Requirements(环境说明)

python > = 3.6

torch > = 1.0.0

numpy > = 1.18.1

sklearn > = 0.23.2

tqdm = = 4.54.0

运行步骤

1.提取标注语料中的关系和实体

```
python build_semeval_dataset.py
```

这里需要在**./data/SemEval2010_task8**文件夹下先放置原始语料文件**TRAIN.TXT**和**TEST.TXT**，如果相应路径下不存在这些文件，会自动从我的github主页下载并放置好相关文件(可能需要使用代理)。之所以不直接从官网下载，是因为官网语料有一些小瑕疵。

完成后会在**./data/SemEval2010_task8/train**和**./data/SemEval2010_task8/test**下生成labels.txt和sentences.txt

2.生成词表

```
python build_vocab.py --data_dir data/SemEval2010_task8
```

完成后会在**./data/SemEval2010_task8**下生成**words.txt**和**labels.txt**

3.训练并评估

```
python train.py --data_dir data/SemEval2010_task8 --model_dir
experiments/base_model --model_name CNN
```

其中参数model_name用于选择模型，共有三种模型可选，对应参数选项分别为“CNN”,"BiLSTM_Att","BiLSTM_MaxPooling",默认为CNN.若输入其他模型参数则会报错。

需要注意的是，**本实验中的模型训练使用预训练的词向量**，也会自动下载至**./data\embeddings**.

文件结构说明

./知识图谱-关系抽取：主目录

./知识图谱-关系抽取/base_model：超参数配置文件以及各模型训练后得到的参数文件

./知识图谱-关系抽取/tools：数据加载和预处理函数，以及其他utils函数

./知识图谱-关系抽取/data/SemEval2010_task8：语料数据

./知识图谱-关系抽取/data/embeddings：预训练词向量

./知识图谱-关系抽取/experiment/model：各模型实现细节

./知识图谱-关系抽取/experiment/(model_name)：以各模型名词命名的文件夹下，存储各模型的网络参数以及实验评估日志