

军队院校数据科学方向研究生课程设置探讨

谢正, 李建平, 粟进英

(国防科技大学 文理学院, 湖南 长沙 410073)

摘要: 数据是信息时代支撑国家安全和军队建设的战略资源, 数据科学与国防事业息息相关, 数据科学高水平人才培养对新形势下的军事革新具有迫切性。通过对比分析斯坦福大学、哈佛大学、麻省理工学院、加州大学伯克利分校的数据科学专业课程设置, 并结合军队院校研究生培养目标, 对新形势下具有国防特色的数据科学研究生课程设置与培养模式进行了思考。

关键词: 数据科学; 课程体系; 专业建设

中图分类号: G642 **文献标识码:** A **文章编号:** 1672-8874(2018)04-0084-04

Discussion on Postgraduate Courses of Data Science in Military Universities

XIE Zheng, LI Jian-ping, SU Jin-ying

(College of Liberal Arts and Sciences, National University of Defense Technology, Changsha 410073, China)

Abstract: In the information era, data are treated as strategic resources, which are important to national security and army building. Data science is closely related to national defense technologies. Researchers mastering in data science can improve the level of defense technology. We analyzed the courses of data science in four universities, which are Stanford University, Harvard University, Massachusetts Institute of Technology and University of California at Berkeley. Based on specific training goals of military universities, we discussed course settings and training patterns of the postgraduates majoring or minoring in data science.

Key words: data science; curriculum system; major construction

一、引言

数据无处不在, 其规模与重要性日益剧增。我们正在对过去数个世纪积累的信息进行数字化, 同时从互联网、移动设备、传感器等现代信息载体中收集各类数据。识别数据中的模式与规律, 可创造商业和社会价值, 可加深对物理、生物、人类社会等领域中的许多复杂系统的理解。在分析理解数据的过程中, 产生了一系列新理论、新方法, 并成功应用于工业、科技、经济、文化等

各个领域, 催生了一门新的交叉科学。它运用统计、计算数学等方法分析各种形式的数据, 从中挖掘知识, 被称为数据科学^[1]。

企业、公共机构以及非营利组织对从事数据分析的专业人员有着越来越大的需求, 但这些人员数量有限。因此, 数据科学专业已成为了目前高等教育的热门方向。许多知名大学设置了数据科学硕士专业, 如斯坦福大学、哈佛大学、麻省理工学院、加州大学伯克利分校等。清华大学、复旦大学等高等院校都设置了硕士学位。2016年以来, 北京大学、中国人民大学等几十余所高校

增设了数据科学与大数据技术本科专业^[2]。这些方向的学生培养依托于与数据科学紧密相关的统计、计算机科学等相关院系,以及院校所成立的数据科学研究机构,如北京大数据研究院。

数据科学与国防事业息息相关。随着信息时代的到来,信息化、智能化成了现代战争的特征。在陆、海、空、天、电、网等多维空间展开的多军兵种一体化行动中产生了规模庞大的多源异构数据,它们是支撑国家安全和军队建设的战略资源^[3-4]。“得失之道,利在先知,谋胜于未胜,慎于未失者,善也。”如何深度挖掘收集到的数据价值是取胜的关键。因此,开展一场新形势下的军事革新,缔造一支打得赢的信息化、智能化人民军队势在必行。具有国防特色的数据科学人才培养是完成这一使命的有力保障,是打造高素质新型军事人才培养和国防科技自主创新高地的重要组成部分^[5-6]。因此军队对数据科学人才的需求十分迫切。

由于军队所关心的数据及其分析目的、分析精度上具有特殊性、其相关人才培养目的与能力要求不同于地方院校。军队所需的数据科学高层次人才能够胜任大规模文本分析、传感器网络数据分析、多源异构军事数据融合分析等众多贴近实战的岗位。而地方院校更关心与学生就业相关的工业与商业等领域中的数据分析,其专业所培养的学生不能完全满足军队需求。

为此,我们调查分析了四所世界一流大学的数据科学专业课程设置与培养方案,重点分析了所开设的课程与国防科技的相关性,在此基础上探讨了军队院校数据科学方向研究生的课程设置与培养模式。

二、课程调研

我们以四所一流院校的数据科学硕士专业为调研对象,分析其课程体系、人才培养目标与培养模式。

1. 加州大学伯克利分校。该校信息学院设置数据科学专业,授予信息与数据科学硕士学位,学制2年。该校网页列出了其课程设置与学生培养方案^[7]。具体而言,培养目标是培养学生具备统计学、信息和计算机科学等方面的知识,以及数据

处理方面的能力。该专业开设以下3类课程:

(1) 基础课程:面向数据科学的Python、面向数据分析的研究设计及其应用、面向数据科学的统计、数据存储与检索、应用机器学习。

(2) 高级课程:试验与因果分析、数据背后的人与价值、大数据、离散响应数据和时间序列及面板数据的统计方法、大规模机器学习、面向自然语言处理的深度学习、数据可视化。

(3) 顶点课程:学生将分成3-4人一组完成一个具有实际背景的数据科学项目,综合考察对真实数据的分析与社会解释能力,以及书面和口头表达能力。

2. 哈佛大学。该校统计系设置数据科学专业,授予数据科学理学硕士学位。学制2年。目标是培养数据科学专家,注重培养学生预测和统计计算、设计实验、建模复杂的依赖关系(网络、基因、数据流)、理解和量化因果关系等方面的能力。该校网页列出了其课程设置与学生培养方案^[8]。

具体而言,该专业开设的课程分为必修课和选修课两类。学生必须在3学期内完成修完12门课程,包括4门技术核心课程、1门关于数据科学思维的课程、1门研究实验型课程,以及6门选修课。这些选修课由计算机科学硕士专业课程(1-4门)与统计硕士专业课程(1-4门),其他相关课程(生物医学信息硕士专业课程,健康数据科学专业硕士课程)(0-4门),研究性课程(0-1门),研讨性课程(0-1门)组成。

必修课程分为4个模块:

(1) 应用计算课程:数据科学1(数据科学导论,技术核心课程)、数据科学2(数据科学前沿问题,技术核心课程)、计算科学与工程顶点项目(研究实验)。

(2) 应用数学课程:高级科学计算(数据分析、推理和优化的随机方法,技术核心课程)。

(3) 计算机科学课程:计算科学的系统开发(技术核心课程)。

(4) 数据科学课程:数据科学中的批判性思维。

选修课分为统计与计算机科学2个模块:

(1) 统计课程:计量经济学、时间序列与预测、线性模型、广义线性模型、实验设计、概率I-II、统计推断I-II、贝叶斯数据分析、序贯

决策。

(2) 计算机科学课程: 数据结构与算法、数据系统、数据可视化、机器学习、人工智能、计算机科学的计算基础、分布式计算导论、大数据系统、高级机器学习、机器学习选讲、面向自然语言处理的机器学习。

3. 麻省理工学院。该校统计中心开设数据科学辅修专业, 授予数据科学理学硕士学位。任何专业的研究生都可以申请辅修。该校网页列出了其课程设置与学生培养方案^[9]。具体而言, 该专业设置了五个模块的课程, 学生需要完成其中的6门课, 目标是培养学生具备概率统计、计算等方面的知识, 以及数据处理方面的能力。

(1) 基础课程(选1): 工科数学之线性代数与常微分方程、电子工程与计算机科学导论 I、Python 中的计算机编程导论、计算思维与数据科学导论、微分方程、线性代数。

(2) 统计模块1(选1): 工程中的不确定性、概率导论 I-II、经济学中的统计方法导论、概率与随机变量。

(3) 统计课程2(选1): 经济学、统计思维与数据分析、应用统计。

(4) 计算与数据分析课程(选2): 工程计算与数据科学、面向力学工程师的数值计算、推断导论、机器学习导论、计算与系统生物学基础、计算机视觉进展、面向脑与认知科学的统计、面向社会学家的数据分析、高等经济学、商业分析中的优化方法、应用概率导论、统计与概率、数据分析中的矩阵方法、信号处理、机器学习、数学在金融中的应用选讲。

(5) 顶点课程(必修): 统计、计算与应用。

4. 斯坦福大学。该校统计系和计算与数学学院联合设置数据科学硕士研究生专业。该校网页列出了其课程设置与学生培养方案^[10]。具体而言, 该专业培养学生的数学、统计、计算与编程能力。课程分为五个模块。毕业需要修完45学分, 36学分学习要获得成绩, 平均学分成绩不低于3.0。毕业没有论文要求, 但需获得导师同意。

(1) 基础课程(需12学分, 3学分一门, 核心课程, 须有考试成绩): 数值线性代数、离散数学与优化算法、工程中的随机方法或者概率分析中的随机算法。

(2) 数据科学选修课(需12学分, 3学分一门, 须有考试成绩): 统计推断导论、回归模型与方差分析导论或者统计模型导论、现代应用统计之学习、现代应用统计之数据挖掘, 或由导师批准的同等级课程。

(3) 高等科学编程与高性能计算核心课程(需6学分, 3学分一门, 须有考试成绩), 包含高等科学编程(需3学分): 如面向科学家和工程师的高级软件开发; 并行与高性能计算(需3学分): 如 MPI, open MP, and CUDA 等并行计算技术导论、分布式算法与优化、数值分析中的并行方法、并行计算、并行计算机体系结构与程序设计、高级多核系统。

(4) 专业选修课(需9学分, 3学分一门): 计算分子生物学的表示和算法、数据驱动医学、现代生物学、现代统计学、社会信息网络分析、机器学习、面向视觉识别的卷积神经网络、大规模数据集挖掘、计算机图形学专题、地质统计学、大数据中的商业智能、人类神经影像学方法、面向数据科学的计算。

(5) 实用性课程(需6学分): 由导师指导的顶点项目, 需学生有较强动手能力的实用性课程等。

三、课程体系分析

数据科学是一个数学、统计学、计算机科学等学科交叉形成的学科。上述四校的课程设置充分体现了这一特点, 其课程可分为基础与应用两类。基础类课程由与数据科学相关的应用数学、计算数学、统计学、计算机科学课程构成。应用类课程由金融数据分析、生物数据分析、自然语言处理、图像处理等课程构成, 涉及商业、社会科学、工程等多个应用领域。

军队院校人才培养目标与国防数据的特点赋予其数据科学课程设置的自身特点。传感器数据分析、多源异构的战场数据融合分析、图像处理、网络舆情分析、军事教育数据分析, 以及相关的数据储存、检索、管理等问题构成了从事数据分析的军校学术型与专业型研究生培养方向。基于这些特点, 以及借鉴上述四校的数据专业课程设置, 可初步得出军校数据科学专业课程设置模

式,包括以下五个课程模块:

(1) 应用与计算数学课程:数值线性代数、离散数学与优化算法、面向数据科学的优化、博弈论、计算方法等。

(2) 概率统计课程:概率、统计、贝叶斯分析、不确定性度量、实验设计等。

(3) 计算机科学课程:面向数据科学的Python、分布式计算、机器学习等。

(4) 应用类课程:面向自然语言处理与面向图像数据的深度学习、数据可视化、社会网络分析等。

(5) 顶点课程:通过组队完成一个研究生课题组项目(例如,大规模文本数据分析等),开展综合实践训练,综合考查学生对真实数据的分析能力。

军队院校数据科学专业研究生培养可采用学分制,并参考以上大学的课程设置方案,设置必修与选修两类课程。必修的核心课程可以包括:面向数据科学的统计、优化、Python以及机器学习。选修课程可以根据学生所在课题组的研究方向针对性开设自然语言处理、图像处理、数据管理、社会网络分析等课程。根据学生的科研选题,由导师指导,组队完成相应的顶点课程。这样的课程设置,保证了对学生基础知识的训练,同时也针对性地加强了学生动手能力。学术型研究生培养偏重理论方面的课程,而专业型研究生培养偏重于针对具体问题的应用类课程。

需要注意的是,数据科学专业课程虽大部分源于应用数学、统计学和计算机科学这三个专业的课程,但内涵改革幅度较大,内容上有其自身特点,需要根据其特点设计教学大纲,强化学生在大规模数据的分析、计算与管理等方面的能力训练。以统计学为例,经典统计学主要研究在信息不确定条件下如何做出推论与决策,而对数据规模所带来的可计算性、复杂性,以及数据质量等问题不做考虑。正是这些问题引导了一系列统计学和计算机科学交叉研究的开展,建立了一系列具有应用价值的理论方法,被归类为数据科学,并在此基础上,构成了数据科学的光明未来。因此数据科学应该是作为一个新增方向来设置,而不是单纯的对原有课程进行内涵改革。

此外,博弈论与军事,以及人工智能关系紧密,可开设面向数据科学的博弈论课程。博弈论在可解释性等方面促进了神经网络,深度学习的发展。深度学习等智能算法又能促进博弈论可计算性的发展。博弈论也成了计算机科学家的主要研究课题,所产生的研究领域融合计算和博弈模型,被称为算法博弈论,发展迅速,在人工智能,无人机等领域有广泛的应用前景。算法博弈论的主要研究领域包括各种均衡的计算、行为预测、资源分配及调度等问题。数据或信息缺失条件下的行为预测是数据科学研究中常需解决的问题。因此面向数据科学的博弈论课程需要包括:信息缺失下的非合作博弈近似Nash均衡智能算法、信息缺失下的决策行为预测、信息缺失下的合作博弈行为演化模型等内容。

四、结论

信息化、智能化的现代战争离不开对战场数据的深度分析。因此,针对国防数据特点,培养数据科学方向高水平人才具有迫切性,是抢占“两个高地”,推动军队信息化、智能化建设,提升军队核心战斗力的有力保障。我们围绕军队对数据科学高层次人才的需求,借鉴了斯坦福大学等四所国际一流院校的数据科学专业设置特点,给出了军队院校数据科学研究生的课程设置与人才培养模式的一些建议。

参考文献:

- [1] Hey T. The fourth paradigm-data-intensive scientific discovery[J]. Proceedings of the IEEE,2011(8):1334-1337.
- [2] 朝乐门,邢春晓,王雨晴.数据科学与大数据技术专业特色课程研究[J].计算机科学,2018(3):1-8.
- [3] 胡晓峰.大数据时代对建模仿真的挑战与思考[J].军事运筹与系统工程,2013(4):5-12.
- [4] 伍文峰,胡晓峰.基于大数据的网络化作战体系能力评估框架[J].军事运筹与系统工程,2016(2):26-32.
- [5] 曾生跃,张立杰,杨俊,等.“双一流”建设背景下研究生课程体系建设的探索与实践——以国防科技大学为例[J].高等教育研究学报,2017(4):25-29.