了解自身学习水平, 明确自身学习弱势, 增强提升自主学习意识、自

表三 学习能力分析工具比较

编号	工具名称	发起者	使用环境	专门程度	使用对象	分析对象	数据支持格式	输出格式	可视化	商业软件/开源	优势与不足
1	ELLlment	英国开放大学的学生项目,后由英国布里斯托尔大学完善	支持 Java 脚本运行的浏览器	专用	大学生及教师	学习者的学习反思过程	文本	文本、图表	是	非商业的专用软件	不可记录全部学习轨迹,只能记录学习反思过程。
2	Enquiry Blogger	英国开放大学的知识媒体研究所与布里斯托尔大学的学习与领导力系统中心合作开发	WordPress 博客网站的扩展功能包	专用	学生及教师	学习者学习任务进展过程	文本	文本、图表	是	开源软件,不收费	能较全面地记录学习者的任务完成情况 & 心理动态,帮助教师及时干预。
3	Socrato	OmniSharp 公司	浏览器	专用	培训机构、学校、教师、学生	学习者测试结果、学习者学习优势劣势	教师可以以 Excel, CSV 格式导入学生成绩	文本、图表、PDF 文件	是	商业软件,收费,但 SAT 及 ACT 考试可长期免费使用	专门针对考试的分析工具,可自动生成各学习维度绩效;但不含考试试题,不能进行自主命题考试。

我学习能力、诊断能力。教师可根据蛛网(雷达图)掌握全班学生的学习能力,依据学习者对自身某一维度或某几个维度的提升意愿,向其提供针对性的建议和策略,以帮助学习者自我调控,提高学习能力。此外,ELLlment 将记录教师的所有指导意见及学习者的反思。

(二) Enquiry Blogger

Enquiry Blogger 是款博客插件。学习者根据教师提供的基于真实情境的学习任务定期撰写博客、利用标签标记其在探究学习过程中的进度、发布心情状态。该工具可收集学生数据并以蛛网形式展现学习者学习能力的七个维度水平、以漩涡图展示学习者任务的进展程度、以折线图展示学习者不同时期的心情状态。教师通过查看数据结果及相关博客,及时了解学习进展情况,发现问题及时干预,对学习者的提供相应指导。

(三) Socrato

Socrato 是在线学习评估平台,提供标准化考试题库与相关学习内容。学生可利用平台模拟各类标准化考试,如 SAT、ACT、SSAT、MCAS 等。平台能迅速反馈诊断性评价。Socrato 可为个人或团体快速提供测试成绩及诊断报告,详细指出学习者各部分考试中各知识点的掌握情况,明确标出学习者的优劣势。教师可利用 Socrato 管理学习者测试成绩、家庭作业,并为学生自动订制个性化的学习计划,发送学习指导寄语。此外,平台还可以记录、跟踪学习者个体、团体学习轨迹,帮助教师生成形成性评价或总

结性评价,通过学习数据洞察学习者个体、团体的学习绩效并为学习者制定个性化学习策略,促进学习者提高考试成绩。

五、学习行为分析工具比较

常见的学习行为分析工具有:Google Analytics、Mixpanel。两种工具的比较见表四。

(一) Google Analytics

Google Analytics 20 是款通用的 Web 数据统计分析工具,用于了解学习者访问各学习页面的频率、停留时间、平台内的移动轨迹、用户参与度等。它利用事件跟踪系统,可跟踪所有重要事件,参与流(engagement flow)可直观显示学习者学习的全过程,包括他们实际查看的页面以及进行的操作。此外,该工具还可统计分析移动终端数据。

(二) Mixpanel

Mixpanel 也是款能提供实时 Web 数据分析服务的通用性工具,可实时监测学习者访问行为,记录、分析学习者特征,追踪评论数。教师可选择任一学习者,查看其在平台中浏览的页面,学习历史记录,了解其行为特征,预测行为趋势等。课程管理者可根据不同维度的监测数据、漏斗分析等衡量学习者在平台的学习体验,如运用“漏斗分析”工具,可查看哪些环节、哪些内容学习者的缺失率在上升,进而针对该环节进行课程改进。当学习者长时间未访问网站时,它可发送消息提醒。此外,Mixpanel 也可统计分析手机 APP 中的数据,还可向手机 APP 用户

表四 学习行为分析工具比较

编号	工具名称	发起者	使用环境	专门程度	使用对象	分析对象	数据支持格式	输出格式	可视化	商业软件/开源	优势与不足
1	Google Analytics	Urchin 公司运营的付费服务, 被 Google 收购后免费开放使用	浏览器 (需进行设置)	通用	学习平台拥有者	用户访问网站的轨迹及相关数据	记录用户行为数据外, 还支持导入 CSV 格式的外部数据	PDF、XML、CSV、TSV	是	商业软件, 收费, 可申请免费账户 (含 80 多个报告)	分析工具丰富, 提供多种功能的 API 接口, 支持内容分析、社交分析等多种统计分析, 主要用于网站数据分析, 记录点击量等。
2	Mix-panel	Mixpanel 公司	浏览器 (需进行设置)	通用	学习平台拥有者	用户访问网站的轨迹及相关数据	自动捕获并记录的用户行为数据	通过 API 接口导出数据	是	商业软件, 按统计数量级收费, 1000 人以内或 25, 000 个行为点内免费。	主要用于商用网站数据分析。可精细统计个人的数据行为, 支持移动数据分析。

表五 综合学习分析工具比较

编号	工具名称	发起者	使用环境	专门程度	使用对象	分析对象	数据支持格式	输出格式	可视化	商业软件/开源	优势与不足
1	WEKA	新西兰怀卡托大学机器学习项目组	Windows/Linux/Mac 操作系统下 Java 环境	通用	机器学习研究者, 科学家, 具有 R 语言或 Java 知识的专业人员	大规模数据	ARFF 文件, CSV	BMP, EPS, JPEG, PNG, ARFF 文件	是	开源软件, 不收费	处理大数据的开源软件, 具有强大的机器学习分析功能; 软件使用门槛很高, 需掌握机器学习等知识
2	SPSS	美国斯坦福大学三位研究生开发并成立 SPSS 公司, 后由 IBM 公司收购	Window/Linux/Mac 操作系统下	通用	粗通统计知识的初学者或熟练者或精通者	社会研究、行业研究	Excel, Dbase, Access, SAV, SYS 等多种格式的数据文件	SAV, SYS, POR, DAT, CSV, DAT, XLS, DBF 等多种格式	是	商业软件, 收费	适用范围广, 略通统计知识即可使用, 针对高阶用户, 还提供拓展功能和语言编程功能。
3	SSAS	微软团队研发 OLAP, 后更为 SSAS	Windows	通用	熟悉 SQL 相关理论、掌握操作技能的专业研究者	多维数据集	Access, TXT, CSV, Excel, ODC 等多种数据库文件	Access, Excel, TXT, CSV, ODC 等多种数据库文件	是	专用商业软件, 收费	目前主要用于分析商业数据, 使用门槛较高, 只能在 Windows 环境下使用

推送“手机问卷”, 收集学习者反馈、记录数据。

六、其他综合性学习分析工具比较

常见的综合性学习分析工具有: WEKA、SPSS、SSAS。三种工具比较见表五。

(一) WEKA

WEKA (Waikato Environment for Knowledge Analysis) 是款基于机器学习的数据挖掘工具。它通过对数据的挖掘获取趋势和模式, 支持多种经典的数据挖掘任务, 可对大规模数据进行预处理、分类、回归、聚类、关联规则分析, 具有可视化工具显示所

有属性的散点图 (二维或 3D) 矩阵, 可视化显示知识流, 可进行贝叶斯网络编辑。该工具也适合开发新的机器学习模式, 但其不能实现多关系数据挖掘, 需转换为单表才能在 WEKA 中实现数据挖掘。WEKA 目前不能实现序列建模。

(二) SPSS

SPSS (Statistical Product and Service solutions) 集数据输入、编辑、统计分析、图表制作、生成结果报表为一体, 可支持从策划到数据收集、分析、报告和部署各环节。该工具包含必须的基础分析模块及 15 个拓展模块。基础分析模块有数据汇总、计数、

交叉分析、分类、描述性统计分析、因子分析、回归及聚类分析等功能。扩展模块可实现高阶的统计分析功能,如分析过程数据、复用小样本数据模拟大样本结果,用二维图与感知图呈现数据关系等进行数据处理与数据挖掘分析。此外,该工具支持 Python 和 R 等语言进行编程二次开发。

(三) SSAS

SSAS (Microsoft SQL Server Analysis Services) 是微软 SQL 服务器中的联机分析处理、数据挖掘与报告工具,广泛用于分析多个数据库或不同类表中的信息处理。SSAS 可使用来自数据集市或数据仓库中的数据进行更深入和更快速的数据分析,以创建多维数据集。针对不同的操作和编程环境,SSAS 可支持不同的 API 接口、目标模型及多种查询语言。在许多不需要实时而具有海量或需要足够灵活的分析模型中,SSAS 有很大优势,比如性能和用户可定制性,可分析多个数据集或不同的数据表。

学习分析工具在学习分析过程中具有举足轻重的作用,不同的工具具有不同的特点,研究者应根据各工具的特点和优势,选择合适的工具并加以有效运用,发挥学习分析技术巨大的应用潜力,实现其学习改进的价值。

[注释]

①Cohere 既可以分析内容,也可以分析网络。它通过对学习内容的相似性判断网络关系。本文把它归类到学习网络分析中。

[参考文献]

- [1]Chen. E., Heritage, M., & Lee, J. (2010). Identifying and monitoring students' learning needs with technology [J]. *Journal of Education for Students Placed at Risk*, (3): 309 - 332.
- [2]Clow, D., & Makriyannis, E. (2011). iSpot Analysed: Participatory Learning and Reputation [EB/OL]. Paper presented at the 1st Learning Analytics And Knowledge Conference, February 27 - March 1, Banff, Canada.
- [3]DeWever, B., Schellens, T., Valleke, M., & van Keer, H. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review [J]. *Computers & Education*. 46 (1): 6 - 28.
- [4]Ferguson, R. (2011). The construction of shared knowledge through asynchronous dialogue. [EB/OL]. [2011 - 11 - 22]. <http://oro.open.ac.uk/19908/>.
- [5]顾小清, 黄景碧 (2010). 让数据说话: 决策支持系统在教育中的应用 [J]. *开放教育研究*. (10): 99 - 105.

[6]胡长爱, 朱礼军 (2010). 复杂网络软件分析与评价 [EB/OL]. *数字图书馆论坛* 2010. 5.

[7]Hansen, D., Rotman, D., Bonsignore, E., Milic - Frayling, N., Rodrigues, E., Smith, M., & Shneiderman, B. (2009). Do you know the way to SNA?: A process model for analyzing and visualizing social media data. U. of Maryland Tech Report [EB/OL]: HCIL - 2009 - 17.

[8]Haythornthwaite, C., & De Laat, M. (2010). Social networks and learning networks: Using social network perspectives to understand social learning. Paper presented at the 7th International Conference on Networked Learning [EB/OL]. January, Aalborg, Denmark.

[9]Huisman, M., & Van Duijn, M. A. J. (2005). Software for social network analysis [A]. In: Carrington, P. J., Carrington, J. Scott, & S. . *Models and methods in social network analysis* [C]. Cambridge University Press, Cambridge: 270 - 316.

[10]Jovanovic, J., Gasevic, D., Brooks, C. A., Devedzic, V., & Hatala, M. (2007). LOCO - Analyst: A Tool for raising teachers' awareness in online learning environments [EB/OL]. In *Proceeding of Creating New Learning Experiences on a Global Scale, Second European Conference on Technology Enhanced Learning*. EC - TEL 2007, September 17 - 20, Crete, Greece.

[11]Jovanovic, J., Gaevic, D., Brooks, C., Devedic, V., Hatala, M., Eap, T., et al. (2008). LOCO - Analyst: semantic web technologies in learning content usage analysis. *International Journal of Continuing Engineering Education and Life Long Learning* [J/OL]. 18 (1): 54 - 76.

[12]Kirschner, A. (2012). Overview of common social network Analysis Software Platforms. San Francisco: Monitor Group [EB/OL].

[13]Knewton (2011). Learning and knowledge analytics [EB/OL]. [2011 - 04 - 17]. <http://www.learninganalytics.net/?p=126>.

[14]Loscalzo, S., & Yu, L. (2008). Social network analysis: Tasks and tools [A]. In H. Liu, J. J. Salerno, & M. J. Young (Eds.) *Social computing, behavioral modeling, and prediction* [C]. New York: Springer.

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

[16]OHalloran, K. (2011). Investigating argumentation in reading groups: combining manual qualitative coding and automated corpus analysis tools [J]. *Applied Linguistics*, 32(2): 172 - 196.

[17]Shum, B. S., & Ferguson, R. (2014). Social learning analytics. Technical report KMI - 11 - 01, Knowledge Media Institute, The Open University/ UK [EB/OL]. [2014 - 03 - 28] <http://kmi.open.ac.uk/publications/pdf/kmi-11-01.pdf>.

[18]Smith, M. A., Shneiderman, B., Milic - Frayling, N., Mendes Rodrigues, E., Barash, V., Dunne, C., Capone, T., et al. (2009). Analyzing (social media) networks with NodeXL. *Proceedings of the fourth international conference on Communities and technologies*, [EB/OL]. New York, NY, USA: ACM. 2 - doi: 10. 1145/

1556460.1556497.

[19]孙洪涛(2013) 开源工具支持下的社会网络分析 - - No-deXL 介绍与案例研究[J]. 中国远程教育 ,(2) :13 - 20.

[20]Suzanne ,L. ,Rebecca ,F. ,& Stefan ,R. (2011). Navigating and discovering educational materials through visual similarity search. In: Ed - Media 2011 - world conference on educational multimedia , hyper-media & telecommunications. [EB/OL] . June 27 - July 1 , Lisbon , Portugal.

[21]The Horizon Report 2011 edition (2011). [EB/OL][2012 - 06 - 25]. <http://wp.nmc.org/horizon2011/>.

[22]The New Media Consortium(2011). Learning Analytics [EB/

OL]. The Horizon Report 2011 edition:28 - 30.

[23]魏顺平(2013). 学习分析技术:挖掘大数据时代下教育数据的价值[J]. 现代教育技术 ,(2) :5 - 11.

[24]王陆(2009). 虚拟学习社区的社会网络结构研究, 博士论文[EB/OL].

[25]闫冠韞, 程伟, 杨寄, 王立波(2013). 中药现代化研究中的网络药理学方法及可视化工具[J]. 哈尔滨医科大学学报 ,(3) :287 - 290.

(编辑:徐辉富)

The Comparison of Learning Analytics Tools

MENG Lingling , GU Xiaoqing & LI Ze

(Department of Educational Information Technology , East China Normal University , Shanghai , 200062 , China)

Abstract: *In recent years , with the rapid development of a smart learning environment , massive , rich , diverse , and heterogeneous data are increasing amazingly. In education field , students' interests , preferences , activities , learning process information , such as the interaction with learning platform , as well as their implicit feedback to the e - learning platform , can all be recorded and traced. How to effectively make use of these data has drawn great concern. The data of a single person seems to be chaotic , but with the data accumulating to a certain extent , it will be presented in an order.*

There are strong or weak relations among the data. For example , what are the characteristics of students in different region or countries? What are the characteristics of learning behavior in different ages? What are the learning habits of different students? Which courses are needed urgently for a successful career? For a special course , which units are needed for review? Which units are needed to be emphasized? Which students encounter difficulties and need help?

Therefore there are amazing insights behind the data. If we extract the rules or determine the relationships among data , tremendous value will be created. Therefore , learning analysis techniques arise.

According to the Horizon Report 2011 in the New Media Consortiums Horizon Project , learning analytics technology will become a hot topic in the next few years. It will contribute to improving the learning process and make the learning more intelligent. As we can imagine , the analytics tools play an important role in the process of learning analytics. Good tools can make the research process more effective.

Many analytics tools have been developed. For example , Nvivo , Atlas. ti can be used to annotate the text and multimedia content. Gephi , JUNG , Guess can be used to analyze learning networks , and SPSS can analyze user data statistics. However , a key issue is how to choose the appropriate tool because different tools will have different functions. Previous studies conducted analysis on the analytic tools , which are fragmented and focused on the comparison of learning network analytics tools. Different from previous works , this paper categorizes the learning analytics tools from multiple perspectives , and compares the tools from the point of developer , application environment , users , the supported data format , visualization , advantages , disadvantages , and so on. Twenty - four tools are discussed in detail in order to provide proper technical support for learning analytics and help to make education more intelligent and effectively.

Key words: *learning analytics; tools; categories; comparison*