

学习分析主题结构研究及可视化分析

黄志南^{1,2} 陆星儿¹ 胡贺宁¹ 李艳燕¹

(1. 北京师范大学教育技术学院智慧学习研究院, 北京 100875;

2. 中央电化教育馆 基础教育教学资源部, 北京 100031)

[摘要] 学习分析作为教育大数据的重要应用领域,自问世以来便迅速得到众多学者的广泛关注。为进一步探析学习分析研究主题,本研究以 Web of Science 核心库为数据来源,以 2010-2015 年间发表的 674 篇学习分析文献为研究对象,采用多种分析工具及可视化技术,从高被引文献、核心作者分布、论文高产机构以及高频关键词方面入手,梳理并揭示了当前学习分析研究现状。研究利用聚类分析等多元统计分析方法,以高频关键词相异矩阵为基础,介绍了学习分析框架模型、驱动力、方法工具、技术支持以及应用研究等五大领域的研究趋向;通过绘制战略坐标图,明确了应用研究在该领域的关键位置,并通过主题演化分析进一步梳理学习分析的研究方向,为后续相关研究提供参考。

[关键词] 学习分析;共词分析;可视化;聚类分析

[中图分类号] G434

[文献标识码] A

[文章编号] 1007-2179(2016)05-0102-10

一、引言

学习分析是通过对学习者及其学习情境进行测量、收集、分析和报告,以达到优化学习环境的目的(Ferguson, 2012)。自学习分析问世以来,国际上众多学者从不同角度对其进行了研究。西蒙斯(Siemens, 2009)提出了改变传统课堂教学的 TEKL (Technologically Externalized Knowledge and Learning) 模型,并创建了具体的学习分析过程模型。沃尔夫冈和亨德里克(Greller & Drachsler, 2012)使用里奇(Richie)的一般形态分析法,对学习分析研究社区里讨论的话题进行归纳总结,形成了学习分析要素模型。韩国檀国大学学者结合在线交互可视化技术,提出了对学习数据进行分析的多维分析法(Kim & Lee, 2012)。此外,普渡大学开发的 Course Signals 工具可以追踪学生的学习过程,能及时向出现学习危机的学生提出预警,并给予正确指导(Iten

et al., 2008)。为了客观、清晰地把握国内外学习分析的研究现状,本研究以 Web of Science 核心库为数据源,以“learning analytics”为主题对 2010-2015 年间发表的文献进行检索,共获得 674 条记录。研究利用 Bicombo 词频分析软件、Citespace 引文分析工具、SCI2 知识图谱构建工具及 SPSS 统计分析软件等多种工具,绘制共被引网络图、关键词社会网络分析图、聚类图、战略坐标图等,从不同视角呈现学习分析研究的历史、现状、领域及发展趋势,为后续学习分析研究提供参考。

本研究还采用基于知识图谱的研究方法,利用可视化技术分析和呈现知识间的网络关系。知识图谱方法主要包括引文分析、共词分析、社会网络分析等方法。其中,引文分析是对文献、期刊等之间的引用与被引用关系进行分析的理论与方法(鲁超等, 2010),利用它可以清晰地发现文献的被引用情况,挖掘学科中的高水平文献与经典言论,进而为后续

[收稿日期] 2016-03-03

[修回日期] 2016-08-20

[DOI 编码] 10.13966/j.cnki.kfjyyj.2016.05.012

[基金项目] 国家自然科学基金项目(61075048)和北京市共建项目专项(BJ20151017)。

[作者简介] 黄志南,中央电化教育馆助理研究员,北京师范大学教育技术学院硕士生;胡贺宁、陆星儿,北京师范大学教育技术学院学生;李艳燕(通讯作者),教授,博士生导师,北京师范大学教育技术学院,研究方向:学习分析技术(liyy@bnu.edu.cn)。

研究服务。共词分析主要用于统计同一篇文章中的不同词组间同时出现的次数,并在此基础上构造共现矩阵、相似矩阵、相异矩阵呈现这些词之间的亲疏关系,最终根据亲疏关系将相似主题的词进行聚类。该方法经多年发展已被广泛应用到信息检索、科学计量、人工智能等领域(储节旺等,2011)。社会网络分析法能将复杂关系进行量化,通过对网络中个体与整体的中心度、聚簇等属性的分析描述不同主体之间的联系,揭示这些关系的结构。

二、引文分析

(一) 高被引文献分析

在一个统计时段内,被引用次数排在学科前列的文献被称为高被引文献。高被引文献分析法,主要用来统计和呈现被引用率较高文献的信息。分析高被引文献,有助于把握学习分析的发展方向,为研究者追踪学习分析提供参考。在数据处理过程中,本研究将下载的文献导入到 Citespace 中,并对该软件的属性进行设置。其中,时间段(Timespan)设置为 2010 - 2015,长度(Slice Length)设为 1,结点类型(Node Types)设置为被引文献(Cited Reference),最终显示每一个 Slice 内被引频次最高的前 50 个结点,得到高引用率文献的可视化图。本研究中显示出来的文献被引频次大于等于 6(见图 1)。

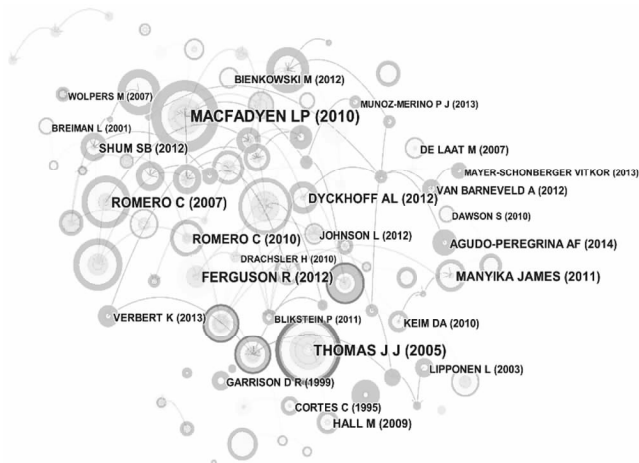


图 1 高引频文献可视化图

由图 1 可以看出,中心度和被引频次均较高的是托马斯(James J. Thomas)2005 年发表的《富有启发的路径:可视化分析的研究和发展议程》。该文较早对可视化分析做了详细定义,指出了可视化分

析流程包括分析推理技术、可视化表现技术和交互技术、数据呈现和转换技术以及支持产品、演示和传播技术四个核心要素,为学习分析可视化发展奠定了理论基础,成为后继研究者对可视化分析学科标准流程研究的范本。加拿大英属哥伦比亚大学麦克法迪恩和道森(Macfadyen & Dawson, 2010)2010 年发表的《挖掘学习管理系统数据为教育者开发一个预警系统:概念证明》一文,对该校本科生的在线学习分析进行实证研究,该研究被引频次 27 次。研究者通过分析 BB 平台学生学习行为的跟踪数据,建立了回归模型预测学生的课程通过情况,从理论上证明了基于学习管理系统进行数据挖掘开发“早期预警系统”的可能性。也有学者从学习分析可用的数据、分析流程、分析工具、利益相关者、制约与限制等角度设计了学习分析整体框架。塔尼亚·埃利安(Tanya Elians)发表的《学习分析:定义、过程与潜力》一文进一步提出由“数据收集(选择与采集)、数据处理(整合与预测)、知识应用(使用和优化)”组成的三个循环阶段的学习分析框架。

本研究还对有较高参考价值的文献在被引频次、中心度、发表时间等信息进行了整理(见表一)。

(二) 高被引机构分析

通过 Citespace 对高被引机构的统计分析,本研究发现主要高被引机构有马德里卡洛斯三世大学、悉尼大学、阿萨巴斯卡大学、国际商业机器公司(IBM Corp)、荷兰开放大学等(见图 2)。高被引机构及其关系图的结点总数为 205 个,边数为 45 条,密度仅为 0.0022。这表明学习分析研究仍处于发展阶段,大的成熟研究团体尚未形成,更广范围的机构合作有待形成。其中比较成熟的研究团队,如以阿萨巴斯卡大学为中心的连线较密,说明其与达尔豪西大学和西蒙弗雷泽大学有着较密切的合作。这三所研究机构均为加拿大著名公立高校,在学习分析研究领域有着较深入的实践经验。

(三) 高被引期刊分析

引量的分析可用于进一步挖掘期刊及收录论文的质量等信息。本研究将数据导入 Citespace 里选择 Cited Journal 命令进行高被引期刊可视化分析(见图 3)。关于学习分析引用较高的期刊和著作主要有《计算机科学讲义》(Lecture Notes in Computer Science)、《计算机与教育》(Computers & Education)、《教

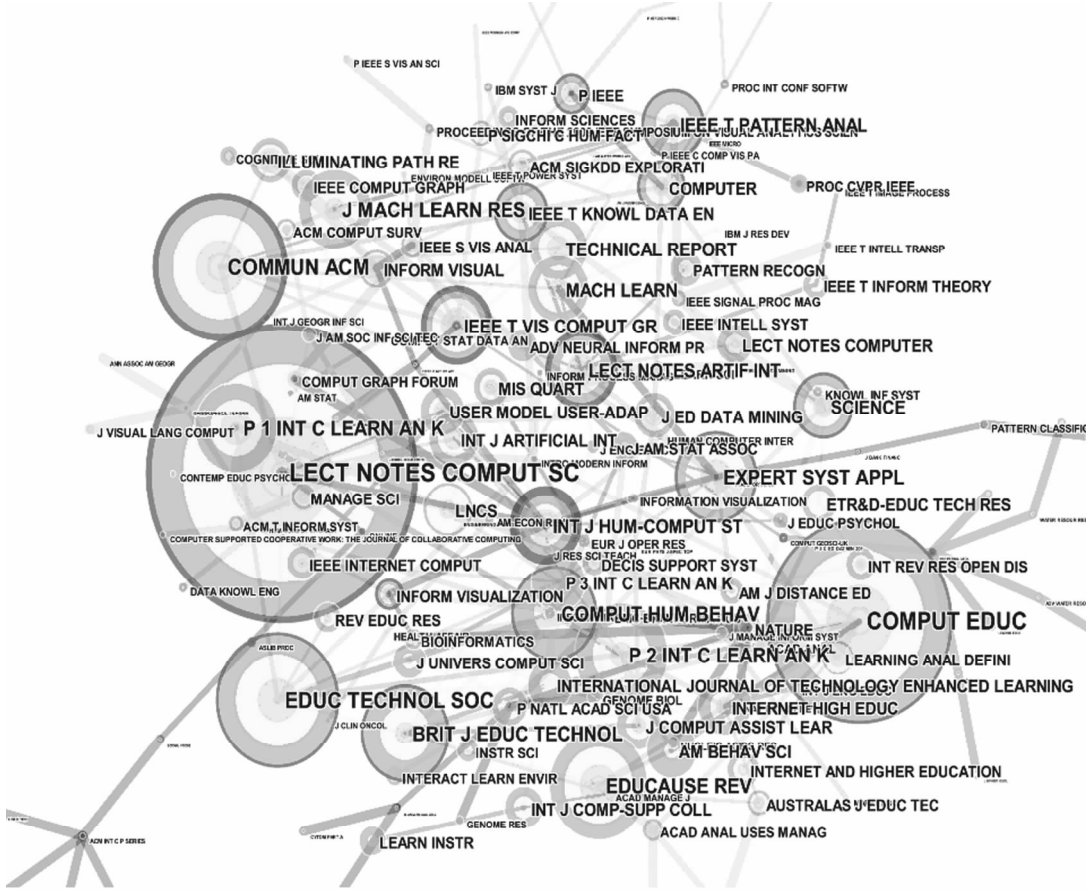


图3 高被引期刊可视化图

据进行合并等语义处理,比如将 Learning Analytics 和 learning analytics 统一转化为 learning analytics,最终得到 2021 个关键词。其中出现 2 次以上的有 556 个,累计频次占 56.7% (见表二)。

(二)高频关键词共现矩阵

本研究使用 Bicomb 软件生成高频关键词的共现矩阵。共现矩阵反映了两个关键词之间联系的紧密程度,可以作为多元统计分析的基础数据。由于原始共现矩阵绝对值的差异会对聚类分析产生影响,因此需要将其转换为斯皮尔曼相关矩阵。斯皮尔曼相关矩阵对角线上的相关系数为 1,表示词自身的自相关程度。非对角线的数字代表横轴和纵轴对应的两个关键词之间的距离,其数值越接近 1,说明关键词越相近。因该矩阵 0 值过多,为方便下一步的聚类分析,需要进一步将相似矩阵转化为相异矩阵,用 1 减去相似矩阵中的数据,得到表示两个关键词相异程度的相异矩阵 (见表三)。相异矩阵中数值越大,代表关键词相似度越小,数值越小,代表

关键词相似度越大。

为了清晰展现关键词之间的亲密度关系,本研究将 36 个高频关键词导入 SCI2 生成共现网络图 (见图 4)。从网络图可以看出,图中共有 36 个结点,各结点之间的边共有 110 条,结点权重最小为 1,最大 21,网络节点连接边数的平均值为 6.11,密度为 0.0873。其中, Big Data、educational data mining、MOOCs、Moodle、visual analytics 等词的中心度较高。可见,基于 MOOCs 等在线教育平台的大数据挖掘和可视化分析是学习分析研究的重要组成部分。

(三)主题聚类分析

为进一步探索学习分析的研究主题,本研究将相异矩阵导入 UCINET6 中对关键词进行聚类,把关系较近的聚成一类用于代表某一领域。根据聚类原理,用来表达某一领域研究主题的关键词在一篇文章中同时出现的频次越高,说明这两个关键词关系越紧密。这类关键词能够聚集成成为同一研究领域的内容 (见图 5)。根据聚类分析的结果,我们将学习

表二 高频关键词列表(词频 > =7)

序号	关键字段	出现频次	序号	关键字段	出现频次	序号	关键字段	出现频次
1	learning analytics	257	8	prediction	26	15	personalized learning	9
2	Big Data	81	9	clustering	22	16	mobile learning	8
3	machine learning	59	10	social networks	17	17	collaboration	8
4	educational data mining	57	11	assessment	17	18	models	8
5	visualization	52	12	MOOCS	13	19	online learning	7
6	visual analytics	40	13	social media	13	20	moodle	7
7	E-learning	27	14	CSCL	11	21	hadoop	7

表三 高频关键词部分相异矩阵

	learning analytics	Big Data	machine learning	educational data mining	visualization	visual analytics	E-learning	prediction	clustering	social networks	assessment
learning analytics	0.00	0.89	1.00	0.42	0.41	0.89	0.48	0.86	0.87	0.47	0.51
Big Data	0.89	0.00	0.48	0.78	0.99	0.97	0.94	0.69	0.99	0.82	0.96
machine learning	1.00	0.48	0.00	0.87	1.00	0.87	1.05	0.52	0.99	0.90	1.05
educational data mining	0.42	0.78	0.87	0.00	0.70	0.90	0.55	0.61	0.94	0.63	0.77
visualization	0.41	0.99	1.00	0.70	0.00	0.49	0.56	0.99	0.94	0.75	0.82
visual analytics	0.89	0.97	0.87	0.90	0.49	0.00	0.87	1.02	0.73	1.00	1.03
E-learning	0.48	0.94	1.05	0.55	0.56	0.87	0.00	0.91	1.00	0.74	0.84
prediction	0.86	0.69	0.52	0.61	0.99	1.02	0.91	0.00	1.03	0.92	0.99
clustering	0.87	0.99	0.99	0.94	0.94	0.73	1.00	1.03	0.00	0.99	0.89
social Networks	0.47	0.82	0.90	0.63	0.75	1.00	0.74	0.92	0.99	0.00	0.81
assessment	0.51	0.96	1.05	0.77	0.82	1.03	0.84	0.99	0.89	0.81	0.00

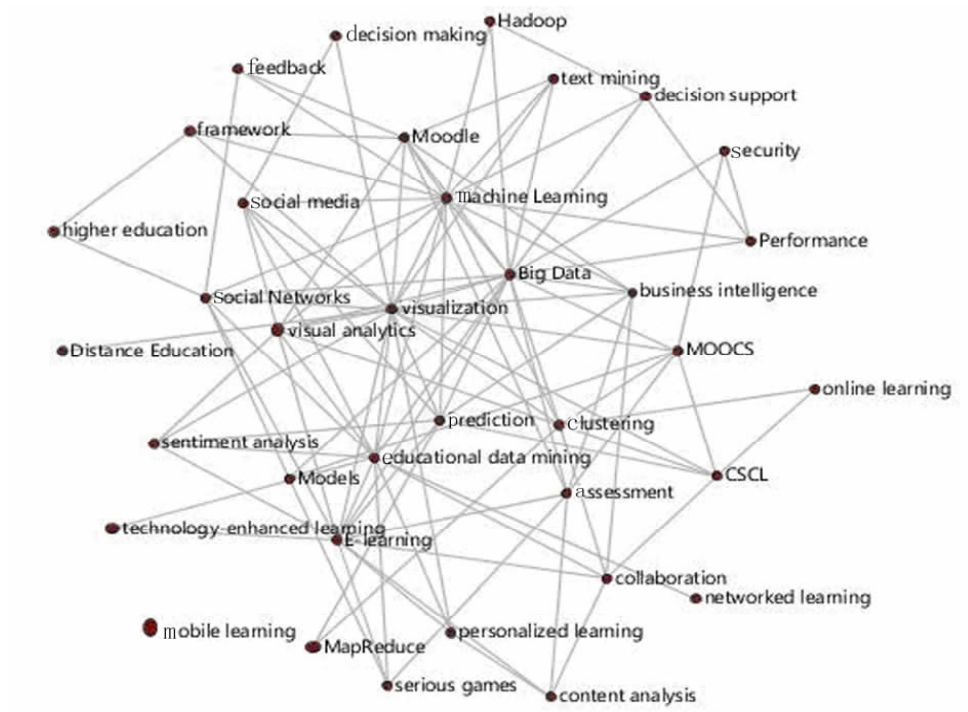


图4 高频关键词共现网络图

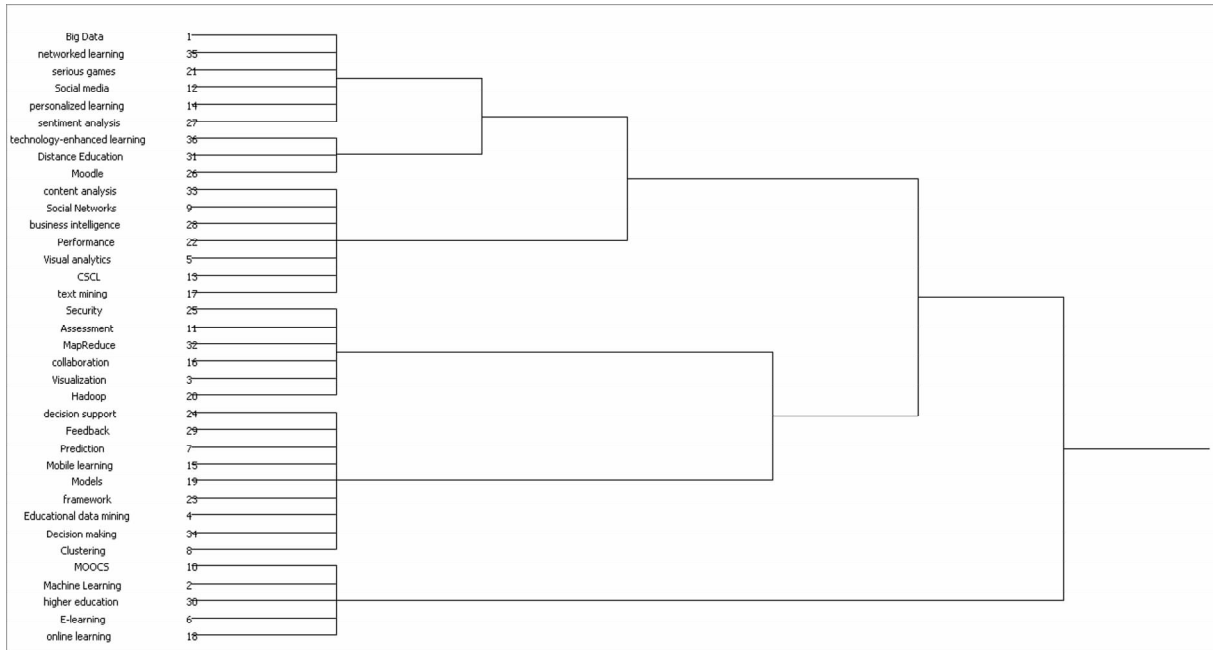


图5 主题聚类图

表四 主题分类

类团	类团名	关键词
1	学习分析框架模型研究	框架;预测;模型;聚类;决策支持;决策;反馈;移动学习;教育数据挖掘
2	学习分析驱动力研究	网络化学习;教育性游戏;社交媒体;个性化学习;情感分析;技术增强的学习;远程教育; Moodle;大数据
3	学习分析方法工具研究	内容分析;文本挖掘;社会网络;商业智能;绩效;计算机支持的协作学习
4	学习分析技术支持研究	MapReduce;Hadoop;协作;可视化;评估;数据安全
5	学习分析应用研究	MOOCS;机器学习;高等教育;数字教育;在线教育

分析分为五个类团(见表四)。

1. 学习分析框架模型研究

学习分析框架模型研究是学习分析研究的基础,它提供了学习分析的总体架构和主要因素。与该主题相关的关键词有:框架、预测、模型、聚类、决策支持、决策、反馈、移动学习和教育数据挖掘等。学习分析的框架模型理论有多种建构方式,但其设计的核心离不开对学生、学习和课程的关系分析。以此为指导设计的多元信息分析框架可以预测学习结果,用聚类等技术分析整合在线学习和移动学习的分析数据,架构了数据、分析工具和环境的有反馈环节的分析系统 (Ifenthaler & Widanapathirana, 2014)。与教育数据挖掘不同的是,学习分析更偏向个人、实时和多元数据分析,更强调社会干预,通过提供决策支持,影响决策结果,以实现对学习过程和学习环境的优化(Siemens & Baker,2012),教育数据

挖掘偏向非实时、机构应用。此类学习分析框架多种多样,一定程度上指导了理论应用,但迄今经过实践验证有效性的框架很少。

2. 学习分析驱动力研究

学习分析驱动力研究是学习分析的主要研究对象之一,与该主题相关的关键词主要有大数据、社交媒体、远程教育、网络化学习、个性化学习、技术增强学习和情感分析等。弗格森(Ferguson)认为学习分析的主要驱动力体现在技术、教育、经济等诸多方面(Media Consortium & EDUCAUSE Learning Initiative, 2013)。在普适计算时代,社交媒体和移动互联网的发展改变了学习方式,催生了各种技术增强型学习,为学习分析提供了丰富的数据来源。远程教育、Moodle等网络化学习存储了大量学习者学习信息,成为学习分析数据的主要来源。在这些数据的驱动下,学习分析技术得到进一步发展。通过情感分析

等技术驱动获取学生的学习状态、学习态度并及时对其做出评估,能够为学习分析提供更加立体多维度的考量(吴永和等,2013)。另外,随着社会对个性化学习需求的增加,学习分析系统带来的直接经济利益是驱动学习分析的另一因素。一方面,学术科研机构可以参考学习分析系统对个人的个性化学习方案的制定提供建议;另一方面,学习分析结果可以作为决策的依据,比如企业可以通过学习分析反馈的与个人学习需求相适应的内容,有针对性地进行基于智能分析决策的学习内容的推荐。这些因素驱动了学习分析研究的发展。

3. 学习分析方法工具研究

学习分析的方法和工具是进行学习分析技术研究的关键环节。与该主题相关的关键词有内容分析、可视化分析、文本挖掘、社会网络、商业智能、表现、计算机支持的协作学习。学习分析方法的功能在于从原始数据中提取可用信息,以便更好地了解学习者和学习环境。其中,内容分析法通过量化的方式从社会学角度分析学习者特征;社会网络分析法关注网络学习中信息分布和学习进展,多角度分析社会结构;文本挖掘情感分析获取学生的学习状态、学习态度并对其表现做出评估(李艳燕等,2012);可视化分析有助于将杂乱无章的数据通过分析推理技术以可视化的方式转换并展现在交互界面上。计算机支持的协作学习因具有协作性和交互性强的特点,在学习分析中常被用于学习者认知工具开发或被融入到学习者协作学习平台的使用过程中。

4. 学习分析技术支持研究

学习分析最关键的因素是数据,如何处理数据的采集、存储和分析关系到学习分析的最终效果。强大的技术支持为实现这一任务提供了可能。与学习分析技术支持研究主题相关的关键词有 MapReduce、Hadoop、协作、可视化、评估、数据安全等。在现有的技术开发领域,大数据关键技术和社会分析技术是两大核心,MapReduce 和 Hadoop 等云计算开源工具是学习分析需要利用的重要分析工具(孟小峰等,2015)。在数据采集过程中,移动互联网时刻链接每个人,云存储等技术则为技术采集和存储提供了出路。它们与云计算工具的协作组成学习分析的重要工具。有学者提出整合型学习分析系统,通

过学习仪表盘等可视化工具实现数据分析,并利用多角度视图提供多方面信息,以利益相关者进行评估(Siemens et al., 2011)。另一方面,学习分析在通过互联网、物联网采集和存储数据的过程中必然会涉及用户隐私。在保密制度和技術尚未明确的今天,进行数据分析需要对用户负责,因此数据安全也是学界的重要关注点(Slade & Prinsloo, 2013)。

5. 学习分析应用研究

学习分析对大数据的处理和指导学习的目的,是其应用的重要依据。与学习分析应用研究主题相关的关键词有:MOOCs、机器学习、高等教育、数字学习、在线教育。国际上对学习分析的定义指出,学习分析区别于其他概念的特点是其关注如何优化学习过程和学习环境。学习分析关注过程分析的特征使其成为一种有效的教学评价模式,这种特征和以MOOCs 为代表的在线课程的规模化、开放性、在线的特点相结合,可以提供多种教学绩效评价模式。学习分析结果的多角度呈现能够为管理者、辅导教师、研究人员提供多方面的数据,建构多种评价模式,这一点在高等教育中体现尤为突出。另外,学习分析技术处理大数据的能力是机器学习的重要基础,将学习分析引入机器学习,能更好地在教育领域利用计算机进行服务,利用智能技术支持学习。

(四) 战略坐标图分析

战略坐标是分别以向心度和密度为横轴和纵轴,用来表示不同领域间的关系程度和领域内部成员的关联强度。战略坐标共有四个象限,整个战略坐标的中心由向心度和密度的平均数构成。第一象限的向心度和密度都高,是研究领域的核心与成熟区域;第二象限的向心度较低、密度较高,是边缘与成熟区域;第三象限的向心度和密度都很低,是边缘与不成熟区域;第四象限的向心度较高、密度较低,是核心与不成熟区域。

根据以上方法,本研究通过计算学习分析领域的每个类团关键词的向心度和密度的平均数得出其向心度和密度,根据所有类团的向心度和密度的加权平均数计算得到学习分析研究领域的战略坐标中心数是(16.54, 15.14),依据战略坐标图确定学习分析研究类团的发展趋势(见图6)。从图中可以看出,“学习分析框架模型”“学习分析驱动力研究”“学习分析方法工具研究”“学习分析技术支持研

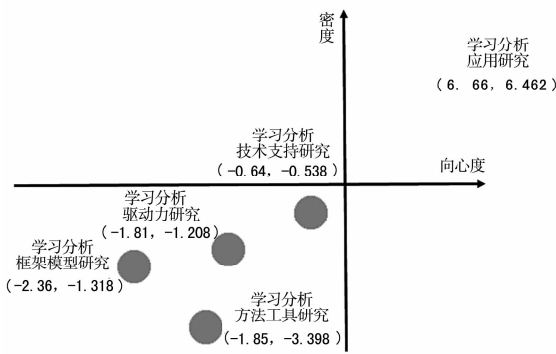


图6 战略坐标图

究”等均在第三象限。从发展状况看,这几类主题
的密度和向心度都较低,内部结构不够紧密,研究主
题不够成熟。但从研究内容看,这些主题均是学习
分析当今时代背景的主题,是学习分析未来研究进
一步发展的方向。未来的研究,一方面要加强研究
主题内部一致性,夯实研究基础;另一方面要抓住时
机,寻找与发展成熟研究主题的结合点。

“学习分析应用研究”类团位于战略坐标图
的第一象限,且密度与向心度较高,表明该主题内部
的关联度较高,同时与外部研究的紧密程度也较强。
通过查阅应用研究相关文献,本研究发现大多数应
用集中在有关学习的推荐系统上,并未对学习分析
做清晰界定,而是在学习分析基础上沿用旧的技术

方法对相关系统搜集的大量数据进行挖掘分析,从
而使学习分析应用领域研究在统计结果上处于较为
突出的位置。可见,针对学习分析的应用研究一直
是该领域学者的关注点。

(五)学习分析主题演化

为了进一步分析学习分析近几年研究重心
的变化过程,本研究采用 NEViewer 社区演化关系分析工
具对近五年学习分析的研究主题演化过程进行可视
化。该方法借鉴冲击图表示主题的演化过程,矩形
色块表示社,越靠近顶端的社区中心度越高,两个时
段的矩形之间的曲线形色块表示演化过程,颜色块
的高度表示社区的节点规模。通过冲击图,我们可
以非常直观地看到学习分析主题研究的进展,每个
社区对应的主题以社区中心度较高的节点(关键
词)表示(见图7)。

从图中可以看到,随着时间的推移,学习分析研
究主题的收缩和扩张现象都存在。

2010-2011年,学习分析研究处于起始阶段,
研究主题较为单一,主要局限在可视化分析和 Web
分析两个领域,侧重于对网络学习数据的分析及其
在教学实践中的可视化应用,并在随后几年一直保
持在靠近顶端的位置,这也是学习分析研究的基础
领域。

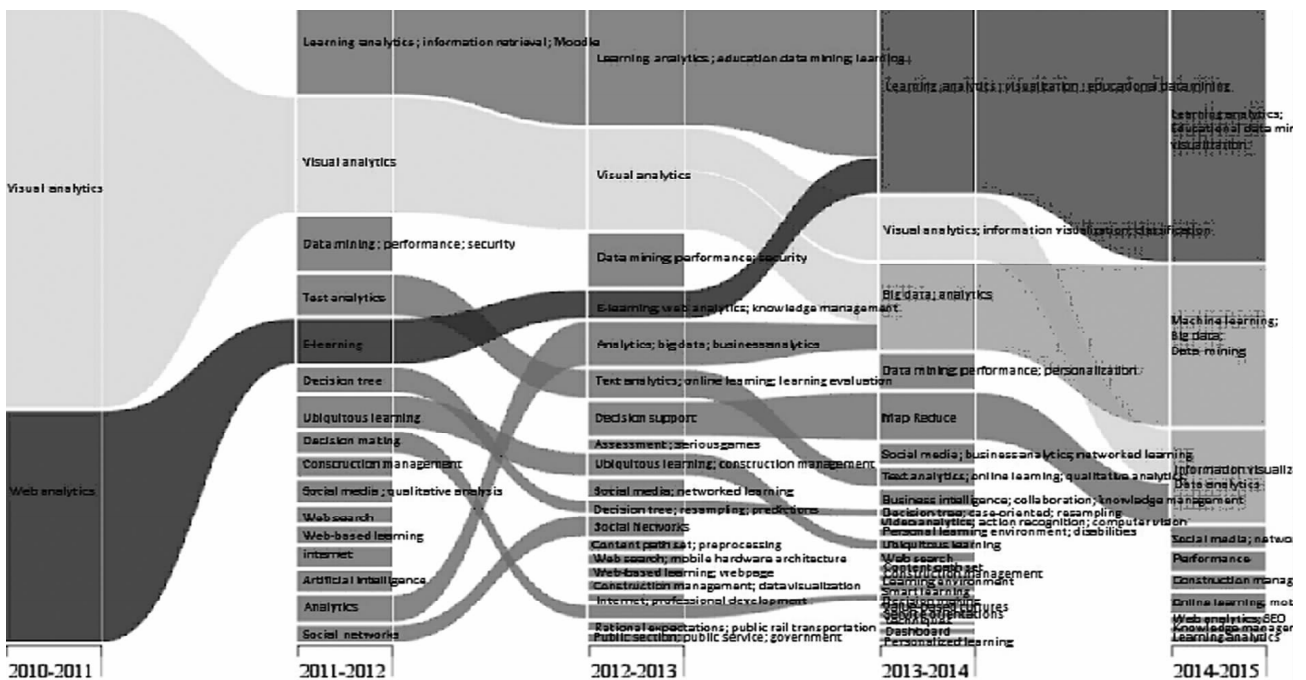


图7 学习分析主题演化冲击图

2011-2014年, 研究内容有了很大的拓展, 研究主题更为丰富。以学习分析为代表的社群开始突起, 并连续四年位居顶端, 成为研究的热点, 且呈逐年扩大趋势, 主要涉及信息检索和教育数据挖掘。Web分析研究虽然逐年收缩, 但逐年受到研究者的关注, 最终与学习分析主题融为一体成为研究的热点。可视化分析则在近三年发生分裂现象, 这反映了学习分析领域研究主题出现了分化, 可视化分析和机器学习成为代表性主题, 并与学习分析一并成为2015年三大核心研究热点。而部分刚兴起的社区在经历短暂时间后没有后继, 这意味着存在研究主题消亡的现象, 如绩效分析、数据安全、决策树、人工智能等。此外, MOOC、在线学习、泛在学习等为代表的社区随着时间的推移, 在随后的时段里排序有些许回落, 这与近五年来学习分析在应用领域研究的快速发展是吻合的。

四、结语

学习分析作为一个新兴的研究领域, 可以有效地优化学习, 受到广大研究者和实践者的广泛关注。本研究采用科学计量的方法, 围绕“学习分析”这一主题, 借助定量统计分析工具对学习分析相关文献进行解读, 揭示了相关研究成果中所隐含的信息, 展示了国际上学习分析的整体研究现状。

从研究结果看, 学习分析在分析模型框架、驱动力、方法工具和技术支持等方面开展了广泛的研究, 不过学习分析技术研究目前还处在起步阶段, 各研究领域不平衡, 也面临着诸多挑战。

在学习分析框架模型研究方面, 目前对数据收集标准的研究尚显不足。随着可用于学习分析的数据类型和数量的不断增加, 收集哪些数据? 如何收集和存储数据? 这些问题需要有相对统一和明确的标准加以规范, 以促进研究成果的推广和交流。

在结果呈现方面, 现有学习分析研究多将计算机分析结果直接呈现, 但这些分析结果大多过于复杂或隐晦, 并不能有效地支持教与学实践。因此, 如何引入专家、学者对分析结果进行专业的筛选和解释, 使分析结果更友好地呈现, 方便教师优化教学过程, 需要进一步探索。

在学习分析驱动力研究方面, 这主要来自数据分析技术的更新。但是, 学习分析研究应关注更多

的教育实践问题, 从教育角度出发应用技术, 而不是推广一些“新概念”, 学习分析要避免成为新技术导向的产物。

在分析工具方面, 数据挖掘技术的发展促进了学习分析研究的繁荣。当前研究主要关注于新方法新技术的应用, 缺少对如何筛选技术以及综合集成多种技术在教育中适用性的探讨。

在学习分析技术支持研究方面, 大部分研究者关注于学习分析工具的开发, 忽视了对工具的评价。如何评价学习分析工具和结果, 如何对学习分析工具及其结果的科学性、适用性和有效性进行评价等亟待研究探讨。

在学习分析应用研究方面, 目前的研究主要以小规模的在线课程作为案例研究, 未来需要面向MOOCs等大规模、跨平台以及线上线下学习方式开展更多的实证研究。同时, 随着以智慧学习环境为代表的新型学习环境的发展, 学习环境日趋变得智能和复杂。新技术和富媒体环境不仅使得教学更加多样和灵活, 也更加容易收集教学中的过程数据。

总体而言, 学习分析在智慧学习环境中将迎来更广阔的发展空间, 应用价值和发展潜力巨大。

[参考文献]

- [1] 储节旺, 郭春侠 (2011). 共词分析法的基本原理及 EXCEL 实现 [J]. 情报科学, 29(6):931-934.
- [2] Ferguson, R. (2012). The state of learning analytics in 2012: A review and future challenges [M]. Technical Report KMI-12-01, Knowledge Media Institute, The Open University.
- [3] Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics [J]. Journal of Educational Technology & Society, 15(3):42-57.
- [4] Ifenthaler, D., & Widanapathirana, C. (2014). Development and validation of a learning analytics framework: Two case studies using support vector machines [J]. Technology, Knowledge and Learning, 19(1-2):221-240.
- [5] Iten, L., Arnold, K., & Pistilli, M. (2008). Mining real-time data to improve student success in a gateway course [C]. In Eleventh Annual TLT Conference. West Lafayette, Purdue University.
- [6] Kim, M., & Lee, E. (2012). A multidimensional analysis tool for visualizing online interactions [J]. Journal of Educational Technology & Society, 15(3):89-102.
- [7] 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):20.
- [8] 鲁超, 刘清 (2010). 引文分析可视化现状 [J]. 情报杂

志, 29(11):48-52.

[9] Macfadyen, 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.

[10] MediaConsortium, N., & EDUCAUSE Learning Initiative (2013). *The NMC Horizon Report: 2013 Higher Education Edition* [J]. Austin, Texas: New MediaConsortium, 12(3):12-32.

[11] 孟小峰, 慈祥. (2015). 大数据管理: 概念, 技术与挑战 [J]. *计算机研究与发展*, 50(1):146-169.

[12] Siemens, G. (2009). Technologically externalized knowledge and learning [J]. *Connectivism: Networked and Social Learning*, 23(2):11-18.

[13] Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Shum, S. B., Ferguson, R., & Baker, R. S. J. D. (2011).

Open Learning Analytics: an integrated & modularized platform[J]. Proposal to design, implement and evaluate an open platform to integrate heterogeneous learning analytics techniques, 27(1):16-27.

[14] Siemens, G., & Baker, R. S. (2012). Learning analytics and educational data mining: towards communication and collaboration [J]. In *Proceedings of the 2nd international conference on learning analytics and knowledge*, ACM, 1(2):252-254.

[15] Slade, S., & Prinsloo, P. (2013). Learning analytics ethical issues and dilemmas [J]. *American Behavioral Scientist*, 57(10): 1510-1529.

[16] 吴永和, 陈丹, 马晓玲, 曹盼, 冯翔, 祝智庭 (2013). 学习分析: 教育信息化的新浪潮[J]. *远程教育杂志*, 31(4):11-19.

(编辑:李学书)

Topic Structure and Visualization Analysis of Learning Analytics Research

HUANG Zhinan^{1,2}, LU Xinger¹, HU Hening¹ & LI Yanyan¹

(1. School of Educational Technology, Smart Learning Institute, Beijing Normal University, Beijing 100875, China; 2. National Center for Educational Technology, Beijing 100031, China)

Abstract: As an important part of educational data mining, learning analytics has drawn extensive attention among international scholars since its origin. To further explore the research on learning analytics, this paper, based on 674 articles on learning analytics between 2010 and 2015 from Web of Science, explored and revealed the international research status about learning analytics.

In this study, we used the “Bicomb” word frequency analysis software, the “Citespace” citation analysis tool and “SCI2” tool and visualization technology to analyze the high-cited articles, the distribution of core authors, high prolific institutions and high frequency keywords. In addition, based on dissimilarity matrix of high frequency keywords, this paper further clarified the research scope using cluster analysis. Furthermore, by mapping out the strategic diagram graph, the trend of the development orientation of learning analytics was further clarified in order to provide reference and suggestions for international research and practices on learning analytics.

Results showed that the research topics had been greatly expanded and were more abundant during the 2011-2014 period. The learning analytics on behalf of communities suddenly rose and ranked to the top for four years, which made it the hottest research area during those period of times. Learning analytics showed an increasing trend year by year, mainly related to information retrieval and education data mining. While, there is a split phenomenon of the visual analytics. It has two successors involving “Visual analytics” and “Machine learning”, respectively as the representative of the theme. The two successors with the learning analytics become the three core research hotspots in 2015.

From the results of the study, we can conclude that the learning analytics technology is still in the initial stage and the field of research is not balanced. Learning analytics research also face many challenges concerning data security, ethics, and privacy.

In general, although learning analytics technology has great application value and development potential, there is still a long way to explore and practice in order for it to be widely applied in learning sciences.

Keywords: learning analytics; Co-word analysis; visualization; cluster analysis