

高校管理人员教育数据挖掘能力培养的实践价值与实施路径*

丁国勇¹,王雪²,秦新国¹

(1.南京审计大学 教务处,江苏 南京 211815;

2.南京审计大学 实验中心,江苏 南京 211815)

摘要:大数据时代,培养高校管理人员教育数据挖掘能力,提高数据使用与数据挖掘的素养,不仅相当必要,而且已经变得特别迫切。高校可以通过教育数据挖掘来处理课程、案例、实习问题,培养管理人员的数据理解能力、数据处理能力、数据建模能力。丰富的数据化的课程资源、共享案例库和高素质的师资是高校管理人员教育数据挖掘能力培养的关键。

关键词:大数据;教育数据挖掘;高校管理人员

中图分类号:G647

文献标志码:A

文章编号:1673-8454(2020)07-0051-04

进入 21 世纪,信息、网络、存储等技术的快速发展以及“互联网+”、云计算、物联网、人工智能、机器学习等技术的大规模应用,在社会生活的许多方面产生了深刻变革,我们已经进入“大数据时代”。大数据时代给高等教育带来了一次教学模式、管理手段、思维方式的全方位变革,这次变革对高校管理人员的数据能力和数据素养提出了新的挑战,培养高校管理人员教育数据挖掘能力是应对大数据挑战、进行教育管理改革与创新的必然选择。

一、教育数据挖掘的缘起及其与教育统计学的关系

1. 教育数据挖掘的起源与发展

教育数据挖掘起源于 Corbett 和 Anderson 于 1995 年提出的贝叶斯知识跟踪(Bayesian Knowledge Tracing)模型。2000 年成立了第一个与教育数据挖掘相关的工作组。2005 年第一次提出教育数据挖掘的概念。2006 年 Romero 等人出版了关于教育数据挖掘的第一本专著《数字化学习中的数据挖掘》。2008 年召开了第一届教育数据挖掘国际会议,之后,该国际会议每年举办一次,目前已举办十届。2009 年出版第一本有关教育数据挖掘的期刊。2010 年,Romero、Baker 等人出版了第一本关于教育数据挖掘的工作手册。2011 年成立第一个国际学术组织“国际教育数据挖掘学会”。相关学术组织和学术研究活动的推动使教育数据挖掘逐渐成为独立的研究领域。

相对于教育学的其他分支领域,教育数据挖掘还属于发展中的研究领域,也没有统一规范的定义。Romero 等(2006)对 1995-2005 年间有关教育数据挖掘的文献进行研究后认为“教育数据挖掘是一个与数字化学习、自适应超媒体、智能导师系统、Web 挖掘、数据挖掘等成熟领域相关的即将到来的研究领域,其应用更多地关注学习者和教育系统的教育方面”。^[1] Baker(2009)认为教育数据挖掘是一个“开发方法、探究教育环境中独特类型数据的新兴学科,通过这些方法来更好地理解学生和他们的学习环境”。^[2]国际教育数据挖掘学会对教育数据挖掘的定义与 Baker 的定义相近,将其定义为“通过开发方法分析教育环境中独特且不断增长的大规模数据的新兴学科”。

国外学者对教育数据挖掘的定义也是多样化的。有的学者侧重于工具和技术的应用,如 Richard(2013)认为教育数据挖掘是“聚焦数据挖掘工具和技术并应用于教育相关数据的领域”^[3],Muna Al-Razgan 等(2014)认为教育数据挖掘是“应用数据挖掘技术分析教育数据的跨学科研究领域”;^[4]有的学者强调方法的开发,如 Alejandro Peña-Ayala(2013)强调教育数据挖掘是“开发模型、任务、方法、算法并探究来自教育环境中的数据”;^[5]有的则强调教育数据挖掘的功能,如 S. Hari Ganesh 等(2015)认为教育数据挖掘“解释教育事实,为教育者向学生提供

* 基金项目:本文系教育部人文社会科学研究青年基金项目“基于教育数据挖掘的高校学生投入影响因素与评价模型研究”(项目批准号:19YJC880016)的研究成果。

高质量教育提供保障……可能对教育研究和实践产生巨大的影响”。^[6]教育数据挖掘不仅适用于改进教师和学生的教与学体验,还可以在支持管理决策科学化、提升学校治理能力上大有作为。教育数据挖掘适应了大数据时代挖掘教育数据的价值、发现教育数据蕴涵规律的现实需要。

2. 教育数据挖掘与教育统计学的关系

教育数据挖掘融合了计算机科学、统计学、教育学、机器学习等多个学科的知识,与多个学科之间都具有交叉关系,是一个复合交叉研究领域,如图1所示。教育数据挖掘的学科属性与教育统计学有较大的相似性,其研究方法限定在数据挖掘,其研究对象限定在教育领域。^[7]从广义上来说,教育数据挖掘的研究领域更为宽泛;运用数据挖掘的专有算法对教育数据进行分析必然是教育数据挖掘;对教育对象进行调查,教育数据进行采集、处理、描述统计、分析等也可以纳入教育数据挖掘范畴;专门为教育领域设计算法、开发工具软件的研究可以纳入教育数据挖掘领域;教育统计学重在抽样,而教育数据挖掘重在全样本,因此在某种意义上教育统计学也可以纳入教育数据挖掘或与教育数据挖掘进行学科融合。

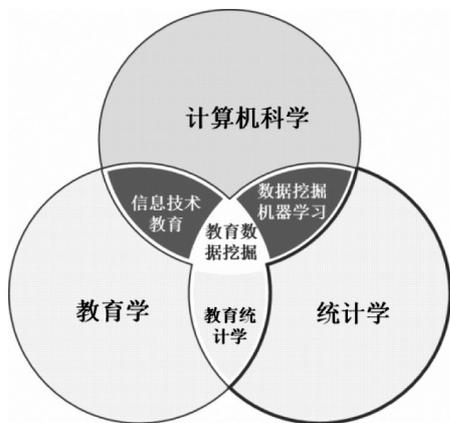


图1 教育数据挖掘的交叉学科属性

二、高校管理人员教育数据挖掘能力培养的实践价值

1. 呼应大数据时代对高等教育管理模式变革的需求

大数据时代高等教育变革体现在两方面,一方面通过MOOC等催生了高校教学模式的变革,以Coursera、edX、爱课程网等为代表的国内外在线课程平台得到全面应用,这些在线课程平台存储了包括学生特征、学习行为过程等相关的海量教学大数据;^[8]另一方面通过智慧校园建设以及各类教育教学管理信息系统的应用,推

动了高校管理方式变革,产生了内容丰富、类型多样、容量巨大的教育管理大数据,在高校治理的背景下,从产生的海量教育管理大数据中挖掘出有价值的信息并使之促进高校决策科学化成为必然要求。面对这样的教学大数据和教学管理大数据的“金矿”,需要大量、高水平的既有坚实高等教育理论基础又有教育数据挖掘能力的“矿工”来进行挖掘。

一项研究表明,国内教育数据挖掘研究者中有78%来自于计算机或数学学科,来自教育学科背景的仅占4%。计算机或数学背景的研究者能熟练运用数据挖掘技术,但缺乏对教育问题的剖析能力;而教育学科背景的研究者拥有坚实的教育理论基础,但缺乏对数据挖掘和分析的能力,结果是国内实证研究能力和数量与国外相比都存在一定的差距。^[9]相比教育数据挖掘的研究者而言,高校中的管理人员拥有更加丰富的教育管理实践经验,是教育数据的直接生产者、管理者和消费者,通过培养他们的教育数据挖掘能力,提升他们对教育数据的使用效率和从教育数据中挖掘教育规律、解释教育现象、制定教育政策的能力,呼应了大数据时代对高校教育管理模式变革的需求。

2. 满足高校对管理人员综合素质的要求

由于历史原因,现在部分高校中还存在着管理工作相对简单机械,管理队伍知识结构不合理、学历层次不高,管理人员现代管理意识缺乏、管理效率低、职业意识不强等一些问题,“什么专业毕业都可以搞高校管理”。高校管理队伍面临着总体素质不高、水平参差不齐的普遍问题,所以经常说高校是“一流的设施、二流的生源、三流的管理”。

随着高等教育的进一步发展,高校规模不断增大,外部环境从单一稳定向多元动态发展,管理工作日趋复杂,内部职责分工将更为明晰,管理工作则必须由专业化管理人员来承担,高校对管理人员的专业素质将会有更高的要求。如同文字处理、表格处理等办公软件的熟练使用早成为管理人员的必备技能,在大数据时代,对数据的理解、采集、处理、分析能力,利用数据提高管理效率、提升问题解决能力和科学决策能力也将会成为每一位管理人员的专业素质的基本组成部分。因此,只有通过培养管理人员的教育数据挖掘能力才能满足高校对管理人员综合素质的要求。

3. 达成高校管理人员对个人职业能力提升的追求

高校管理人员中,学科背景存在着文科生居多、理工科偏少的问题,这种分布特征在某种意义上导致管

理人员数学和信息处理能力相对较弱,对方程、统计、建模等偏理科内容的掌握不足,甚至存在某种“惧怕感”。教育数据挖掘与教育统计相比更加容易理解与操作,不需要去学习过于繁琐的数学公式方程和编程代码,使用现有的成熟的教育数据挖掘软件,完全可以进行全程可视化的无代码操作,简而言之就是“容易学、上手快”。

高校管理人员通过教育数据挖掘来进行相关课程的学习,不仅提升数据分析与处理能力,同时在课题研究和问题解决过程中,还可以提升协调与合作能力、沟通与交流能力以及对高校教育资源与环境的熟悉程度。对这些教育数据进行挖掘和分析后,高校管理人员可以撰写研究报告、申报课题、发表相关研究论文,从而提升职业能力。

三、高校管理人员教育数据挖掘能力培养的实施路径

1. 高校管理人员教育数据挖掘能力的构成要素

(1)教育数据理解能力。教育数据理解能力是指教育数据及其存在样态的认知能力。教育数据的类型及存在形式是多种多样的:从类型来看,可以分为基础数据、状态数据、资源数据和行为数据。基础数据包括保障教育活动的各类基础信息,状态数据包括教育装备、教育环境、教育业务的运行状态,资源数据包括教育过程中建设生成的教学资源,行为数据包括教学活动中教师的教学行为数据和学生的学习行为数据等。^[10]从存在样态来看,包括结构化的关系数据库中的数据和非结构化的音频、视频、文字、社交媒体等方面的数据。教育数据理解能力是教育数据挖掘能力中最基础的一项构成要素,在对需要被研究的教育问题进行充分调研后,要能够从繁杂多样的教育数据中界定出与教育问题可能相关的数据,并为后续的采集处理做好准备。

(2)教育数据处理能力。教育数据处理能力是指对教育数据进行采集、清洗以及初步描述的统计分析能力。教育数据处理是进行教育数据挖掘的基础,在一个完整的教育数据挖掘流程中,教育数据理解和教育数据处理可能要占到总时间的30%~75%,经过处理的教育数据的质量是教育数据挖掘能否成功的关键。^[11]教育数据处理过程中需要应用到不同的工具,主要包括数据库、数据仓库及相关工具软件。如处理结构化的教育数据,则操作关系数据库是必备的技能,特别要熟悉结构化查询语言(Structured Query Language,SQL);而对教育数据进行描述统计,则可以采用SPSS或Tableau等数据可视化分析工具。教育数据处理能力实际是在对

教育数据充分理解的基础上,操作系列相关工具软件的能力。

(3)教育数据建模能力。教育数据建模能力是指选择合适的教育数据挖掘工具及算法,对经过处理的数据进行挖掘,抽象出一定的规则表达或数学表达的能力。教育数据挖掘的建模方法包括统计学建模、聚类建模、分类建模和关联规则建模等,其操作也是依靠相关的工具软件。依据功能不同,将教育数据挖掘工具软件分为通用工具软件与专用工具软件两种。通用的工具软件包括SPSS Modeler、RapidMiner和Weka等,专用工具软件包括用于文本挖掘的LIWC、WMatrix、Coh-Matrix等,用于社会网络分析的Gephi、EgoNet、NodeXL、Pajek、SNAPP等,用于贝叶斯知识跟踪的BKT-BF、BKT-SM等,这些通用和专用分析工具提供的算法和建模框架,可以用于建模及挖掘预测教育数据中的关系。^[12]教育数据建模能力中还包括对所建立的模型进行评价、优化,以及结合模型解释教育问题、提出政策建议的能力。

2. 高校管理人员教育数据挖掘能力培养的主要介质

(1)基础课程。教育数据挖掘能力的培养介质之一是开设相关的在职培训课程,主要内容应该包括但不限于数据科学、数据库、数据挖掘等,该课程学分一般可设为3学分,采用必修方式,具体教法应该侧重于基本理论、原理和算法思想,不必过度聚焦于具体技术细节,主要培养管理人员的教育数据理解能力。

(2)案例教学。教育数据挖掘能力培养的是应用能力,其培养介质之二是增加丰富的案例教学,在案例中学会相关工具软件的使用。常用的工具软件可以包括数据库工具如SQL Server、Access等关系数据库,SPSS等常用的统计分析软件,SPSS Modeler、RapidMiner等常用数据挖掘软件。^[13]案例教学采用问题导向形式,每一个案例选定教育领域的一个专门问题,从问题理解、数据理解、数据处理、数据建模、模型评价等,完成教育数据挖掘的一次执行过程。一个案例教学可以设置为1~2学分,采用选修的方式,通过案例教学培养管理人员的教育数据处理与教育数据建模能力。

(3)实践演练。实践演练是管理人员教育数据挖掘能力培养的最有效介质。在掌握一定的理论基础与操作能力后,基于本校或其他高校,就教学管理、学生管理、科研、资产等某一方面具体的问题,在教师和相关实务人员的指导下,进行教育数据挖掘的实际操作。实践演练处理的是实体环境下的真实案例,在检验并提升管理人员教育数据挖掘能力的同时,培养其环境适应能力、

沟通交流能力等综合能力。

3.高校管理人员教育数据挖掘能力培养的保障因素

(1)充分的课程资源。教育数据挖掘能力培养需要丰富的课程资源支持,但目前国内教育数据挖掘方面相关资源并不丰富。一是专著较少,只有葛道凯编写的《教育数据挖掘:方法与应用》(教育科学出版社2012年版)等少数几本,专门教材几乎没有;^[14]二是与利用教育数据挖掘方法进行研究的学术、学位论文也不多;三是在线课程资源也不多,相关的课程只有在中国大学MOOC上由上海大学叶志明教授开设的《大数据与高等教育》(2018年10月第一次开课)和浙江工业大学江波副教授等开设的《教育数据挖掘技术与应用案例》(2017年9月第一次开课,16课时),在教育数据挖掘课程资源建设上还有许多工作要做。

(2)完善的案例库。丰富的教学案例库是教育数据挖掘能力培养的有力保障。教育数据挖掘案例应该主要来源于高校的实务部门,案例主题应该涵盖高校教学、管理的每一个环节,每一个教育数据挖掘教学案例包括业务理解、数据理解、数据处理、数据建模等主要模块,并提供经过过去隐私化处理的完整数据。同时建立案例共享机制,利用“中国专业学位教学案例中心”等平台进行案例分享或校际共享。

(3)优质的师资队伍。教育数据挖掘的课程资源、案例开发和教学应该由三支队伍共同完成:具有深厚高等教育学理论功底 of 的导师,具有数学、统计学、计算机、教育技术背景的导师以及来自高校实务部门的教师或管理干部。三支队伍在理论、技术和实践上各自发挥优势,互助互补,以具体问题的解决为导向,不断开发课程资源、更新教学案例。

总之,当高校面临新技术革命特别是现代信息技术革命的挑战时,我们需要来一场迈克尔·富兰所说的“突破”。^[15]在《国家大数据战略》和《高等学校人工智能创新行动计划》的指引下,大数据、人工智能将与教育更加深度融合,高校教育教学管理模式必须再来一场真正的“突破”,否则,“时代抛弃你时,连招呼都不会跟你打”。高校管理人员教育数据挖掘能力的培养只是这场“突破”的试水性探讨。

参考文献:

[1]C Romero,S Ventura.Educational data mining: A survey from 1995 to 2005[J]. Expert Systems with Applications,2007,33(1):135-146.

[2]RYAN S.J.D. BAKER.The State of Educational Data Mining in 2009: A Review and Future Visions [J]. Journal of Educational Data Mining,2009,1(1):1-16.

[3]RA Huebner.A Survey of Educational Data-Mining Research[J].Research in Higher Education Journal,2013(19):13.

[4]Muna Al-Razgan.etc. Educational Data Mining: A Systematic Review of the Published Literature 2006-2013 [C].Proceedings of the First International Conference on Advanced Data and Information Engineering(DaEng-2013), 2013:711-719.

[5]A Peña-Ayala.Educational data mining: A survey and a data mining-based analysis of recent works[J].Expert Systems with Applications,2014,41(4):1432-1462.

[6]S Hari Ganesh,AJ Christy.Applications of Educational Data Mining: A survey[J]. International Conference on Innovations in Information,2015:1-6.

[7]任训学.论教育统计学的学科性质——兼谈教育统计学教学与发展的若干问题[J].湖北大学学报(哲学社会科学版),1997(2):83-86.

[8]新华网.教育部推出首批490门国家精品在线开放课程[EB/OL]. http://www.xinhuanet.com/politics/2018-01/15/c_1122261678.htm.

[9]李婷,傅钢善.国内外教育数据挖掘研究现状及趋势分析[J].现代教育技术,2010(10):21-25.

[10]杨现民,唐斯斯,李冀红.发展教育大数据:内涵、价值和挑战[J].现代远程教育研究,2016(1):50-61.

[11]Jing Luan.Data Mining and Its Applications in Higher Education[J].New Directions for Institutional Research,2002(113):17-36.

[12]Slater, S.,Joksimovic,S.,Kovanovic, V.Baker,Ryan S.,Gasevic,D.Tools for Educational Data Mining: A Review [J].Journal of Educational & Behavioral Statistics, 2016,42(1):85-106.

[13]薛薇.基于SPSS Modeler的数据挖掘(第二版)[M].北京:中国人民大学出版社,2014:23-25.

[14]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.

[15]迈克尔·富兰等.突破[M].北京:教育科学出版社,2009.

(编辑:王天鹏)