

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

**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”和“[cls]答案[sep]评语”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，并采用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0.830，改进后为0.838。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。

**荣誉证书**

**实习经历**

* 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NLP关系抽取、文本分类、实体识别、答案评分等任务，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fairseq工具包，熟悉在集群中训练网络模型。
* 熟练使用Excel、PPT、Word等办公软件。

**技能情况**

**教育背景**

**2017.09 – 2021.06 哈尔滨理工大学 自动化 一本**

**2021.09 – 至今 哈尔滨理工大学 电子信息 硕士在读**

185 8305 3106

cxzhou7@163.com

周晨星

年龄：24

哈尔滨理工大学

**项目经历**

**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”和“[cls]答案[sep]评语”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，并采用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0.830，改进后为0.838。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。

185 8305 3106

cxzhou7@163.com

晨星

年龄：24

哈尔滨理工大学