

# 高等教育大数据分析:机遇与挑战

李馨

(中央电化教育馆,北京 100031)

**[摘要]** 高校师生网上互动产生了大量有用的数据痕迹,这些数据可以用来帮助师生更好地适应教与学。数字革命与新技术的发展,改变了高等教育的教学模式,高等教育机构也因此面临着日益复杂的挑战。本文探讨了大数据解决教与学问题的潜能,提出了高等教育大数据分析的概念框架和数据框架;从教师、管理者和学生三种用户和描述性、预测性和规定性分析等方面概述了实施高等教育大数据分析的机遇,提出了大数据有助于高等教育机构产生良好的绩效和过程结果,同时阐明技术、数据存储、隐私安全等方面存在的挑战,提出了高等教育大数据分析未来的发展和实施方向。

**[关键词]** 大数据分析;高等教育;知识发现和数据挖掘;数据仓库;学习分析

**[中图分类号]** G434

**[文献标识码]** A

**[文章编号]** 1007-2179(2016)04-0050-07

过去几十年里,无处不在的计算设备、灵活的课堂设计和大规模开放式网络课程等,从根本上改变了学和教的模式和可访问性,明显地影响了高等教育的学术研究和教学。高等教育正在接受新的类格式和技术,以满足不断发展的学生需要(Mok, 2005)。本文首先阐述了影响高等教育机构发展的全球趋势,探讨了潜在的大数据分析;其次,概述了高等教育中实施大数据管理的机遇和挑战;最后,剖析了高等教育机构采用大数据分析的发展和实施方向。

## 一、知识发现、数据挖掘与大数据解读

### (一)知识发现与数据挖掘

知识发现(Knowledge Discovery,简称KDD)是人工智能、机器学习、模式识别、统计学、数据库和知识库、数据可视化等众多学科相互交叉、融合形成的新兴且有广阔前景的领域,是从大型数据集中识别和提取有用、有意义部分的方式方法。数据挖掘是应用数据分析和发现算法产生特定计数模式分析数据,是知识发现的主要方面。多年来,聚类、关联、分类算法、回归模型、预测方法和因子分析等关键方法,已经在数据挖掘研究中占主导地位,实现的功能

有数据总结、分类、聚类、估计与预测、关联和序列发现(Baker & Yacef, 2009; Romero & Ventura, 2010)。

### (二)大数据解读

#### 1. 大数据的出现

2012年,联合国发布的大数据白皮书《大数据促发展:挑战与机遇》指出,大数据时代已到来,它的出现将会对社会各领域产生深刻影响(徐鹏等, 2013)。许多组织正在充分利用数据,以更好地决定它们的战略和业务方向。利用数据做出决策并不新颖,20世纪90年代初,很多企业就已对大量数据进行分析 and 存储。研究发现,大部分企业运用的是结构化数据(数量和类型)(Basu, 2013)。然而,这不能反映现有企业数据未被开发的复杂商业价值。根据国际商业机器公司(IBM)数据组织统计,80%数据是非结构化的,如文本、视频、音频、图像和图表格式等。分析工具是把这种复杂的数据转换成有意义的模式和价值。

#### 2. 大数据相关概念

大数据基本上描述为太大的、超过传统数据库系统处理能力的数据库。它还包括创新的技术,以捕捉、存储、分发、分析和处理不同结构的数据集。批评者质疑,“大”这个词本身有误导性,它不是反映

[收稿日期] 2016-05-27

[修回日期] 2016-06-27

[DOI编码] 10.13966/j.cnki.kfjyyj.2016.04.007

[作者简介] 李馨,副编审,中央电化教育馆,《中国电化教育》杂志社副主编/副社长,研究方向:信息化教学的理论和实践(geri521@sina.com)。

数据大小,而是指其复杂性。研究者指出,大数据的定义不是指数据本身,而是指突发的适应技术,能以较快的速度处理各种类型的数据(Yang,2013)。一般来说,大数据有多个基本特性,其中关键特性如下:

1)信息量大,往往在存储、处理、传输、分析和保存上有挑战性。

2)传输速度与提高组织内信息流的速度有关(例如,组织处理财务信息且有能力处理)。

3)数据的准确性指数据的偏差、噪声和异常。它着眼于数据如何被存储和有意义地挖掘,也包括信任和不确定问题。

4)形式不同的数据有不同的格式,如结构化和非结构化。

5)验证指数据的验证和安全。

6)最重要的是,组织内的数据被用来产生有价值的见解、利益和业务流程等。

也有研究者归纳大数据最常见的性能,如数据规模的扩大、产生的增长率、表示的累积范围及数据的有效性(准确性和波动性)(Douglas,2001)。除上述大数据特性外,高等教育机构还需要考虑解锁大数据价值的措施,包括数据采集、数据分析、可视化和应用、数据仓库。

#### 1)收集

数据收集是大数据价值累积的第一步。确定数据,可以揭示有用和有价值的信息。数据必须经过过滤,因为绝大多数数据不可用。

#### 2)分析

数据要呈现为可用的形式,必须加以分析,以产生可操作的信息。但是,随着数据性质的日益多样化,管理和分析不同的数据集变得复杂,需要连接不同的数据集,以掌握大量的信息,并通过这些数据进行传达。因此,这又被称为“复杂”的大数据。

#### 3)可视化及应用

可视化及应用属于大数据应用最后阶段,即分析数据并进行可视化显示,然后被解释和集成现有流程的形式供用户使用,最终用于指导决策。

#### 4)数据仓库

海量的数据正被收集和存储在各种机构的数据仓库中,比如在线知识库、教育数字图书馆以及相关的工具中。所有这些数据都可以被用来作为改变实践的重大催化剂。数据如被存储在学生信息系统,就

是一个关于学生社会化媒体使用、学习管理情况、学生图书馆使用情况、个人计算机和学习途径等的信息管理系统。

数据以不同格式(如文本、图片、音频、视频)不断增长。当学习者与数字设备进行交互时,这种相互作用的数据可以很容易地被捕获或“记录”,并供后续分析。

在教育领域,工具和数据的数据分析和处理技术很有吸引力。增加注意力分析也推动了计算能力的进步。例如,智能手机已超过了台式电脑的计算能力,功能比大型机计算机强大,可以完成几年前不可能完成的任务。计算能力的提升,能支持研究人员分析大量的数据。此外,新的系统和分析工具的出现是数据分析研究最重要的进步。

## 二、高等教育大数据分析

大数据正使人类的思维、生活和工作发生变革,也对整个教育系统产生极大冲击,逐渐成为推动教育系统改革和创新的颠覆性力量。有学者探讨了教育变革的趋向、大数据时代背景下教育模式的转变、学习方式的变革、教育研究的新范式等(Long & Siemen,2011)。还有学者探究了教育领域有普遍使用价值的大数据技术,如教育数据挖掘、学习分析等(杨现民等,2016)。

### (一)教育大数据的含义

教育大数据有狭义和广义两种。狭义的教育大数据指学习者的行为数据,主要来自于在线学习平台、学生管理系统和课程管理平台等;广义的教育大数据指来源于日常教育活动中人类所有的行为数据(胡弼成等,2015)。

### (二)教育大数据的构成

总体而言,教育大数据可分为四个层面,分别是基础层、状态层、资源层、行为层。基础层是存储基础性的国家教育数据,状态层存储的是各种教育环境、教育装备以及教育业务的运行状态信息,资源层存储的是教育过程中建立或生成的各种教学资源,行为层存储的是与广大教育者相关用户(教研员、教师、学生、教育管理者等)的行为数据。

### (三)高等教育大数据分析框架

大数据可以影响高等教育实践,丰富学生的经验,提高学术研究水平,以有效的证据为基础做出战

略决策,将复杂化、非结构化数据转化为可操作的信息。有研究者指出,大数据提供了具有成本效益的前景,有利于促进个性化学习,提高教育决策(姜强等,2015)本研究建立一个概念性研究框架,来描述高等教育大数据的四部分组成(见图1)。

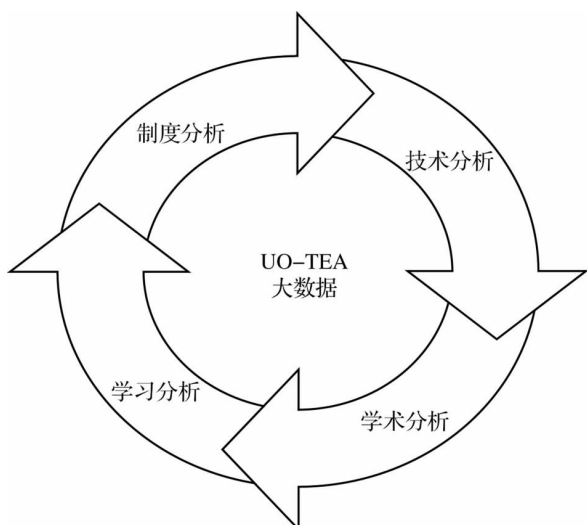


图1 高等教育大数据概念框架

### 1. 制度分析

制度分析包括评估政策分析、教学分析、结构分析,主要利用报告、数据仓库和数据仪表盘,为机构提供及时的数据驱动决策能力。

### 2. 信息技术分析

该分析有助于监测开发或部署技术,开发数据标准、工具和流程,组织协同和政策落实。它旨在整合各种系统的学生信息数据、学习管理和校友录系统,以及系统管理课堂外学习经验。分析结果可以用来制定严格的数据建模,以便学生访问和提高系统的可用性,并评估教学干预,揭示学习障碍。学习系统可以通过细化相关业务流程帮助收集关键数据,如能捕获不被显示在单独系统中的数据可能会非常有用。

### 3. 学术分析

学术分析是未来高等教育管理、资源分配重要组成部分。它封装了所有学习活动进程(Tulasi, 2013),提供了具体的方案和如何应对挑战的信息系统;反映了在制度层面数据分析的作用,以及学习分析中的学习过程(包括分析学习者、内容、制度和教育家之间的关系)(Long & Siemen, 2011)。它结合了大数据集的统计技术和预测模型。所提供的数

据,可供管理人员使用,支持战略决策过程以及与其他机构分析对比。学术分析有利于制定学习规划,以有效的方式衡量、收集、解释、报告和分享数据,从而使学术规划和学生的长处、弱点相结合。

### 4. 学习分析

随着高等教育机构越来越多地采用混合式教学方法,基于网络环境和平台的学习越来越多。教育数据挖掘社区和学习建模社区由此可以探索跟踪学生行为记录的变量,如点击次数和耗费在网页上的时间,以及越来越微妙的信息如概念的柔性和记忆。具体的行为数据包增加了越来越多与学生相关的信息库。学习分析是个新兴的领域,目的是访问和理解其中的行为数据。新媒体联盟(New Media Consortium)将学习分析定义为:利用松散耦合的数据分析技术与收集工具,研讨分析学习者的学习参与、学习表现和学习过程的相关数据,从而对课程、教学和评价进行实时修改(王良周等,2015)。学习分析关注测量、采集、分析和对学习及其背景的数据报告。更广泛地说,学习分析软件和技术通常用于改善工作流程,测量教育机构的数据,并提高机构运行的有效性。

通常情况下,学习分析与学术分析紧密相关。学习分析是在机构开展的教和学活动中进行的,很大程度上有助于学习者学习的成功(Jones, 2012)。与学习分析相关的是教师分析。有学者扩展了学习分析范围,认为其包括教学分析,如利用数字图书馆和在线资源,分析教师的网上行为;利用教育数据挖掘技术,确定不同的教学架构和不同的在线行为(Xu & Recker, 2012)。

图2展示了彼此链接的数据系统。该系统描述

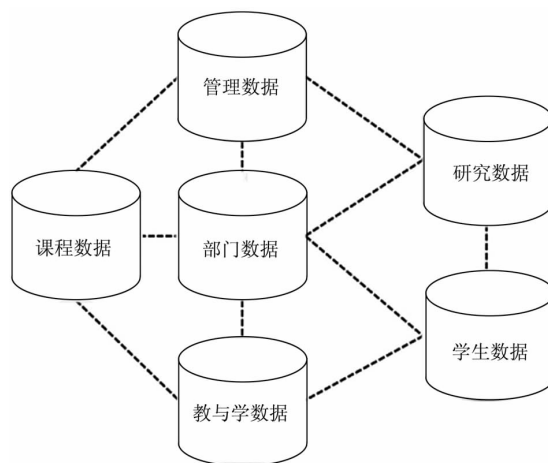


图2 理想的数据框架

了行政和业务数据的采集过程,针对教学中出现的问题进行研究等。

### 三、机遇和挑战

#### (一)机遇

大数据分析可以改变现有的管理、教学、学习、学术工作,并帮助解决高等教育面临的挑战;为个别学生提供预测工具,以提高其学习成绩,并确保学术课程符合高品质标准;为学生定制学习过程每一步模板,有效满足学生需求;收集作业及课程中数据,形成学习知识树,将产生更好和更丰富的学习反馈。

高等教育可以利用大数据方法分析个别学生的技能水平,推进个性化学习。有效的大数据,可以帮助教育机构丰富教学管理经验,提高学生成绩,减少辍学率和增加毕业生数(见图3)。

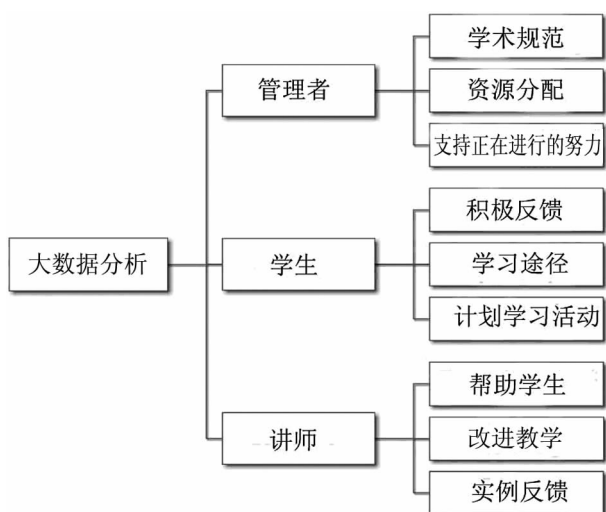


图3 高等教育大数据为三种用户提供机会

大数据分析的主要贡献,将取决于应用程序的三种数据模型(描述、关系和预测)和实用程序,以更好地引导决策(见图4)。

#### 1. 描述性分析

描述性分析的目的是描述和分析学生、教学、研究、政策和其他行政过程。目标是识别模式样本,如学生报到人数、毕业率等。描述性分析可以为高等教育机构提供互动数据,为教学和研究发现趋势和机会,促进师生就当前和未来的重要问题进行对话。具体而言,运用描述性分析,比如通过登录次数、页面视图、访问次数、完成率、内容被

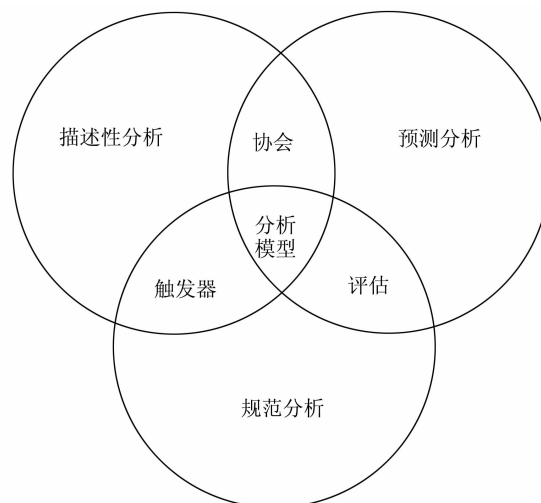


图4 高等教育数据分析模型

访问次数等,教育机构可以通过学习管理系统调查了解学习者。

#### 2. 预测分析

预测分析可以为机构提供更好的决策和基于数据的可操作见解。目的是通过寻找趋势和确定相关问题的关联,预测未来的风险或机会,估计未来学习事件发生的可能性。它可以揭示描述性模型中隐藏的数据关系,如人口统计和完成率;帮助学期初表现出风险行为的学生,使他们不至于弃学或不及格;帮助教师预测课程完成率。

#### 3. 规定性分析

规定性分析有助于高等教育机构评估现状,在有效和一致的预测基础上做出明智的选择。规范分析使决策者了解未来关键任务流程,发现机会(问题)以及提出好的行动路线。

总之,大数据分析为高等教育机构提供利用现有数据和收集丢失的数据,帮助决策者针对不同的结果做出决策(见图5)。

大数据有助于高等教育机构产生如下绩效:

#### 1) 绩效结果

- ①更好地了解教育机构;
- ②更好地理解被分析的有效数据;
- ③更好地利用大数据的坚实基础;
- ④更好地改进标准化和简化的数据处理;
- ⑤更好地利用数据分析,提高理解精度并实施决策;
- ⑥更好地推动数据驱动决策和实践;

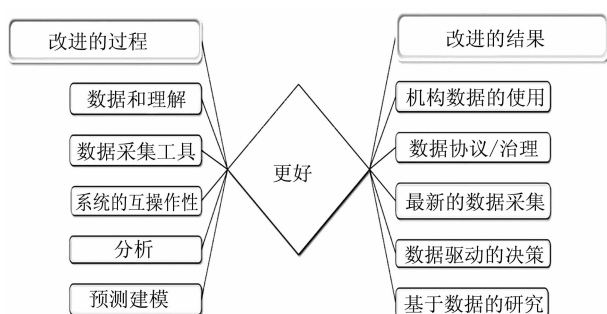


图5 大数据分析结果

⑦更好地检验假设、网络试验、场景建模、模拟、敏感性和数据挖掘。

#### 2) 过程结果

- ①开发数据收集、处理、分析和解释的工具；
- ②促进数据系统的互联互通和系统连接；
- ③建立数据分析和预测模型；
- ④实时绘制分析学生和教师的表现；
- ⑤提供部门之间可靠和可比的性能指标；
- ⑥利用历史数据作出明智的决定；
- ⑦探索、开发、利用数据，预测可能的教学结果。

#### (二) 挑战

首先，技术的挑战。诚如邬贺铨(2013)指出：“从目前看，大数据技术的应用仍存在挑战和困难，主要表现在大数据挖掘的数据收集、数据存储、数据处理和结果的可视化呈现四个环节中，使结果更加直观，便于观察。”

其次，数据存储技术及应用于数据分析和处理的技术挑战，包括超级计算机算法技术、计算机硬件数据处理能力等。使用收集、存储和开发的算法挖掘数据，往往耗时又复杂，投入巨大。第一，大多数机构的数据系统是不可互操作的，因此聚合管理数据以及课堂和在线数据可能带来额外的挑战。第二，数据的兼容性挑战中，最突出的是数据集成，不同的数据存储系统数据格式和编码不统一，造成在不同系统间的数据，特别是不同来源的结构化和非结构化数据共享困难。此外，数据清洗时，进行整合的结构化与非结构化数据可能丢失。大数据分析产生的信息质量完全依赖于所收集的数据质量和所使用的指标，缺乏标准化的措施和指标，可能导致收集和报告数据质量存在很大挑战。第三，数据采集与问题解决分析技术的挑战。学者们一致认为，大数

据的价值会越来越凸显，能为后续政策和战略的提出贡献更多力量(Schleicher, 2013)。还有学者指出，可以倡议高等教育机构跟企业合作增加大数据的应用，帮助所有群体迎接学习过程所表现的数据持久性、所有权等的挑战(Wagner & Ice, 2012)。也有学者认为，学习分析应该是透明和灵活的，教育工作者容易获得的(Dringus, 2012)。

另外，大数据在高等教育机构也存在挑战。美国《通过教育数据挖掘和学习分析促进教与学》报告的受访专家指出，可以通过基础设施建设和技术研发克服大数据教育应用中所面对的技术挑战，但是大数据教育应用中80%以上的限制和挑战来源于“人”。比如，在如何提取数据、什么数据可用和哪些数据必需，以及如何更好地使用数据之间存在鸿沟，所有这些使教育机构间进行合作产生困难。有学者也承认沟通这些想法的重要性，指出教育规划应分析数据负面或中性的影响(Macfadyen & Dawson, 2012)。他们主张研究社会学习技术，确保激发参与者协同学习的动力，激励组织采纳大数据，形成应用数据的文化。美国教育部报告认为，高等学校大数据的成功实施取决于机构各部门之间的协同行动。例如，服务部门参与、计划的数据收集和使用被认为是至关重要的。

最后，隐私与伦理道德限制的挑战。如灾难恢复计划、强大的密码策略、防火墙、加密和防病毒软件技术等，对减少损失或操纵大数据风险的成效仍在调查中。此外，高等教育的数据保护、隐私风险和安程序等许多制度还缺失。例如，大多数高等教育机构似乎都有规范和管理知识产权的政策，保护数据隐私和控制对数据的访问，但这些政策可能不足以应对那些因数据使用而产生的伦理道德的挑战(Slade & Prinsloo, 2013)。比如，使用的数据量的不断增加促进了良好的学习管理、可靠数据仓库和管理，灵活而透明的数据挖掘和提取，及准确而负责的报告的形成，但这在很大程度上又会增加学生活动的透明性，带来一系列社会伦理道德问题，需要引起相关教育机构的重视(赵慧琼等, 2016)。

## 四、总结与展望

### (一) 未来前景

未来高等教育的趋向是个性化学习。在高等教

育个性化时代,课程材料将适应性地满足学生学习的个性化需求,电子教育、网络教育、主动学习是其显著特点。目前教育机构积累了大量非结构化和结构化数据,利用教育数据挖掘技术,能使研究者有更多机会研究学生的学习。根据数据挖掘信息,高等教育机构可以建构并利用教育大数据库,帮助教师理解学生,调整教育方案,提供个性化学习模式。

## (二)发展方向

数据采集、处理、传输和存储能力的进步,使许多组织将数据集成到数据仓库。高等教育的数据在不断增长,但大部分分散在各部门,难以检索或巩固。为了有效利用这些数据,分析不同信息集的能力是必要的。管理和控制数据的同时保护敏感信息的数据库,是高等教育大数据处理的关键要求。换句话说,数据仓库是一种解锁高等教育大数据价值的有效方法。虽然结合各种无关系系统中的数据集非常困难,但它提供了更全面的见解,必然导致预测模型能力的改进(徐鹏等,2013)。

此外,数据仓库实行标准化的数据格式。每个部门都产生符合标准的数据格式,从而使数据表示更准确。数据仓库可以存储大量的历史数据,可以容易地实验和分析不同的时间周期和趋势,使未来的预测更准确。大数据能够提供新见解,例如,学习分析作为高等教育大数据的组成部分,为研究人员实时分析学习提供机会。学习分析尽管有很大的不确定性,但它的持续增长意味着需要考虑的不仅是大量的机会,还要探讨伦理挑战。探索高等教育大数据的管理和治理工作仍在进行中。本研究希望发展高等教育大数据分析的基础理论,制定关键绩效指标,提出采集、处理和显示数据以及开发诊断工具的数据分析框架。

未来的工作将涉及确定和制定政策,详细说明如何对机构的数据和信息负责,包括其准确性、可访问性、一致性、完整性和可维护性。政策在制定过程中,需要考虑数据和信息如何被存储、归档、备份和保护以及开发使用的数据标准和规范,对如何使用数据和信息的授权人员则需要实施一套审核控制程序,以确保符合政府法规和工业标准。伴随着大数据时代来临,社会科学领域开始从关注宏观整体到关注微观个体,对于教育领域来说,大数据应用让专门培养和针对性训练成为可能,让研究个体学习成

为可能,从而比任何时候都更能促进人类自身的发展。

## [参考文献]

- [1] Baker, R. S. J. D. ,& Yacef, K. (2009). The state of educational data mining in 2009:A review and future visions[J]. *Journal of Educational Data Mining*,1(1):3-17.
- [2] Basu, A. (2013). Five pillars of prescriptive analytics success [J]. *Analytics*, (3):8-12.
- [3] Douglas, L. (2001). 3D data management: controlling data volume, velocity and variety. [EB/OL]. [2013-12-30]. <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>.
- [4] Dringus, L. P. (2012). Learning analytics considered harmful [J]. *Journal of Asynchronous Learning Networks*,16(3):87-100.
- [5] 胡弼成,王祖霖(2015).“大数据”对教育的作用、挑战及教育变革趋势——大数据时代教育变革的最新研究进展综述[J]. *现代大学教育*, (4):98-104.
- [6] 姜强,赵蔚,王朋娇,王丽萍(2015).基于大数据的个性化自适应在线学习分析模型及实现[J]. *中国电化教育*, (1):85-92.
- [7] Jones, S. (2012). Technology review: The possibilities of learning analytics to improve learner-centered decision-making[J]. *Community College Enterprise*, 18 (1):89-92.
- [8] Long, P. ,& Siemen, G. (2011). Penetrating the fog: Analytics in learning and education[J]. *Educause Review*, 46 (5):30-40.
- [9] Macfadyen, L. P. , & Dawson, S. (2012). Numbers are not enough. Why e-learning analytics failed to inform institutional strategic plan[J]. *Educational Technology & Society*,15(3):149-163.
- [10] Mok, K. H. (2005). Fostering entrepreneurship: Changing role of government and higher education governance in Hong Kong [J]. *Research Policy*, 34 (4):537-554.
- [11] Romero, C. R. ,& Ventura, S. (2010). Educational data mining: A review of the state of the art [J]. *IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews*, 40 (6): 601-618.
- [12] Slade, S. ,& Prinsloo, P. (2013) Learning analytics: Ethical issues and dilemmas [J]. *American Behavioral Scientist*, 57 (10): 1509-1528.
- [13] Schleicher, A. (2013). Big Data and PISA. [EB/OL]. [2013-08-04]. [http://oecdeducationtoday.blogspot.co.nz/2013/07/big-data-and-pisa.html? m = 1](http://oecdeducationtoday.blogspot.co.nz/2013/07/big-data-and-pisa.html?m=1).
- [14] Tulasi, B. (2013). Significance of Big Data and analytics in higher education [J]. *International Journal of Computer Applications*, 68 (14):23-25.
- [15] Wagner, E. ,& Ice, P. (2012). Data changes everything: Delivering on the promise of learning analytics in higher education [J]. *EDUCAUSE Review*,7 (8):33-42.
- [16] 王良周,于卫红(2015).大数据视角下的学习分析综述

- [J]. 中国远程教育, (3):31-37.
- [17] 邹贺铨(2013). 大数据时代的机遇与挑战[J]. 求是, (4): 47-49.
- [18] Xu, B., & Recker, M. (2012). Teaching analytics: A clustering and triangulation study of digital library user data[J]. *Educational Technology & Society*, 15 (3):103-115.
- [19] 徐鹏,王以宁,刘艳华,张海(2013). 大数据视角分析学习变革——美国《通过教育数据挖掘和学习分析促进教与学》报告解读及启示[J]. *远程教育杂志*, (10):11-17.
- [20] Yang, L. (2013). Big Data Analytics: What Is the Big Deal? [EB/OL]. [2013-12-30]. <http://knowledge.ckgsb.edu.cn/2013/12/30/technology/big-data-analytics-whats-big-deal/>.
- [21] 杨现民,唐斯斯,李冀红(2016). 发展教育大数据:内涵、价值和挑战[J]. *现代远程教育研究*, (1):50-61.
- [22] 赵慧琼,姜强,赵蔚(2016). 大数据学习分析的安全与隐私保护研究[J]. *现代教育技术*, (3):5-11.
- (编辑:李学书)

## Big Data Analytics in Higher Education: Opportunities and Challenges

LI Xin

(National Center for Educational Technology, Beijing 100031, China)

**Abstract:** Higher education is operating in an increasingly complex and competitive environment. It is under increasing pressure to respond to national and global economic, political and social changes such as the growing need to increase the proportion of students in certain disciplines, embedding workplace graduate attributes and ensuring that the quality of learning programs are both nationally and globally relevant. In addition, different stakeholders expect higher education institutions respond to these demands in a timely manner, albeit with declining government funding, declining support from business and private sectors, growing regulatory demands for transparency and accountability, and declining admissions rates due to increasing tuition, upsurge in high schools dropout and increasing operational costs. How can higher education respond effectively and timely to global changes that affect their environment? The decisions required for dealing with these rapid changes are complex and many are made without recourse to vast data sources that have been generated but are not available to those entrusted to make relevant and timely decisions. These data can play a major part in how we understand the often contested nature of higher education governance to ensure that institutions are not only able to respond effectively to changes happening within and outside of them, but also remain pertinent to their purposes in the societies that they serve. This paper first identified the key global trends affecting higher education and explored the potentials of Big Data and Analytics in addressing these changing trends. Secondly, the paper outlined opportunities and challenges associated with the implementation and governance of Big Data in higher education. The paper concluded by outlining future directions relating to the development and implementation of institutional project on Big Data. The paper also opened up new research areas that can be explored to enrich our understanding of the role of Big Data in higher education.

**Key words:** big data analytics; higher education; knowledge discovery and data mining; data warehousing; learning analytics