

关系抽取实验

基于卷积神经网络的关系抽取

实验环境

只要当前环境能运行即可，除非运行报错，否则不需要更换下面的环境

```
python3.8  
torch 1.6.0+cu10.2  
numpy 1.18.5
```

参考论文: <https://aclanthology.org/C14-1220.pdf>

实验中用到的word embedding在<https://pan.baidu.com/s/1Nm1dcad7h3Y6xVQ8TK9KDw>下载，提取码qn5x，并放在 ./CNN/embedding 文件夹下

实验要求

- 根据论文补充 ./CNN/model.py 的 forward 部分(101行)，跑通训练代码
- 运行 run.py 训练模型并预测结果
- 对预测结果进行评估，要求复现结果F1值大于82.7(原论文)*0.9
- 进行消融实验，尝试去除PF(Position features)重复实验，并报告结果
- 提交要求：在实验报告中报告复现的F1值，同时报告消融实验的F1值；提交代码文件./CNN/model.py

远程监督关系抽取

实验环境

参考开源库<https://github.com/thunlp/OpenNRE>进行环境配置

```
git clone https://github.com/thunlp/OpenNRE.git --depth 1  
pip install -r requirements.txt  
python setup.py install
```

参考论文: <https://aclanthology.org/P16-1200v1.pdf>

实验要求

- 安装开源库OpenNER，并运行 ./benchmark/download semeval.sh 下载相关数据集
- 使用cnn作为编码器，设置 --aggr 为 att，也就是使用句子级注意力，训练以及推理，报告AUC以及F1值
- 使用cnn作为编码器，设置 --aggr 为 avg，也就是使用句子平均向量，训练以及推理，报告AUC以及F1值
- 使用pcnn作为编码器，设置 --aggr 为 att，也就是使用句子级注意力，训练以及推理，报告AUC以及F1值

- 使用pcnn作为编码器，设置 `--aggr` 为 `avg`，也就是使用句子平均向量，训练以及推理，报告AUC以及F1值
- 提交要求：在实验报告中给出四种设置的AUC(accuracy)以及F1

预训练模型关系抽取

实验环境

```
python3 -m pip install -r requirements.txt
python3 -m spacy download en_core_web_lg
```

参考论文: <https://arxiv.org/pdf/1906.03158>

参考开源代码: <https://github.com/plkmo/BERT-Relation-Extraction?tab=readme-ov-file>

实验要求

- 运行 `main_task.py` 代码，要求复现 `accuracy>0.74`
- 模型默认使用了论文中的ENTITY MARKERS+ENTITY START，修改成ENTITY MARKERS+[CLS]并重复实验。（代码位于 `./src/model/ALBERT/modeling_albert.py` 的590~599行），并报告 `accuracy`
- 提交要求：将两种设置的accuracy报告在实验报告中，提交代码文件 `modeling_albert.py`