



周晨星

年龄：24

185 8305 3106

哈尔滨理工大学

cxzhou7@163.com

## 教育背景

2017.09 – 2021.06

哈尔滨理工大学

自动化(自动化学院)

一本

2021.09 – 至今

哈尔滨理工大学

电子信息(自动化学院)

硕士在读

## 技能情况

- ◆ 熟练掌握 python 基础知识，熟悉 linux 指令，具有良好的面向对象编程思想。
- ◆ 熟悉 pytorch 和 Tensorflow 深度学习框架，并对网络训练流程有深入理解。
- ◆ 熟悉 CNN、RNN、GRU、LSTM、Transformer、BERT、HMM、CRF、InstructGPT 等深度学习和机器学习算法的应用和原理，并在项目中熟练运用。
- ◆ 熟练掌握脑电信号采集、预处理、特征提取等操作，并有相关的实战经验。
- ◆ 有复现开源社区代码能力，熟练使用 Vscode、Pycharm 等开发工具，熟练使用 git 命令。
- ◆ 熟练使用 MNE、EEGLAB 工具包，熟悉在集群训练网络模型。
- ◆ 熟练使用 Excel、PPT、Word 等办公软件。

## 荣誉证书

硕士一年级一等学业奖学金

硕士二年级一等学业奖学金

cet6

## 实习经历

科大讯飞股份有限公司

助理研究算法工程师

2023.1.3-2023.4.3

### 英语口语考试评分项目

**项目描述：**通过使用深度学习算法，搭建端到端系统。实现输入为学生作答的口语考试答案，输出为学生的考试分数。

**主要技术：**pytorch + Huggingface + fairseq + pandas + 集群训练

**责任描述：**1、对历史模型进行改进，历史模型采用 pipeline 形式，先训练由学生作答到给出人工评语（主语缺失、谓语形式错误等），再通过人工评语得到最终得分。首先将人工评语从中文改成更专业的英文人工评语；使用 “[cls]问题[sep]答案[sep]评语[sep]” 和 “[cls]答案[sep]评语[sep]” 来替换只使用评语作为输入；再通过历史数据训练基底模型，并在各省份的定标集上进行 finetune。2、使用端到端的系统完成口语考试的评分，在 huggingface 上调研适合的模型，使用 “[cls]问题[sep]学生作答” 作为模型的输入，直接输出考试得分。使用数据增强的方法扩充数据，采用 bert 的 mask 方法对学生作答进行 mask 操作以降低过拟合，并在集群上进行训练。

**工作结果：**1、历史模型改进后有提升但不明显，在 finetune 前改进前相关度 0.793，改进后为 0.818。2、端到端系统改进后的模型在 finetune 前，bert-large 的相关性从 0.866 到 0.877，albert-large 从 0.887 到 0.891，roberta-large 从 0.868 到 0.876。结果表明端到端系统相关性更高，且经过数据增强后相关性有提升。

## 项目经历

### “贴吧文本关系抽取”项目

**项目描述：**通过深度学习算法对百度贴吧文本中的主体、客体以及他们之间的关系进行学习，从而实现对未见过的句子进行三元组关系抽取。详细步骤已上传博客：

[https://blog.csdn.net/weixin\\_49327481/article/details/128092238?spm=1001.2014.3001.5502](https://blog.csdn.net/weixin_49327481/article/details/128092238?spm=1001.2014.3001.5502)

**主要技术：**pytorch + Hugging face + Self-Attention + Bert + 服务器训练

**责任描述：**1、自定义 Dataset 函数实现对数据的**批量提取**，方便后续训练。2、使用 transformers 库中 BertTokenizerFast 模块对文本序列进行中文分词，返回分词后文字 id 值和偏移量，从而解决在输入文本中的中英文混用情况时**分词不可逆问题**。3、对预测值的损失设置不同的平衡权重系数，避免**数据不平衡**带来的训练问题。4、在预测主体时利用自注意力机制来增强上下文语义信息，提高主体识别准确度。5、在 kaggle 云服务器上使用 GPU 加速网络模型的训练。

**工作结果：**1、对预测出来的损失值使用不同的权重系数后，f1 分值从 0.27 提升至 0.62。2、通过分词时的偏移量信息可以正常返回原文中的位置。

### “运动想象脑电信号分类”项目

**项目描述：**通过使用深度学习算法，搭建端到端系统。实现输入为脑电信号，输出运动想象的类别。

**主要技术：**pytorch + MALTAB + mne + Self-Attention + 服务器训练

**责任描述：**1、使用 EEGLAB 工具箱对脑电数据进行读取和预处理。2、设计深度学习网络模型，采用时空频三条路来进行模型的训练（利用短时傅里叶变换得到脑电信号的时频图、利用 EEGLAB 工具箱生成脑电地形图、再叠加原始信号）。3、采用将同种类别的脑电信号进行随机权重的加权平均来扩充训练样本数量。4、调研深度学习分类算法，在公开脑电数据集 BCIIV2a 上进行基底模型训练。4、在自己采集的数据集上进行网络的 finetune。

## 校园经历

2017.09 – 2021.06

学习委员

2023.08 – 至今

深信服 24 秋招校园大使