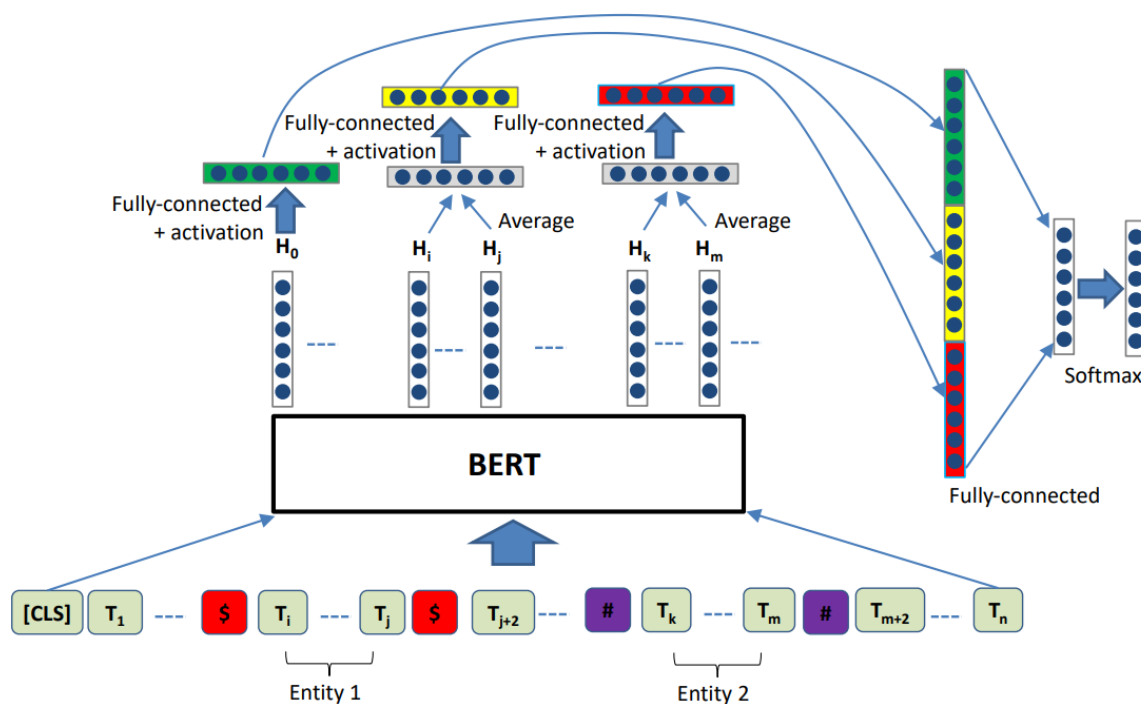


R-BERT复现

(非官方) 使用pytorch复现 R-BERT: [Enriching Pre-trained Language Model with Entity Information for Relation Classification](#)



使用须知

1. 需要python>=3.6环境依赖
2. 使用命令行 `pip install -r requirements.txt` 安装包
3. 本仓库所有文件一共500MB左右，运行一次程序后保存的模型参数大小约为1GB
4. 使用RTX3080运行一次（5个epoch）需要约5分钟

运行方式

使用命令 `python main.py` 或者使用IDE（如VScode）打开 `main.py` 运行

运行 `main.py` 后，程序会新建 `logs` 文件夹生成运行日志和结果，新建 `run` 文件夹记录运行后模型的参数。多次训练记得处理旧文件，以免占用太大空间。

仓库详情

- code：运行程序的所有代码、数据集和预训练模型
 - model：**bert-uncased-base**预训练模型。详见[Hugging Face](#)
 - SemEval2010_task8_all_data：**SemEval-2010 Task 8**原始数据集。地址[Github](#)
 - *.py：源码文件，具体思路详见[note.md](#)
- `Enriching Pre-trained Language Model with Entity Information for Relation Classification.pdf`：本项目复现的目标论文

- `requirements.txt` : python包依赖
- `note.md` : 记录了coding的心路历程, 编写逻辑和运行原理

特点

- 在 `TextDataset.__init__` 中使用正则表达式提取目标文本
- 在model中使用 `torch.mul()` 对应元素相乘 (又称 *element-wise product*、*element-wise multiplication* 或 *Hadamard product*), 借助**broadcast机制**在batch中提取**entity**向量, 避免使用循环 (现在还没找到batch的map函数)

致谢

- 感谢MEDAI3给我机会复现这一篇论文, 虽然很辛苦, 但这又一次显著提升了我的代码能力。
- 感谢chatgpt的代码思路支持, 虽然问题由代码编写转变为如何准确描述任务, 也有不少难度。
- 感谢中国人民大学明德地下机房提供的RTX3080显卡支持。
- 感谢自己的认真、坚持和毅力