



面向机器学习课程的案例教学

滕月阳^{1*} 沈婷婷²

1. 滕月阳, 东北大学, 研究方向: 机器学习, tengyy@bmie.neu.edu.cn

2. 沈婷婷, 东北大学, 研究方向: 机器学习, 2071232@stu.neu.edu.cn

*通讯作者: 滕月阳

摘要: 案例教学主要针对现阶段东北大学医学与生物信息工程学院机器学习课程理论与实践脱节的问题。通过增加实验学时和建设各种实验资源, 使学生对学习产生兴趣, 达到学生自主学习目的, 同时又可以锻炼动手操作能力。要将教师和学生放在同等的地位, 围绕项目展开合作, 形成以项目为驱动、以产业为导向的实验模式和教学体系。

关键词: 机器学习, 企业合作, 实验学时, 自主学习

2017年7月, 由国务院发表的《新一代人工智能发展规划》中提出, 要将中国的人工智能推向全球最高点^[1]。为落实这一举措, 次年4月教育部又提出《高等学校人工智能创新行动计划》, 通过校企联动, 推进校企深度融合的方式进行学习^[2], 目的在于通过高校的人工智能教育培养出能够推进社会发展的科技创新型人才^[3]。随后, 党中央、国务院加强了学习人工智能的力度。目前人工智能的发展热潮也推动着作为其核心技术—机器学习—成为各高校开设的热门课程, 并且各个领域、各个学科都设计了与机器学习相结合教学的案例。

1 机器学习的基本概念

机器学习发展的速度与使用的广泛性来源于它的强大作用, 其通过模仿人类的学习行为构造模型来解释分析历史经验数据, 从中学习到合理有用的特征和知识结构, 进而对真实世界中即将发生的事情做出预判^[4]。以机器学习中的深度学习为例, 首先要对获取到的数据进行预处理, 再利用这些处理好的数据输入网络训练模型, 接着用得到的模型对部分数据进行测试输出结果, 最后对测试结果进行优化。其解决问题的流程如图

2789-5521/© Shuangqing Academic Publishing House Limited All rights reserved.

Article history: Received September 13, 2022 September 20, 2022 Available online September 21, 2022

To cite this paper: 滕月阳, 沈婷婷 (2022). 面向机器学习课程的案例教学. 教育研究前沿进展. 第2卷, 第2期, 1-8.

Doi: <https://doi.org/10.55375/jerp.2022.2.7>

1 所示。

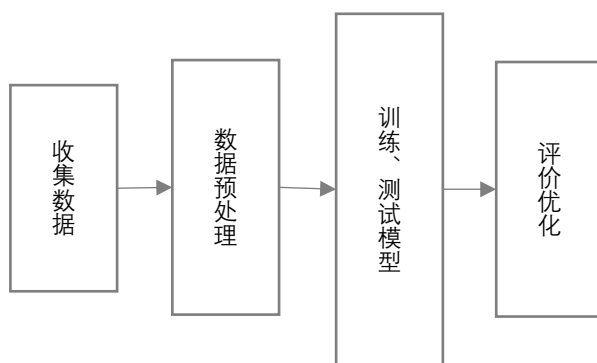


图 1 机器学习解决问题的一般流程

该课程涉及高等数学、概率论、计算机、信息论等多门学科，并且还在不断发展的过程中汲取其他新学科的知识进行结合，已经成为人工智能这一领域的主要创新源之一。机器学习涉及领域的多样性导致学生需要掌握很多繁杂琐碎的知识点。首先要掌握细节的理论基础知识，其次要有一定的理解分析思维能力，可以自己设计学习方法和解决办法，最重要的是要有一定的编程能力。所以在设计本课程教学方案时要围绕这几个方面来开展授课教学^[6]。

目前作者所在学院机器学习教学的主要内容围绕基础算法开展，例如朴素贝叶斯算法(Naive Bayesian)、k 近邻算法(K-Means)、AdaBoost 算法(Adaptive Boosting)、支持向量机(SVM)等^[6]。课程内容包含机器学习领域的基本概念、理论和算法，介绍特征提取、特征选择、分类、聚类等基本概念及方法。虽然学好理论是基础，但仅凭借理论知识的学习不足以深刻理解这门学科。机器学习的强大是在实践中体现出来的，要想学好这门课程必须要有足够的历史经验并将理论与实际结合起来。目前，本学院现有的实验课程安排不能满足教师教学需求，实验学时不足、实验环节设置不成熟并且无法与课程内容相匹配，实验部分往往只是授课教师演示其做过的一些项目成果，学生大多时候只是照本宣读，没有真正理解机器学习经典算法的根本原理并不能追根究底，也不具备用机器学习解决实际问题的能力。综合以上的问题，机器学习作为一门正在发展中的热门学科，它的教学模式还是不太成熟，所以改革必不可少并且刻不容缓。

2 机器学习与开设课程的关系

本学院是生物学、工程学、医学紧密结合的交叉学科^[7]，随这些学科的发展持续更新，十分依赖科技的进步，目前机器学习技术也成为了数学信息和计算机科学的研究热点^[8]。基于东北大学生物医学工程专业的培养方案和学生未来就业发展规划，学院设立生物医学工程专业(中外合作办学)、智能临床医学专业，并将机器学习纳入专业课程大纲。

2.1 专业介绍

生物医学工程专业。学生主要学习医学基础、医学影像学等方面的知识，能够运用工程手段解决生物、医学问题。

智能临床医学专业。学生主要学习计算机科学与技术、数据处理和分析等方面的知识，培养具有优秀且全面的智能医疗健康数据管理、智能影像诊断等科学素养的高端专业人才。

2.2 机器学习在专业中的应用

结合生物医学工程的发展方向，在处理医学数据时，例如测量病人心跳、血压等身体特征，或者使用机器拍摄患者影像数据，需要建立模型以便于应用到医学数据的预测、分类、分割和诊断等，都可以使用机器学习的方法对其进行处理。要想熟练掌握并且运用这些方法，学生必须要结合相关项目多动手处理实际案例，但实验课时较少、实践不足也是本课程目前存在的最大问题。

3 机器学习教学现状

机器学习课程的开设是为了让学生紧跟社会发展的步伐，了解当前社会关注的热门，熟练掌握机器学习的基本知识，能够运用学校给与的平台将所学知识学以致用，使用机器学习算法解决实际问题，以顺应未来职业发展大方向。但是目前教学现状还存在诸多问题需要改进：

3.1 教学方式传统、教材单一

从现有的课程教学方式来看，主要还是以教师课上讲解理论知识、学生被动接受为主，由于理论知识相对比较难懂，课上需要学生积极配合努力思考，学生在理解过程中就会感觉比较枯燥乏味，所以这种形式在一定程度上容易形成学习疲劳，也缺少必要的方法创新和学生参与度。

此外，市面上关于机器学习的教材很多，课程的内容选择是开放性的，并且不同的学科在不同领域中发挥的作用也不一样。由于课程内容的广泛性就会导致专业性下降以及内容模块化，很难将相邻模块之间联系起来，这样教师讲解的内容就会显得很琐碎，缺乏系统性。

3.2 教学内容多且难度大

由于机器学习是一门涉及多学科知识并且不断发展的新兴学科，对于各高校开设的课程需要包含很多方面的知识点，不仅需要考虑开设的课时情况也要结合学生的自主学习能力。再加上目前教案的参考材料不多，无论从老师还是学生的角度这都是一项挑战，需要在摸索中找到前进的方向。根本上说，机器学习难度大体表现在以下几个方面：

涉及内容的广泛性。机器学习包含的内容十分广泛，主要可概括为三大板块：有监督学习、无监督学习、特征学习，内容结构如图 2 所示。每个板块的内容又可以展开为很多个小节，不仅要理解基础概念，还必须熟练掌握才能够得以运用，这对很多学生来说都是一个不小的挑战。

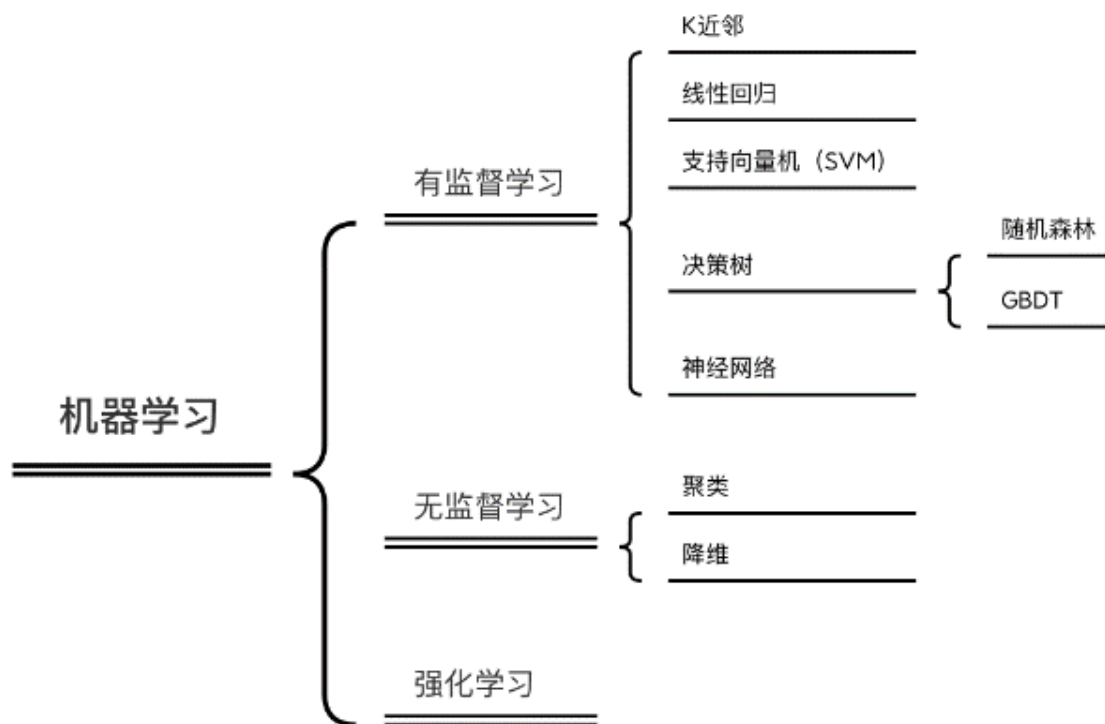


图 2 机器学习内容结构

多学科知识融合的复杂性。学习机器学习需要有强大的基础理论支持，该学科与多种其他学科交叉应用，不仅需要有很好的数学思维能力，做到熟悉很多基础的数学知识，如概率论、数理统计等，同时需要有一定的编程能力。对于很多学生来说，这些基础课程的学习和掌握就比较困难，机器学习还需要将它们结合起来进行运用就显得更为复杂。

3.3 基础知识体系不完善

学生在学习本课程之前会学习一些基础课程为后续学习做铺垫，如线性代数、统计学、概率论、编程语言、程序设计等。这些课程与机器学习没有一同开设，往往会间隔较长时间，所以会导致学生由于基础知识的遗忘而造成学习后续知识的压力。不仅需要学习新知识还要加强对基础知识的复习，同时需要锻炼自己的编程能力，多重压力可能会使学生的学习热情不高，这对于教师的教学是很不利的。

3.4 没有做到产学研医相结合

本课程的综合度比较高，要求教师在教授的过程中将理论与实践相结合，使学生们学有所用，但目前存在的课程设置中理论课程占比很大，仅有理论的学习不加以运用不是医工学科的学习目的。主要问题存在于以下几个方面：

实验课时短缺。学院现有的《机器学习》课程 16 学时，全部为理论学时。可以看到实验学时明显不足，甚至一些专业没有实验学时。现有的培养方案由于缺乏实验课或者实验课时不足，无法体现机器学习课程在本院课程中的重要作用，也无法保证学生达成培养目标。

实验数据缺乏。有实验课时的学生在跟随老师做实验过程中使用的是老师所给的数据，这些数据大多是教师之前所做项目使用过的数据，没有及时更新，也无法适用于所有专业学生的需求。

来自产业的师资队伍匮乏。没有来自产业的专家进行指导，比如产业项目的负责人或医院的医生，这会

导致在实验过程中学习的知识相对产业发展滞后，学生不能及时学习到最新的实验手段和应对实验错误的解决方法，也缺少与实战经验丰富专家交流学习的机会。

3.5 课程考核方式比较单一

本课程的考核方式主要还是以期末考试为主，部分专业没有实验课或者开设实验课但占比不大，所以会导致很多学生没有在真正理解后运用理论知识，只是“应付式”考试。这种单一的卷面考核不能够带动学生积极动手的实践能力，没有贯穿到整个课时学习过程当中，缺少综合能力的考察。

3.6 实验设施匮乏

目前本学院只能给学生提供公共空间实验，而公共实验室的电脑都设置开机自动更新，不能够保留每节课学生的实验结果，机器学习代码的实现需要在平台上建模，仅凭借课上的时间远远不够。

此外，算法功能的实现对于配置的硬件条件也有要求，需要考虑电脑的 GPU 速度、内存等各个方面。一方面受限于机器学习课程知识多、实践少，以至于学生动手能力被忽视；另一方面学生对于教师的讲解内容一知半解，再加上不能够很好的运用，就使得学生的能力无法提高，达不到设置机器学习这门课程的理想教学效果。

4 课程建设内容细节

目前关于机器学习的教学案例很少或者不成熟，本科生的机器学习教学也是近几年才逐步发展，存在多方面的问题需要解决，针对本学院在机器学习课程中所涉及问题，建设细节如下：

4.1 合理选择教学内容

针对课程内容繁多且复杂的问题，学院在设置上做到有取舍，主要讲授核心课程如基本算法思想、各种算法的实战案例等。课前要讲授的课件需要提前发至学习群，让学生预习。同时增加实验课时，在实践中教会学生如何搭建算法模型，利用算法解决实际项目中存在的问题。

4.2 注重培养学生自主学习能力

改变只有“教师主动，学生被动”的传统授课方式，把书本上的知识分章节安排给各个学生，让学生课前准备查阅相关资料，课上分享，教师提问之后进行补充，使得每个学生在课上都有课堂参与感，也达到学生主动学习的目的。具体教学流程如图 3 所示。

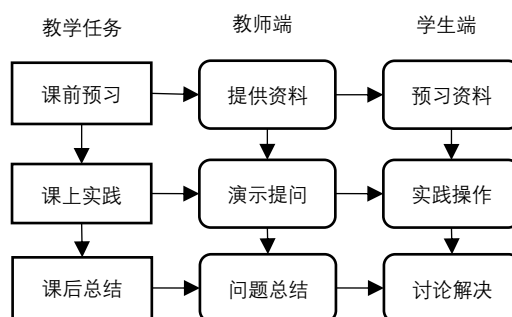


图 3 机器学习教学流程图

对于实践部分来说,在实践课时充足的前提下一个课时完成一个问题,不对学生限制实验平台(如 Alteryx、阿里云机器学习平台 PAI、Amazon SageMaker 等)、编程框架(如 Tensorflow、Pytorch、Keras 等)和编程语言(如 Python、C、Matlab 等),鼓励学生发挥主观能动性,对于同一问题提供不同的解决方案,这样提供不同解法的学生可以相互分享,教师在授课过程中也会受益良多。

4.3 以项目为驱动、以产业为导向

要使学生真正理解、运用知识必须要结合具体项目,了解日后的职业规划,明白现在所学知识对以后就业的影响,围绕项目展开合作,形成以项目为驱动、以产业为导向的实验模式和教学体系。具体可以分为以下几个方面:

加入实战应用案例分析。在实验课中插入学科前沿的项目,例如使用 WDBC 乳腺癌数据集^[9]进行良性肿瘤患者和恶性肿瘤患者的分类。WDBC 数据集共包含 569 个肿瘤患者的数据,在训练时需要先去除不必要的信息,比如患者 ID (Identity document),将关键性特征如诊断结果输入神经网络进行训练,再利用得到的模型参数对新的数据进行预测分类计算准确度。具体流程如图 4 所示。

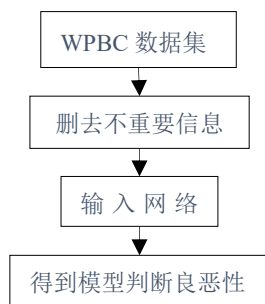


图 4 实战案例流程图

常用数据积累。机器学习实验的开展离不开对大量历史数据的分析,平时不仅要收藏公开数据集,也要学会自己制作数据集。例如深度学习中的有监督学习需要给网络输入成对的数据(数据和标签),但是一般在医学图像中很少能找到成对的数据,所以就要自己制作数据或者标签。

邀请相关专家指导。一方面,作为生物医学工程专业的学生,与医学的联系必不可少,但学院在教学中缺乏医生指导。日后需要与医院建立密切联系,从医生角度针对性教学;另一方面,学生需要接触实战项目,需要项目负责人指导,在实战中进行学习。

进入专业领域实践。本学院本科阶段机器学习教学采用实践和实习相结合的方式。部分专业设立实验课进行课堂实践,针对不同专业日后的就业发展需要增加实验课时,另外设置了暑假小学期给学生提供实习机会。在此基础上学院或者授课教师需要给学生创造合适的实习环境,例如本院学生可以与东软医疗系统有限公司以及各大医院合作,在医院采集数据后使用在公司参加的项目中。

5 教学评价与考核

5.1 现行的评价与考核制度

在做好相关课程改革的同时也需要完善课程考核制度,现有的机器学习课程考核评价方法主要有平时签

到、实验、笔试、课程设计等。考核评价各方面的占比如下：其中生物医学工程专业签到 10%，课上教师提问回答 20%，笔试包含期中与期末考试一共 70%；智能临床医学专业签到 10%，实验 35%，笔试 55%。期末考试一般以百分制闭卷形式开展，无论哪个专业在期末考试上的占比都是很大，并且具有“一票否决权”。这在一定程度上不是很合理，所以教师们一直在寻找更加合理有效的考核方法。

5.2 建立评价指标

各高校考核方法大致相同，不同的就是在不同评价体系中改变各个指标的占比。改变以往唯“期末分数”论的方式，把重点转移到实践上来。

对于生物医学工程专业没有实验课时的问题，应当安排课时，需减少 4 个理论课学时，增开 6 个实验学时，过程中提交一次实验的课程作业，无期中考试，将理论课与实践课穿插教学，每 3 节理论课后安排 2 个学时的实验课。具体课程安排及考核方式如表 1 所示。

表 1 生物医学工程考核制度

形式	时间（学时）	权重
理论课	1-4, 7-10, 13-16	5%
实验课	5-6, 11-12, 17-18	5%
课程作业	5-18	45%
期末考试	18	45%

对于智能临床医学专业来说需要改变的就是在实验中的占比，由原来的 35%提升到 45%，笔试的占比由原来的 55%下调到 45%，其他各项考核指标不变。

最终综合成绩的评价标准为所有占比的总和，公式如下：

$$\text{总成绩} = \text{签到} \times 10\% + (\text{实验} + \text{期末}) \times 45\%$$

6 结语

本文分析了本科机器学习的现状和现阶段该课程存在的一些问题，给出了面向本科生开展机器学习课程的相关措施和解决办法。

总的来说，课程改革的有效实施需要学院、老师和学生的共同配合，学院提供硬件基础，老师改善教学方式，学生发挥主观能动性，多方共同努力才可以为学生以后的科研和工作打下坚实基础。

参考文献:

- [1] 蒋良孝.机器学习课程教学的实践探索[J].新课程研究,2019(23).
- [2] 关汉男,万昆,吴旻瑜. 校企深度融合:中国高校发展人工智能的“关键一招”——《高等学校人工智能创新行动计划》解读之二[J].远程教育杂志, 2018.
- [3] 赵雪峰,施珺.面向本科生机器学习课程的教学探索[J].计算机教育,2021(2).
- [4] 周夏冰,陈飞.机器学习课程教学探索[J].电脑知识与技术,2020 (30).
- [5] 徐新洲,魏建香,李燕.机器学习大学本科课程的教学活动开展[J].科技创新导报,2020 (10).
- [6] 徐洪学,孙万有,杜英魁,汪安祺.机器学习经典算法及其应用研究综述[J].电脑知识与技术,2020 (33).
- [7] 李俊,周宇葵.数据挖掘在生物医学工程文献检索中的应用[J].图书馆学研究,2008(1).
- [8] 杨秀锋.基于机器学习的生物医学数据处理方法研究[J].中国科学院大学, 2014.
- [9] UCI Machine Learning Repository [DB/OL].<http://archive.ics.uci.edu/ml/datasets/>.