# 命名实体识别

### 实验目的

* 了解神经网络中的基础模型

神经网络是指一系列受生物学和神经科学启发的数学模型。主要是通过对人脑的神经元网络进行抽象，构建人工神经元，并按照一定拓扑结构来建立神经元之间的连接。前馈神经网络（FNN）是最早发明的简单人工神经网络；卷积神经网络（CNN）是一种具有局部连接、权重共享等特性的深层前馈神经网络；循环神经网络（RNN）是一类具有短期记忆能力的神经网络。

* 了解深度学习框架Pytorch的使用

PyTorch使用python作为开发语言，近年来和TensorFlow, keras, caffe等热门框架一起，成为深度学习开发的主流平台之一。PyTorch的基本元素包含张量(Tensor)、变量(Variable)、神经网络模块(nn.Module)等。

* 了解使用深度学习解决序列标注任务基本流程

序列标注是NLP领域的基础问题之一，涵盖范围非常广泛，如分词、词性标注、命名实体识别、关系抽取等等，本质上是对线性序列中每个元素根据上下文内容进行分类的问题。命名实体识别是指识别文本中具有特定意义的实体，主要包括人名、地名、机构名、专有名词等，以及时间、数量、货币、比例数值等文字。

以PyTorch为例，一个常规的序列标注任务代码开发流程是：安装并导入相关的深度学习库、定义标签集合(Label set)、数据获取和预处理、定义神经网络、定义损失函数(loss function)和优化器(optimizer)、训练网络和测试网络。

### 实验环境

python3 + jieba + pytorch1.0.0 + numpy

pip install -r requirement.txt

* python3

除了高性能外，拥有NumPy、SciPy等优秀的数值计算、统计分析库。TensorFlow、Caffe等著名的深度学习框架都提供了Python接口。

* jieba

jieba是一款优秀的Python第三方中文分词库，支持三种分词模式：精确模式、全模式和搜索引擎模式。

* PyTorch

PyTorch是一个针对深度学习，并且使用GPU和CPU来优化的tensor library，它是一个以Python优先的深度学习框架，不仅能够实现强大的GPU加速，同时还支持动态神经网络。

* NumPy

NumPy是Python语言的一个扩展程序库，支持大量的维度数组与矩阵运算，此外也针对数组运算提供大量的数学函数库。

### 实验数据

数据集用的是论文ACL 2018Chinese NER using Lattice LSTM (<https://github.com/jiesutd/LatticeLSTM>) 中收集的简历数据，数据的格式如下，它的每一行由一个字及其对应的标注组成，标注集采用BME，B表示实体开头，M表示实体中间，E表示实体结尾，句子之间用一个空行隔开。该数据集就位于目录下的ResumeNER文件夹里。如下为数据及标签示例：

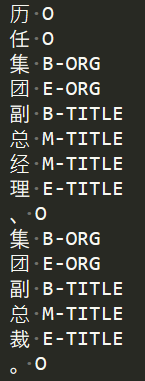


图 1 数据集示例

### 快速开始

* 配置模型相关超参数及预训练词向量路径：

/models/config.py



图 2 配置文件

* 模型训练及测试：



其中：

[模型]：cnn 或lstm

--crf ：增加crf层

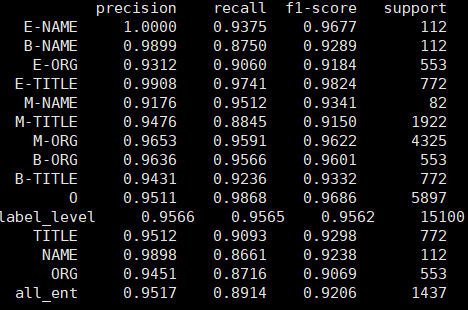
--use\_w2v ：使用预训练词向量

### 实验步骤

* 分别使用BiLSTM和BiLSTM + CRF进行NER任务，对比实验结果。

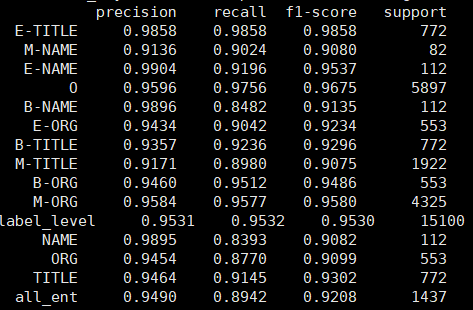
1. 阅读LSTM模型代码，运行并得到结果





1. 向模型添加CRF层，运行并得到结果

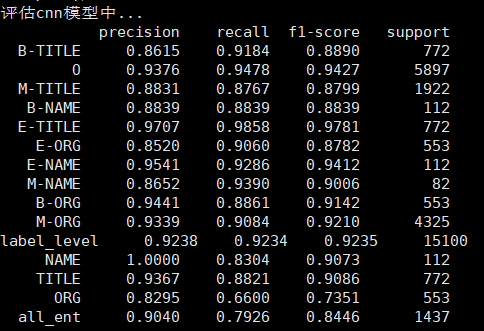


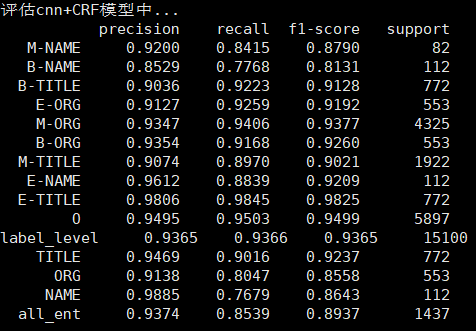


* 分别使用BiLSTM+CRF和CNN + CRF进行命名实体识别（Named Entity Recognition, NER）任务，对比实验结果。

1. 将原先的BiLSTM模型替换为CNN模型







2. 调整模型的卷积核尺寸和filter/channel参数，运行并得到结果

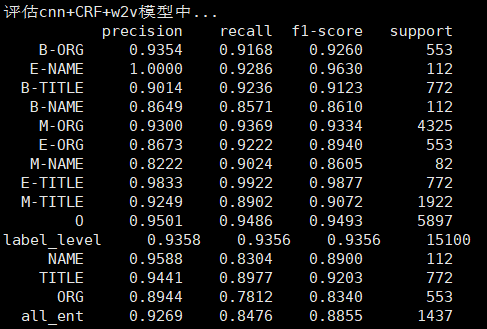
* 分别在使用预训练词向量和不使用预训练词向量的情况下进行NER任务，对比实验结果。

1. 下载预训练词向量文件，或者自己使用语料训练词向量

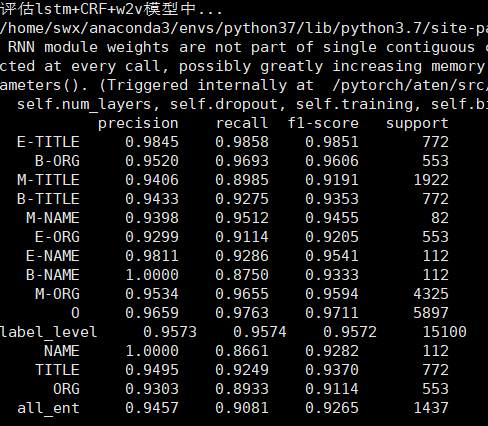
2. 编写对应的词向量初始化函数，载入已有词向量文件/随机初始化

3. 修改词向量使用参数，分别运行并得到结果









### 提交时间

**11月24号截止**

由吴雨欣同学负责收集，姓名-学号-第3次实验

### 实验要求

* 完成所有实验内容
* 良好的代码风格
* 完整的实验报告

### 参考资料

1. [Bidirectional LSTM-CRF models for sequence tagging](https://arxiv.org/pdf/1508.01991.pdf)
2. [**A survey on deep learning for named entity recognition**](https://ieeexplore.ieee.org/abstract/document/9039685/)

**附件1：**

**武汉大学国家网络安全学院**

**实验报告**

**课程名称 自然语言处理**

**专业年级 2019**

**姓 名**

**学 号**

**协 作 者 无**

**实验学期 2021-2022　 学年 第一 学期**

**课堂时数 课外时数**

**填写时间 2021 年 11 月 15 日**

|  |
| --- |
| **实验介绍** |
| **【实验名称】：自然语言处理第三次实验** |
| **【实验目的】：**   1. 了解神经网络中的基础模型 2. 了解深度学习框架Pytorch的使用 3. 了解使用深度学习解决序列标注任务基本流程   **【实验环境】：**   1. Windows 10 家庭中文版 2. Anaconda Navigator 1.9.12 3. conda 4.8.3 4. Python 3.8.11 5. torch 1.9.1 6. sklearn 1.0 7. numpy 1.21.2 8. matplotlib 3.4.3   **【参考文献】：**  **[1]** 多分类模型Accuracy, Precision, Recall和F1-score的超级无敌深入探讨  https://zhuanlan.zhihu.com/p/147663370 |
| **实验内容** |
| **【实验方案设计】：**  实验衡量指标：  本次实验主要使用了以下三种用于分类问题的衡量指标：precision、recall、f1-score。  在以下内容中，TP表示预测为正的正样本，TN表示预测为负的负样本、FP表示预测为正的负样本、FN表示预测为负的正样本。   1. precision     Precision，即精确率，表示预测结果为正的样本中正样本的比例。当产生FP的代价很高，即需要尽可能避免将负样本预测为正样本的情况时，应当注重提高precision，而不是下文提到的recall。   1. recall     Recall，即召回率，表示所有正样本中被成功预测为正样本的比例。当产生FN的成本很高，即需要尽可能避免将正样本预测为负样本的情况时，应当注重提高recall，而不是上文提到的precision。   1. f1-score     F1-score是Precision和Recall的综合考量，同时兼顾了精确率和召回率两个衡量指标，与上述两个衡量指标相同，f1-score应当尽量处于较高值。  实验步骤：   1. 分别使用BiLSTM和BiLSTM + CRF进行NER任务，对比实验结果。 2. 阅读LSTM模型代码，运行并得到结果   使用命令python main.py --model lstm 或python main.py --model bilstm  运行结果如下所示     1. 向模型添加CRF层，运行并得到结果   使用命令python main.py --model bilstm --crf  运行结果如下所示    对比添加CRF层前后的各标签的f1-score值，大部分标签的f1-score值在添加CRF层后都得到了提升，因此可以初步认为BiLSTM + CRF比BiLSTM性能更好。   1. 分别使用BiLSTM+CRF和CNN + CRF进行命名实体识别NER任务，对比实验结果。 2. 将原先的BiLSTM模型替换为CNN模型   使用命令python main.py --model cnn  运行结果如下所示    使用命令python main.py --model cnn --crf  运行结果如下所示    对比添加crf前后的f1-score数值，可以发现在大部分的分类结果中增加crf层后f1-score数值更高，模型效果更好。   1. 调整模型的卷积核尺寸和filter/channel参数，运行并得到结果   卷积核尺寸：  取kernel\_size = 3、padding = 1时结果如下所示    取kernel\_size = 5、padding = 2时结果如下所示    取kernel\_size = 7、padding = 3时结果如下所示    取kernel\_size = 9、padding = 4时结果如下所示    kernel\_size取不同值时的epoch-loss折线图可见实验结果分析。  filter / channel参数：  取hidden\_size = 50时结果如下所示    取hidden\_size = 100时结果如下所示    取hidden\_size = 200时结果如下所示    取hidden\_size = 300时结果如下所示    取hidden\_size = 400时结果如下所示    hidden\_size取不同值时的epoch-loss折线图可见实验结果分析。   1. 分别在使用预训练词向量和不使用预训练词向量的情况下进行NER任务，对比实验结果。 2. 下载预训练词向量文件，或者自己使用语料训练词向量   已下载预训练词向量文件。   1. 编写对应的词向量初始化函数，载入已有词向量文件/随机初始化   data.py中存放词向量初始化函数，并且使用nn.Embedding.from\_pretrained方法载入预训练词向量。  3. 修改词向量使用参数，分别运行并得到结果  cnn + CRF + w2v    cnn + CRF    lstm + CRF + w2v    lstm + CRF    以上四种情况的epoch-loss折线图可见实验结果分析。  **【实验结果分析】：**  kernel\_size分别取值3、5、7、9、11时loss折线图如图所示。可以发现，当kernel\_size取值3、5、7、9时，随着kernel\_size数值的增大，loss折线的收敛速度更快，并且数值更小。当kernel\_size取到11时，loss折线并没有变得更加优秀，相对而言表现不如kernel\_size取值9时的情况。因此，从loss折线图的角度看，kernei\_size取值9是一个不错的选择。    以衡量指标f1-score为评价标准时，可以发现模型kernel\_size取值为5时比kernel\_size取值为3时f1-score数值普遍更高，kernel\_size取值为7、9时比kernel\_size取值为5时f1-score数值普遍更高。kernei\_size取值7、9时f1-score表现差别不大，相对而言kernel\_size取值为7时性能会更好一些。  hidden\_size分别取值50、100、200、300、400时loss折线图如图所示。可以发现当hidden\_size取值为100时，loss折线图的收敛速度以及数值都比取更大的hidden\_size时效果更好。    以f1-score为评价标准时，得到模型hidden\_size取值50、400时效果更好，其次是hidden\_size取值为100的模型，hidden\_size取值为200、300时差别不大。此时可能存在由于训练次数不足，使得模型尚未达到最优状态；或者可能存在过拟合情况。综合考虑模型的hidden\_size可以取值为100。  使用不同参数时loss折线图如图所示。从loss折线图的角度进行分析，可以发现四种情况差别不算太大。其中lstm模型比cnn模型在接近收敛时的loss值会更低一些，使用预训练词向量与否在loss折线图中效果表现得不太明显。因此使用loss折线图分析该模型性能并不是一种有效的方法。    以f1-score作为评价标准分析时，可以发现模型lstm与模型cnn相比，其大部分标签的f1-score值更高，因此从该数据可以在一定程度上说明模型lstm在该测试集上的表现更好。而根据f1-score数值可以说明使用预训练词向量并没能使模型性能提升较大，在训练集中部分标签上不使用预训练词向量的模型表现更好。 |
| **【实验总结】：** |
| **评语及评分（指导教师）** |
| **【评语】：**      **评分：**  **日期：** |

**附件2：**

**实验报告说明**

**1．实验名称：**要用最简练的语言反映实验的内容。

**2．实验目的**：目的要明确，要抓住重点。

**3．实验环境**：实验用的软硬件环境（配置）。

**4．实验方案设计（思路、**步骤和方法等**）**：这是实验报告极其重要的内容。包括概要设计、详细设计和核心算法说明及分析，系统开发工具等。应同时提交程序或设计电子版。

对于**设计型和综合型实验**，在上述内容基础上还应该画出流程图、设计思路和设计方法，再配以相应的文字说明。

对于**创新型实验**，还应注明其创新点、特色。

**5．实验结果分析：**即根据实验过程中所见到的现象和测得的数据，进行对比分析并做出结论（可以将部分测试结果进行截屏）。

**6．实验总结：**对本次实验的心得体会，所遇到的问题及解决方法，其他思考和建议。

**7．评语及评分：**指导教师依据学生的实际报告内容，用简练语言给出本次实验报告的评价和价值。