

◆ 教改纵横

文章编号: 1672-5913(2019)09-0154-04

中图分类号: G642

基于项目实践的机器学习课程改革

赵卫东, 袁雪茹

(复旦大学软件学院, 上海 200433)

摘要: 分析传统机器学习教学中的问题, 提出将机器学习理论教学与企业项目实践融为一体的培养方式, 介绍基于项目实践的机器学习教学改革措施, 总结基于实践的机器学习培养方式中的潜在问题。

关键词: 实践教学; 机器学习; 培养方式; 项目驱动

DOI: 10.16512/j.cnki.jsjy.2019.09.034

0 引言

当前, 随着信息时代的快速发展, 银行、投资、零售、互联网甚至传统的制造业都产生了大量数据。IDC 在《2020 年的“数字宇宙”》研究报告中预测, 到 2020 年, 全球数据量将达到 40ZB, 中国在全球数据市场的份额也将由目前的 13% 上升到 22%^[1]。各行各业开始逐步应用机器学习算法分析数据, 以便在海量数据中提取有效的规律, 辅助决策。这种发展趋势使就业市场对数据科学、机器学习人才的需求不断扩大, 同时对人才的多元化和综合实践能力提出了新的要求。

随着数据分析相关行业的快速发展, 数据分析在各个领域都得到了很多成功的应用, 企业和政府期望在各个业务岗位都有数据分析能力强的人胜任, 而不是将数据分析作为一项专业技能, 更期望员工能够探索有效的数据分析方法, 并根据实际数据场景分析结果作出决策, 将分析和处理数据作为日常工作流程的一个环节。同时, 随着数据种类和数量的爆炸式增长, 市场对毕业生的数据分析和处理能力提出了更高的要求, 需要有数据分析技能的人才预测行业前景, 及时抓住发展机会, 形成独有的竞争优势。高校的基本职能是培养人才, 为了使学生更好地适应现代工作场所和终身发展, 需要认真思考如何培养应用型人才, 以适应当前的就业环境。机器学习相关专业以培养数据分析师、算法工程师、大数据工程师等高度数据分析、应用型人才为目标, 这些工作不仅要求学生理解算法本身, 更需要培养学

生跨学科的实践能力, 将算法逻辑应用到实际生产、生活场景以解决现实问题。

企业对数据分析人才的数量和质量的高要求导致了巨大的大数据技术、人工智能人才缺口, 而目前高校的机器学习教学偏向理论化, 更注重算法本身, 缺乏完善的实践教学体系和教学资源。学生的课堂学习只是面对多种专业知识的组合, 缺少真实项目的实践过程, 学生不能有效地将学习内容应用到实践过程中, 这与应用型人才的培养目标存在一定的偏离, 毕业生不足以适应竞争激烈的就业市场。因此, 高校需要考虑就业环境与学生的真实需求, 对传统的教学模式进行变革, 掌握数据科学时代的新技术和新应用, 在遵循教育规律的基础上, 将实际项目实践与理论教学融为一体, 逐步调整课程内容, 培养学生自主思考与解决实际问题的能力, 从而提高竞争优势。

1 传统机器学习教学的困境

如何在教学过程中结合项目实践, 已成为各高校关注的话题。目前各高校分别建立了机器学习、数据科学与大数据相关专业, 而机器学习作为其中的重要课程, 具有高度实践性且涉及的算法众多。如何结合真实项目以培养应用型人才, 是当前人才培养的一个难题。

1) 机器学习技能难以短期速成。

机器学习(包括深度学习)涉及的算法众多, 而掌握这些算法要求学生在理论学习基础上, 将算法结合企业实际项目, 对理论知识有扎实的理

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等, wdzhao@fudan.edu.cn。

解。这一过程需要学生接触实际问题,经过长期训练达到将理论知识输出为实际应用的水平。尽管目前网络上可接触许多精品课程,但这些课程大多以知识体系的教学为主,注重介绍算法本身、介绍机器学习中常用的理论,较少涉及实际应用。同时,新技术进入课堂教学需要过程,这对教师提出了更高的要求。教师需要积极获取企业先进技术与解决方案,将实际生产流程归纳成适应于教学氛围的学习资源,而新技术快速迭代和教材易老化成为机器学习教学的一大痛点,教师需要不断积累和沉淀,高效创建出合适的教学大纲和教学内容。因此,机器学习的技能培养与企业人才需求存在差距,这种技能需要通过解决实际问题来不断摸索和积累经验,难以通过短时间授课的方式获得。

2) 教学过程缺少实际数据和案例。

机器学习相关专业的学生倾向于选择机器学习工程师、数据分析师等高度应用型工作。为了让学生更好地掌握机器学习算法,很多人工智能、数据挖掘课程都包含了实验教学部分。在实验过程中学生需要自主实现算法,培养动手操作能力。然而,教学实验内容往往是理想化的,由于企业的真实数据难以获取,在校学生接触到的数据普遍是二手数据。更多时候,为了实现某一特定算法,实验内容往往经过特殊处理。当面对一个企业实际项目时,在与客户的交流过程中,客户的实际需求、数据代表的真实含义,可能由于客户自身缺乏相关的技术知识背景难以表述明确,需要数据分析人员从庞杂的数据中自主降噪、挖掘有效信息,这与课程实验存在较大反差。此外,由于传统的机器学习教学偏理论,将更多的精力放在对算法本身的理解上,缺少具有实际意义的应用案例。为快速把握企业的真正痛点,要求学生对相关领域有一定程度的认知。机器学习的教学集中在理论方面,学生难以深入理解真实场景下的企业项目。因此,在数据、问题和性能要求等方面都与企业实际存在偏离,对从课堂转向企业实际问题的学生来说,尚且存在很多挑战。

3) 人才培养体制不够完善。

在当前高校教学评价体制下,教师可能忽视了或不愿意参与企业真实调研及与企业合作项目。学生通过课堂和实验室的知识积累,可以完成图像识别、声音模拟,通过网络可以较为容易地获取很多开源的代码,这些技术已经得到比

较广泛的应用,易于实践和理解。然而数据分析方法和算法的应用具有很强的艺术性,数据分析的每个阶段都没有固定的答案,需要结合实际场景进行调整和优化,并考虑实际生产环境的可行性。与教学实验不同,面对一个新的实际问题,条件苛刻且结果未知,学生无法直接获得解决问题的思路,而现实问题是无法规避的,且这些问题难以在校园环境中为学生所接触。因此,面向项目实践的机器学习教学需要在实践中摸索,与应用结合,培养人才的多学科性。强调项目实践不等于否认理论教学,由于实际问题的复杂多变,反而要求学生具备扎实的理论基础。

传统的机器学习教学在技能培养、数据与实际案例的选择以及人才培养体制上仍存在提高的空间。这与新时代机器学习人才发展需要存在一定偏离,有必要对现有的人才培养方式进行与项目实践相结合方向上的探索,尝试新的满足社会发展需要的实践教学模式,为培养具有专业素质和创新能力的机器学习人才奠定坚实的基础。

2 基于项目实践的机器学习教学改革措施

项目实践教学能够培养学生的实践能力和创新能力,提高人才培养质量。针对机器学习这一高度实践性课程,很多高校提出了丰富教学内容、加强实用型师资队伍、优化教学条件等一系列措施,逐步建立并完善专业培养目标和培养模式的实践教学体系。然而,如何更有效地提高实践教学质量并与传统的理论授课相结合,是实现基于项目实践的机器学习教学目标的关键。

2.1 理论知识与应用场景认知培养

由于机器学习的相关知识在实际应用中是灵活多变的,需要根据特定场景对算法进行调整和优化,对多种经典算法进行取舍。因此,实际工作对算法本身的理解提出了更高的要求,需要在坚实的理论基础上引导实践,从而在实践中加深对理论的理解。针对现阶段人才需求目标,在机器学习教学初期,理论基础教学应注重完整机器学习过程的每一个环节,使学生掌握从数据获取、数据预处理到模型评估与选择以及模型优化与结果分析的每个步骤,形成完整的机器学习思路。最有效的方法是教师身体力行,积极参与企业合作的科研项目,在教学过程中不仅注重实现算法本身,也可以尝试从大数据竞赛网站了解真

实的问题场景,并选择现阶段成熟的开源工具实现数据处理,如微软 Azure 平台、腾讯云平台,这些平台提供现成可用的解决方案与经典案例的实现步骤,允许学生自主创建项目并进行机器学习算法取舍。

此外,由于机器学习的概念范围十分广阔,而具体的应用范围可能包括舆情分析、异常检测、人脸识别、自然语言处理等,机器学习工程师等相关岗位都有着不同的技能侧重要求。可介绍机器学习每一步骤对应的典型工作岗位与职责,让学生在教学场景中理解和体验所学知识的真实应用过程,培养学生的职业意识。例如,通过引入作者与京东合作的发现好货、会买专辑等写作机器人项目,可以详细介绍文本分析的过程,从商品数据和用户评论的获取(OCR技术的应用)、分词与语义的识别,到用户情感分析、长短期记忆网络的组词造句、文本质量的评价,都有具体的项目内容对应。这些内容基本覆盖了文本分析章节的大部分内容,可以使学生在一定程度上沉浸在项目中,对各个阶段遇到的问题,讨论解决方法,开阔思路。

2.2 实践项目的验证与设计培养

现阶段各高校的机器学习相关课程普遍开设了实验环节,但主要内容是实现课本中的算法,学生通过实验能够掌握很多经典算法,提高动手能力。但是,机器学习技术更新快,教材内容易老化,学生的实验场景、问题、数据、解决方案都偏于理想化,与真实的企业场景存在偏离,仅仅掌握实验内容难以胜任实际工作。

为了使学生更好地适应就业环境需要,在直接接触企业实际项目前,可在实践项目的验证与设计两个环节对学生进行培养。高校可邀请企业中机器学习相关方向的专业人员,教授当前最新项目实施思路与过程,与教师和学生一起讨论在企业数据处理中使用的各种算法。教师在熟悉企业场景及解决方案的基础上,制订实践培训的内容、教学形式和培训过程,尽量在教学过程中营造真实的企业项目发展氛围。在与京东公司上述的合作项目中,笔者不仅把项目成果引入课程,介绍项目实施过程中的具体思路和辛酸苦辣,还将部分问题作为课程研讨的选题。

在学生理解算法原理的基础上,可采用灵活的模块化教学方法培养学生对实际应用场景的认知。以企业项目为驱动,项目案例分为验证项目和项目设计两个部分^[2]。基于高校现有的多媒体

实验环境,有实践经验的教师或企业专家在讲解一些知识点和分析相关案例后,结合案例程序展示其应用,然后结合教学进度提出一些问题;学生通过模仿实现一个类似的验证型实验项目,该项目作为设计实验项目的原型,学生可访问、分析其功能、代码并测试其效果。随后,以此为基础做扩展实践,学生可以模仿教师提供的案例,通过创新自主设计并实现一个相对完整的项目,深化巩固所学的知识,锻炼整体考虑的问题能力,提高灵活的应用知识能力和创新能力。

2.3 校企合作完成企业实践项目

应用型人才需要具有扎实的理论基础、科学研究方法以及跨领域的视野和解决生产问题的能力,更需要具备实际操作的技能。在理论基础与了解企业相关领域的基础上,高校可以与企业联合组建合作项目和教学团队,结合企业需求,开展各种形式的培训和实践,不断完善实践教学内容。真实的企业实践项目能够将学生所学知识应用于解决实际问题,形成一个完整处理问题的思路^[3]。

在数据获取上,高校可利用校企合作关系,在教学内容中加入企业实践学时,教师可根据班级人数将学生分为若干调研团队,使学生能够与企业进行直接交流^[4]。在调研过程中学生能够密切关注企业实际生产,接触企业的真实数据。企业数据真实庞大,且大多质量糟糕,例如存在很多噪声以及数据不一致等一系列现实问题,学生需要思考如何提取数据,消除噪声,达到可用的目的。这一过程学生可对数据建立更加感性的认识。

在问题解决上,由于企业面对的很多问题并不能直接交由机器处理,数据的筛选、特征提取以及算法的整合与取舍在实验的层面可能是需要技巧的;同时,企业实践项目真实灵活且与当前研究热点紧密相关,在项目解决方案的探讨中学生会面对很多瓶颈,例如样本的不平衡、算法存在的某些缺陷等,这些瓶颈不能直接从课堂或其他途径获取到有效的解决方案,需要学生总结经验,在现有的思路上进行调优,从而解决算法缺陷,自主发现一些原有教学中被忽略的难点。在真实的企业环境中,客户的要求十分苛刻,可能存在阶段性变化,这需要机器学习相关领域工作人员掌握与客户沟通的软技能,需要对所在领域有一定程度的认知,从可行性的角度上理解业务。

企业实践项目具有一般知识和实践技能的

特点,能够加深学生对数据分析算法的理解和解决问题的经验,也是培养学生实践能力和创新意识的一个重要手段。高校可与企业合作建立实验室、培训基地,开展合作项目或联合研发,使企业直接参与课程策划,资助课外活动中的科技创新,建立俱乐部实现对接。实际的应用型教学能够激发学生兴趣,使其对原有的理论知识有更深入的理解。笔者最近几年与10多家企业开展了制造企业质量管理、银行贷后催收语音合成、上海图书馆自动标注、妇婴用品电子推荐、在线学习学生行为分析等项目,与这些企业建立了长期的合作关系,可以供学生调研实习,参与联合研发项目。

3 基于实践的机器学习培养方式中的潜在问题

基于实践的机器学习教学,对教师和学生提出了一些挑战,其中会遇到一些突出的问题需要解决。

1) 实践项目的选择难题。

任课教师在企业合作项目的选择上需要深入研究行业整体,同时根据学生水平进行课堂内容调整,设计适合不同层次学生的实践项目,并通过研讨了解学生自身实际编程基础和分析能力,这对教师提出了更高的要求。

2) 实践过程的管理难题。

机器学习实践教学的目的是培养学生的自主分析和实际应用能力,加深对理论知识的理解,验证课堂教学理论,从而与职业发展平稳接轨。这一过程中的每个环节都不同于单纯的课堂和实验室教学,而是需要真实的实践环境,需要与企业沟通以获得实践环境以及必要的师资指导支持。

如何培养有较强应用能力的师资是需要深入探讨的问题,从企业引进有项目经验的师资或培养积极参与企业合作项目的师资都是必要的。这需要高校的薪酬和职称制度相应的配套措施改革。

3) 实践思路的引导难题。

企业实践项目不同于常规教学实验,在大多数传统教学方法中,学生按照已有步骤和原则进行规范化的实验,往往可以获得满意的结果。然而,在整个过程中,学生很少独立思考,往往不会考虑这些步骤背后的思路,很难培养独立解决问题和创新的能力。在实际应用中,可能同时面临很多新的问题,例如数据是否足够、数据是否满足客户问题的分析等,这些现实的问题与课程教学、实验存在较大差距。因此,有实际项目经验的教师在对学生的思路引导方面也是重要的^[5]。

4 结 语

大数据时代对机器学习方面的人才提出了更高的实践性要求,而高校机器学习课程主要以理论知识教授为主,这与实践性人才培养的目标存在一定的偏离。笔者最近几年与京东、上海图书馆等多家企业合作了机器学习项目,通过抽象和提炼这些项目的成果,形成第一手的案例和实验素材,引入机器学习课程,不仅使课程的内容更容易理解,而且锻炼了学生的数据思维。这并没有耽误理论课的教学,反而可以作为学习和检验理论的重要手段。

高校对人才的培养不能脱离行业实践,需要积极与企业建立项目合作,理解企业所需的技能,培养和引进有实践项目经验的师资,这样才能培养学生自主思考与解决实际问题的能力,从而提高竞争优势。

参考文献:

- [1] 比特网. IDC 最新调研报告: 2020年的“数字宇宙”[EB/OL]. (2017-03-10)[2019-04-08]. <http://info.chinabyte.com/487/12496987.shtml>.
- [2] Feng X, De G, Wei Z. Research on teaching reform and practice of engineering training in application-oriented University under the background of large engineering[C]. Proceedings of the 2017 International Conference on Education and E-Learning, ACM, 2017: 73-76.
- [3] Hong-Mei S, Rui Sheng J. Research on case teaching of software development comprehensive practice based on project driven[J]. Procedia Engineering, 2012(29): 484-488.
- [4] Li L, Qi J. The feasibility and prospect of university education reform based on the integration of education and industry[C]// International Conference on Computer Science & Education. Washington D C: IEEE, 2017: 604-607.
- [5] 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(6): 58-61.

(编辑: 孙怡铭)