



Weekly Work Report

Name

NER & RE

May 1, 2022

1 工作进展

1、瑞金瑞金医院基于命名实体识别的关系抽取代码，两份代码都跑通了，其中涉及的都是一些细节的问题。

2、命名实体识别，没有正确的使用所里的数据

2 问题

数据部分的格式问题，因为其中主要是集成的包，暂时没有理解数据的输入格式问题

3 RE

模型输入包括两个实体的句子，输出句子中两个实体的类别，即对句子进行多分类。其中两个实体的关系类别一共有 11 类，包括赛题中的 10 类实体关系（正样本）和 None 关系（负样本）。

队伍二的代码比较简单，需要注意的问题是：个别语句因为版本变动，一些参数名称发生了变化。

队伍一的代码整体比较复杂，训练时间也更长一些，其他的语句问题也和队伍二一样，发生了一些变化。

A 队伍 2

```
# 利用候选 relation 所在的句子训练字符级别的字向量
word2idx = {'<pad>': 0, '<unk>': 1}
word2idx, index_to_key, w2v_embeddings = train_word_embeddings(
    entity_pairs=chain(train_entity_pairs, test_entity_pairs),
    word2idx=word2idx,
    vector_size=100,
    epochs=1
)
```