文章编号：1672-5913(2019)06-0139-04
中图分类号：G642

数据分析类课程的技能培养方法探讨

赵卫东，蒲　实
（复旦大学 软件学院，上海 200433）

摘　要：针对高校培养的数据分析人才不能满足工业界需求的现状，分析高校人才培养的问题，阐述数据分析技能的基本要求，结合“行动中学习”，提出在实践中培养数据分析技能的培养方式，最后讨论在数据分析技能培养过程中需要注意的潜在问题，并给出相应的解决方法。

关键词：数据分析；技能培养；行动学习；实践
DOI:10.16512/j.cnki.jsjjy.2019.06.034

0　引　言

人类社会已经迈入大数据时代，各个行业都不断地产生大量的数据，如何在海量的数据中发现知识、获取价值已成为学术界和工业界炙手可热的话题，这也催生了“数据科学”新兴交叉学科和一批专业的数据分析人员 [1]。

数据分析是一个较为宽泛的领域，传统的数据分析包括统计分析和数据可视化等，后来融入了数据挖掘和机器学习的技术。随着人工智能的发展，深度学习也逐渐成为处理图像、音频和文字等数据分析的热点和有效手段。大数据在多个领域应用成功后，各个行业对数据分析人才的需求日益增长。数据分析相关从业人员逐年增加，但招聘市场仍出现“人才荒”的现象 [2]。国际数据公司（IDC）预测到 2020 年时仅美国就需要超过 190 000 名的数据分析人员 [3]；在中国，2017 年中国大数据产业规模为 4 700 亿元人民币，但大数据人才远不能满足发展需要，大数据人才队伍建设亟须加强 [4]。

当前工业界有许多从事大数据和人工智能的企业，在招聘市场上数据分析领域有较多的职位，例如数据科学家、大数据分析师、算法工程师、机器学习工程师、数据挖掘工程师等。对数据分析人员职位的划分没有一个确定的标准，各类公司在招聘时职位需求也常有重叠交叉之处。总的来说，可以从“纵向”和“横向”的角度看待各类职位的区别，“纵向”代表着数据分析角色的专业化，例如专注数据收集或清洗的数据工程师、专注算法建模的机器学习专家；“横向”代表着具有多个行业知识、有统筹安排整个数据分析流程的能力、足以指导不同专家团队协作的数据科学家 [5]。深入分析两种角度的数据分析人员的职位需求，可以倒推出企业对数据分析人才的能力要求。数据分析理论知识是数据分析人员从事分析工作的基础，但这不足以在竞争中脱颖而出。企业对数据分析的技能有更高的要求。

数据分析的技能包括业务理解能力、数据探索能力、数据建模能力以及项目管理能力等。以业务理解应用能力为例，这是数据分析中基础且至关重要的环节，是数据分析人员的核心能力之一 [6]，不仅要求数据分析人员掌握扎实的数据分析方法理论，还要有某个特定领域的知识，能够理解特定领域的业务问题并将其转化为数据分析的技能。

数据分析技能难以在短期内培养，需要参与大量的实践才能逐步提高，这也是导致数据分析人才培养跟不上市场需求的一个重要原因 [7]。数据分析技能的培养已成为数据分析人才培养的痛点，值得高校数据分析类课程任课教师的关注和探讨。

1　数据分析技能培养的困境

面对当前各行业激增的数据分析人才需求，

基金项目：2018 年复旦大学本科教学研究与改革实践项目（A 类重点项目）。
第一作者简介：赵卫东，男，副教授，研究方向为商务数据分析和机器学习，wdzhao@fadan.edu.cn。

高校作为数据分析人才培养的基地，担负着向社会源源不断地输送优秀人才的重任。教育部在2017年发布了“新工科建设复旦共识”，不少高校在近几年也逐渐进行教学改革，建立了一批与大数据、人工智能等有关的专业和学院，还有一些高校联合企业成立了实验室，这些举措都与数据分析息息相关。在线教学平台上的数据分析类课程也一直炙手可热。但是纵观大多数高校的数据分析人才培养计划以及相应的各类数据分析课程，发现一个普遍的问题是偏重数据分析知识的培养，而缺乏数据分析技能的培养。

现在不同层次学校的数据分析课程主要处于知识传授的阶段[8]。相当一部分课程以理论教学为主体，教师多局限于传授数据分析相关的理论知识。在课堂上教师容易对数据分析的算法和简单的二手案例进行讲解，对涉及的数理基础进行归纳，但是学生的工程实践机会较少，教学难以延展到数据分析的技能培养，有些学校单纯依赖企业承担实验和实训教学也不是长久之计。

部分高校的教师对数据分析课程进行改革，引入课程Project、课程实验等来平衡理论教学与实践。但是这种项目教学法一般是抽象的数据分析问题，例如对某家商务酒店的用户评价进行情感分析、对某种商品的销售数据进行分析等。这类课程项目有处理好的数据、明确的需求和考核目标，学生需要考虑的因素较少，通常集中在课堂所学知识的简单应用上。但这远远不够，课程项目与企业项目有非常大的差别。企业项目没有明确地分析问题，需要分析人员从复杂的业务背景和问题中予以提取。企业项目的数据非常杂乱，需要大量的数据预处理工作。企业项目有近乎严苛的验收标准，伴随着紧张的项目排期和时间压力。在企业项目中，很多问题以及解决思路在教科书和课堂上未曾提及，需要在项目实践中通过不断地试错和创新去解决。因此，使用传统的项目教学法培养的学生在面对一个实际的企业项目时，可能非常茫然、不知所措。

当前数据分析领域不再是学术界一枝独秀，工业界的贡献与日俱增。近年来优秀的数据分析工具和算法也多源于企业，例如IBM的Watson、Google的Tensor Flow、Microsoft的CNTK等。因此高校的课程体系需要涵盖数据分析领域庞大且日益更新的内容体系，建立从理论基础到应用实践的多层次教学。当前部分高校的课程内容却还停留在数年甚至十年前，培养的学生跟不上市场的技能需求。有限的课程时间也阻碍了从理论到实践的教学过程。

综上所述，学生的工程实践能力得不到充分的锻炼，学生数据分析的技能水平与企业的需求有较大的脱节，导致了数据分析技能培养困境的出现。

2 数据分析技能的有效培养方式

为了探讨数据分析技能更有效的培养方式，我们对一些成才的学生和工业界的成功人士做了跟踪调查，并结合在复旦大学软件学院实施的“基于项目沉浸式的数据分析类课程教学”的实践[9]，发现有效培养数据分析技能在于“行动中学习”。

在数据分析领域有著名的1万小时定律，换算后是5年左右的时间，意味着一个人能游刃有余地解决实际的数据分析业务问题需要5年左右的学习和实践。其中，不必刻意学完所有的知识点再去实践，而是在实践中学习，只有这样才能深刻体会理论知识的内涵，并在实践中与技能共同学习。如果数据分析的理论知识不在实践中使用，很难真正地理解和掌握。这也符合当前教育界著名的“行动学习”理论。“行动学习”是在20世纪中期由英国Reg Revan教授提出，他认为行动学习为L=P+Q，其中P代表结构化的知识，Q代表质疑性洞察。行动学习是一个“知行合一”的循环学习过程。首先，建立基本的结构化知识体系，这是P阶段的任务；然后在实践中将知识加以应用，发现问题，并在归纳总结中加深对知识的理解，获得洞察，这是Q阶段的工作。P阶段和Q阶段交错循环，构成了从行动学习认知世界和改造世界的基本规律。

将行动学习理论应用到数据分析技能的培养，由两个培养阶段组成。第一阶段是对基本理论、方法和工具的学习，对应行动学习的P环节；第二阶段是在仿真项目、比赛项目以至正式项目中的实践，对应行动学习的Q环节。在第二阶段，学生可以将第一阶段学习到的知识加以应用，理论就不再局限于书本，可以将知识转化为洞察；同时可以发现理论学习的不足，从而反哺到第一阶段。学习在两个阶段不断交替进行，第

一阶段学习为第二阶段实践提供基础，实践又反馈到理论学习中，由此形成良性的循环，在不断的实践和总结反思中获得数据分析的技能。

上述两个阶段的学生可细分为下述 4 个具体可操作的步骤。

（1）掌握较坚实的数据分析理论知识。这是数据分析技能培养周期的第一个 P 阶段。行动学习首先要掌握一定的理论知识，这是后续实践与反思的基础。强调数据分析技能与实践，并不是意味着理论知识不重要。相反，理论知识是数据分析技能的奠基石。没有数学基础，理解算法一定有困难，更别提熟练运用算法。对算法理解不深，就不能得心应手地选择算法，参数调优也可能收效甚微。行动学习的主体是学生，但是教师需要为学生构建起步阶段的环境，包括需要针对数据分析理论体系合理设计教学方案，重在传授学习的方法，为学生建立数据分析的认知体系，使得学生掌握数据分析整个流程以及每一步的方法技巧。同时，在数据分析理论的学习中，初期不用拘泥在代码实现上，教师应指导学生将重点聚焦于分析问题的思路，可以直接调用一些易用的开源框架的 API 尝试算法的应用，强化对数据的理解以及数据分析的思维方法。因此，在学习的初期，提倡学生使用一些可视化的、组件式的数据分析工具进行学习，例如 IBM SPSS、腾讯 TI-One、华为 FusionInsight 等工具，所见即所得的学习方式也能提高学生学习的兴趣。

（2）学习优秀数据分析师的思路。在该阶段中行动学习进入了 Q 环节，学生需要在实践中将 P 环节的知识加以应用，并开始行动学习的循环。数据分析包括业务理解、数据采集、数据预处理、建模分析、结果评估和建议，这是一个完整的流程。流程的每个环节都对应了大量的理论知识，如何将理论知识融入到数据分析的整体思路中，如何在分析的每个阶段能有正确的思路是值得考虑的问题。初学者比较有效的方法是“模仿”。本阶段教师可以整理已参与完成的数据分析项目，包括但不限于比赛项目、实际企业项目等。将这些项目整理成案例和实验文档，将整个项目的实现完整地展示出来，其中需要突出项目中遇到的疑难点。然后将数据和文档交给学生，让学生来重演整个项目。对于比赛项目的优胜者，还有优秀的分析思路，涵盖了各式各样的数据分析方法。学生在复现这些项目时可以将理论加以映射，同时可以看看其他人是如何思考的，在数据分析的每个环节选择了什么方法，思考能不能做进一步的优化。学生在第一次进行到 Q 环节开始“模仿”时，P 环节的知识就能在实践中融入到学生的分析中，逐渐掌握数据分析的技能，这就是从知识到洞察的学习过程。同时学生可能会发现理论学习的不足，这就为行动学习的下一个循环提供了条件，下一循环中 P 环节的知识就来源于上一循环中 Q 环节的问题。

（3）参与新项目。这是行动学习循环的一个提升。在初期的简单复现时，学生技能水平还比较弱，但是在后期不断重复后技能会不知不觉提高，对实际业务问题的理解也会有一个更高的层次。这时候就可以面对新的问题，这个阶段无论是 P 环节的知识储备，还是 Q 环节的实践要求，与第二阶段的“模仿”相比都可能有了质的提升。新问题没有他人的思路参考，学生需要去学习新的知识。因此这个阶段更考验学生的自学能力。在该阶段的项目实战中，学生需要将新知识、技能与已有的知识体系融会贯通，并尝试性摸索数据分析的方法、思路。这个阶段学生行动学习的主体意识应当更强，教师不再需要编写完整的项目案例或文档，将 P 环节的学习留与学生自主完成，着重点转移到行动学习的 Q 环节中。教师可以指导学生参加一些大赛，例如阿里天池、KDD-Cup、Kaggle 等。新项目的难度参差不齐，教师需要根据学生当前行动学习的所处阶段和学生能力情况推荐合适的项目。在学生项目实践中，教师重在对学生方法和思路给予针对性建议或指导，并组织鼓励学生不断反思总结。新项目能显著提高学生分析问题和应用知识的能力，实战中学生数据分析的技能也会积累到理想水平。

（4）参与企业实际业务项目。这个阶段是对行动学习最终成果的检验。在一定的学习积累后，需要参与企业实际项目。与企业项目相比，企业实际业务问题难度更大。用户在企业项目开展初期，可能只提出简单的业务问题或目标，需要将其抽象出合适的数据分析问题，这需要熟稔业务领域。企业的数据可能分散在各个数据源中，不像比赛会提供一个完整的数据集，因此提取哪些数据、如何提取数据都是必须考虑的问题。企业问题会有严格的业务审核，数据分析的结果必须要达到一定的性能要求。与比赛相比，这对数据分析的质量要求会更高。企业项目一般

还有时间的限制，相应给数据分析人员带来更多的压力。学生在该阶段面临的困难和挑战更多，需要有项目经验的教师给予积极的指导。针对学生的问题，教师还需要积极与企业的专家合作，利用企业专家丰富的经验为学生提供切实可行的建议。能够解决企业实际业务问题是数据分析技能的高级培养目标。学生需要经过行动学习 P 环节和 Q 环节的反复实践，在上述几个具体步骤中一步一步稳扎稳打地获得数据分析的技能。

3 数据分析技能培养的潜在问题

3.1 一些企业项目的对接问题

很多高校教师没有在工业界工作的经历，也没有充分认识到工业界和学术界的差异，因此一些教师会找不到企业的项目，没有和企业合作的机会；一些教师可能轻视企业项目的难度，投入精力较少使得项目停滞。因此教师需要加强和企业的对接能力。

当前企业的需求非常多，教师可以通过讲座、技术论坛、数据分析竞赛等方式得到与企业交流的机会，教师需主动走出去。值得注意的是，企业问题都有严格的检验标准，所以需要老师投入更多的时间。在教学中引入企业的问题，对学生的技能提高有一定的效果，花费时间也是值得的。

由于学生的学习水平不同，且前期缺乏项目经验，不管参加大型比赛还是企业项目，都有一定的难度。老师可以把以前的项目资料收集起来供学生研究，或者与有经验的企业专家一起指导学生攻克关键问题，并潜移默化地带领学生进步。

3.2 教师的主导作用

数据分析技能的培养强调行动中学习，学习的主体是学生，教师需要在掌握数据分析技能的基础上，身体力行地参与到学生学习的每个环节中，教师应更多地激励、驱动学生的自主学习。教师和学生都需要认识到数据分析技能培养是一个循序渐进的过程，不是一朝一夕可以学习完成的。教师需要引领学生认识数据分析的技能体系，通过积极参与项目实践，从中攻克数据分析的难题。

4 结 语

当前工业界对数据分析人才需求旺盛，但是高校人才培养却跟不上市场需求。高校的数据分析类课程更侧重理论知识的传授，学生在实际项目中的实践技能还需要大幅提高。数据分析人才技能培养周期长、对技能要求高，需要从模仿项目、比赛项目到企业项目的长时间积累。数据分析技能培养对高校教师和学生而言都是挑战，教师首先需要积极参与实践，积累必要的数据分析技能，并在与企业的合作中，培养出真正能解决实际问题、为社会所用的数据分析人才。

参考文献:

[1] Dhar V. Data science and prediction[J]. Communications of the Acm, 2013, 56(12): 64-73.
[2] Jensen S. Integrating big data Services into an undergraduate MIS curriculum[J]. International Joural of Systems and Service-Oriented Engineering, 2017, 7(2): 58-73.
[3] Market Analysis Perspective. Worldwide developer demographics communities and skills [EB/OL]. [2018-11-17]. https: //www. idc. com/getdoc. jsp?containerId=US42054117.
[4] 大数据白皮书(2018年). 中国信息通信研究院[EB/OL]. [2018-11-17]. http: //www. catr. cn/kxyj/qwfb/bps/201804/t20180426_158555. html.
[5] Saltz J S, Grady N W. The ambiguity of data science team roles and the need for a data science workforce framework[C]// IEEE International Conference on Big Data. IEEE, 2017: 2355-2361.
[6] Mauro A D, Greco M, Grimaldi M, et al. Beyond data scientists: A review of big data skills and job families[C]// International Forum on Knowledge Asset Dynamics, Ifkad. 2016.
[7] Mikalef P, Giannakos M N, Pappas I O, et al. The human side of big data: Understanding the skills of the data scientist in education and industry[C]// IEEE EDUCON 2018 Global Engineering Education Conference. IEEE, 2018.
[8] Huang X, Qin N, Zhang X, et al. Experimental teaching design and practice on big data course[C]// International Conference on Computer Science and Education. IEEE, 2017: 566-569.
[9] 赵卫东, 赵洪博. 基于项目沉浸式的数据分析类课程教学研究[J]. 计算机教育, 2017(06): 58-61.

（见习编辑：郭安琪）