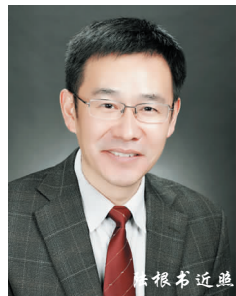


# 大数据在高等教育领域中的应用 及面临的挑战



陆根书

(1. 西安交通大学 人文学院, 西安 710049; 2. 西安交通大学 中国西部高等教育评估中心, 西安 710049)

**摘要:**对大数据概念与内涵的界定主要有3种方式,即生产导向的方式、过程导向的方式和认知导向的方式。大数据在高等教育领域的应用对完善学校规划、促进学校发展,感知教学现实、提升教学效能,优化学习经历、提高学习质量以及促进科学研究、推动跨学科发展具有重要意义。教育大数据分析主要采用两种技术,即教育数据挖掘与学习分析。在高等教育的教与学领域中,这两种技术常用的方法包括分类与预测、聚类以及异常值检测等。大数据在高等教育领域应用面临认识论、方法论、审美、技术、法律和伦理以及政治和经济的两难挑战。

**关键词:**大数据;教育数据挖掘;学习分析;高等教育

[中图分类号]G646 [文献标志码]A [文章编号]16738012(2022)04003408

## 一、大数据的概念与内涵

大数据是当今学术界谈论的一个时髦名词,许多人都在通过电子邮件、社交媒体、金融交易以及教学、科研、医疗和商业平台等各种方式,以文本、图像、音频和视频等形式或以这些形式组合生成的数据为大数据做着贡献。大数据常被应用于不同的方面,如社会现象、基础设施、信息资产、数据集以及数据的存储、加工与分析技术等。因此,对大数据概念与内涵的界定在学术界并没有达成共识,大致可以分为如下3种不同的界定方式<sup>[1]</sup>:

一是生产导向的方式,强调大数据的各种信息特征,特别是它的大容量、高生产速度和多样性等特征。比较流行的大数据定义大多属于这一类。2001年,Laney在讨论企业如何应对21世纪初电子商务的发展面临的数据管理挑战时,曾提出了一个包括数据容量(volume)、速度(velocity)和多样性

修回日期:20220525

基金项目:陕西本科和高等继续教育教学改革研究重点攻关项目“高校毕业生高质量就业创业研究”(21ZG004)

作者简介:陆根书,男,江苏溧阳人,西安交通大学人文学院教授,博士生导师,中国西部高等教育评估中心主任,主要从事教育经济学、大学学习理论和高等教育评估研究。

引用格式:陆根书. 大数据在高等教育领域中的应用及面临的挑战[J]. 重庆高教研究,2022,10(4):3438.

Citation format:LU Genshu. Applications and challenges of big data in higher education[J]. Chongqing higher education research, 2022,10(4):3438.

(variety)的三维分析框架,并呼吁企业在作战略决策时需要采取新的分析架构和解决方案<sup>[2]</sup>。虽然这一分析框架没有明确提到大数据,但后来被称之为大数据定义的“3V”模型。之后,一些学者又增加了价值(value)和准确性(veracity/validity)等特征。其中,容量是指随着海量数据的产生和收集,数据的规模变得越来越大<sup>[3]</sup>。这是大数据最明显的一个维度,因为有了大量的数据才产生了分析和处理的需求。在处理大数据时,存储和分析数据的速度也是一个关键因素,因为要在尽可能短的时间内或实时地创造最大价值。多样性维度是指数据来源与格式的多样化。除了高度结构化的数据外,还包括一些半结构化以及音频、视频或文本文件等非结构化的数据。数据的多样性和大容量使得大数据分析更具挑战性。价值维度是指数据能够给组织增加的价值,即通过从数据中识别、转换和提取信息,从而为业务增加价值。准确性是指数据统计的一致性和数据的可信性,包括数据的来源、采集和处理的方法以及信息基础设施是否便捷、可靠等。

二是过程导向的方式,强调处理大量数据背后的技术要求。例如,大数据采集、储存、处理和 analysis 的方法以及大数据系统基准分析等<sup>[4]</sup>。美国国家标准与技术研究院(NIST)在定义大数据时就强调了对高效存储、处理和分析数据的可扩展架构的需求<sup>[5]</sup>。有研究者将大数据与某种阈值的跨越联系起来,认为当数据超过了传统数据库系统的处理能力时,它就是大数据,因而需要用不同的方式处理它。在这种意义上,大数据是指任何一组大到或复杂到传统信息技术无法处理的数据,是不能用常规方式存储、不能存储在单个服务器中的庞大的或非结构化的数据。

三是认知导向的方式,关注大数据对人类认知能力的挑战以及大数据对经济社会发展的影响。由于大数据之间的相互作用非常复杂,人类可能因认知能力限制而无法完全理解它们。Boyd 和 Crawford 认为,大数据不仅数据量大,而且涉及搜索、汇集和交互引用大型数据集的能力<sup>[6]</sup>。他们把大数据定义为一种文化、技术和学术现象,认为大数据依赖于技术的相互作用(如最大化计算能力和算法精度)与分析(如识别大数据的形态),也是一种神话(如相信大数据能够提供一种拥有事实、客观性和准确性光环的更高形式的智力)。

由此可见,对大数据的界定有不同的方式,仅仅从其中任何一个单独的视角来理解大数据可能是不全面的。笔者认为,大数据是具有大容量、高速度和多样性等特征的一种信息资产,要让这种信息资产产生经济和社会价值,需要采用特定的技术和分析方法。

## 二、大数据在高等教育领域的应用及其重要意义

随着大数据的发展,其在高等教育领域的应用也日渐广泛,正成为高等教育领域快速发展的研究前沿。在中国知网以主题词“大数据”加“高等教育”对2010—2021年的研究文献进行检索,并在英文数据库 ScienceDirect 以“big data”和“higher education”为主题对同期英文文献进行检索,可以了解国内外大数据在高等教育领域的一些应用情况。虽然中英文文献近10年来均有快速增长,但相比于英文文献,中文文献的数量相对较少。在中国知网中,2011年才出现有关研究文献,且2011年和2012年的文献数量很少,从2013年开始有较快增长,到2019年达到450篇,之后趋于下降;英文文献比中文文献多很多,2010年在 ScienceDirect 中有3106篇,之后其数量一路上升,到2021年达到14725篇。

大数据在高等教育领域的应用涉及许多方面<sup>[7]</sup>。例如,大数据可以提升高校理解复杂环境的能力,为其识别早期发展风险、有效处理各种复杂问题、创新体制机制、改善决策和资源管理、改进学校规划和提升管理与决策效能提供帮助;可以持续跟踪、监测和评估教师和学生的教与学状态,并通过有针对性的反馈帮助教师和学生改进教与学的策略,为提高教育教学质量和优化学生的学习经历提供帮

助;可以为高校科学研究尤其是跨学科研究的发展提供重要机遇,为提升科研生产率等提供重要支持。

### (一)完善学校规划,促进学校发展

大数据可以帮助高校通过追踪和分析本校与其他高校在一些关键指标上的差异,用证据来说明学校在哪些方面取得了成功,哪些方面还需要改进,进而更好地评估自己的优势、劣势、机会和挑战,明确学校办学目标和人才培养目标及其定位,规划发展战略,更加积极主动地追求进步,促进学校发展。

### (二)感知教学现实,提升教学效能

大数据可以帮助学校更好地了解教师和课程教学的实际,进而更好地理解、分析和预测教师教学的状态。大数据还可以让教师获得即时、客观的反馈,对其课程及教学过程的有效性进行评估,了解教学过程中的优点和不足,进而为教师改进课程与教学提供支持。大数据还可以帮助管理者和教师监控学生的学习过程,跟踪他们的学习参与、师生互动和合作学习等情况,判断他们成功与失败的可能性,并及时回应他们的需求,为其提供形成性评价,进行个性化指导。这为提升教学效能提供了可能性。

### (三)优化学习经历,提高学习质量

大数据及其基础设施的发展,可以为师生提供更加适宜、高效和有效的教与学环境,为师生交流、知识获取与创造提供更便利的条件。学校可以应用大数据跟踪、分析学生的学习过程,利用大数据挖掘分析方法获得有意义的信息,为优化学生的学习经历、提高学习质量、促进学生发展提供支持。

### (四)促进科学研究,推动跨学科发展

大数据的出现也引发了科学研究范式的变革。从历史发展看,科学研究最早以描述自然现象的实验科学(experimental science)为基础,之后出现了探求各种规律与定理的理论科学(theoretical science)。为了解决实际问题,之后又出现了计算科学(computational science),通过仿真模拟各种复杂现象以寻求解决方案。大数据的出现则催生了第4种研究范式,即数据密集的科学(data-intensive science),研究者直接从许多不同来源的大数据中挖掘所需的信息与知识<sup>[8]</sup>。大数据除了推动科学研究范式的变革和人们思考方式的更新之外<sup>[9]</sup>,还促进了跨学科研究的发展与应用。例如,推动了以大数据为研究对象,综合应用信息科学、数学、社会科学、系统科学、心理学和经济学等学科理论与方法,从数据中提取知识的跨学科的学科——数据科学的发展。

由于大数据在高等教育领域的应用十分广泛,本研究主要讨论两种教育大数据分析技术,即教育数据挖掘和学习分析在高等教育的教与学领域中的应用情况,以及大数据在高等教育领域应用面临的一些两难挑战。

## 三、教育数据挖掘与学习分析方法及其应用

在高等教育领域,教育大数据分析主要采用两种技术,即教育数据挖掘和学习分析。对这两者的关系,不同的学者有不同的解读。有的认为学习分析是教育数据挖掘的一个分支,有的则认为教育数据挖掘是学习分析的一个分支。无论如何,教育数据挖掘和学习分析技术的发展都促进了高等教育研究中定量研究能力的发展,适应了高等教育政策和实践中日益增长的循证分析的需求。

国际教育数据挖掘协会(The International Educational Data Mining Society)认为:“教育数据挖掘是一门新兴学科,它关注开发探索来自教育系统的独特类型数据的方法,并使用这些方法来更好地理解学生以及他们所在的学习系统。”<sup>[10]</sup>教育数据挖掘是计算机科学、教育学和统计学等领域相互交叉形成的一门新兴学科,它综合应用这些学科的理论与方法来理解教育系统中的教学、学习、管理和科学研究问题。教育数据挖掘应用这些不同学科领域的理论与技术研究、开发和实施软件工具以识别

教育大数据的形态与特征。

学习分析研究协会(The Society for Learning Analytics Research)将学习分析定义为:“测量、收集、分析和报告关于学习者及其所处的情景的数据,以便理解和优化学习者的学习及其发生的环境。”<sup>[11]</sup>学习分析的目标是对大数据进行分析并向教师和学生提供反馈。例如,通过分析学生活动和同伴互动所获得的信息,提出用于改进学生学习或课程教学的建议。

虽然学习分析和教育数据挖掘有许多相似的目标和共同的兴趣,但它们之间也有一些差异。学习分析起源于语义网络、智能课程和系统干预,而教育数据挖掘则起源于教育软件、学生建模和预测课程结果等。进行学习分析的研究者会更频繁地使用统计、可视化、社会网络分析、情感分析、影响分析、话语分析和概念分析等方法,而进行教育数据挖掘的研究人员则更多地依赖分类、聚类、关系挖掘和模型发现等。学习分析较侧重对数据和结果的描述,而教育数据挖掘则更侧重程序和技术。

应用教育数据挖掘和学习分析方法进行教育大数据分析的基本过程包括多个不同环节,如数据的采集、报告、预测、行动和完善等(如图 1)。

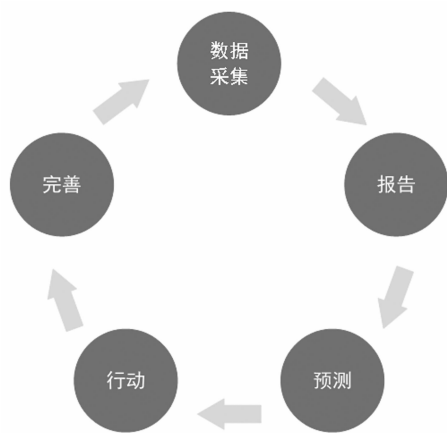


图 1 教育大数据分析的基本过程

教育大数据分析的基本取向可以分为两类:一类是数据驱动的分析取向(如图 2),另一类是内容或需求驱动的分析取向(如图 3)<sup>[12]</sup>。数据驱动的分析取向是从拥有的关键数据资源,如各类政策与管理文件、学校发展战略、招生数据、课程数据、师生背景信息、校友数据、管理数据和督导反馈数据等出发进行分析。内容或需求驱动的分析取向则强调从学生的需求出发,并将不同类型的变量、学生需求和教育质量指标联系起来。

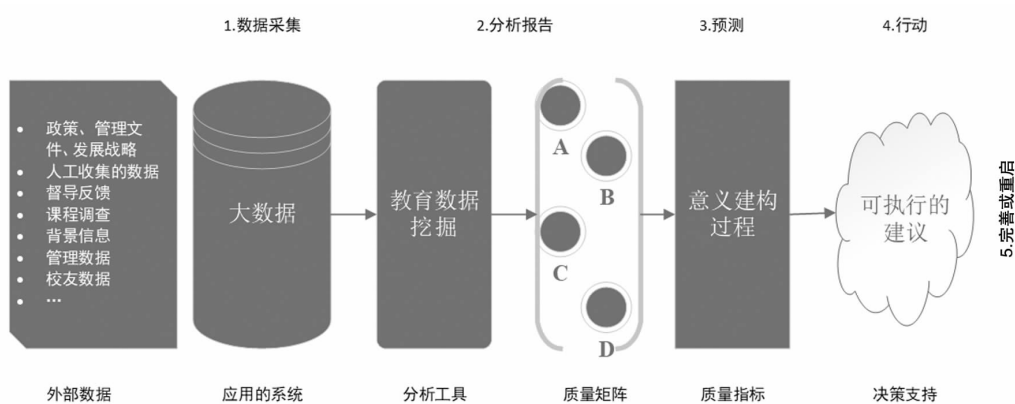


图 2 数据驱动的教育大数据分析取向

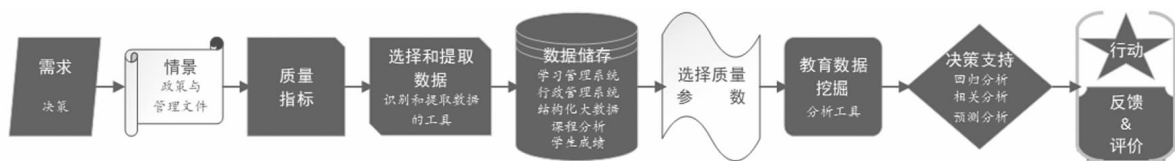


图3 内容或需求驱动的教育大数据分析取向

大数据在高等教育领域的应用涉及许多方面,如知识建模、行为建模、经历建模、主题建模和课程分析等,其中在教与学领域常用的一些教育数据挖掘和学习分析方法包括以下几个方面<sup>[13]</sup>:

### (一)分类与预测

分类与预测是指根据数据的过去状态来搜索和识别自变量和因变量之间关系的数据挖掘方法。在大数据分类和预测分析中,采用的算法或方法有很多<sup>[14]</sup>。例如,Sivasakthi 应用学生人口统计数据以及编程导论课程的成绩,采用 multilayer perception, Naïve Bayes, SMO, J48 和 REPTree 等分类算法对学生进行分类,并利用基于上述分类算法的预测数据挖掘模型,对修读编程导论课程的一年级本科生的成绩进行预测<sup>[15]</sup>。

### (二)聚类

聚类是指识别一个完整的数据集里在某些方面相似的数据点,以便将其划分为不同类别的小数据集。例如,Ding 等利用西安理工大学管理系统大数据中的学生出席率、阅读数量、利用图书馆的时间、学习时间、加权平均成绩和课程通过率等指标,应用优化的 K-means 算法对学生的行为进行了聚类分析,以识别不同的学生群体及其行为特点,进而为学校更好地了解和管理学生提供支持<sup>[16]</sup>。大数据聚类涉及单机和多机的不同聚类技术,其中单机聚类方法包括基于样本的技术和减维技术等不同方法,多机聚类方法包括并行聚类方法(parallel clustering)和基于 MapReduce 的聚类方法等<sup>[17]</sup>。

### (三)异常值检测

异常值检测方法是指识别与其他数据有显著差异的数据点的过程。被检测为异常值的数据点通常比其他数据要大或小许多。我们曾基于 2008—2017 学年全国高校博士研究生学位授予大数据,采用异常值检测方法,构建了基于博士研究生学位平均攻读年限的学位授权点监测预警模型,筛选学位攻读年限处于异常状态的学位授权点进行监测预警<sup>[18]</sup>。

### (四)关系挖掘

关系挖掘是指发现数据集里变量间的关系,并将其编码为规则以便今后使用的过程。有多种不同的关系挖掘方法,比较流行的是关联规则挖掘。例如,Wang 等提出了一种基于深度学习的“指导者—被指导者”关系的识别方法,探索了从 PhDtree 项目中爬取“指导者—被指导者”配对关系,并通过匹配作为实验数据集的 DBLP 数据集来提取他们的发表信息的方法<sup>[19]</sup>。

### (五)社会网络分析

社会网络分析将个体视为网络中的节点,将个体之间的关系视为节点之间的连接。社会网络分析的目的是确定与理解网络环境中个体之间的关系。例如,我们曾利用原美国科学情报研究所高被引数据库(ISI Highly Cited Database)收录的具有国际声誉的高被引学者数据,应用社会网络分析方法,从学科角度分析了 2 636 名高被引学者的受教育机构与其工作机构构成的社会网络类型及其特征<sup>[20]</sup>。

### (六)过程挖掘

过程挖掘可以对信息系统中事件日志提供的有用知识进行可视化表达。过程挖掘包括过程发现、过程与质量标准一致性评估、过程强化与支持等<sup>[21]</sup>。例如,Goel 等利用互动和比较过程挖掘技术,分析了澳大利亚某大学博士生实际学习历程的类型,可视化分析了实际过程与要求过程之间的差

异,比较了不同组别学生的表现并分析了出现不利结果的原因<sup>[22]</sup>。

### (七)文本挖掘

文本挖掘是指用于分析非结构化文本和从原始文本中获得高质量信息的过程。文本挖掘包括文本分类、聚类、名词提取、情感分析和实体关系建模等。例如, Gurcan 等应用文本挖掘技术分析了2008—2018年发表的27 735篇期刊论文,以揭示远程学习研究的主题和发展趋势<sup>[23]</sup>。

### (八)为作判断提取数据

为作判断提取数据是指使用可视化和交互界面来表示数据以便让人能够快速识别或分类数据特征的过程。例如,研究者可以用这种方法分析教育大数据,以便帮助人们识别学生的学习活动或行为模式<sup>[24]</sup>。

### (九)利用模型发现

利用模型发现是一种技术,它使用先前验证过的现象模型(用预测、聚类或知识工程方法获得)作为其他分析(如预测或关系挖掘)的一个组成部分。例如,在高等教育领域,利用模型发现可以用来发现学生行为与其特征或学习情境变量之间的关系,进而将相关模型应用于机器学习中<sup>[25]</sup>。

上述教育数据挖掘和学习分析方法,有些是在教育大数据分析中专门使用的,有些在常规的高等教育研究中也会使用。尽管如此,常规的高等教育研究方法 with 教育大数据分析仍然存在着一些差异,表1呈现了两者之间的差异<sup>[26]</sup>。

表1 标准的研究方法与大数据研究方法的差异比较

特征	常规研究方法	大数据研究方法
数据	为研究而创建	之前就存在
数据采集机制	为特定目的建立数据采集机制	挖掘先前的数据,动态地积累数据
数据清洗	精心检验数据质量	接受数据质量方面的不确定性
数据采集干涉性	高	低
人均数据采集成本	高	低
假设规范性	先验的假设规范	后验的假设规范
伦理要求	通常要求被试许可	分析用的数据已经存在,可能是开放获取的;需要考虑二手数据使用和原始数据采集政策方面的一些事项
方法论	确定需要哪些数据	了解有哪些数据
分析	SPSS/NVivo等一些现有方法	根据不同的大数据解析和报告需要哪些潜在的新工具
样本特性	严格界定	比较宽松地界定
情景	一种或较少	很多
可复制性	维持或再创造研究情景、资源可能比较困难	接触和分析数据相对容易,设施可能是开放的
样本量	可变的,一般较小	可变的,一般较大
时间性	静态的最终报告	动态的仪表盘显示

## 四、大数据在高等教育领域应用面临的两难挑战

大数据在高等教育领域的应用,虽然为描述、分析与解决高等教育问题提供了前所未有的大样本,但也面临许多挑战。首先,我们需要面对复杂的数据对象。在高等教育领域应用大数据,不仅需要面对高等教育大数据收集、处理与分析的安全性、隐私性和相关伦理要求,而且需要解决如何描述高等教育

大数据的复杂性特征这一根本问题。其次,大数据计算的复杂性也是我们需要面对的问题。大数据的多源、海量、快速变化等特征,使得传统的计算方法难以适应大数据分析的要求,因此需要创新处理与分析高等教育大数据的方法,开发面向高等教育大数据的高效算法。第三,在高等教育领域应用大数据还需要大数据基础设施和处理系统的支持。大数据分析既需要解决计算的复杂性问题,也需要满足处理的实时性要求,要求有相应的基础设施支持。这不仅对高等教育大数据的系统架构、计算框架和处理系统的设计提出了新的挑战,而且对它们的运行效率和能源消耗提出了严格要求。

Ekbia 等曾将大数据应用面临的两难挑战概括为如下几个方面<sup>[1]</sup>:一是认识论的两难挑战。例如,在大数据分析中,是强调因果联系还是统计相关?是强调探求因果机制还是基于表象预测发展趋势?二是方法论的两难挑战。这涉及定量与定性方法应用的长期争论。例如,在收集、清洗数据时,是强调计量还是不强调计量?在分析数据时,是否需要以及如何选择统计显著性水平?三是审美的两难挑战。例如,进行数据可视化分析时,是强调呈现的精确性还是强调呈现的美观性?四是技术的两难挑战。例如,计算机系统是强调连贯性还是创新性?是强调机器自动化还是人的作用?五是法律和伦理的两难挑战。例如,如何保护隐私?是参与还是不参与大数据的收集与分析?是开放还是保护大数据的知识产权?六是政治和经济的两难挑战。例如,大数据作为一种信息资产,是为其形成积极地做贡献还是只考虑如何利用它?对由大数据塑造的一些社会形象,是简单服从还是抵制?这些两难挑战,也是大数据在高等教育领域应用时会面临的问题。要推动大数据在高等教育领域应用的发展,必须认真处理这些问题。

## 参考文献:

- [1] EKBLIA H, MATTIOLI M, KOUPER I, et al. Big data, bigger dilemmas: a critical review[J]. Journal of the association for information science and technology, 2015, 66(8): 1523-1545.
- [2] LANEY D. 3D data management: controlling data volume, velocity and variety[J]. META group research note, 2001, 6(70): 1.
- [3] CHEN M, MAO S, LIU Y. Big data: a survey[J]. Mobile networks and applications, 2014, 19(2): 174-209.
- [4] CHONG D, SHI H. Big data analytics: a literature review[J]. Journal of management analytics, 2015, 2(3): 175-201.
- [5] NIST big data public working group. Big data interoperability framework: volume 1, definitions[R]. Gaithersburg: National Institute of Standards and Technology, 2015: 47.
- [6] BOYD D, CRAWFORD K. Critical questions for big data: provocations for a cultural, technological and scholarly phenomenon[J]. Information, communication and society, 2012, 15(5): 662-679.
- [7] COOPER A. A framework of characteristics for analytics[J]. CETIs analytics series, 2012, 1(7): 4-17.
- [8] TOLLE M K, TANSLEY S W D, HEY J G A. The fourth paradigm: data-intensive scientific discovery[J]. Proceedings of the IEEE, 2011, 99(8): 1334-1337.
- [9] 维克托·迈尔-舍恩伯格, 肯尼思·库克耶. 大数据时代: 生活、工作与思维的大变革[M]. 盛杨燕, 周涛, 译. 杭州: 浙江人民出版社, 2013: 79, 172-3.
- [10] Educational data mining[EB/OL]. [20220210]. <https://educationaldatamining.org>.
- [11] What is learning analytics? [EB/OL]. [20220210]. <https://www.solaresearch.org/about/what-is-learning-analytics>.
- [12] NAZARENKO M A, KHONUSOVA T V. Big data in modern higher education: benefits and criticism[C]//2017 international conference "quality management, transport and information security, information technologies". Saint Petersburg: IT&QM&IS, 2017: 676-679.
- [13] RAY S, SAEED M. Applications of educational data mining and learning analytics tools in handling big data in higher education[M]. Cham: Springer, 2018: 135-160.
- [14] BAI X, ZHANG F, LI J, et al. Educational big data: predictions, applications and challenges[J]. Big data research, 2021, 26: 100270.
- [15] SIVASAKTHI M. Classification and prediction based data mining algorithms to predict students' introductory program-

- ming performance [C]//2017 international conference on inventive computing and informatics. Coimbatore: ICICI, 2017: 346350.
- [16] DING D, LI J, WANG H, et al. Student behavior clustering method based on campus big data [C]//2017 13th international conference on computational intelligence and security. Hong Kong: CIS, 2017: 500503.
- [17] SHIRKHORSHIDI A S, AGHABOZORGI S, WAH T Y, et al. Big data clustering: a review [C]//International conference on computational science and its applications. Guimaraes: ICCSA, 2014: 707720.
- [18] 李珍艳, 陆根书, 高玉建, 等. 基于大数据的博士学位攻读年限监测预警分析 [J]. 研究生教育研究, 2020, 59(5): 1523.
- [19] WANG W, LIU J, XIA F, et al. Shifu: deep learning based advisor-advisee relationship mining in scholarly big data [C]//Proceedings of the 26th international conference on world wide web companion. Perth: ACM, 2017: 303310.
- [20] 彭正霞, 陆根书, 康卉. 高被引学者的受教育机构与工作机构的社会网络类型与特征 [J]. 高等教育研究, 2014, 35(11): 1930.
- [21] GARCIA C D S, MEINCHEIM A, JUNIOR E R F, et al. Process mining techniques and applications: a systematic mapping study [J]. Expert systems with applications, 2019, 133: 260295.
- [22] GOEL K, LEEMANS S, WYNN M T, et al. Improving PhD student journeys with process mining: insights from a higher education institution [EB/OL]. [20220210]. <https://eprints.qut.edu.au/226154/1/bpmi2021goel.pdf>.
- [23] GURCAN F, CAGILTAY N E. Research trends on distance learning: a text mining-based literature review from 2008 to 2018 [J]. Interactive learning environments, 2020: 422.
- [24] MAZZA R, MILANI C. Gismo: a graphical interactive student monitoring tool for course management systems [EB/OL]. [20220210]. <https://repository.supsi.ch/9832/1/TEL04.pdf>.
- [25] BIENKOWSKI M, FENG M, MEANS B. Enhancing teaching and learning through educational data mining and learning analytics: an issue brief [R]. Washington, D. C.: Office of educational technology, U. S. department of education, 2012: 1412.
- [26] ELLAWAY R H, PUSIC M V, GALBRAITH R M, et al. Developing the role of big data and analytics in health professional education [J]. Medical teacher, 2014, 36(3): 216222.

(编辑:张腾 校对:杨慷慨)

## Applications and Challenges of Big Data in Higher Education

LU Genshu

- (1. School of Humanities and Social Science, Xi'an Jiaotong University, Xi'an 710049, China;
2. West China Higher Education Evaluation Center, Xi'an Jiaotong University, Xi'an 710049, China)

**Abstract:** There are mainly three approaches to define the concept and connotation of big data, namely the product-oriented perspective, the process-oriented perspective and the cognition-oriented perspective. The application of big data in the field of higher education is of great significance for improving university strategic planning and promoting university development, perceiving teaching reality and improving teaching efficiency, optimizing the learning experience and improving learning quality, promoting scientific research and promoting interdisciplinary development. The analysis of educational big data mainly adopts two techniques, namely educational data mining and learning analysis. In the field of teaching and learning in higher education, these two techniques employ the methods such as classification and prediction, cluster and outlier detection. The application of big data in higher education is facing epistemological, methodological, aesthetic, technical, legal and ethical, as well as political and economic dilemmas.

**Key words:** big data; educational data mining; learning analysis; higher education