

序

FOREWORD

数据科学与工程核心课程的系列教材终于要面世了,这是一件鼓舞人心的事。作为华东师范大学数据学院的发起者和见证人,核心课程和系列教材一直是我心心念念的事情。值此系列教材出版发行之际,我很高兴能被邀请写几句话,做个回顾,分享一些感悟,也展望一下未来。

借着大数据热的东风,依托何积丰院士在 2007 年倡导成立的华东师范大学海量计算研究所,2012 年 6 月在时任 SAP 公司 CTO 史维学博士(Dr. Vishal Sikka)的支持下,我们成立了华东师范大学云计算与大数据研究中心。2013 年 9 月,学校发起成立作为二级独立实体的数据科学与工程研究院,开始在软件工程一级学科下自设数据科学与工程二级学科,开展博士研究生和硕士研究生的培养工作。在进行研究生培养的探索过程中,我们深切感受到计算机类的本科生人才培养需要反思和改革。因此,到了 2016 年 9 月,研究院改制成数据科学与工程学院,随后就开始招收数据科学与工程专业本科生,第一届本科生已于 2020 年毕业,这就是我们学院和专业的简单历史。经过这么几年的实践和思考,我们越发坚信当年对“数据科学与工程”这一名称的选择,“数据学院”和“数据专业”已经得到越来越多的认可,学院的师生也逐渐接受“数据人”这一称呼。

这里我想分享以下几方面的感悟:为什么要办数据专业?怎么办数据专业?教材为什么很重要?对人才培养有什么贡献?

为什么要办数据专业?数据是新能源,这是大家耳熟能详的一句话。说到能源,我们首先想到的是石油,所以大家就习惯把数据比喻成石油。但是,在我们看来,“新能源”对应的英文应该是“New Power”。“Data is Power”,这是我们的基本信念,也是我们要办数据学院的根本动机。数据是人类文明史上的第三个重要的 Power,之前的两个 Power 是蒸汽能(Steam Power)和电能(Electric Power),它们分别引发了第一次和第二次工业革命。如果说蒸汽能和电能造就了从西方世界开始的两百多年的工业文明,数据能(Data Power)将把人类带入数字文明时代。数据是数字经济发展的要素,这个生产要素不同于土地、劳动力,也不同于资本、技术。如果要给数据找一个恰当的比拟物,也许只有 19 世纪末伟大的发明家尼古拉·特斯拉发明的交流电。数据是新时代的交流电,就像 20 世纪,交流电给世界带来的深刻变化一样,随着人们对数据能(Data Power)认识的提高,我们将进入一个“未来已来,一切重构”的时代。数据学院就像一百多年前的电力学院或电气学院。

怎么办数据专业?我们数据学院脱胎于软件工程学院,在此以前还有计算机科学与工程学院,数据相关的研究和偏向管理的图书情报方向的信息系统学科及专业也密切相关,应用数学、概率统计更是数据分析和处理的理论基础,不可或缺。到底什么样的专业才算是数据专业?起初的时候,这对我们来说基本上可以说是一个“灵魂拷问”。为此,我们发起成立了由国内十五所高校三十多位知名教授组成的“高校数据科学与工程专业建设协作组”。并且以协作组成员为班底,成立了数据科学与工程系列教材编委会,除了协作组成员,还邀请了多位有丰富教材编写经验的华东师范大学教师加入编委会,共同策划教材的内容安排。我们相信,有了先进的理念,再加上集体的力量,数据专业建设的探索之路就能走通。截至 2020 年 11 月,协作组已经召开了四次研讨会,确定了被称为 CST 的专业建设路线图,C 代表

Curriculum(培养计划),S 代表 Syllabus(课程大纲),T 代表 Textbook(教材建设)。在得知我们的工作后,ACM/IEEE 计算机工程学科规范主席约翰·因帕利亚佐(John Impagliazzo)教授邀请我们参与了 ACM/IEEE 数据科学学科规范的制定。协作组经过讨论达成共识:专业课程分为基础课、核心课、方向课三类,核心课是体现专业区分度的一组课程。与数据专业(DSE)最相近的专业就是计算机科学与工程(CSE)及软件工程(SE)两个专业,我们确定的第一批 DSE 区别于 CSE 和 SE 的 8 门核心课程是:数据科学与工程导论、数据科学与工程数学基础、数据科学与工程算法基础、应用统计与机器学习、当代数据管理系统、当代人工智能、分布式计算系统、云计算系统。随后我们又确定两门课纳入这个系列,分别是:区块链导论——原理、技术与应用,数据中台初阶教程。数据专业作为一个新专业,三类课程的边界还不清晰,我们将关注重点放在核心课程上面,核心课有遗漏的知识点可以纳入基础课或方向课。这样可以保证知识体系的完整性,简单起步,快速迭代。随着实践和认识的深入,逐渐明晰三类课程的边界,形成完善的培养计划。

教材为什么很重要?建设好一个专业,确定培养计划和课程体系固然很重要,但落实在根本上是教材。一套好的教材是建成一个好的专业的前提。放眼看去,无论是国内还是国外,无论是具体某个高校还是国家区域层面,这都是不争的事实,即好的专业都有成体系的好的教材。当然,现在的教材已经不仅仅指单纯的一本教科书,还有深层次的内容,比如说具体的教学内容和教学方式。我们都知道,教材是知识的结晶,是站到巨人肩膀上的台阶。在自然科学领域,确实如此,一百年前我们民族的仁人志士呼唤“赛先生”,在中华大地上科学的传播带来了翻天覆地的变化。在更广泛的领域,教材也还是技术、工艺和文化的传承,是产业发展的助推器。拿信息技术来举例,技术的源头和产业的发祥地都在美国 and 欧洲,像 IBM、Lucent、Oracle 等跨国企业在我国商业上取得的巨大成功无一不与他们重视教材开发密切相关。试想一下,我们的学生在课堂上学的都是他们研究和研发的东西,等走上工作岗位,自然会对熟悉的技术和系统有亲近感,这应该是产业或产品生态最重要的一个环节。21 世纪以来,随着互联网的蓬勃发展,人们已经深刻认识到,互联网改变世界。在人类的文明史上,没有任何一项科研成果像互联网这样深刻地改变人、改变世界。互联网之所以能改变世界,是因为它真正发挥了数据的威力。互联网实现了信息技术发展从“以计算为中心”到“以数据为中心”的路径转变。用“昔日王谢堂前燕,飞入寻常百姓家”来形容很多我们以前甚至当前教材上的一些内容,可以说毫不为过。以互联网为代表的新型产业的发展,极大地推动了技术的进步,我们已经到了可以编写自己的教材,形成自己的技术体系和科学理论体系的时候了。我们是现代科学的后来者,已经习惯了从科学到技术再到应用的路径,现在有了成功的应用,企业也发展出了领先的技术,学界可以在此基础上发展出技术体系和科学理论体系,应用、技术和科学的联动才是真正的创新之路。

对人才培养有什么贡献?在信息技术领域,迄今为止我们更多地是参考或沿袭了西方发达国家的培养计划和教材体系。在改革开放以来的四十年,这种“拿来主义”的做法很有效,培养了大量的人才,推动了我国的社会经济发展。但总的来说,我们的高校在这一领域更像是在培养“驾驶员”,培养开车的人,现在到了需要我们来培养自己的造车人的时候了。技术发展趋势如此,国际形势也对我们提出了这样的要求。我们处在一个大变局的时代,世界充满不确定性,开放和创新是应对不确定性的不二之选。创新成为人才培养的第一性原理,更新观

念、变革教育、卓越育人是我们华东师范大学新时期人才培养的基本理念。人才培养是大学的第一要务,科学研究、社会服务和文化传承是大学的另外三大职能,大学通过这三大职能的实现可以更好地服务于人才培养。人工智能时代最稀缺的是想象力,想象力是比知识更重要的东西。如何在传播知识、传承文化的同时,保护和激发学生的想象力,这也许是当前教育需要关注的。激发想象力,培养创新能力,这是数据专业核心课程系列教材建设的指导思想,我们愿意为之付出,久久为功地建设这套数据专业核心课程系列教材就是我们践行以上认识和理解的一个具体行动。

最后,要特别表示感谢。感谢华东师范大学出版社和高等教育出版社的支持和鼓励,感谢数据科学与工程专业建设协助组的各位老师的通力协作和辛勤劳动,也要感谢数据学院师生的信任和付出。心有所信,方能行远;因为相信,所以看见。希望作为探路者而付出的所有艰辛能够成为我们学术和事业生涯中的一笔重要财富。

“The best way to predict the future is to invent it.” —— Alan Kay

“Imagination is more important than knowledge. For knowledge is limited to all we now know and understand, while imagination embraces the entire world, and all there ever will be to know and understand.” ——Albert Einstein