

机器学习科研实践课程建设

中国石油大学(华东)计算机与通信工程学院 王雷全 吴春雷 郭晓菲

【摘要】针对本科机器学习课程中的教学实际,分析在教学过程中遇到的问题,结合目前机器学习领域的研究热点,提出在本科机器学习教学中以深度学习作为课程体系的主导,引入Tensorflow框架,重新整合教学资源、建设相关实验内容,构建以深度学习为教学内容核心、以Tensorflow为实践平台的机器学习科研实践课程体系。

【关键词】机器学习;深度学习;Tensorflow;课程建设

机器学习是当前的标志性技术,且已成为电子信息类课程中最为重要的一个专业方向^[1]。近年来,深度学习作为机器学习的一个分支,在图像和语音识别等领域取得了重要突破,是当前计算机领域最具影响力的热点研究方向。掌握“深度学习”的基本知识与实践技术是当前对本科生机器学习课程的基本要求。

1. 传统机器学习课程面临挑战

传统机器学习课程的授课内容主要以经典机器学习算法为主^[2],包括支持向量机、逻辑斯谛回归和隐马尔可夫模型等经典算法。随着深度学习的迅速发展,传统机器学习的教学内容受到了挑战,在本科机器学习教学过程中增加“深度学习”的相关内容,已是大势所趋。然而,将深度学习融入到机器学习课程中却面临两个问题。一方面,在大学本科阶段涉及深度学习知识的课程还比较少,教学案例严重缺乏,尤其是针对深度学习的实践类指导资源更为稀少。另一方面,目前主流的深度学习框架较多,在实际科研与开发中,Caffe、Theano、Torch等不同开源深度学习框架种类繁多。然而,对于课程内容设置来说,很难面面俱到地涉及到所有框架,需要选择一种主流开源框架为核心,作为课程的主要线索加以展开。为了解决上述问题,机器学习科研实践课程将机器学习课程与科研实践课程相结合,并将深度学习框架Tensorflow纳入到课程体系中。

2. 机器学习科研实践课程

《科研实践》课程是针对我校计算机与通信工程学院本科专业第7学期开设的一门科研实践类课程,共6周,合96学时。该课程是一门以科研实践为主导,采取小班研讨、启发式教学、结合实际科研案例的综合性实践类课程。

《科研实践》课程的教学过程划分为理论讲解、主题研讨和实验验证三个部分。该教学方式的要点包括三个方面:

①案例引导的理论知识铺垫。把深度学习的基础理论融入到项目案例的细节中,以技术发展的过程重新组织并介绍理论,增强教学效果;

②面向科研主题的研讨交流活动。为了充分调动学生学习的积极性和主动性,并培养学生总结科研进展的能力,该课程增设了基于主题的研讨式教学环节,通过启发式教育激发学生的科研创新能力。

③科研项目驱动的完整实践过程。建立了以科研创新为主线的实践流程,目标是使学生能够以行业先进的技术 with 平台进行科研训练,验证科学想法,规范化、系统化地培养学生科研实践能力。

根据目前的科研与应用热点,结合机器学习课程的现状,课程组将机器学习方向引入到科研实践课程中,并将深度学习作为机器学习科研实践课程的主导项目。通过机器学习与科研实践课程相结合,课程组在课程目标与培养标准、知识体系域教学内容、开发平台与实践流程、案例建设与共享交流等方面进行改革,进一步提升机器学习课程的教学质量与本科学生的科研实践能力。

3. 课程改革实施方案

针对本科生科研能力薄弱的特点^[3],课程组认为科研案例与实践指导资源的缺乏,严重制约了本科生机器学习实践层次的提升。目前面向人工智能方向课程主要集中在研究生阶段、企业内部以及各类IT培训机构。在大学本科阶段开设该类专门实践课程的还非常少,尤其是深度学习方向的科研实践类课程更不多见。因此,一方面,课程组

重新审视课程定位,完善知识体系和教学内容;另一方面建设以Tensorflow为核心的科研创新与应用平台,从而提升科研实践教学水平。

(1) 重新定位课程目标与培养标准

根据工业界与学术界对于人才的实际需求,从知识、能力、素质和长期发展规划等多个角度对培养目标进行分解,形成具有高度可操作性的细化培养标准。修订科研实践课程教学大纲,深化深度学习的教学内容,并将Tensorflow基础纳入教学大纲,贯穿整个课程教学与实践体系。课程组借鉴国内外相关优秀课程,走访国内相关高校与知名企业,调研深度学习的现状与前景,完善课程的培养目标、知识体系和教学内容。目前机器学习科研实践课程教学大纲分为四部分内容:

机器学习基础与深度学习(24学时)

Tensorflow基础与实践(16学时)

深度学习经典案例研讨与Tensorflow实践(24学时)

以Tensorflow为平台的科研实践课程设计(32学时)

(2) 优化课程知识体系与授课内容

借鉴国内外著名大学的课程资料如CS231n、CS224d等,优化当前教学大纲中的知识体系,使知识体系更完整和更具时效性。目前,传统机器算法已经不能满足本科生掌握最新机器学习知识的需要,在机器学习科研实践课程中明确以深度学习为核心教学内容,重新分配教学内容和学时,并编写配套的参考案例和课件。

根据行业的最新应用需求与科研方向,科研实践课程将深度学习作为教学的核心内容。基础理论部分包括梯度下降、BP算法、CNN与LSTM等。在此基础上,课程为学生提供若干个热点科研主题供学生自主选择,如物体检测、机器翻译、语音识别等,然后围绕最新科研项目组织教学,将学生的综合能力融入到一个完整的科研实践过程中,在该过程中从不同的侧面介绍理论体系、指导科研实践,重点培养以升学为发展方向的学生科研实践能力。

(3) 规范实践平台、完善实践流程

目前深度学习框架较为混杂,包含了caffe2、mxnet、theano等许多平台。新的机器学习实践教学中确定以Tensorflow框架为核心实践平台,在课程理论环节的案例教学、实践环节的实验教学与科研实践中均统一应用,对照新的课程开发计划,增加Tensorflow平台操作练习题,设置深度学习关键技术练习题(如BP、Dropout、Batch Normalization等),建立以Tensorflow为核心的实践平台,增加Tensorflow基础的实践流程,在此基础上指导学生完成科研项目驱动的完整实践过程,以提高学生的科研实践能力。

此外,我校最新一版的电子信息类本科生培养方案加大了人工智能课程群的课时,尤其是实验课时。例如,《智能计算基础》的实验课时设置为8学时;《机器学习》的实验课时设置为16课时。我们希望通过本项目的实施,能够指导其它人工智能方向(包含研究生课程)的教学内容与实践项目的改革,最终实现以Tensorflow为核心的人工智能课程群的实验与实践创新平台。

(4) 加强典型教学案例建设

深度学习教学案例的缺乏是当前深度学习方面课程建设的主要问题。课程组成员跟踪最新深度学习内容,充分利用CVPR、ICLR等知名国际会的最新进展,搜集适于本科教育的具有纵向发展的主流科研方向作为面向机器学习科研实践的深度学习教学案例,主要案例包括:卷积神经网络、物体检测、CNN-LSTM以及对抗生成网络等。

在学生科研项目选题时,课程组结合实际科研动向提供科研案例以供选择或启发学生自行立意。同时,将优秀的学生科研成果加入教学案例库。此外,课程组在课程建设中探索了具有石油特色的机器学习教学案例。石油是中国石油大学的办学核心,近年来我校多个

基金项目:教育部高等学校计算机类专业教学指导委员会专业课程改革项目(编号:MXF2016-2-1);中国石油大学研究性教学改革项目(编号:YK201612)。中央高校基本科研业务费专项资金资助(17CX02041A)。

相关专业不断尝试如何将石油技术与最新的机器学习方法相结合（比如，深度学习与地质勘探的结合、深度学习与海上溢油区发现的结合）。课程组将该类成果通过Tensorflow平台转化出来，丰富具有石油特色的深度学习教学案例，服务于石油石化行业。

4. 课程特色与亮点

目前面向机器学习的课程主要集中在研究生阶段^[4]、企业内部以及各类IT培训机构。在大学本科阶段开设该类专门实践课程的还非常少，尤其是深度学习方向的科研实践类课程更不多见。本课程具备以下特色：

（1）夯实数学理论基础，培养数学应用能力

考虑到电子信息类专业的特点，绝大多数实践类课程以工程锻炼为主，缺少理论应用的训练^[5]，导致大多数学生难以理解前期设置的数学类公共基础课程的作用。因此，课程组认为需要着眼于学生职业生涯的长远发展，在科研实践中强调基础理论的作用，提升学生的理论水平和科研实践能力。

（2）实时结合业界热点，弥补教学科研鸿沟

近年来，机器学习与模式识别发展迅速，尤其以深度学习为代表，在工业界与学术界取得了里程碑式的发展。机器学习与模式识别方向，无论在工业界或是学术界都是热点，且新的成果不断涌现。在目前的国内本科生培养方案中，学生难以接触到学科的最新科研成果，使得教学与科研严重脱节。因此，需要在授课过程中，实时地将该方向的最新研究成果引入到机器学习科研实践课程中，培养本科生掌握人工智能的相关思维方法和技术手段，并将之应用于以后的科研工作之中。

（3）借鉴国外名校课程，编写深度学习实践指导

经典的深度学习课程，如美国斯坦福大学的CS231n、CS224d等，为深度学习的发展起到了尤为关键的作用。然而，目前国内深度学习课程并不多见，大多数学校仍然以经典机器学习算法为主要教学内容，尤其缺少适于本科教学使用的深度学习实践指导。Tensorflow在图像分类、音频处理、推荐系统和自然语言处理等多种场景下都有丰富的应用，是目前业界最为流行的深度学习框架。因此，建立以Tensorflow为核心的深度学习实践指导也是本项目的一个特色与亮点。

（4）实施本科生创新教育，鼓励学生将成果转化科研论文与专利

年轻人思维活跃，是人工智能社区的生力军，在人工智能的科研论文中占有一定优势。受益于当前中国的教育体制，我国高校本科生具备深厚的数学功底，这对于从事机器学习领域尤为重要。课程组希望通过机器学习科研实践课程的实施，为学生创造条件，鼓励他们学习成果转化科研成果，实施本科生创新教育。

5. 总结

通过机器学习科研实践课程体系改革，为深度学习的教学积累了丰富的经验与经典案例。机器学习科研实践课程对知识体系的先进性、实践平台的规范完整性、项目案例的示范性、交流共享平台的实用和开放性等方面进行了深化改革，为机器学习实践教育提供了借鉴和示范作用。

参考文献

- [1] 闵锋, 鲁统伟. “机器学习”课程教学探索与实践[J]. 教育教学论坛, 2014(53): 158-159.
- [2] 李勇. 本科机器学习课程教改实践与探索[J]. 计算机教育, 2015(13): 63-66.
- [3] 胡雪蕾, 孙明明, 孙廷凯, 等. 研究生“机器学习”课程教学改革实践与探讨[J]. 煤炭高等教育, 2012(1): 118-121.
- [4] 曲衍鹏, 邓安生, 王春立, 等. 面向机器学习课程的教学改革实践[J]. 计算机教育, 2014(19): 88-91.
- [5] 曾宪华, 李伟生, 于洪. 智能信息处理课程群下的机器学习课程教学改革[J]. 计算机教育, 2014(19): 60-62.

作者简介:

王雷全（1981-），男，辽宁丹东人，博士，计算机与通信工程学院，实验师，研究方向：计算机视觉。

吴春雷（1980-），男，河南台前人，博士，计算机与通信工程学院，副教授，研究方向：软件工程。

郭晓菲（1980-），女，山东平度人，硕士，计算机与通信工程学院，实验师，研究方向：通信工程。

（上接第49页）

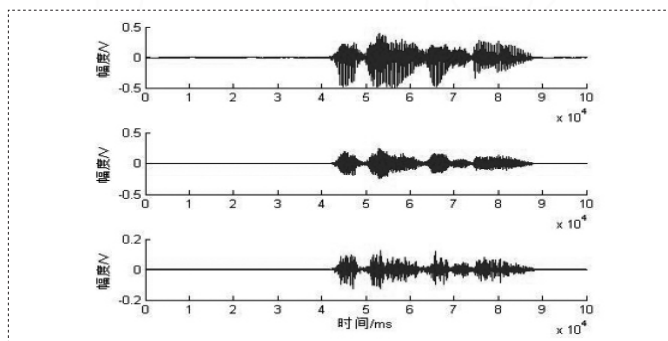


图5 去混响前后语音时域波形对比图

表1 去混响性能改善评测表

| 去混响方法 | 评测方法 | 客观评测方法 | | | 主观评测方法 | |
|--------|------------------|------------------|------------------|-----|--------|--|
| | 时域评测 | 频域评测 | | | | |
| | Δ SRNR/dB | Δ LPCC/dB | Δ MFCC/dB | 清晰度 | 自然度 | |
| 复倒谱去混响 | 1.9828 | -0.3038 | -11.5919 | 中 | 差 | |

本文采用主观和客观两类评测方法，客观评价方法包含时域和频域两类评测指标，分别是信混噪比改善 $\Delta SRNR$ ，这是一种时域的评测标准，还有线性预测倒谱系数距离改善 $\Delta LPCC$ 和 MEL 频

率倒谱系数距离改善 $\Delta MFCC$ ^[3]，这是两种频域评测指标。其中， $\Delta SRNR$ 反映信号在时域波形上的失真度，能够很直观的反映问题。 $\Delta LPCC$ 和 $\Delta MFCC$ 是从频域方面模拟人耳的听觉特性。

表1是采用复倒谱法去混响实验中应用两类评测方法所得出的结果，客观评测指标数值越高表明去混响效果越好，主观评测指标则是按照优、良、中、差、劣五个等级来说明，可以看出语音去混响有一定效果。

5 结束语

研究了混响是如何产生的以及混响的特性，论述同态信号处理的原理，说明了复倒谱法去混响的原理和运算方法，依据在复倒谱域上混响信号与语音信号所在位置不同的特征，设计出了适用的低通滤波器来减弱房间冲击响应对声音信号的干扰。分析了评价去混响效果的主观和客观的评测指标，并给出了评测结果，证明复倒谱去混响的方法起到了一定效果。

参考文献

- [1] 张德会, 陈光治. 复倒谱域语音信号去混响研究[J]. 声学技术, 2009, 28(1).
- [2] 张雪英. 数字语音处理及MATLAB仿真[M]. 北京: 电子工业出版社, 2016.
- [3] 宋知用. MATLAB在语音信号分析与合成中的应用[M]. 北京: 北京航空航天大学出版社, 2013.