

美国数据科学硕士专业培养要求 与课程设置研究^{*}

曹高辉¹ 胡紫祎¹ 郭家乐² 王艳凤²

(1. 华中师范大学信息管理学院, 武汉 430079; 2. 华中师范大学教育信息技术学院, 武汉 430079)

摘要: 本文选取美国18所开设数据科学硕士学位的高校作为研究对象, 采用文献研究法、网络调研法与内容分析法, 围绕培养要求、课程结构、课程设置进行统计分析, 总结美国高校在数据科学硕士课程设置及专业建设方面的特点, 包括各高校开设课程数量差异较大, 必修课程注重技术与应用, 选修课程面向特定领域开设, 跨学科与多学科交叉培养, 以及教学方式多样。结合国内硕士研究生教育现状, 从人才培养模式、课程体系建设、教学与考核方式、校企合作四个方面为我国数据科学硕士专业建设提供建议。

关键词: 数据科学; 课程设置; 人才培养; 硕士专业; 专业建设

中图分类号: G642.3

DOI: 10.3772/j.issn.1673-2286.2018.05.006

信息技术与经济社会的交汇融合引发数据迅猛增长, 数据已成为国家基础性战略资源, 世界各国都将发展大数据产业作为重要的经济发展战略。2015年8月, 国务院发布《促进大数据发展行动纲要》, 将大数据纳入国家战略, 其中明确指出“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”^[1], 大数据相关专业人才培养成为当前高校的一个重要任务。叶鹰等^[2]指出伴随大数据的热潮, 数据科学正成为继OA、iSchool运动之后又一影响信息科学的热流。图书情报档案相关专业如何结合自身专业特色、学科特色开设大数据相关专业, 形成新的学科发展增长点, 成为一个值得重点关注的问题。

1 相关研究

国内学者对数据科学教育的研究主要集中在两个方面。①对比分析国内外高校针对数据科学在课程设置、教学方式上的差异, 为国内高校的数据科学人才培养提供建议。如迪莉娅^[3]探讨数据科学家的含义及所具备的素质, 分析国内外高校在数据科学专业硕士课程设

置上的特点, 从人才培养模式、人才培养方向和授课方式三个方面为我国高校数据科学硕士课程设置提供建议; 许嘉等^[4]从教师授课、课程实验、课后作业、课程项目、课外阅读五个方面探讨哈佛大学数据科学人才培养体系。②对比分析高校对数据科学人才培养的概况与工业领域对数据科学人才的实际要求, 为数据科学人才培养提供建议。如朱扬勇等^[5]认为目前的数据科学家分为“从事商业数据分析的人”“从事科学数据分析的人”及“研究数据的人”, 针对每一类的数据科学家给出与之对应的知识结构与培养方式; 马海群等^[6]将数据人才分为数据技术人才、数据管理人才、数据安全人才、数据分析人才、数据政策人才、数据开放人才与数据科学家7类, 认为数据科学家至少具备上述2项基本技能, 并以贵州省为例, 从政策扶持、高校人才培养、产学研结合三个方面探析大数据时代我国数据人才的培养路径, 同时, 就图书情报学科领域如何培养数据人才问题提供思路; 陈振冲等^[7]调查了世界范围内优秀高校数据科学人才培养的概况, 并与工业界对数据人才的实际要求进行对比, 根据两者间差异, 建议国内高校在数据科学人才培养方面应当大力支持基础学科的学习, 并在硕

^{*}本研究得到湖北省高等学校省级教学研究项目 (编号: 2017039) 和中央高校科研业务费专项基金项目 (编号: CCNU16A02035) 资助。

士阶段开设专门的数据科学硕士培养计划。

本文将参考借鉴已有研究,重点对所调查高校的数据进行更细粒度分析,并结合国内硕士专业申请及人才培养的实际需求,尝试为国内高校数据科学专业的建设提供参考借鉴。

2 调查对象与研究方法

为积极应对当前社会对大数据领域人才的需求,全球范围内已有200多所高校开设了与大数据相关的博士和硕士专业^[8]。美国2012年启动“大数据研究与开发计划”,旨在加强其在数据科学领域的研究进程,此外,美国国家科学基金会已经实施一系列措施来鼓励相关机构资助高校开设数据科学专业课程^[9]。在此背景下,美国多所高校相继开设数据科学专业的硕士课程。基于美国高校数据科学专业办学早、课程设置较成熟等方面考虑,本文选取美国开设数据科学硕士课程的高校进行调查研究。

北卡罗莱纳州立大学高级数据分析研究院统计了全美100多所高校的大数据相关硕士学位,并将其分成4类:①数据分析硕士学位(Master of Science in Analytics);②数据科学硕士学位(Master of Science in Data Science);③商业数据分析硕士学位(Master of

Science in Business Analytics);④其他学科的硕士学位(Other M.S. Programs, Tracks and Concentrations),但主修方向是数据分析^[10]。本文将聚焦数据科学专业建设及人才培养,因此,仅选取其提供的18所设有数据科学硕士学位的高校作为研究对象,具体包括纽约大学^[11]、伊利诺斯理工学院^[12]、弗吉尼亚大学^[13]、伍斯特理工学院^[14]、圣彼得大学^[15]、路易斯大学^[16]、圣托马斯大学明尼苏达分校^[17]、南达科他州立大学^[18]、艾姆赫斯特学院^[19]、罗切斯特大学^[20]、印第安纳大学伯明顿分校^[21]、北卡罗莱纳大学夏洛特分校^[22]、哥伦比亚大学^[23]、南卫理公会大学^[24]、威斯康星大学^[25]、加州大学伯克利分校^[26]、卡内基梅隆大学^[27]和加州大学圣地亚哥分校^[28]。对18所高校网站进行逐一访问,主要采集关于培养要求、课程设置方面的信息,对此进行分类、统计、分析与总结。

3 培养要求

研究生培养方案是研究生教学管理的准绳和质量评估的重要依据^[29];而培养要求又是培养方案必不可少的组成部分,它明确了学位获取必须要满足的条件,包括学制、学分、学时和学费等。如表1所示,18所高校对数据科学硕士学位的培养要求。

表1 18所高校培养要求一览表

学校名称	学 制	学 分	学时/月	估计学费/美元
纽约大学	FT/PT	36	24	54 000
伊利诺斯理工学院	FT/PT/O	34	12~24	42 500
弗吉尼亚大学	FT	36	11	25 500 (r)、40 600
伍斯特理工学院	FT/PT	33	≥12	45 100
圣彼得大学	FT/PT	36	18	35 600
路易斯大学	FT/PT/O	36	18~24	28 300
圣托马斯大学明尼苏达分校	PT	36	8~18	33 600
南达科他州立大学	FT/PT	30	≥12	9 900 (r)、17 600
艾姆赫斯特学院	O	30	24	23 500
罗切斯特大学	FT	30	12	43 300
印第安纳大学伯明顿分校	FT/PT/O	30	16	11 500 (r)、35 500
北卡罗莱纳大学夏洛特分校	FT/PT	33	15~24	20 600 (r)、39 100
哥伦比亚大学	FT/PT	30	≥9	21 772 (r)、53 300
南卫理公会大学	FT/PT/O	31	18~24	52 800
威斯康星大学	O	36	≥12	29 700
加州大学伯克利分校	FT/PT/O	27	12~20	60 000
卡内基梅隆大学	FT	144*	16	65 000
加州大学圣地亚哥分校	PT	38	24	36 500

注: FT表示全日制教育, PT表示在职教育, O表示网络在线教育, r表示州内居民。卡内基梅隆大学每门课程为12学分, 其他高校每门课程为2~3学分。

学制分为全日制教育、在职教育和网络在线教育3种模式,其中,77.8%的高校提供全日制教育,72.2%的高校提供在职教育,38.9%的高校提供网络在线教育,27.8%的高校同时提供3种模式供学生选择;在学分的设置上,88.9%的高校为30~38学分,仅有加州大学伯克利分校为27学分,卡内基梅隆大学为144学分;在学时的设置上,88.9%的高校学时为12~24个月,仅有哥伦比亚大学(≥ 9 个月)和弗吉尼亚大学超过(11个月)的学时低于12个月;数据科学硕士专业的学费普遍在2万美元以上,同时,部分高校(弗吉尼亚大学、南达科他州立大学、印第安纳大学伯明顿分校、北卡罗莱纳大学夏洛特分校、哥伦比亚大学)针对生源不同有不同的收费标准。

4 课程设置

4.1 课程数量

美国数据科学硕士课程由专业必修课和专业选修课构成。以下从4个方面剖析18所高校的课程数量及必修课和选修课的配置。

(1) 课程总数量。艾姆赫斯特学院网站没有提供选修课课程列表,卡内基梅隆大学的选修课无法准确统计,其他16所高校共开设课程650门。其中,纽约大学开设的课程数量最多(236门),哥伦比亚大学开设的课程数量最少(9门)。除上述4所高校外,平均每所高校开设的课程为28.9门,其中62.5%的高校课程数量分布在10~20门。

(2) 必修课数量。必修课的课程设置方式有两种:一是高校开设固定数量必修课,要求学生全部修满;二是高校提供模块化或多个研究方向的课程,学生需从不同模块或方向中选择课程进行学习。如印第安纳大学伯明顿分校的必修课程由数据分析与统计、数据生命周期、数据管理与基础构建、领域应用四个模块组成,卡内基梅隆大学的必修课程分为系统方向、分析方向、以人为中心的方向,伍斯特理工学院的必修课程由数据课程理论、数理分析、数据处理与数据管理、数据分析与数据挖掘、商业智能与案例分析五个模块组成。调查的高校共开设必修课程202门,除选择第二种必修课课程设置方式的印第安纳大学伯明顿分校(44门)、卡内基梅隆大学(25门)和伍斯特理工学院(14门)外,平均每所高校开设的必修课程为7.9门。

(3) 选修课数量。调查的高校共开设选修课程482

门,其中,纽约大学开设的选修课程数量最多(231门),圣彼得大学、南达科他州立大学、威斯康星大学没有设置选修课程。除6所高校(纽约大学、圣彼得大学、南达科他州立大学、威斯康星大学、艾姆赫斯特学院、卡内基梅隆大学)外,平均每所高校开设的选修课程为20.9门。

(4) 必修课与选修课的配置。56.3%的高校选修课程数量多于必修课程数量;31.3%的高校选修课占比集中在50%~60%;25.0%的高校选修课占比超过80.0%,分别是纽约大学、罗切斯特大学、伍斯特理工学院和伊利诺斯理工学院。

4.2 必修课设置

专业必修课的教学内容能较好地反映专业设置及专业教学的主要特点,为深入了解18所高校数据科学硕士专业的特点,本文将从课程类别及课程内容两个维度进行分析。

4.2.1 必修课程类别分析

目前国内外学者已对数据科学课程体系进行了探讨,如Cleveland^[30]认为数据科学课程应包含6类,并给出各个类别的资源配置情况,即跨领域科学研究(25%)、数据建模和方法(20%)、数据计算(15%)、教学方法(15%)、工具评价(5%)和理论(20%);朱扬勇等^[5]将数据科学的学科结构分为数据科学理论研究、数据探索、科学研究的数据方法、数据技术研究四个部分;陈振冲等^[7]根据数据科学的定义,将数据科学硕士培养分为相关基础学科学习、知识发掘方法学习、数据科学理论在大数据背景下的应用及数据科学在不同行业中的应用四个部分。本文参考以上文献,并结合美国数据科学硕士专业课程的实际设置情况,将必修课程分为4类。①理论类,包括数学与统计学原理、数据科学概论、计算机科学概论等基础理论课程。②技术类,包括系统构建、数据计算,以及数据采集、存储、组织、检索、应用等过程所需的技术。③应用类,指数据科学的跨领域应用研究课程。④实践类,包括课程实验、专题研讨与实用技能培养相关的课程。

不同类别课程数量在一定程度上反映高校对该类课程的重视程度、细分程度,以及课程设置体系的侧重点^[31]。调查的高校共开设必修课程202门,其中,理论类课程47门、技术类课程122门、应用类课程14门、实

践类课程19门。其中, 55.6%的高校理论类课程占比不低于30%, 83.3%的高校技术类课程占比不低于50%。由此表明, 美国数据科学硕士专业重视基础理论知识和学生技术能力的培养。应用类及实践类课程在数据科学专业人才培养中具有重要作用, 但相对来说, 基础理论知识(如统计学、数学等)方面的训练是实践应用的重要支撑, 由于学分等方面的限制, 本文所调研高校的应用实践类课程普遍开设较少。

从课程体系来看, 4所高校(弗吉尼亚大学、圣彼得大学、印第安纳大学伯明顿分校、威斯康星大学)的课程体系较为完善, 覆盖理论类、技术类、应用类与实践类课程。如印第安纳大学伯明顿分校开设必修课程44门, 其中, 理论类课程9门、技术类课程22门、应用类课程2门、实践类课程11门; 此外, 大部分高校的课程仅覆盖理论类与技术类课程; 北卡罗莱纳大学夏洛特分校仅设置了技术类课程。

4.2.2 必修课程内容分析

必修课程的设置体现高校对数据科学专业知识的基础要求, 直接影响学生的知识结构。本次调查未能获取到每门必修课程的具体教学大纲, 以下仅对18所高校必修课程的课程名进行汇总分析, 统计结果见表2。

18所高校都非常重视数学与统计学原理、机器学习、数据挖掘、数据分析等课程。其中, 77.8%的高校开设数学与统计学原理的相关课程, 内容包括数学与统计学方法、统计学概论、应用统计学、概率论等; 另外, 72.2%的高校开设机器学习课程, 66.7%的高校开设数据挖掘课程和数据分析课程, 50.0%的高校开设大数据课程。

4.3 选修课设置

对提供完整选修课课程数据的13所高校进行分析发现, 高校的选修课程类目多、涉及面广、层次清晰且细分程度高, 课程涵盖数据科学在生物学、医学、地理学、社会科学、天文学、城市科学、商业、电气等领域的应用。其中, 61.5%的高校开设数据科学在生物学领域的应用课程, 如纽约大学开设的“生物信息学: 序列分析”课程涵盖核酸与蛋白质序列的计算机表示, 核酸与蛋白质序列的可用数据库、数据库搜索方法, 核酸与蛋白质序列的翻译和密码使用, 原核生物与真核生物的基

因组织与基因结构, 可用程序和Web资源的概述, 旨在利用数据分析方法与工具对大量的序列数据进行处理。生命科学领域的数据结构复杂多变, 包括基因序列、分子结构、不同物种等, 如何对复杂的数据进行分析处理, 使其更好地为人类健康服务是生命科学领域关注的话题。另外, 38.5%的高校开设数据科学在商业领域的应用课程, 如伍斯特理工学院开设的“数据挖掘在业务分析中的应用”课程为学生提供了将原始数据转化为商业智能数据的关键方法和工具, 解决现实业务问题, 包括客户细分、客户偏好及最新的信用评分、流失管理和欺诈检测等问题。选修课中涉及数学、计算机科学与数据科学的课程, 大多是必修课的补充或延伸课程, 帮助学生更深入地掌握数据科学领域的知识。

4.4 社会实践课程

88.9%的高校要求学生在结业之前完成类似于国内的毕业设计、毕业实习的顶点项目(capstone), 学生以小组或个人的形式直接参与项目, 待项目完结后上交项目报告并进行项目汇报。如纽约大学要求学生工业、政府机构、非政府机构或学术界存在的真实问题进行研究, 学生需要参与项目的整个流程, 从数据采集、数据处理、数据分析到提供解决方案。同时, 学校为每个团队配备1名项目咨询老师, 并且大部分的原始数据由项目赞助人(企业、机构等)提供。南卫理公会大学和加州大学伯克利分校为学生提供3~4天的深入课程(immersion course), 在此期间, 学生将有机会与行业领导者、数据科学家及数据科学领域的教师深入地交流与学习。

4.5 课程设置特点分析

(1) 各高校开设课程数量差异较大。18所高校平均开设课程数量为28.9门, 但课程数量不低于40门的高校有5所, 其中2所高校的课程数量超过100门; 课程数量不高于15门的高校有8所, 而不高于10门的有2所, 可见美国各高校在数据科学课程数量的设置上存在较大差异, 分析发现, 这与数据科学专业的设立时长、学校对数据科学的重视程度及学校的综合实力等相关。

(2) 必修课程注重技术与应用。学校重视培养学生对数据处理、数据分析与数据管理技术的学习, 典型的技术类课程包括机器学习、数据库、数据挖掘和数据

表2 18所高校必修课程统计表

学校名称	必修课程名称
纽约大学	数据科学入门, 数学与统计学, 知识表示与推理, 机器学习与计算统计, 大数据
伊利诺斯理工学院	数理统计, 应用统计学, 面向机器学习的统计学, 项目管理, 高级数据库, 密集型数据计算, 机器学习, 数据分析, 参与公众活动的数据科学家
弗吉尼亚大学	面向数据科学的线性模型, 计算机基础, 面向数据科学的系统编程, 面向数据科学的统计计算, 数据挖掘, 机器学习, 数据科学实践与应用I, 数据科学实践与应用II, 大数据伦理学I, 大数据伦理学II
伍斯特理工学院	数据科学理论(数据科学导论), 数理分析(面向数据科学的统计方法), 线性回归, 应用多变量分析, 数据处理与数据管理(数据库管理系统原理), 数据库应用与发展, 高级数据库系统, 大数据管理, 数据分析与数据挖掘(知识发现), 数据挖掘, 大数据分析, 机器学习, 商业智能与案例分析(商业智能), 数据挖掘在商业领域中的应用
圣彼得大学	数据科学导论, 数据分析与决策建模, 大数据与数据管理, 统计编程: SAS与R, 数据挖掘, 大数据分析, 机器学习, 数据可视化与商业智能, 预测分析与金融建模, 市场分析与运筹学, 数据科学中的法律, 伦理与隐私
路易斯大学	面向数据科学家的数学, 统计学概论I, 数据挖掘与数据分析导论, 统计编程, 加密与认证, 数据可视化, 大数据存储系统设计与应用, 机器学习
圣托马斯大学明尼苏达分校	软件开发基础, 软件工程, 数据库管理系统与设计, 数据分析基础, 数据挖掘与预测分析, 大数据架构, 大数据管理, 机器学习, 数据分析与可视化, 数据仓库
南达科他州立大学	现代应用统计学I, 现代应用统计学II, 时间序列分析, 运筹学, 非参数统计, 统计编程, 大数据分析, 数据挖掘, 数据分析与编程, 预测分析
艾姆赫斯特学院	面向数据科学的定量方法, 面向数据科学的高阶分析方法, 数据仓库, 数据挖掘项目, 大数据高级编程语言, 高级数据挖掘与分析, 数据挖掘与商业智能
罗切斯特大学	统计计算概论, 统计计算方法, 数据挖掘, 数据库系统
印第安纳大学伯明顿分校	数据分析与统计(算法设计与分析), 基于神经科学与基因学的人工智能, 数据挖掘, 计算机符号建模, 人工智能专题, 面向信息学的编程, 信息学专题研讨之数据可视化, 信息学专题研讨之关系概率模型, 信息检索, 信息可视化, 信息与图书馆学专题研讨之大数据背景下的信息学, 统计学概论, 探索性数据分析, 统计学习与高维数据分析, 数据生命周期(人工智能原理), 机器学习, 应用机器学习, 大数据的管理, 访问与使用, 信息学专题之复杂的系统, 数据管理专题研讨, 学术交流, 语义数据, 数字图书馆, 数据管理与基础构建(分布式系统), 基于知识的人工智能, 高级数据库概论, 系统专题研讨之云计算, 数据库理论与系统设计, 数据库与信息系统专题研讨之大数据管理, 操作系统, 计算机网络, 网络安全, 大数据软件开发与应用, 信息安全与协议, 信息学专题研讨之复杂网络及其应用, 数据库, 领域应用(Web挖掘), 高性能计算, 生物信息学导论, 大数据应用与分析, 机器学习在生物信息学中的应用, 组织信息学与经济安全, 信息学专题研讨之医学中的大数据
北卡罗莱纳大学夏洛特分校	数据分析与可视化, 机器学习, 数据库系统, 商业竞争中的大数据分析(分析工具), 商业智能分析, 商业分析与决策, 消费者行为分析
哥伦比亚大学	概率论, 统计推断与建模, 数据科学算法, 计算机系统, 机器学习, 探索性数据分析与可视化
南卫理公会大学	实验统计I, 实验统计II, 数据科学概论, 抽样统计, 文档组织与数据库管理, 数据与网络安全, 数据挖掘, 信息可视化, 量化的世界: 数据分析方法与技术
威斯康星大学	数据科学基础, 统计方法, 数据科学与战略决策, 计算机编程语言: Python与R, 数据仓库, 大数据: 高性能计算, 数据挖掘, 数据可视化与非结构化数据的分析, 规范分析, 数据交流, 数据科学伦理学
加州大学伯克利分校	统计学, 数据存储与检索, 应用机器学习, 数据分析, 数据可视化与交流
卡内基梅隆大学	系统方向(操作系统), 数据库, 并行计算机体系结构与编程, 云计算, 分布式系统, 高级云计算, 高级数据库, 存储系统, 多媒体数据库与数据挖掘, 高级分布式操作系统, 移动计算与普适计算, 分析方向(机器学习), 文本挖掘, 高级机器学习, 机器学习与大数据, 智能信息系统设计, 云计算, 信息系统项目, 搜索引擎, 多媒体数据库与数据挖掘, 机器学习与大数据, 大数据分析, 多媒体信息分析, 以人为中心的方向(行为研究方法), 人机交互
加州大学圣地亚哥分校	概率统计, 数据分析, 数据管理系统, 机器学习, 数据分析: Hadoop与Spark, 数据集成, 数据科学案例研究

建模,旨在帮助学生掌握Python/Java/Pig/Hive等编程语言,提高学生应用分布式文件系统工具的能力,提高学生的数据分析能力。仅从必修课程的设置情况来看,18所高校设置必修课程202门,其中技术类课程数量为122门,占比达60.4%。

(3) 选修课程面向特定领域开设。18所高校都十分重视数据科学理论、方法、技术在特定领域的应用,选修课程涉及的领域包括生物学、医学、地理、社会科学、天文学、城市科学、商业、电气等,这类细分课程适应数据科学相关职位对特定领域知识的需求,有利于培养更专业、更具职业竞争力的人才。其中,罗切斯特大学的课程涉及地理、生物学、医学、商业、社会科学等多个领域。

(4) 跨学科与多学科交叉培养。数据科学涉及的方法来源于多个学科领域,是统计学、计算机科学、信息管理等学科领域原有方法的继承、扩展或创新,数据科学虽然在研究方法上的表达和应用有一定差异,但具有极强的跨学科和多学科特性已成为学界共识^[32]。基于此,许多学院选择联合办学模式来培养数据科学人才,尤其以计算机学院、数学与统计学院、信息管理学院、商学院、工程学院间的联合办学最多。同时,学院鼓励学生跨学科、跨专业选修,从不同的学科角度共同构建数据科学专业的知识结构。纽约大学的选修课涉及生物学、统计学、商学、政治科学、城市科学等多个学科领域。

(5) 教学方式多样。美国高校的数据科学硕士课程教学方式多样,包括案例分析、专题研讨、前沿讲座和实践项目。案例分析与专题研讨课就某一特定问题组织学生进行讨论,充分激发学生的学习热情,调动学生的积极性和主动性;前沿讲座邀请领域内的专家学者讲授,让学生了解专业的前沿进展;实践项目要求学生对企业、政府机构、非政府机构或学术界存在的真实问题进行研究并提出解决方案,有助于提高学生的应用能力。

5 对我国高校开设数据科学硕士专业的启示

(1) 探索多层次、多类型的数据科学专业人才培养模式。大数据产业的发展,加大了市场对数据科学人才的需求。在2016年中国大数据产业峰会上,清华大学武永卫教授指出:未来3~5年,中国需要180万名数据人才,但目前只有约30万人^[33]。数据科学人才短缺的问题

亟需解决,而作为向社会培养和输送人才的基地,高校理应承担起培养合格、专业的数据科学人才的重任,探索多层次、多类型的数据科学专业人才培养模式。纵观国内高等教育,本科教育阶段通常更重视学生基础能力的培养,硕士阶段更重视对学生应用能力与学术能力的培养,基于此,建议国内高校在本科阶段适当开设与数据科学相关的通识课程,在硕士阶段设置专门的数据科学学位。同时,各高校应结合自身实力及社会需求,选择适合高校发展的学制模式(全日制教育、在职教育、网络在线教育),提供学位证明或相应的资格证书,以培养不同层次的数据科学人才。此外,高校应当针对不同类型的人才制订培养方案,对于学术型人才,应重视数据科学理论与应用能力的培养;对于应用型人才,应加强数据挖掘与数据分析等技术能力及实践应用能力的训练。

(2) 构建大数据时代新型课程体系。我国数据科学专业建设仍处于探索阶段,一方面,国内开设数据科学专业的高校较少;另一方面,数据科学专业的课程数量少且大多没有设置选修课程,导致国内很难培养出具有复合型特点的数据科学人才。在大数据时代背景下,国内高校应科学规划以数据驱动为背景,知识发现为目标,跨界复合为模式,数据挖掘、数据分析等专业技能为核心,数据管理、数据统计、数据分析等职业能力为重点的新型课程体系,以期培养出具有深厚理论基础、全面专业知识、开放国际视野、丰富人文情怀的数据科学专家。

鉴于数据科学是一门结合应用数学、统计、模式识别、机器学习、数据可视化、数据仓库及高性能计算的综合性学科,建议国内高校将数据科学硕士课程体系大致分为基础学科课程、知识挖掘方法、数据科学理论在大数据背景下的应用,以及数据科学在不同领域中的应用四个部分。数据科学的相关基础学科课程应包括高等微积分、数理统计、概率论等数学方面的内容,以及数据科学概论、计算机概论等计算机方面的内容;知识挖掘方法课程应包括统计推理、机器学习、数据挖掘和数据分析等;数据科学理论在大数据背景下的应用课程应包括大数据的高性能计算、并行计算、分布式计算、数据仓库、数据库管理等,以及Spark、Hadoop等大数据处理平台的内容;数据科学在不同领域中的应用课程可以结合学校已有的特色专业,如伍斯特理工学院和北卡罗莱纳大学夏洛特分校将数据科学与商业分析决策结合,罗切斯特大学专注于医疗分

析、认知科学和人工智能领域。

(3) 结合课程特点,建立多元化的教学与考核方式。数据科学是一门强调技术与应用的学科,建议高校在教学过程中注重将理论与实践相结合。首先,升级大数据实验设备,搭建大数据实验教学平台及虚拟化的Hadoop环境,为学生提供条件优良的实验环境。其次,结合课程特点创新教学模式。理论类的课程可通过课堂讲解、专题研讨、前沿讲座等形式开展,课堂讲解有助于学生构建新学科的认知体系;专题研讨能激发学生的学习热情,调动学生的积极性和主动性;前沿讲座邀请信息科学、数据科学、计算机科学等领域的专家学者讲授,让学生了解专业的前沿进展。部分课程可采用社会实践与科研实践等方式开展,社会实践活动能够帮助学生从就业角度了解专业定位,加深学生对数据科学相关基础理论和基础知识的理解,提高学生分析和解决实际问题的能力;科研实践能让学生学习科研方法、培养探索和创新的能力。最后,根据理论与实践教学的特点建立多元化考核方式。基础理论课程采用书面形式进行考核;实验课程重点考察学生对实验原理的理解和实际操作能力;社会实践类课程要求学生以小组形式参与项目,集体完成项目并撰写项目报告,以汇报、答辩的形式进行考核。

(4) 迎合社会需求,增强校企合作。数据科学的一个关键方面是对数据分析思维的培养,如今越来越多的企业是基于数据分析驱动的,具有较强的专业化优势,许多传统行业中的企业正在为取得竞争地位而探索新的和已经存在的数据资源^[34]。目前,我国开设数据科学硕士课程的高校较少,且高校没有充分利用政府、企业的信息资源来培养学生的实践能力。鉴于此,建议高校设立专门的校企合作部门,积极主动地开展学院与企业间的合作模式。通过校企合作,一方面可以让高校教师深入相关的企业学习,掌握更切合实际的大数据分析技术和实战能力,建立专门的数据科学专家型教师团队;另一方面,可以让企业的数据科学家深入学生课堂,学生也可以借此直接接触从事大数据工作的相关人员,掌握前沿的大数据分析技术。此外,学生还可以直接参与企业数据科学的相关工作,培养学生的实战能力,为就业积累工作经验。

6 结语

本文以美国18所开设数据科学硕士专业的高校为

调查对象,围绕培养要求、课程结构、课程层次设置进行深入对比分析,总结美国高校数据科学硕士课程设置的特点,同时为我国数据科学硕士专业的开设与发展提供建议。虽然数据人才短缺的问题在短期内无法得到彻底解决,但作为人才培育的主要阵地,高校应该义不容辞地负担起这份艰巨的任务,适时改革人才培育体系,积极探索校企合作新模式,努力实现数据人才的可持续发展。传统的图书情报领域应重视自身在这项任务中的角色,正视大数据时代带来的机遇与挑战,以在数据人才培养的新时期新任务中有所作为^[6]。

本文虽对采集的数据进行了分析总结,但仍存在不足:首先,课程名称的翻译上可能不够准确;其次,对必修课程的分类,本文仅从课程名与学校提供的课程简介判断课程所属类别,可能导致部分课程分类不够准确;最后,本研究仅以美国的部分高校作为研究对象,样本数量不大,还需要进一步采集更多高校数据科学硕士专业的课程,从而更深入地分析跟踪数据科学硕士专业的特点及发展趋势。

参考文献

- [1] 中华人民共和国国务院. 促进大数据发展行动纲要 [J]. 成组技术与生产现代化, 2015, 32 (3): 51-58.
- [2] 叶鹰, 马费成. 数据科学兴起及其与信息科学的关联 [J]. 情报学报, 2015 (6): 575-580.
- [3] 迪莉娅. 高校数据科学专业硕士课程设置研究 [J]. 教学研究, 2014, 37 (6): 39-43.
- [4] 许嘉, 吕品. 哈佛大学数据科学课程教学初探 [J]. 教育界: 高等教育研究, 2015 (5): 109-110.
- [5] 朱扬勇, 熊贲. 大数据人才培养的基础条件初探 [J]. 大数据, 2016, 2 (5): 107-114.
- [6] 马海群, 蒲攀. 大数据视阈下我国数据人才培养的思考 [J]. 数字图书馆论坛, 2016 (1): 2-9.
- [7] 陈振冲, 贺田田. 数据科学人才的需求与培养 [J]. 大数据, 2016 (5): 95-106.
- [8] SWANSTROM R. Colleges With Data Science Degrees [EB/OL]. (2012-04-09) [2018-03-17]. <http://101.datascience.community/2012/04/09/colleges-with-data-science-degrees/>.
- [9] 何海地. 美国大数据专业硕士研究生教育的背景、现状、特色与启示——全美23所知名大学数据分析硕士课程网站及相关信息分析研究 [J]. 图书与情报, 2014 (2): 48-56.
- [10] Institute for Advanced Analytics [EB/OL]. [2018-03-14].

- http://analytics.ncsu.edu/?page_id=4184.
- [11] NYU. MS in Data Science [EB/OL]. [2018-03-14]. <http://cds.nyu.edu/academics/ms-in-data-science/>.
- [12] Illinois Institute of Technology [EB/OL]. [2018-03-14]. <http://science.iit.edu/programs/graduate/master-data-science>.
- [13] UVA DSI [EB/OL]. [2018-03-14]. <http://dsi.virginia.edu/academics>.
- [14] Worcester Polytechnic Institute [EB/OL]. [2018-03-14]. <https://www.wpi.edu/academics/departments/data-science>.
- [15] Saint Peter's University [EB/OL]. [2018-03-14]. <http://www.saintpeters.edu/data-science-and-business-analytics/>.
- [16] Lewis University [EB/OL]. [2018-03-14]. <http://www.lewisu.edu/academics/data-science/>.
- [17] University of St. Thomas [EB/OL]. [2018-03-14]. <http://www.stthomas.edu/gradsoftware/programs/masters/msdatascience/>.
- [18] South Dakota State University [EB/OL]. [2018-03-14]. <http://www.sdstate.edu/mathematics-statistics/masters-data-science>.
- [19] Elmhurst College [EB/OL]. [2018-03-14]. http://www.elmhurst.edu/data_science.
- [20] School of ARTS & SCIENCES [EB/OL]. [2018-03-14]. <http://www.rochester.edu/data-science/degrees/MSdetails.html>.
- [21] Indiana University Bloomington [EB/OL]. [2018-03-14]. <http://www.soic.indiana.edu/graduate/degrees/data-science/graduate/>.
- [22] UNC CHARLOTTE [EB/OL]. [2018-03-14]. <http://analytics.uncc.edu/academics>.
- [23] Columbia University [EB/OL]. [2018-03-14]. <http://datascience.columbia.edu/master-of-science-in-data-science>.
- [24] SMU [EB/OL]. [2018-03-14]. <https://datascience.smu.edu/>.
- [25] University of Wisconsin [EB/OL]. [2018-03-14]. <http://datasciencedegree.wisconsin.edu/data-science-program/>.
- [26] Berkeley School of Information [EB/OL]. [2018-03-14]. <https://datascience.berkeley.edu/academics/>.
- [27] Carnegie Mellon University [EB/OL]. [2018-03-14]. <http://mcids.cs.cmu.edu/learn-us>.
- [28] UC San Diego Jacobs School of Engineering [EB/OL]. [2018-03-14]. <http://jacobsschool.ucsd.edu/mas/dse/>.
- [29] 葛敬民, 韩正彪. 高校图书馆硕士点研究生培养机制浅析 [J]. 图书馆理论与实践, 2009 (2): 65-68.
- [30] CLEVELAND W S. Data science: an action plan for expanding the technical areas of the field of statistics [J]. International Statistical Review, 2015, 7 (6): 420-422.
- [31] 司莉, 王思敏. 美国iSchools课程设置与能力培养的调查与分析 [J]. 大学图书馆学报, 2014, 32 (1): 102-109.
- [32] 王曰芬, 谢清楠, 宋小康. 国外数据科学研究的回顾与展望 [J]. 图书情报工作, 2016, 60 (14): 5-14.
- [33] 刘春媛. 专家: 我国数据人才缺口约150万 [EB/OL]. [2018-03-19]. <http://news.sina.com.cn/o/2016-05-31/doc-ifsqyku0026616.shtml>.
- [34] PROVEST F, FAWCETT T. Data science and its relationship to big data and data-driven decision making [J]. Big Data, 2013, 1 (1): 51-54.

作者简介

曹高辉, 男, 1980年生, 博士, 副教授, 研究方向: 信息检索、信息服务, E-mail: ghcao@mail.ccnu.edu.cn。

胡紫祎, 女, 1993年生, 硕士研究生, 研究方向: 信息服务。

郭家乐, 女, 1993年生, 硕士研究生, 研究方向: 科学教育。

王艳凤, 女, 1965年生, 副研究馆员, 研究方向: 信息服务。

Research on Training Requirement and Curriculum of Master's Degree in Data Science in United States

CAO GaoHui¹ HU ZiYi¹ GUO JiaLe² WANG YanFeng²

(1. School of Information Management, Central China Normal University, Wuhan 430079, China;

2. School of Educational Information Technology, Central China Normal University, Wuhan 430079, China)

Abstract: This paper chooses 18 colleges and universities with a master's degree in data science in the United States as the research object, and uses the literature research method, the network research method and the content analysis method, analyzes the training requirement and curriculum, and summarizes five features: the number of courses varies greatly, compulsory courses focus on technology and application, elective courses for specific areas, interdisciplinary and multidisciplinary cross-training, a variety of teaching methods. From the four aspects of personnel training mode, curriculum system construction, teaching and assessment methods, school and enterprise cooperation for China's data science master professional construction to provide recommendations.

Keywords: Data Science; Curriculum; Personnel Training; Master's Degree; Major Construction

(收稿日期: 2018-04-23)