

命名实体识别

1. 数据集来源

人民日报

18000+句子, 300+样本用于测试

2. 模型

BiLSTM+CRF

3. 运行环境：

- python 3.6
- tensorflow 1.7

4. 训练步骤：

- 生成词典, 保存在 `data/word2id.pkl` 中。

```
python dict_prepare.py
```

- 开始训练, 默认训练 30 个epoch, batch_size=64：

```
python main.py --mode=train
```

训练得到的模型保存在 `data_save/{time}/results` 中。

5. 训练结果：

数据集中标注的有3中实体：人名（PER），地名（LOC），组织名（ORG）：

accuracy: 98.01%; precision: 90.57%; recall: 83.71%; FB1: 87.01。

LOC: precision: 88.15%; recall: 81.51%; FB1: 84.70

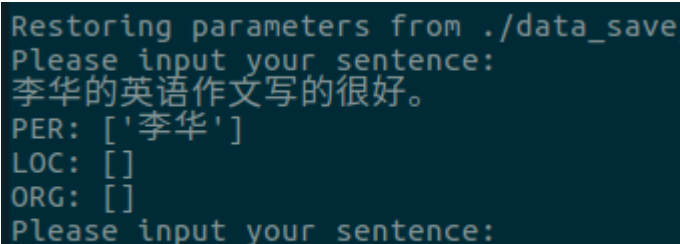
ORG: precision: 86.89%; recall: 76.81%; FB1: 81.54

PER: precision: 97.42%; recall: 94.03%; FB1: 95.70

6. 使用训练好的模型测试

```
python main.py --mode=demo
```

- 输入中文句子, 输出识别结果



```
Restoring parameters from ./data_save
Please input your sentence:
李华的英语作文写的很好。
PER: ['李华']
LOC: []
ORG: []
Please input your sentence:
█
```