

多“源”异“构”培养大数据 创新型人才

吴力波

复旦大学大数据学院, 上海 200433

摘要

大数据科技与产业的迅猛发展对人才培养提出了巨大的挑战。数据科学家需要具备从巨量数据中挖掘商业价值的能力, 因此要进行多学科交叉培养。从大数据人才培养目前进展和需求来探讨多“源”异“构”大数据创新型人才的培养。在人才培养中要打造核心知识体系, 使未来的数据科学家能够具备持续创新的基础。大数据人才培养必须要有数据环境, 要打通政府、企业与高校之间的数据壁垒, 面向真实数据培养针对性的应用技能。大数据人才培养要打破学科壁垒, 从人才培养方案设计到教学、实习、毕业、考核各个环节标准都要体现交叉特色。

关键词

大数据; 人才培养; 学科交叉

中图分类号: TP3

文献标识码: A

doi: 10.11959/j.issn.2096-0271.2016057

Cultivating big data talent by combing various disciplines and utilizing multiple resources

WU Libo

School of Data Science, Fudan University, Shanghai 200433, China

Abstract

Rapid development of big data science and technology put forward great challenges to talent cultivation. Data scientist should have the capability of coax commercial value from tremendous data and require multi-disciplinary training. Based on the progress and demand of cultivating big data talent, the patterns of cultivation big data talent by combing various disciplines and utilizing multiple resources were discussed. In the patterns of talent cultivation, core knowledge system should be built and the basis of innovation capability should be solid. The talent cultivation should have data resource. Therefore, breaking through the data open barriers of government, private firms and universities and providing real world data is essential for talent education. To strengthen the interdisciplinary features, talent cultivation should remove the barriers of disciplines and build up systematic standards through the teaching, internship, graduate and evaluation.

Key words

big data, talent cultivation, inter-discipline

1 数据科学与数据人才的需求特征

大数据是否会带来一场技术和工业革命, 仍然有待观察。但毋庸置疑的是, 大数据的开发与应用确实在改变着当今社会信息流动、能源流动的方式, 而这一特征, 恰恰契合了历次工业革命的基本特征: 通信技术与能源利用技术变革结合, 改变了支撑社会进步的核心要素的配置方式, 从而带来经济发展方式的根本转型。巨量数据的开发应用给计算机科学、统计学、计算数学等基础学科带来巨大挑战: 大数据的存取、交换、分析、应用无一不涉及基础理论和应用技术的创新, 从硬件到软件、从存储到超算、从数据库到数据安全、从网络传输到并行计算、从数据分析到统计建模、从科学计算到优化方法等。而大数据的发展还具备一个鲜明特征: 与各个学科领域的深度融合。大数据在商业、金融、医疗、能源、传媒等各个领域都有着越来越广泛的应用, 甚至开始颠覆这些学科在原有统计样本支撑下形成的理论体系和应用架构。“数据科学”不再仅仅是数学家或者统计学家的专属领域, 而成为站在计算机科学、统计学、应用数学等学科巨人肩膀上, 与经济学、金融学、医学、生物学、新闻学、社会学等多学科高度交叉的21世纪的“创新型科学”。

数据科学是一个新的学科, 具有高度的学科交叉特性, 同时又高度面向产业应用。数据科学植根于数学、统计学、计算机科学等学科, 但是在研究对象、方法论、学科体系等方面又与这些学科有显著不同。数据科学的内涵包含了两个层次, 第一个层次是以来源多样、结构各异、规模巨大、传输高速、应用广泛的大数据为

研究对象, 解决大数据在获取、处理、分析、展示与应用领域的理论与实践问题, 如数据挖掘、机器学习、人工智能、数据库、统计计算等领域; 第二个层次则是以大数据为研究手段的数据交叉科学, 如生物信息、精准医疗、电子商务、大数据金融、智能电网、智慧城市等领域, 大数据分析技术为这些学科提供了新的研究范式, 也在解决这些学科计算复杂性问题的过程中获得进一步的发展。由此可见, 数据科学的内涵已经超出了传统学科的范畴, 通过将统计分析、系统计算、交叉科学等有机整合, 形成一套面向大数据分析全流程、大数据应用全产业链的完整知识体系。

发展数据科学的目标之一是要面向数据产业的发展, 培养具备大数据思维的人才, 而这种人才的培养与主流的大学人才培养模式存在着显著的差异, 需要有新的人才培养理念、方式和标准。2012年10月, 美国哈佛商业评论对“数据科学家”的价值进行了阐释: 数据科学家是可以从看似杂乱、无规则的数据中提炼财富的职业, 这类人才既要具备全面的数据分析能力, 还要具备敏锐的市场嗅觉, 能够以价值创造为目标, 对数据进行各种形式的分析, 对看似无关的数据进行关联、解构。与很多传统学科的发展不同, 发展大数据科学与技术、培养数据科学家这类特殊人才的强烈需求是来源于市场, 由于与互联网、物联网、社交媒体深度融合的企业积累了大量数据资源, 这些数据资源已经成为企业创造财富的新动能, 因此对于能够实现这一过程的人才就提出了强烈需求。因此大数据人才培养的主基调是鲜明而富有活力的, 只有具备将科学逻辑与应用价值有机结合的全新知识体系, 数据科学家才能够各个行业的大数据浪潮中成为应对自如的弄潮儿。

2 国内外大数据人才培养进展

在大数据浪潮的席卷之下,美国大量名校都开设了大数据人才项目。2013年起,美国纽约大学和英国邓迪大学设立“数据科学”科学硕士学位,美国芝加哥大学开设了夏季培训课程,美国弗吉尼亚大学和美国华盛顿大学也设置数据分析科学硕士学位,美国西北大学设立了分析科学硕士学位,美国纽约大学还设立了商业分析科学硕士学位,后加州伯克利分校、麻省理工学院、卡内基梅隆大学等也开始了相关的人才培养计划。这些人才培养计划从本科到硕士、博士课程,课程设置各有千秋,在计算机科学、统计学和计算数学类基础课程的设置上有不同的侧重,在与其他学科的交叉方面也各具特色,有侧重于与管理学、经济学交叉的商业分析类,也有侧重于与医疗、交通等交叉的公共管理类,还有与环境、能源、电力、气候、生物等交叉的科学类。对学生的培养目标也有不同,有些项目侧重于掌握全面数据分析、可视化技术的大数据产业领域的工程师,有些项目则侧重于培养具备数据分析与决策能力的领导者。

我国的大数据人才培养尚未形成可比拟美国的气候。虽然互联网商务发展迅猛,大数据的应用在这些行业中也具备了相当规模,但是我国的大数据发展生态链还未形成,可供大数据人才培养的创新型环境还有很多短板需要补全。一方面是大数据企业与大学之间的合作还未全面展开。美国的很多大数据人才培养项目都是企业与大学合作建设,例如领英-全球最大的职业社交网站就与加州圣芭芭拉大学联合建设大数据商务分析硕士项目。国内校企之间的合作往往局限于研发项目合

作,人才培养项目方面的合作还很少。另一方面是受限于我国大学的资源配置模式。真正实现学科交叉、面向产业应用的人才培养项目还刚刚起步。“学科建设”是我国大学发展的主要路径,由此形成的学科体系尽管在不断地随着科学技术的发展而更新、分化,但是在交叉学科的发展方面则显示出固有的缺陷性。由于国内的科学研究、人才培养的资金流动体系都是与学科高度挂钩,导致交叉学科的发展很难获得足够的空间和资源配置,更勿论形成全面、完整的人才培养体系,进行交叉学科的人才培养。大数据人才的培养同样由于这样的背景原因而受到了一定的限制,一般往往是大学中计算机、数学、统计学哪个学科更强,哪个学科就举起大数据的大旗,其他学科只能有限参与,难以形成深度的交叉融合。

要推动大数据的发展,就必须开展数据科学与大数据技术专业人才的系统培养。从人才培养的纵向结构来看,要建立从基础人才到专业技术人才再到高端科技创新人才的完整培养体系。也就是说,大数据基础人才是适应各行业的大数据相关业务拓展的基础应用型人才;大数据专业技术人才是集中面向大数据产业的较高层次的应用拓展型人才;而高端创新人才则是从事大数据高端技术研发、重大理论创新攻关的高层次人才。从人才培养的学科广度来看,则要重基础、宽视野。力争在夯实计算机、统计学、应用数学基础的同时,拓宽数据应用类知识体系,使不同层次的学生都具备一定的适用于本层次人才的应用技术创新能力。

3 我国大数据人才需求分析

培养大数据人才,是响应国家大数据

产业加快发展的号召,为大数据催生的新产业、新业态、新经济培养具备扎实数据分析和系统建构能力、了解数据应用场景和价值创造机制的高端人才。

(1) 大数据产业发展的迅猛态势对大数据人才提出了强烈需求

随着社会的进步和科技的发展,人类已经迈入了“大数据时代”。大数据的应用已经涉及众多领域,成为“互联网+”与“工业4.0”时代推动社会进步的重要引擎。国务院《促进大数据发展行动纲要》中指出要“加强专业人才培养,创新人才培养模式,建立健全多层次、多类型的大数据人才培养体系。要大力培养具有统计分析、计算机技术、经济管理等多学科知识的跨界复合型人才”。在经济、政治、文化、生物、医学、能源等多个领域,大数据科学已经被广泛运用。我国大数据产业近年来蓬勃兴起,无论是典型的以百度、阿里、腾讯为代表的互联网企业,还有一批正在崛起的大数据分析技术企业、大数据交易、大数据银行、智能机器人等新兴产业都需要大量的大数据人才。而传统产业与大数据的融合更是开启了大数据人才需求的巨大闸门,以3D打印为代表的智能制造对数据分析提出了全新的要求,基因诊疗、在线诊断、药物筛选等新模式的出现也使得大数据成为医疗行业的新宠。这些行业的人才需求是传统的数学、统计和计算机科学等学科人才培养体系无法满足的。构建复合型的知识培养体系以及专注于大数据的研究、应用和商业价值的核心竞争力的培育,是社会对高等学校提出的现实需求。

(2) 政府提高公共治理效率、开放、开发、应用公共数据的需求日益强烈,对相关人才的需求十分迫切

国务院《促进大数据发展行动纲要》明确提出要推动大数据在政府治理中的

应用。在未来5~10年打造精准治理、多方协作的社会治理新模式,建立运行平稳、安全高效的经济运行新机制,构建以人为本、惠及全民的民生服务新体系。政府数据涵盖了公共交通、教育医疗、智慧城市、养老助困、人口环境等很多领域,是一个尚未开发的宝藏,既可以提升政府的公共治理效率,更可以成为服务社会经济发展的巨大财富。而目前政府数据的开发利用受限于人才技术瓶颈而进展缓慢。近几年来,上海市政府主动开放政府数据,经过脱敏后供大学生研究开发各类应用,上海开放数据创新应用(SODA)大赛受到了很多高校学生的关注,也培养了不少大数据工作团队,甚至有团队成为市场追捧的新宠。随着政府公共数据开放机制的健全,这类资源不仅可以成为人才培养的实战基地,也可以更多地被合理化应用,带来相关产业的发展。大数据人才培养应该以提升政府、社会公共治理能力为己任,将大数据公共治理的相关课程纳入课程体系,确保人才培养可以为政府管理、民生工程、智慧城市建设等多个领域提供支撑。

(3) 大数据人才市场需求巨大,但相关人才却十分稀缺

大数据作为新兴的战略产业,市场需求巨大。然而由于技术门槛较高,目前社会上的人才稀缺。大数据研究和运用需要具交叉学科背景的专业人才,兼具计算机科学、统计学、数学、经济学、生物科学等多学科知识。随着大数据研究和应用的推进,对于大数据人才的需求也在急剧增加。根据目前统计,国内只有北京大学、中南大学设立了数据科学与大数据技术本科专业,复旦大学设立了数据科学与大数据技术方向“2+2”本科项目,并在2017年本科大类招生中面向全国招生。其他如中国人民大学、上海财经大学等高校的大数据研究院都只有研究生项目,一般归入应用

统计方向招生, 尚未形成大数据独立的人才培养体系。这样的发展态势显然无法满足社会对相关领域人才的巨大需求, 因此亟需在有基础的高校设立数据科学与大数据技术相关的本科专业, 为大数据研究和运用以及推动社会发展提供相应的专业人才。我国实施创新驱动战略需要加强创新型人才的培养, 要能够积极应对全球工业4.0时代所特有的以数据为导向的制造模式、流通模式、消费模式、商业模式的变革, 大数据人才是面向这些新机遇、新挑战、新应用的高端专业数据人才。

4 多“源”异“构”培养大数据创新型人才

大数据的人才培养要秉承多“源”异“构”的思想, 利用多种本源学科, 打造具备差异化知识结构的人才。国外相关人士对数据科学家应该具备的技能进行了整理, 这些技能主要包括: 数学和统计学方面的知识、计算机编程和数据库知识与技能、领域知识和软技能以及沟通和可视化技能等。因此大数据的人才培养绝不是单一学科可以完成的, 也不是仅仅在大学中就可以学习到数据科学家应该具备的各项技能的, 大数据人才的培养模式必须要进行创新, 要整合优势学科资源, 打通产学研合作的壁垒, 方能培养出从巨量数据中榨取价值的创新型人才。

(1) 在本科教育中要把数据科学作为通识教育的基础内容, 让数据科学成为渗透各个学科的创新思维

数据思维是当今各个学科发展都需要具备的, 当考古学、语言学、地理学、新闻学等社会科学大量使用数据挖掘技术的时候, 谁还能够否认数据思维应该成为渗透各个学科的创新思维? 因此要在本科教育

中把数据科学作为通识教育的基础内容, 让不同学科的学生都能够懂一些数据科学、掌握一些数据科学的基本原理和方法论, 为大数据交叉人才的培养育好苗。

(2) 大数据人才培养必须要打造核心知识体系, 使未来的数据科学家能够具备持续创新的基础

尽管数据科学家们在不断研究拓展大数据分析的理论内涵与外延, 在应用层面各种创新也层出不穷, 区块链、人工智能等炙手可热的技术成为大数据人才出奇制胜的新宠。但是大数据人才培养必须要打造核心知识体系, 包括统计推断、数学分析、数据库、机器学习、大规模分布式系统等知识在内的核心知识体系培养要能够做到理论与应用相结合, 方能使之成为未来数据科学家持续创新的基础。

(3) 大数据人才培养必须要有数据环境, 要打通政府、企业与高校之间的数据壁垒, 面向真实数据培养针对性的应用技能

大数据科学在很大程度上是一门实验科学, 数据在量变的过程中推动理论和应用技术的质变。因此数据资源始终是大数据人才培养中最为重要的教学资源。空谈理论、脱离数据、脱离应用环境的大数据人才培养是无法培养出真正意义的数据科学家的。因此要集合公共资源, 建立能够面向科学人才培养的大数据实验场, 通过真实、模拟等各类数据源对人才的数据分析能力提供实战机会, 这样的人才进入产业界时, 才能够真正被市场所接纳。

(4) 大数据人才培养要打破学科壁垒, 从人才培养方案设计到教学、实习、毕业、考核各个环节标准都要体现交叉特色

交叉学科的人才培养关键是打破学科壁垒, 建立起一套能够体现本学科特色的人才培养方案并且在各个环节中以交叉为标准来进行过程管理。交叉学科不是拼盘学科, 不能有多重学科标准, 而是要建立起

完整的学科体系和人才培养标准。只有这样，才能确保大数据学科的发展始终面向学科交叉，而不会成为传统学科“新瓶装旧酒”的牺牲品。

5 结束语

我国的大数据人才培养已经有了良好的开端，教育部批准设立数据科学与大数据技术专业激发了各个高校建设大数据相关专业的热情。复旦大学率先在全国成立

了大数据学院，打造从本科、专业硕士到博士的完整的人才培养体系，同时通过大数据实验场的建设为交叉学科人才培养提供数据平台。北京大学的大数据人才培养非常注重与业界的合作，已经在大数据产业领域集聚了很多优势资源。这些进步都预示着大数据人才培养的美好未来。大数据人才培养是市场需求力推动的结果，同时得到了政府的大力支持，相信未来中国本土培养的数据科学家们能够成为大数据产业发展与科技创新的脊梁！

作者简介



吴力波(1974-),女,复旦大学大数据学院教授、副院长、博士生导师,国家社科重大项目“基于大数据的宏观经济现时预测理论与方法研究”首席科学家,主要研究方向为能源大数据分析、宏观经济建模分析、能源-环境-经济多系统建模与政策评估。

收稿日期:2016-08-20