

2023级新生入门任务

前言

2023级新生入门任务为期八周，包含必须完成学习的基本知识，以及需要自己长时间持续学习的知识。所有学习的内容必须有**==相关笔记==**（不限格式，可 .docx、.md）以及**==代码==**的输出，每周学习得到的笔记和代码使用git远程仓库管理，以周为单位保存，如：

```
name
|-- week1
|   |-- 李宏毅视频笔记
|   |   |-- xxx.md 或 xxx.docx
|   |-- git工具使用笔记
|   |   |-- xxx.md 或 xxx.docx
|-- week2
|   |-- 李宏毅视频笔记
|   |   |-- xxx.md 或 xxx.docx
...
```

每周周末填写**周报**，并发送到邮箱：xxx，邮件中添加excel周报附件，以及git远程仓库的地址。大家学习中的疑惑也可以记在笔记或者邮件中，各位老师师兄看到都会回复。

必须完成学习的知识包括：

1. 深度学习入门

主要参考资料：

- [李宏毅2021/2022春机器学习课程](#)
(注，本视频集完整记录了李教授2022春包括tutorial的所有课程，除了任务安排指定，课程中选修部分都不做硬性要求；)
(作业部分统一完成2022年版的，2021年的有兴趣的自行查看)
- [跟李沐学AI](#)
- [【斯坦福CS224N】2023公认最通俗易懂的《深度学习自然语言处理》课程](#)

2. Linux 基本命令

3. git 工具

(注，涉及有关项目实践和任务输出的内容，都使用gitee进行版本和内容管理)

4. 循环神经网络 (Recurrent Neural Network, RNN)、Transformer、部分预训练模型 (Pre-trained Model)、大语言模型 (Large Language Model) 入门

5. pytorch

主要参考资料：

- [PyTorch 1.4 教程&文档 \(apacheecn.org\)](#)

需要长时间学习的知识包括：

1. Python

2. 循环神经网络 (Recurrent Neural Network, RNN) 、Transformer、部分预训练模型 (Pre-trained Model) 、大语言模型 (Large Language Model) 入门
3. [跟李沐学AI的个人空间 哔哩哔哩 bilibili](#) 论文精读系列

==注意，不要花费过多的时间在各种环境配置，以及扩展插件等功能上，所有环境都只是工具，为实现代码而服务==

第一周 (6.19-6.25)

任务

1. 神经网络与深度学习相关概念以及Pytorch基础操作
 - [第一节 - 机器学习基本概念简介](#)第一节内容：P1-P4, P6, P7, P9-P11
深度学习：P13-P14
2. git 代码仓库的学习
 - [Git使用教程,最详细, 最傻瓜, 最浅显, 真正手把手教 - 知乎 \(zhihu.com\)](#)
 - [收藏了! Git 核心操作图解 \(qq.com\)](#)
 - [GitHub仓库快速导入Gitee及同步更新 - Gitee.com](#)
 - [Gitee 帮助中心 - Gitee.com](#)

输出

1. 在[Gitee](#)上，建立自己的笔记仓库，利用git工具进行自己笔记的版本管理，学会分支、合并等基本操作
2. 视频课程学习笔记并将自己的笔记上传到在1.上建立的自己的gitee笔记仓库中，作业部分不用完成（可以根据视频讲解和博主提供的git/gitee仓库中的内容学习）

第二周 (6.26-7.02)

任务

1. 机器学习和深度学习任务
 - [第二节 - 机器学习任务攻略](#)
2. Linux环境配置以及基本命令
 - 在windows上运行WSL2
 - [适用于 Linux 的 Windows 子系统文档 | Microsoft Docs](#)
 - [Windows10/11 三步安装wsl2 Ubuntu20.04 \(任意盘\) - 知乎 \(zhihu.com\)](#)
 - 在windows子系统WSL2中建立Ubuntu环境
 - 了解Linux基本操作
 - [第一章：引言 · The Linux Command Line 中文版 · 看云 \(kancloud.cn\)](#)

输出

1. 视频课程学习笔记并上传到自己的gitee笔记仓库中，章节作业部分不用完成
2. 完成linux的配置以及基本操作的学习

第三周 (7.03-7.09)

任务

1. 卷积神经网络、循环神经网络
 - [第三节 - 卷积神经网络\(CNN\)](#)
 - [第4讲 \(选修\) : RNN\(Part I & Part II\)](#)
2. 在WSL2下安装anaconda或miniconda（功能更轻量级，占用空间更小）并配置PyTorch环境（PyTorch有GPU版和CPU版，个人笔记本上使用CPU版就行，后续进入实验室后会分配有GPU的服务器）
 - 安装方法自行搜索，例如[linux安装anaconda及配置pytorch环境](#)
3. 在WSL2中安装jupyter notebook，在windows中打开可视化界面，并熟悉jupyter使用方法
 - 安装方法自行搜索，例如[搭建 Python 轻量级编写环境 \(WSL2+Jupyter 自动开启本地浏览器\) - 知乎 \(zhihu.com\)](#)
 - 使用方法：
 - [Jupyter Notebook介绍、安装及使用教程 - 知乎 \(zhihu.com\)](#)
 - [最详尽使用指南：超快上手Jupyter Notebook - 知乎 \(zhihu.com\)](#)
4. PyTorch包学习
 - [什么是PyTorch? \(apacheecn.org\)](#)
5. PaddlePaddle学习
 - [PaddlePaddle与PyTorch的转换](#)
 - [飞桨PaddlePaddle-源于产业实践的开源深度学习平台](#) 第一、二、五、六、七章

输出

1. 视频课程学习笔记，以及[作业HW1](#)、[作业HW2](#)、[作业HW3](#)的试验
2. 完成环境配置以及对PyTorch的初步了解
3. 初步了解PaddlePaddle，以及与PyTorch的转换关系

第四周 (7.10-7.16)

任务

1. 神经网络训练任务
 - [第四节 - 自注意力机制\(Self-attention\)](#)
 - [第五节 - 类神经网络训练不起来怎么办](#)
 - 主要涵盖Seq2Seq、encoder-decoder框架及Transformer相关知识
2. PyTorch入门中剩下部分的学习
 - [PyTorch 1.4 教程&文档 \(apacheecn.org\)](#)
3. PaddlePaddle包配置尝试
 - [使用指南-使用文档-PaddlePaddle深度学习平台](#)

输出

1. 视频课程学习笔记以及作业
2. 完成PyTorch教程上的代码教程
3. 完成环境配置以及对PaddlePaddle的初步实践

第五周 (7.17-7.23)

任务

1. NLP相关基础知识

1. 自然语言处理导引及词向量[Lecture 1 - Introduction and Word Vector](#)

1. Word2vec参考资料

1. [Efficient Estimation of Word Representations in Vector Space](#)
2. [Distributed Representations of Words and Phrases and their Compositionality](#)
3. [word2vec Parameter Learning Explained](#)

2. 依赖分析[Lecture 4 - Dependency Parsing](#)

3. 语言模型[Lecture 5 - Language Models and RNNs](#)

4. 预训练模型[Lecture 10 - Transformers and Pretrained Model](#)

2. 实战演练-阶段一

1. [nlp](#)
2. [midlevel_data](#)
3. [nlp_dive](#)

输出

1. 视频课程学习笔记以及作业
2. 阶段一涵盖NLP中模型构建的基本步骤，记录实验结果

第六周 (7.24-7.30)

任务

1. NLP相关任务

1. 翻译任务[Lecture 7 - Translation, Seq2Seq, Attention](#)

这部分内容中包含的序列模型，注意力机制在第四周的神经网络训练任务部分已经讲述过，所以该部分可以略看。

2. 问答任务[Lecture 11 - Question Answering](#)

3. 自然语言生成任务[Lecture 12 - Natural Language Generation](#)

4. 共指消解任务[Lecture 13 - Coreference Resolution](#)

2. 实战演练-阶段二

1. [词嵌入 TransE实现](#)
2. [词嵌入 Glove实现](#)
3. [序列模型与Attention](#)
4. [情感分析模型](#)
5. [关系抽取模型](#)

输出

1. 视频课程学习笔记以及作业
2. 阶段二主要涵盖词嵌入学习及基础模型应用两部分内容完成模型调优，记录实验结果

第七周 (7.31-8.06)

任务

1. 大语言模型相关

[T5 and Large Language Model](#)

[【生成式AI】ChatGPT 原理剖析](#)

P1-P3、P7-P14

2. 实战演练-阶段三

若对预训练模型微调部分感兴趣，可以参考huggingface主页。

1. [文本分类](#)

2. [问答1](#)

3. [问答2](#)

以下内容选择性完成

1. [BERT微调](#)

(本地或教程中提及Colab中执行)

2. [RoBERTa微调](#)

输出

1. 视频课程学习笔记

2. 阶段三涵盖预训练模型微调的基本过程，主要用意为熟悉huggingface平台，记录实验结果

第八周 (8.07-8.13)

任务

1. 其他

1. 知识增强[Lecture 15 - Add Knowledge to Language Model](#)

2. 伦理道德[Lecture 16 - Social & Ethical Consideration](#)

3. 模型分析与可解释性[Lecture 17 - Model Analysis and Explanations](#)

4. NLP未来[Lecture 18 - Future of NLP](#)

2. 实战演练-阶段四

1. [语义检索系统](#)

2. [智能问答系统](#)

3. [情感分析](#)

输出

1. 视频课程学习笔记以及作业

2. 阶段四涵盖预训练模型微调的基本过程，主要目的为熟悉paddleNLP平台，记录实验结果