

“数据标注实操”课程设计与教学改革

李 晓 张登辉

浙江树人学院信息科技学院 浙江 杭州 210015

摘要：伴随信息技术飞速发展，新工科、新数据科学的高速融入，社会众多行业对新一代信息技术人才的需求不断逐步扩充，包含大数据、人工智能、云计算等。截止到2023年共有937所高职高专院校开设大数据技术专业，有众多高等院校逐步开设数据科学与大数据专业。面对社会的数据人才缺口，需要增加数据标注的基础教学。本文对新课程《数据标注实操》进行课程设计和实施，实现为人工智能、机器学习提供数据标注基础。

关键词：数据标注；人工智能；课程设计

引言：

随着人工智能、机器学习的发展，数据标注行业也异军突起。《数据标注实操》课程是计算机科学与技术、数据科学与大数据、人工智能专业等专业的选修课，数据标注是人工智能项目处理基础环节。采用由浅入深的教学，由数据标注概念开始，系统梳理标注数据的分类和标注流程，对数据标注质量检测和管理，根据数据标注的应用实现案例式教学，让学生掌握通用数据标注技术。

课程介绍数据标注基本概念以及发展趋势，系统地梳理数据标注分类及数据标注流程，对数据标注质量检验和数据标注管理介绍，学习在热门行业中数据标注的应用，通过数据标注实战案例让学生掌握数据标注技术。课程目标是培养学生掌握数据标注技术，独立完成数据标注处理工作。培养学生进行特定领域的应用，能够成为具体系统级水平的数据标注工程师。本文结合教学实际，从教学设计思路、课程改革等方面进行探讨，以期为该课程的教育教学提供参考。

一、数据标注实操课程的设计思路

1. 数据标注概念和方式

数据标注是标记数据集中的每一项数据，以更好地被人工智能算法和机器学习算法识别和理解。数据标注可以是人工标注、自动标注和半自动标注，常见的标注方式有文本标注、图像标注、视频标注等。其中人工标注：通过人工的方式，人们根据预先定义的标准规则，对数据进行标注^[1]。这需要专业的标注员或团队完成，对图像进行物体检测或图像分类的标注，对文本进行情感分类的标注等。自动标注利用一些现有的算法或模型，对数据进行自动标注。利用计算机视觉算法对图像进行标注，运用自然语言处理技术对文本进行标注。半自动标注结合人工和自动标注的方式，人工标注员通过一些辅助工具或算法进行标注。标注员通过使用指定的标注工具，结合自动标记的结果，对数据进行进一步验证和修正。数据标注的目的是收集和标记数据，更好支持机器学习和人工智能算法的应用。

2. 突出课程重点

在数据采集、清洗、标注到模型训练和部署的关键环节中，数据标注（Data Annotations）是指对收集到的、未处理的原始数据或初级数据进行加工处理，包括文本、语音、图片、视频等类型的数据，并转换为机器可识别信息的过程。重点介绍数据标注的知识、实际操作流程案例^[2]。

3. 确定教学目标

目标可分为：知识、能力和素质。知识层面包含数据标注的理论知识，了解数据标注的概念、作用、多种数据，掌握数据标注的操作流程。让学生对于数据标注的基本概念和实操技能，具备考取数据标注师的证书的能力。

1.4 确定课程内容

根据目标设计课程的教学内容。为学生清晰化、系统化科学掌握数据标注知识和技能，课程按照数据的类型进行案例实操教学，以文本、语音、图片、视频等不同类型的开展^[3]。

针对不同类型的数据选取软件及实操案例。文本数据用Notepad++软件，案例是蛋糕店评价；语音标注用精灵标注实操软件，案例是文本情感；图片标注用labelme软件，案例是人脸图像；视频标注用labeling软件，案例是道路场景。

二、数据标注实操课程设计

1. 教学理念和设计方法

本课程遵循以学生为中心的教学理念和教学方法。学生学习理念和实操流程，针对不同的类型的数据标注软件，提高自身操作和小组团队协作能力，树立数据标注团队合作的操作意识。

（1）充分进行理论教育

目前国内外高校拥有数据科学与大数据技术专业的有：北京大学、清华大学、浙江大学、国防科技大学、北京航空航天大学、北京邮电大学、哈尔滨工业大学、

上海交通大学、南京大学、华中科技大学,美国数据科学专业有哈佛大学、芝加哥大学、哥伦比亚大学等,多数成为经典教程便于反复学习,针对实时性较强的数据标注实操类课程一直更新^[4]。

全国高校依照资源共建、信息共享原则合作建设网络视频教学平台,名校教师认真讲解、深度剖析,形成经典公开课,充分满足数据标注的学习需求。本课程将充分利用数字化技术、网络资源制作丰富的教学和辅导材料,构建良好的网络学习方式,加强直观教学和实操,强化对数据标注的理解。

(2) 增加实践机会

学校作为教学的主体,在教师的带领指导下,合运用数据标注基础、常识以及软件进行数据标注实际操作^[5]。在数据标注的学习过程中,首先需要掌握数据标注的规则,然后使学生学会实用标注步骤对软件进行实操。

通过课堂理论学习与多种数据标注实践操作,全面实现对数据标注理解和操作。在学习及实操过程中,建立数据标注团队,促使学生自主地对标注项目需求进行体验,提高学生的综合素质。

(3) 数据标注职业发展定位

教师在教学中需要有机地融入数据标注师职业发展定位,结合社会实际工作需求,指导学生对未来数据标注的规划和定位。针对目前信息化发展的长三角发展需求,数据标注行业的发展需,形成的数据标注职业需求较多。根据实际调查发现,数据标注项目工程师、数据标注团队管理师、数据标注实际操作师。学生可以定位数据标注实际操作师岗位,积极学习然后发展为团队管理师,成熟后可以管理数据标注实际项目。在课程教育中,形成对相关知识技能的引导,引导学生对于数据标注职业有良好定位,除此之外还有数据科学、人工智能对于数据标注的需求^[6]。教师不仅要传播正确的数据标注知识,还需要重视智慧场景的需求,实现数据标注与

人工智能、机器学习的关联性理解,明确数据标注的重要性,对不同类型的数据标做操作进行掌握。

2. 课程考核评价方式

数据标注实操的考核分为理论和实操两部分。学生能够针对理论问题的实际回答,主要包含:数据实操的基本概念和操作步骤。其次,学生需要通过数据标注实操的学习,能够完数据标注团队的搭建和分工协作^[7]。

针对数据标注实操的考核评价环节,主要依托于文本、语音、图片、视频等类型的案例数据实操,学生需要对自己的数据标注操作记录和案例总结,期末考试学生需要对实际操作总结个人感受进行汇总,需要多方面综合评价。

3. 课程持续优化措施

第一,充分理解数据标注应用领域

时刻关注数据标注、人工智能、机器学习模式的相关技能知识,实时引入课堂。从理论中的知识到实际操作,能使学生更好地理解数据标注的基础性,增强对于人工智能、机器学习的数据需求理解。

第二,实现线上和线下教学结合

线上利用学习总结,熟悉基本操作步骤;线下利用软件安装、样本数据、实际操作的形式使学生充分理解掌握数据标注,同事对数据标注职业进行定位,通过相关行业的职业需求,能够获取数据标注职业化证书^[8]。实现学生对于数据标注分工协作的职位意识。

第三,实现教学内容的实时更新

信息技术的更新速度比较快,教学内容需要根据数据标注的应用场景进行更新总结,能够及时提出相关的应用策略。需要快速了解社会领域的发展方向,对数据标注的场景应用做出成熟的应对方案,高校的数据标注进行职业化定位具有一定的就业价值,对于区域化经济发展的需求能够快速适应。

结语:

“数据标注”是新时期信息科技教育教学发展的重要基础方向,评价方式主要依托于理论和实操共同评价,为学生的职业范围和能力保准提出较为合适的教学方案,将培养和评价系统的定位与课程体系,实现数据标注能力培养在课堂上完成,报职业能力标准融入教学内容,能够为数据标注行业提供合适的人才。

参考文献:

- [1] 何友,朱扬勇,赵鹏等.国防大数据概述[J].系统工程与电子技术,2016,38(6):1300-1305.
- [2] LI A,ZANG Q,SUN D,et al.A text feature-based approach for literature mining of IncRNA-protein interactions[J].Neurocomputing,2016,(7):1417-1421.

[3] 李明,李秀兰.基于结果模式的 Deep Web 数据标注方法[J].计算机应用,2011,31(7):1733-1736.

[4] 孙丽娟.大数据时代河南语言服务产业发展的路径研究[J].决策探索(下),2021(5):17-18.

[5] 李贞,刘海燕,庞禄申.基于众包模式的数据采集和标注系统研究[J].大众科,202123(10):15-18.

[6] 轩中.人工智能行业中隐藏的“富士康”式劳动密集型产业[J].互联网周刊,2018(21):26-27.

[7] 数蔡莉,王淑婷,刘俊晖,等.数据标注研究综述[J].软件学报,2020,31(2):302-320.

[8] 崔爽.数据标注师:人工智能背后的人工力量[J].科学中国人,2019(10):72-73.