requirements

```
numpy==1.19.3
python-crfsuite==0.9.6
six==1.15.0
sklearn-crfsuite==0.3.6
tabulate==0.8.3
torch==1.1.0
tqdm==4.31.1
boto3==1.16.40
requests==2.25.1
regex==2020.11.13
sentencepiece==0.1.91
scikit-learn==0.23.2
```

BILSTM

预处理数据

```
from BILSTM.util import combie_mini_data,combie_data_BIO
# 组合生成全集数据
combie_data_BIO(raw_data_file='./data/source.txt',raw_BIO_file='./data/BIO.txt')
# 生成mini数据
combie_mini_data(raw_data_file='./data/source.txt',raw_BIO_file='./data/BIO.txt'
,num=20000)
```

训练

```
from BILSTM.BILSTM import bilstm
# 设置超参数
model=bilstm(epoch=30,learning_rate=0.01,batch_size=64,print_step=5)
# 读取数据
model.read_data(test_data='./data/test.txt',dev_data='./data/dev.txt',train_data
='./data/train.txt')
# 开始训练
model.train()
```

BERT

使用crf模型

```
from BERT.ner_crf import run_ner_crf
run_ner_crf()
```

使用softmax模型

```
from BERT.ner_softmax import run_ner_softmax
run_ner_softmax()
```

使用span模型

```
from BERT.ner_span import run_ner_span
run_ner_span()
```

CRF模型

训练

```
from CRF.model import CRF_Model
from CRF.utils import read_data

# 训练数据
sent_data, tag_data =
read_data(sent_data='./data/source.txt',tag_file='./data/BIO.txt', lines=10000)

sep_line = 8000
train_sent_data = sent_data[:sep_line]
train_tag_data = tag_data[:sep_line]
test_sent_data = sent_data[sep_line:]
test_tag_data = tag_data[sep_line:]

model = CRF_Model()
model.train(train_sent_data, train_tag_data)
```

测试

```
from CRF import ner as crf_ner
crf_ner(sentence)
```