基于TensorFlow构建中国传统文化机器学习课程教学大纲

二〇二二年六月

一、课程简介

1.课程基本信息

|  |  |
| --- | --- |
| 课程名称（中文） | 基于TensorFlow构建中国传统文化机器学习课程 |
| 课程名称（英文） | Building Chinese Traditional Culture Based on the TensorFlow Machine Learning Framework |
| 适用专业 | 计算机科学与技术、人工智能 |
| 总学时 | 24学时 |
| 先修课程 | 高等数学、线性代数、Python程序设计 |

2.教学目标

中国有五千年的文明史，历史悠久，蕴藏丰厚，其传统文化丰富多彩，涉及诗词、绘画、 保健、中药等，中国传统文化是东方文化的代表。如何利用现代机器学习将中国传统文化进行淋漓尽致的呈现，是值得我们探索的方向。本课程旨在利用TensorFlow这一强大的开源机器学习框架，深入探索和构建针对中国传统文化（如手写汉字识别、古诗词生成、中药饮片识别、中药方剂靶点预测等）的机器学习模型。

通过课堂教学和实践教学相结合，本课程将探讨如何将 TensorFlow 融入到中国传统文化的机器学习课程中，使学生能够深入理解和掌握机器学习方法，培养具有一定传统文化底蕴的机器学习人才。具体课程目标如下：

知识目标：理解并掌握TensorFlow的基本原理和编程基础。学习机器学习的基础理论，包括神经网络、卷积神经网络、循环神经网络和生成对抗网络、大模型微调等。掌握如何针对不同的应用场景选择、构建和应用深度学习模型。了解中国传统文件的特点和处理需求。

能力目标：能够使用TensorFlow建立和训练神经网络模型。能够在处理真实图像、文本等数据时，运用所学知识提高模型性能。能够针对中国传统文件开发出有效的机器学习应用。培养独立解决人工智能领域问题的能力。

德育目标：培养对传统文化的尊重和传承意识。增强对人工智能技术的责任感，理解其在社会和文化领域的作用。培养团队合作精神和批判性思维。

二、选用教材及教学资源

**1．选用教材**

《机器学习实战：基于Scikit-Learn、Keras和TensorFlow》，2020年10月第2版，奥雷利安·杰龙，北京：机械工业出版社

**2．主要参考书目**

（1）《统计学习方法》，2012，李航，北京：清华大学出版社

（2）《机器学习》，2016，周志华，北京：清华大学出版社

（3）《An Introduction to Statistical Learning》，2021Second Edition，Gareth James

**3．其他学习资源**

（1）2022吴恩达机器学习Deeplearning.ai课程，吴恩达，哔哩哔哩

（2）An Introduction to Statistical Learning，https://www.statlearning.com/

三、教学内容和教学要求

**第一章 基于Tensorflow卷积神经网络中文手写数字体识别**

【目的要求】

掌握

1. 卷积神经网络（CNN）的基本原理

2. TensorFlow框架的使用

3. 中文手写数字体识别的流程和方法

熟悉

1. 数据预处理和增强技术

2. 卷积神经网络的架构设计

了解

1. 深度学习在图像识别中的应用

2. 模型优化和调参技巧

【教学内容】

第一节 卷积神经网络（CNN）的基本原理

1. 卷积层的作用
2. 池化层的作用
3. 全连接层的作用
4. 激活函数的选择

第二节 TensorFlow框架的使用

1. TensorFlow的安装与配置
2. 模型的定义、编译和训练

第三节 中文手写数字体识别的流程和方法

1. 数据集的准备
2. 数据增强技术
3. 模型的构建和训练
4. 模型的评估和预测

第四节 数据预处理和增强技术

1. 图像的归一化
2. 数据增强的方法
3. 数据生成器的使用

第五节 卷积神经网络的架构设计

1. 卷积层的设计
2. 池化层的设计
3. 全连接层的设计

Dropout层的使用

第六节 模型训练与评估

1. 训练集、验证集和测试集的划分
2. 损失函数和优化器的选择
3. 模型评估指标

第七节 模型优化和调参技巧

1. 学习率的调整
2. 批次大小的选择
3. 超参数的调整

【教学方法和时数】

教学方法：课堂讲授，案例法，互动问答法，实验操作；

学时安排：4学时

**第二章 基于RestNet中药饮片识别**

【目的要求】

掌握

1. 残差网络（RestNet）的基本原理

2. TensorFlow框架的使用

3. 中药饮片识别的流程和方法

熟悉

1. 数据预处理和增强技术

2. 残差网络的架构设计

了解

1. 深度学习在图像识别中的应用

2. 模型优化和调参技巧

【教学内容】

第一节 残差网络（RestNet）的基本原理

第二节 TensorFlow框架的使用

第三节 中药饮片识别的流程和方法

第四节 数据预处理和增强技术

第五节 残差网络的架构设计

第六节 模型训练与评估

一、训练集、验证集和测试集的划分

二、损失函数和优化器的选择

三、模型评估指标

第七节 模型优化和调参技巧

【教学方法和时数】

教学方法：课堂讲授，案例法，互动问答法，实验操作；

学时安排：4学时

**第三章 基于Tensorflow RNN生成古诗词**

【目的要求】

掌握

1. 循环神经网络（RNN）的基本原理

2. TensorFlow框架的使用

3. RNN生成古诗词的方法和流程

熟悉

1. 文本数据预处理技术

2. RNN的架构设计

了解

1. RNN在文本生成任务中的应用

2. 模型优化和调参技巧

【教学内容】

第一节 循环神经网络（RNN）的基本原理

第二节 TensorFlow框架的使用

第三节 RNN生成古诗词的方法和流程

第四节 文本数据预处理技术

第五节 RNN的架构设计

第六节 模型训练

第七节 模型优化和调参技巧

【教学方法和时数】

教学方法：课堂讲授，案例法，互动问答法，实验操作；

学时安排：4学时

**第四章 基于Pix2Pix 素描生成中国山水画**

【目的要求】

掌握

1. Pix2Pix网络的基本原理及应用流程。

2.生成对抗网络（GAN）在图像风格转换中的应用。

3. 基于TensorFlow和Colab的深度学习模型训练流程。

熟悉

1. 数据预处理与增强技术，特别是图像边缘检测方法。

2. Pix2Pix网络的生成器与判别器架构设计。

了解

1. 深度学习在图像风格迁移和艺术生成中的应用。

2. 生成对抗网络的优化与调参技巧。

【教学内容】

第一节 生成对抗网络（GAN）基本原理

1. GAN 的基本概念与 Pix2Pix 的网络架构。

2. 生成器和判别器的设计及其交互方式

第二节 边缘检测与数据预处理

1. 图像边缘检测的重要性及其在山水画生成中的应用。

2. 数据增强技术：旋转、翻转、缩放等方法的实现。

3 使用 OpenCV 和 TensorFlow 进行边缘检测处理。

第三节 Pix2Pix 网络架构设计

1. Pix2Pix 的生成器（Generator）和判别器（Discriminator）模块。

2. 生成网络的上下采样模块设计。

3. 判别网络的图像分类模块及对抗损失。

第四节 模型训练与评估

1. 数据集划分：训练集、验证集、测试集的划分。

2. 损失函数与优化器：对抗损失、L1损失；Adam 优化器的应用。

3. 模型评估指标：使用生成图像的视觉质量及其相似性指标。

第五节 Pix2Pix 生成模型的调参与优化

1. 关键超参数的调节方法，如学习率、L1损失权重等。

2. 使用 Colab 调试工具分析模型训练过程。

3. 讨论模型优化技巧，提高生成效果。

第六节 应用与扩展

1. 山水画风格生成的实际应用场景讨论。

2. 探索生成模型在其他艺术风格生成中的潜在应用。

【教学方法和时数】

教学方法：课堂讲授，案例法，互动问答法，实验操作；

学时安排：4学时

**第五章 基于双线性注意力的中药方剂靶点预测研究**

【目的要求】

掌握

1. 图卷积神经网络等模型组件的基本原理

2. 双线性注意力机制在中药方剂-靶点预测中的应用

3. 当归芍药散作为案例的靶点预测研究

熟悉

1. 药物-靶点对数据集的处理和特征提取

2. 预训练蛋白质语言模型的概念

了解

1. 中药方剂靶点研究的概念和意义

2. 深度学习模型的训练和评估流程

【教学内容】

第一节 图卷积神经网络基础

一、图结构数据的表示方法

二、图卷积中的消息传递机制

第二节 双线性注意力机制原理

一、注意力机制的基本概念

二、双线性注意力在药物-靶点预测中的应用

第三节 模型搭建流程

第四节 预训练蛋白质语言模型

第五节 数据来源和预处理方法

第六节 模型训练与评估

一、损失函数和优化器的选择

二、模型评估指标

第七节 当归芍药散案例研究

【教学方法和时数】

教学方法：课堂讲授，案例法，互动问答法，实验操作；

学时安排：4学时

**第六章 基于Gemma微调的中成药问答**

【目的要求】

掌握

1、自然语言处理 (NLP) 基本概念和常见任务

2、Gemma 模型的基本原理和使用方法

4、中成药问答系统的构建流程

熟悉

1、数据收集和标注方法

2、文本预处理技术

3、评价指标和评估方法

了解

1、NLP 在中医药领域的应用

2、模型优化和调参技巧

【教学内容】

第一节 自然语言处理 (NLP) 基本概念

第二节 Gemma 模型介绍

第三节 中成药问答系统概述

第四节 数据收集和标注

第五节 文本预处理

第六节 Gemma 模型微调

模型选择和配置

训练过程和参数调整

模型评估和优化

第七节 未来发展趋势

【教学方法和时数】

教学方法：课堂讲授，案例法，互动问答法，实验操作

学时安排：4学时

1. 课程考核评价

《基于TensorFlow构建中国传统文化机器学习课程》为考查课。采用综合考核的方法，通过多种形式对学生在本课程的学习中进行评价。即以实验操作考核为主，并结合考勤、自主学习、作业、上机实验报告等方面的考查，综合评定成绩。其中，期末项目考核占总评的50%，考勤、作业、自主学习、实验报告占总评的50%。

**各教学评价环节成绩占比**

|  |  |
| --- | --- |
| 评价环节 | 占比（%） |
| 1.考勤 | 10 |
| 2.作业 | 15 |
| 3.自主学习 | 10 |
| 4.实验报告 | 15 |
| 5.期末考试成绩 | 50 |