###

运行环境：

pytorch 单卡3090：5个小时左右

训练运行：

pip install -r requirements.txt

python main.py --seed=2021

###

方案关键和创新：

1、模型medbert，添加注意力交互层进一步交互信息，并设计软通道注意力降维代替平均池化

2、针对文本长度过长难题，选取文本长度1024，应用拼接512的方式实现

3、针对数据量少问题，进行数据增强，通过随机拼接不同诊断疾病文本，实现多样化文本数据

4、ema训练，不增加训练和推理时间下实现精度提升

###

预训练模型clip和medbert采用公开的模型，链接如下：

https://huggingface.co/trueto/medbert-base-chinese/tree/main

###