

# 学习分析技术：挖掘大数据时代下教育数据的价值\*

魏顺平

(国家开放大学 现代远程教育研究所, 北京 100039)

**【摘要】**当前,大数据时代已经来临,教育领域同样积累了海量数据。教育领域已经部署了众多的学习管理系统,在这些软件系统中存储着海量的学习者信息及学习过程数据。如何利用这些数据,使这些数据转变为信息、知识,并为教学决策、学习优化服务,已成为教育工作者以及学习者所关注的内容。学习分析技术有助于发挥学习过程数据的价值,使数据成为审慎决策、过程优化的重要依据。该文介绍了国内外学习分析技术研究现状,归纳出学习分析技术的关键技术及分析模式,并以实例从不同用户视角包括管理者、辅导教师、学习者展示了学习分析技术在网络学习过程分析中的应用过程。

**【关键词】**大数据时代;教育数据;学习分析;关键技术;分析模式

**【中图分类号】**G40-057 **【文献标识码】**A **【论文编号】**1009—8097(2013)02—0005—07 **【DOI】**10.3969/j.issn.1009-8097.2013.02.001

刚刚过去的 2012 年,大数据(big data)一词越来越多地被提及,人们用它来描述和定义信息爆炸时代产生的海量数据,并命名与之相关的技术发展与创新<sup>[1]</sup>。国际数据公司(IDC)的研究结果表明,2008 年全球产生的数据量为 0.49ZB,2009 年的数据量为 0.8ZB,2010 年增长为 1.2ZB,2011 年的数量更是高达 1.82ZB,相当于全球每人产生 200GB 以上的数据。最早提出“大数据”时代到来的是全球知名咨询公司麦肯锡,麦肯锡称:“数据,已经渗透到当今每一个行业和业务职能领域,成为重要的生产因素。人们对于海量数据的挖掘和运用,预示着新一波生产率增长和消费者盈余浪潮的到来。”在教育领域,《2013NMC 地平线报告(高教版)》(预览版)非常有预见性的认为“大数据和学习分析”将在未来 2 至 3 年成为主流技术。

## 一 核心概念界定

在首届“学习分析和知识国际会议”上,与会者一致认为: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。《2012NMC 地平线报告(高教版)》也给出了近似的定义,即学习分析技术是对学生生成的海量数据的解释和分析,以评估学生的学术进展,预测未来的表现,并发现潜在的问题。从这些定义可以看出,学习分析技术分析的对象是学生及其学习环境,目的是评估学生、发现潜在问题、理解和优化学习,基础是海量数据。

## 二 国内外研究现状

2011 年 2 月底,首届“学习分析技术与知识国际会议”在加拿大的阿尔伯塔省班芙市举行,主题之一就是学习分析技术<sup>[2]</sup>。美国新媒体联盟发布的 2010 年度和 2011 年度《地平

线报告》均预测学习分析技术将在未来的四到五年内成为主流<sup>[3]</sup>。可见,学习分析技术已逐渐成为教育中的一项新兴技术。

事实上,在“学习分析技术”概念出现之前,与之相关的技术、工具及其应用研究已经开展起来。2004 年在高等教育中出现的“智能导师系统”和“人工智能系统”掀起了“教育数据挖掘”研究的热潮,也促使学术分析技术这一关注学习者行为的分析技术的诞生<sup>[4]</sup>。Romero&Ventura(2007)以及 Baker & Yacef(2009)对 10 余年的教育数据挖掘研究进行分析,归纳出 5 类教育数据挖掘方法,它们是统计分析(可视化;聚类(聚类、离群点分析);预测(决策树、回归分析、时序分析);关系挖掘(关联规则挖掘、序列模式挖掘、相关挖掘);文本挖掘<sup>[5]</sup>。另外,随着网络学习及相关学习管理系统的不断普及,数据挖掘方法在学习管理系统中得到应用,并开启了利用网络分析技术对学习行为加以分析的研究,Romero,C.等人(2005)对 Moodle 学习平台的日志分析是这方面研究的典范<sup>[7]</sup>。记录在学习管理系统中的学习者行为数据,经过聚集、分类、可视化以及关联规则分析等操作,生成实时的数据报告,或者利用从行为数据中反映出的常模来生成预测模型。随着学习分析研究与实践的不断深入,除了数据挖掘方法外,一些原本属于社会科学领域的方法如社会网络分析法、话语分析法、内容分析法等也得到成功应用并成为学习分析的关键技术,这些方法的典型应用之一是师生交互行为分析。

国内外多位研究者的研究实践证明,学习分析技术对于学生、教师、管理人员、研究人员以及技术开发人员均具有重要价值:(1)对于学生而言,学习分析技术可以从学习者行为角度了解学习过程的发生机制,并用来优化学习,以基于学习行为数据的分析为学习者推荐学习轨迹,开展适应性学习、自我导向学习<sup>[3]</sup>。(2)对于教师和管理人员而言,学

习分析技术可用来评估课程和机构,以改善现有的学校考核方式,并提供更为深入的教学分析,以便教师在数据分析基础上为学生提供更有针对性的教学干预<sup>[5]</sup>。(3)对于研究人员而言,学习分析技术可作为研究学生个性化学习的工具和研究网络学习过程和效用的工具<sup>[9]</sup>。(4)对于技术开发人员而言,借助学习分析技术发现的学习管理系统各模块使用频次以及使用路径,优化学习管理系统界面设计,并且根据其他人员开展学习分析的需要优化学习管理系统日志功能。

另外,学习分析技术在网络高等教育领域大有可为,有助于监控学习过程,保障教育质量。学习分析的开展有赖于系统化、结构化的海量数据,恰好网络高等教育由于其全面采用学习管理系统和教育管理信息系统,已经积累了大量的系统化、结构化数据。自1998年教育部开展现代远程教育试点工作以来,我国网络高等教育得到了迅猛发展。据教育部统计数据,2010年网络高等教育本、专科招生数达166万人,在校生数达453万人,分别占当年全国高等教育招生数和在校生数的16%和14%,由此可见网络高等教育在我国高等教育中的重要地位。近年来出现的网络高等教育学生考试集体作弊事件让社会对网络高等教育质量提出质疑,因此,着力提高网络高等教育的质量应该放在更为突出的位置。应用学习分析技术,调查网络高等教育中学习者网络学习的现状,包括学习者的主要特征、网络学习行为特点、学习行为的影响因素及其所带来的学业结果,有助于掌握成人网络学习规律,从而优化学习过程,改进学习效果,提升教育质量。

### 三 关键技术与分析模式

#### 1 学习分析关键技术与主要工具

学习分析关键技术涉及内容分析、话语分析、社会网络分析、系统建模等技术以及统计分析与可视化、聚类、预测、关系挖掘、文本挖掘等一系列数据挖掘方法。

常见的学习分析工具有很多,如支持对原始帖子进行标注或编码、交叉引用和简短评论的工具包括Nvivo、Atlasti;支持基本的基于词典的文本分析的工具,如CATPAC、LIWC;专门的内容分析工具,如北京师范大学知识工程研究中心开发的智能化内容分析工具VINCA;专门的社会网络分析工具,如UCINET;用于系统建模的工具,如Coordinator系统建模工具;专门的数据挖掘工具,如SQL SERVER 2005 Analysis Service(SSAS)、Weka、SPSS等等。

#### 2 学习分析技术应用情境分析与学习分析模式构建

学习分析的对象自然是学生的学习。通过对学习过程各要素的分析,我们可以进一步细化学习分析的对象,从而界定出各种应用情境,并构建特定的应用模式。我们采用格语

法中的格框架方法,以“学习”为中心谓词进行语义格标识。常见的语义格包括施事格、工具格、承受格、使成格、处所格、客体格等。所得结果如图1所示。

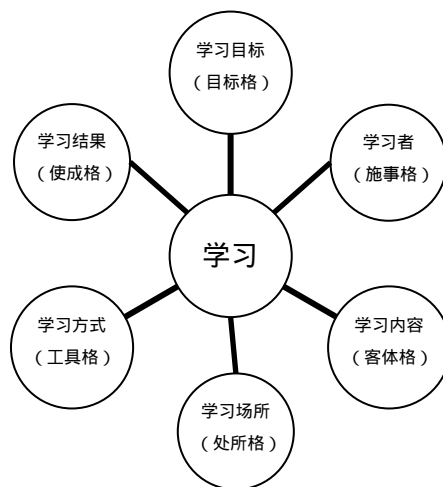


图1 “学习”的语义格图

如图1所示,采用学习分析技术对“学习”的语义格包括“学习者”、“学习内容”、“学习方式”、“学习结果”等进行挖掘分析,将使格框架中的中心谓词“学习”的细节更加清晰,从而形成对学习现状的更加完整的认识,能够回答“谁在学、学什么、怎么学、学的结果如何”等一系列问题。针对解决不同类型的问题,将产生三类学习分析技术应用情境,即用于回答“谁在学”的学习者特征分析,用于回答“学什么、怎么学”的学习过程分析以及用于回答“学的结果如何”的学习结果分析,从而构建出三种学习分析模式,可为他人完成类似的挖掘任务提供参考。

学习分析模式由“学习分析流程”、“工具与算法”及“数据与信息”三要素构成。“学习分析流程”包括数据收集、数据预处理、分析、预测、应用等环节,根据应用情境的不同“学习分析流程”各环节有所区别;“工具与算法”为“学习分析流程”提供支撑,它们从“学习分析关键技术与主要工具”中选取;“数据与信息”既有向“工具与算法”输入的数据,又有从“工具与算法”输出的信息或知识。

一种用于学习过程分析任务情境的学习分析模式如图2所示<sup>[8]</sup>。

在数据挖掘工作流的核心环节,主要的数据挖掘任务有学生登录行为分析、学生资源浏览模式分析、师生交互论坛分析和学生行为影响因素分析,所用到的工具和算法则有SSAS的聚类分析与顺序分析、聚类分析、关联规则等以及UCINET的网络图绘制等。

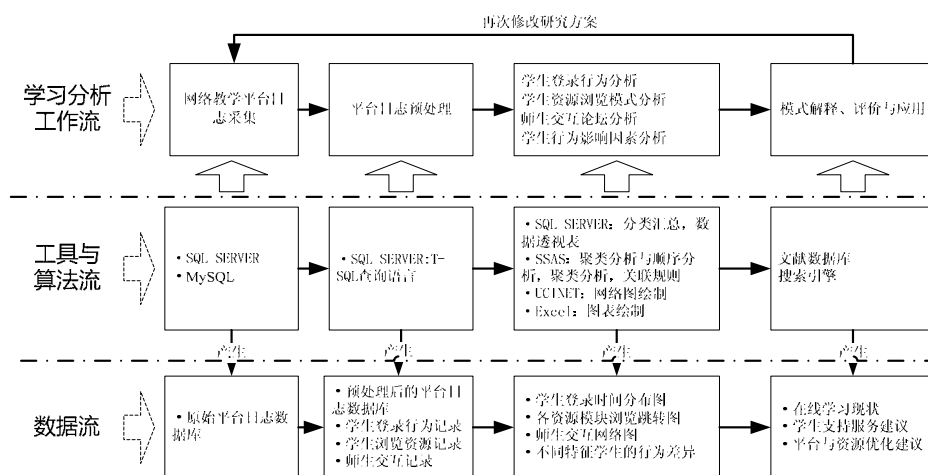


图2 用于学习过程分析任务情境的学习分析模式图

## 四 基于用户视角的学习分析技术应用

前文已经提到，学习分析技术对于网络学习有独特价值。由于笔者所在单位是 68 所现代远程教育试点院校之一——中央广播电视大学，可以较为方便的获取该校的学生数据，并且该校的办学规模在整个网络高等教育中具有独特地位，其招生数和在校生数都远高于其他远程教育试点。自 2004 年以来，中央广播电视大学所举办的网络高等教育（一般称作“中央电大开放教育”，后文在谈及中央电大网络高等教育时将采用这一称法）始终占有全国网络高等教育 60%~70% 的份额。本文将选取中央广播电视大学的学习管理系统，采用各类学习分析工具，基于学习管理系统的日志数据以及教育管理信

息系统有关学籍、课程、教师的数据，从管理者、辅导教师以及学习者等多种用户视角深入分析成人学习者网络学习现状及其影响因素，以优化成人学习者学习过程。

### 1 管理者视角的学习分析技术应用

作为教学管理者，他们一般采用学习分析技术了解某个办学机构的整体教学情况，如师生数量、学生各类活动总量及平均情况、教师各类活动总量及平均情况，主要用到“统计分析与可视化”这一类方法。

为了有效监控地方电大网上教学情况，对地方电大网上教学工作进行有针对性的指导和帮助，中央广播电视大学每年要组织对地方电大日常网上教学进行跟踪检查。常见的检查内容如表 1 所示（该表由成都广播电视大学改造完成）。

表 1 中央广播电视大学网上教学检查基本内容表

办学单位网上教学业绩										
上网情况						网上学习情况				
上网学生数	上网学生数/在校生数	学生在线时间	生均在线时间	学生登录次数	学生登录次数/上网学生数	新登录学生数/总登录学生数	浏览资源次数	生均浏览资源次数	论坛发帖总数	生均发帖数

下面对表 1 中的一些重要指标进行说明。

通过“上网学生数”、“上网学生数/在校生数”两项指标反映办学单位组织、引导、督促学生上网学习的情况。这两项指标是对办学单位组织、引导、督促学生网上学习情况进行后续分析的前提。

通过“学生登录次数”、“学生登录次数/上网学生数”两项指标反映学生网上学习活动的频度。进一步对学生登录的时间段分析，可以反映学生网上学习的进度和网上学习是否均衡。“新登录学生数/总登录学生数”以当月登录学生数和登录次数分别占全年登录学生数和登录次数的比例，说明学生网上学习的持久性和稳定性。

通过“学生在线时间长度”反映说明学生参加网上学习的深度。

通过“浏览资源次数”、“生均浏览资源次数”、“论坛发帖总数”、“生均发帖数”等指标反映学生在网上点击

资源和发帖交流的情况。该指标是评估办学单位组织和实施网上学习的重要指标。

### 2 辅导教师视角的学习分析技术应用

对于辅导教师而言，一般以所负责的某门课程为单位开展教学分析，分析内容包括学习者网络学习表现统计描述（从学习时间投入、学习活动频度、学习资源浏览、考试成绩等多个方面）、学习者自主学习路径分析、师生交互行为分析、影响成人学习者网络学习表现的各种因素分析，等等。所用的学习分析技术可以很丰富，如师生发帖的内容分析、师生交互的社会网络分析以及统计分析与可视化、聚类、关联规则、序列模式挖掘等一系列数据挖掘方法。

作者选取中央广播电视大学开放教育学生的入学课程《开放教育学习指南》课程作为样本来说明辅导教师视角的学习分析技术应用特点<sup>[9]</sup>。

《开放教育学习指南》是中央广播电视大学开放教育专

科和专科起点本科各专业学生的必修课，是开放教育学生的入门课程。以《开放教育学习指南》网络课程的在线学习过程数据作为研究样本，这些数据包括学生在使用电大在线学习平台时产生的登录、资源浏览、学习体验等数据以及学生在使用形成性考试平台时产生的考试成绩数据。为了开展影响因素分析，作者还从中央电大教务管理平台中采集学生基本信息数据。最终，作者获得了来自江苏电大参加 2010 年春季《开放教育学习指南》网络课程试点的 9369 名学生的基本信息数据、登录数据、资源浏览数据和形成性考试数据等四份数据。作者即以这 9369 名学生对应的基本信息、登录、资源浏览和考试成绩等四类数据作为分析对象，应用学习分析技术了解该课程中课程模块浏览情况以及学生自主学习路径。

(1) 课程模块浏览情况分析

《开放教育学习指南》网络课程包含“课程各章节”、“体验区”、“视频资源库”、“问题库”、“资料库”和“常用工具库”等内容模块。如表 2 所示，是网络课程各模块的一个说明。

表 2 《开放教育学习指南》课程各模块说明表

模块名称	模块说明
课程各章节	网络课程主体模块，包含有 61 个章节内容页面。
体验区	模拟网络平台使用过程，有 7 个过程模拟页面。
视频资源库	列出了课程各章节模块出现的视频资源。
问题库	列出了网络学习中存在的常见问题。
资料库	列出了开放教育的一些教学管理文件。
常用工具	列出了网络学习中可能用到的一些工具

表 3 网络课程各模块浏览情况表

模块名称	被浏览的资源页面个数	浏览频次	单个页面浏览频次均值	浏览学生数	学生参与率	人均浏览频次	单个资源人均浏览频次
课程各章节	103	809797	7862.11	9230	0.99	87.74	0.85
体验区	9	206720	22968.89	9049	0.97	22.84	2.54
视频资源库	37	36978	999.41	4321	0.46	8.56	0.23
问题库	6	34889	5814.83	3262	0.35	10.70	1.78
资料库	49	15843	323.33	2761	0.29	5.74	0.12
常用工具库	11	634	57.64	382	0.04	1.66	0.15

如表 3 所示，是《开放教育学习指南》网络课程各模块的浏览情况。

各个统计项说明如下：

- 1 被浏览的资源页面个数：通过页面浏览日志数据统计得到的各模块页面个数，可能少于各模块实际包含的页面数。
- 2 浏览频次：即一个页面被点击打开的次数。
- 3 单个页面浏览频次均值：由“浏览频次/被浏览的资源页面个数”算式得来。主要用来反映该模块各资源页面的使用程度。
- 4 浏览学生数：通过页面浏览日志数据统计得到的浏览各模块的学生数。
- 5 学生参与率：由“浏览学生数/学生总数”算式得来。用来反映该模块被学生关注的程度。
- 6 人均浏览频次：由“浏览频次/浏览学生数”算式得来。用来反映各学生在该模块的学习程度。
- 7 单个资源人均浏览频次：由“人均浏览频次/被浏览的资源页面个数”算式得来。用来反映各学生在该模块各资源页面的学习程度。

如表 3 所示，从模块的被使用程度和被学生关注的程度来看（观察“浏览频次”列和“学生参与率”列），由高到低依次是课程各章节、体验区、视频资源库、问题库、资料库、常用工具库，这恰好是这些模块在课程首页由上而下呈现的顺序。说明模块在首页的布局反映了学生的学习习惯，

也可能是这种布局对于模块的被使用程度产生了影响。对于网络课程界面设计的意义则在于，要想使某个模块得到学生关注，应该将其置于首页的显著位置。

如表 3 所示，从各学生对不同模块各个资源页面的学习程度来看（观察“单个资源人均浏览频次”），由高到低依次是体验区、问题库、课程各章节、视频资源库、常用工具库、资料库。学生对体验区、课程各章节各资源页面的学习程度高说明作业的设置产生了明显的导向作用。相比之下，体验区的学习程度更高一些，每个学生对体验区的每个页面的浏览次数平均为 2.54 次，而课程各章节的单个资源生均浏览频次则在 1 次以下。产生这一差异的原因可能是：体验区采用一种模拟环境，以互动的方式让学生掌握网上学习技能，更能引发学生的学习兴趣，吸引学生反复学习。对于其他四个非作业要求的模块来说，学生对问题库的学习程度要显著高于其他三个模块，说明学生有较强的通过常见问题库来解答初入学各种疑问的需求。今后应该进一步丰富问题库，并将问题库的检索框置于首页的显著位置，以便及时、方便地解答学生的各种疑惑，帮助学生尽快进入开放教育之门。

(2) 自主学习路径分析

通过“课程模块浏览情况分析”，可以帮助辅导教师了解学生学习过程的静态信息，而进一步开展“自主学习路径分析”，则可以帮助辅导教师进一步了解学生学习过程的动态信息。



这里采用 Microsoft 顺序分析和聚类分析算法，从模块访问跳转的角度分析学生的学习路径。数据来源则是用户每天浏览课程页面产生的过程数据，如表 4 所示。

表 4 课程浏览样例表

浏览日期	学号	浏览页面 所属模块	浏览顺序号
04 27 2010	1032001200001	课程章节	1
04 27 2010	1032001200001	课程章节	2
04 27 2010	1032001200001	体验区	3
04 27 2010	1032001200001	体验区	4
04 27 2010	1032001200001	体验区	5

在构建挖掘模型时，以“浏览日期+学号”作为键值，以“浏览顺序号”作为序列键值，以“浏览模块”作为预测值来构建挖掘结构，并应用 Microsoft 顺序分析和聚类分析算法，得到如表 5 和图 3 所示结果。

表 5 是各模块直接的转换概率。在概率的计算方法上，以值“课程章节，课程章节”为例，它的概率通过“课程章节，课程章节”这一转换组合的频次除以所有转换组合的频次得来。从中可以了解到，学生在登录网络课程后，最有可能先浏览的是“课程章节”模块，其次是“体验区”模块，极少从其他模块开始浏览。

图 3 是展示各模块间跳转情况的状态转换图。这里模块间的转换概率计算方法与表 5 又有不同，这里给出的是条件概率。以“问题库-课程章节”这一转换组合为例，图中线条中标明的转换概率由“问题库，课程章节”转换组合的频次除以“问题库”这一模块出现的频次得来。所以大家会发现，在表 5 中，“问题库，课程章节”转换值的概率是 0.007；而在图 3 中，“问题库→课程章节”的转换概率是 0.23。

表 5 各模块（PAGETYPE）之间的转换概率表

变量	值	概率
PAGETYPE.转换	[开始]->课程章节	68.525%
PAGETYPE.转换	课程章节，课程章节	67.734%
PAGETYPE.转换	[开始]->体验区	29.346%
PAGETYPE.转换	体验区，体验区	16.624%
PAGETYPE.转换	课程章节，视频资源库	2.406%
PAGETYPE.转换	视频资源库，课程章节	2.241%
PAGETYPE.转换	问题库，问题库	1.988%
PAGETYPE.转换	资料库，资料库	0.845%
PAGETYPE.转换	体验区，课程章节	0.832%
PAGETYPE.转换	视频资源库，视频资源库	0.822%
PAGETYPE.转换	[开始]->问题库	0.806%
PAGETYPE.转换	问题库，课程章节	0.731%
PAGETYPE.转换	[开始]->资料库	0.632%
PAGETYPE.转换	[开始]->视频资源库	0.532%

从图 3 中可知，学生一旦进入“课程章节”或“体验区”模块，则主要是在本模块活动，而在中间几乎不去访问其他模块直至离开。“问题库”、“资料库”、“常用工具库”等学习辅助模块在课程主体部分即“课程章节”学习过程中几乎没有用到（学生偶尔从这三个模块进入，然后转入“课程章节”模块，而不是反过来）。这反映了学生学习方法的重要特点，也反映课程的链接设计还有改进的空间。

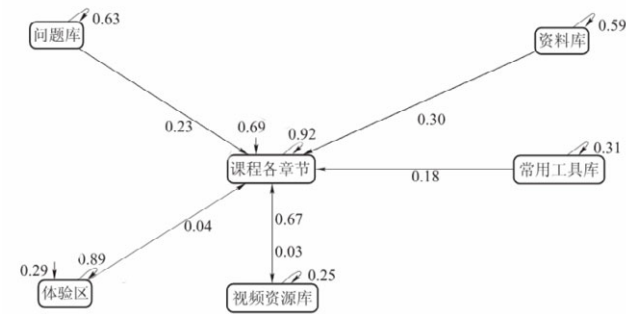


图 3 各模块的跳转情况图

### 3 学习者视角的学习分析技术应用

对于学习者而言，了解自己的行为表现及与他人的关系将是学习分析技术的主要用途。

选取中央广播电视大学主办的网络教育从业人员培训班作为研究对象<sup>[10]</sup>，该培训班依托 Moodle 平台开设（网址：<http://etutor.crtvu.cn>）。Moodle 平台具有完善的 web 日志功能，记录了用户对每个模块的访问及行为方式（浏览、增加、修改等），它的“报表”模块可让学习者了解自己一段时间的行为概况及行为。一个典型的学习者活动报告（概要报告）如图 4 所示。“活动报告”有多种呈现方式，如“概要报告”，展示一个学生在课程各章节的行为表现，“所有日志”将展示一个学生沿着时间维度的行为频次分布情况，并列出所有行为记录。基于这些行为记录，学习者可以了解自己在 Moodle 平台中各类功能模块的活动频次和活动类型。如表 6 所示，是该门课程学习过程中，师生访问平台模块的频次统计结果。

从统计结果可知，学习者可以知道，自己最常访问的模块是“讨论区”，说明课堂讨论是主要的学习活动。相比之下，“资源”模块的访问频次要低很多，这就使学习者反思，自己是否是在充分自主学习的基础上参与讨论的。

学习者在 Moodle 平台中的学习不是孤立的，他也想知道与辅导教师、同学之间的关系，这种关系即有无向关系，如对某项资源的共同关注，也有有向关系，如师生在论坛中的回复与被回复的关系。这种关系的呈现可以借助社会网络分析方法来完成，常用社会网络分析工具有 UCINET。例如，一位学习者在学完一门课程后，想了解自己在论坛中的回复和被回复情况，此时可以做一个以该学习者为中心的师生交互网络分析。



图 4 Moodle 平台中的“活动报告”界面图

师生交互网络节点由师生群体构成，成员之间的关系通过回复讨论区中不同成员所发表的言论而得以建立，属于信息传播型社会网络。在整个培训班的社会关系矩阵中，矩阵中的行和列分别代表参与网上讨论的师生， $i$  行与  $j$  列对应的值  $Z_{ij}$  表示成员  $i$  回复成员  $j$  言论的次数。其矩阵样例如表 7 所示。

表 6 某学生访问平台模块情况表

模块中文名称	模块英文名称	频次
讨论区	forum	111
课程	course	50
Wiki	wiki	33
作业	assignment	16
资源	resource	14
用户消息	user	13
上传	upload	9
博客	blog	3
标签	label	1
小组	groups	1

表 7 某培训班师生交互网络的关系矩阵（部分）表

回  成	成 复 员	48	49	88	94
48		1		2	1
49					
88		4		1	
94		4	1	6	4

采用 UCINET 中 netdraw 工具，打开如表 7 所示矩阵，可绘制以某学习者为中心的师生交互网络。如图 5 所示，即是以学习者“116”为中心的师生交互网络，箭头的指向表明“回复—被回复”关系，线条的粗细表明“回复—被回复”的频次。

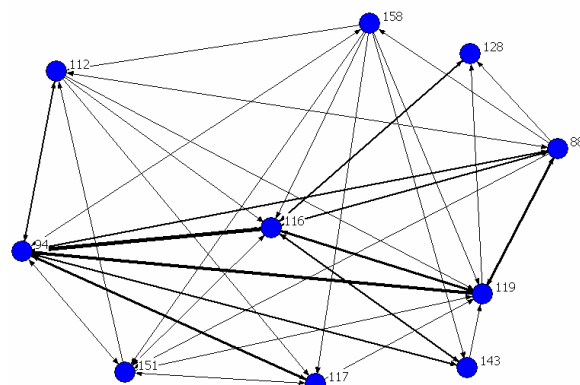


图 5 以编号“116”为中心的交互网络关系图

在 UCINET 中，沿着 Networks-Centrality-Degree 对如表 7 所示矩阵进行中心度分析，结果发现学习者“116”出度为 11，入度为 14，即学习者“116”回复了 11 人次，并得到 14 人次的回复。由图 5 可知，学习者“116”主要回复了“94”、“88”两位位成员，得到了“158”、“112”、“151”三位成员的回复，与成员“128”、“143”有较为频繁的交互。

## 五 小结

21 世纪以来，随着信息化进程在教育领域的推进，特别是数字化校园建设和网络高等教育的大力推进，教育领域已经部署了众多的学习管理系统，在这些软件系统中存储着海量的学习者信息及学习过程数据。如何利用这些数据，使这些数据转变为信息、知识，并为教学决策、学习优化服务，而不至于“淹没在数据的海洋中，却又忍受着信息的饥渴”，已成为教育工作者以及学习者所关注的内容。此时，致力于“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”的学习分析技术有助于发挥学习过程数据的价值，使数据成为审慎决策、过程优化的重要依据。面向未来，本文认为相关研究还可从以下三个方面进一步推进：（1）学习分析技术涉及多种方法和工具，有待整理和比较，以发现不同的方法和工具的适用场合。（2）应用学习分析技术可解决的问题多样，有待对学习分析技术

典型应用进行梳理,归纳出若干典型任务,并整合学习分析技术的若干方法和工具,构建学习分析模式。(3)学习分析技术的应用领域有待进一步拓展,如面向整个网络高等教育领域,这样可以发挥更大价值,既为微观决策服务,也为宏观决策服务。

## 参考文献

- [1]黄荷.今日谈:大数据时代降临[J].半月谈,2012,(17).
- [2]Siemens,G.1<sup>st</sup> International Conference on Learning Analytics and Knowledge 2011[EB/OL].  
<<https://tekri.athabasca.ca/analytics/about.>>
- [3]Johnson,L.,Adams,S.,and Cummins,M.(2012).The NMC Horizon Report: 2012 Higher Education Edition.Austin,Texas:The New Media Consortium.
- [4]Baeppler, P. & Murdoch, C. J.(2010). Academic Analytics and Data Mining in Higher Education. International Journal for the Scholarship of Teaching and Learning, 4(2). 170-178.
- [5]Chen,E.,Heritage,M.&Lee,J. Identifying and Monitoring Students' Learning Needs With Technology[J].  
Journal of Education for Students Placed at Risk,2010 (3) :309-332.
- [6]Romero & Ventura. Educational Data Mining:A Survey from 1995 to 2005[J]. Expert Systems with Applications.2007 , (33):125-146.
- [7]Romero, C.,Ventura, S.,& Garcia, E. (2005). Data mining in course management systems: Moodle case study and tutorial. Computers & Education,51(1),368-384.
- [8]葛道凯,张少刚,魏顺平.教育数据挖掘:方法与应用[M].北京:教育科学出版社,2012:29-30.
- [9]顾小清,张进良,蔡慧英.学习分析:正在浮现中的数据技术[J].远程教育杂志,2012,(1):18-25.
- [10]魏顺平.在线学习行为特点及其影响因素分析研究[J].开放教育研究,2012,(4):81-90.
- [11]魏顺平.Moodle 平台数据挖掘研究——以一门在线培训课程学习过程分析为例[J].中国远程教育,2011,(1):24-30.

## Learning Analytics: Mining the Value of Education Data under the Big Data Era

WEI Shun-ping

(Institute of Open and Distance Education, the Open University of China, Beijing 100039, China)

**Abstract:** Currently, the big data era has arrived and the field of education also has accumulated huge amounts of data. In the field of education, more and more software systems have been deployed and there stores massive educational data. How could we fully use the massive educational data and transfer the data into useful information and knowledge in order to make scientific educational decision and optimize instruction has become the content concerned by educators and learners. Learning analytics could be used for realizing the value of learning process data. In this paper, the author introduces domestic and foreign current studies of learning analytics, summed up the key technology and analytical model of learning analytics, and with three examples respectively from different users perspective including administrators, tutors and learners, demonstrates application process of learning analytics in analyzing online learning process.

**Keywords:** big data era; education data; learning analytics; key technology; analytical model

\*基金项目:本文系中央广播电视大学2012年课题“学习分析技术及其在远程开放教育中的应用研究”(项目编号:Q0012A-Y)的研究成果。

作者简介:魏顺平,博士,国家开放大学(中央广播电视大学)现代远程教育研究所副研究员,研究方向:教育数据挖掘研究。

收稿日期:2012年12月27日

编辑:宋树