**一、模型简介**

        在序列标注任务（中文分词CWS，词性标注POS，命名实体识别NER等）中，目前主流的深度学习框架是BiLSTM+CRF。其中BiLSTM融合两组学习方向相反（一个按句子顺序，一个按句子逆序）的LSTM层，能够在理论上实现当前词即包含历史信息、又包含未来信息，更有利于对当前词进行标注。BiLSTM在时间上的展开图如下所示。

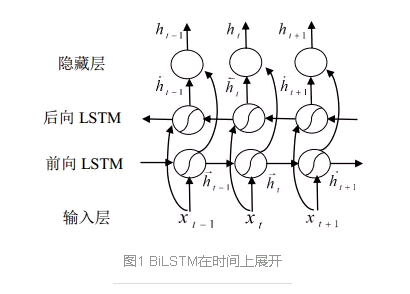


图1 BiLSTM在时间上展开

        若输入句子由120个词组成，每个词由100维的词向量表示，则模型对应的输入是（120，100），经过BiLSTM后隐层向量变为T1（120，128），其中128为模型中BiLSTM的输出维度。如果不使用CRF层，则可以在模型最后加上一个全连接层用于分类。设分词任务的目标标签为B（Begin）、M（Middle）、E（End）、S（Single），则模型最终输出维度为（120，4）的向量。对于每个词对应的4个浮点值，分别表示对应BMES的概率，最后取概率大的标签作为预测label。通过大量的已标注数据和模型不断迭代优化，这种方式能够学习出不错的分词模型。

        然鹅，虽然依赖于神经网络强大的非线性拟合能力，理论上我们已经能够学习出不错的模型。但是，上述模型只考虑了标签上的上下文信息。对于序列标注任务来说，当前位置的标签L\_t与前一个位置L\_t-1、后一个位置L\_t+1都有潜在的关系。

        例如，“我/S 喜/B 欢/E 你/S”被标注为“我/S 喜/B 欢/B 你/S”，由分词的标注规则可知，B标签后只能接M和E，因此上述模型利用这种标签之间的上下文信息。因此，自然语言处理领域的学者们提出了在模型后接一层CRF层，用于在整个序列上学习最优的标签序列。添加CRF层的模型如下图所示。

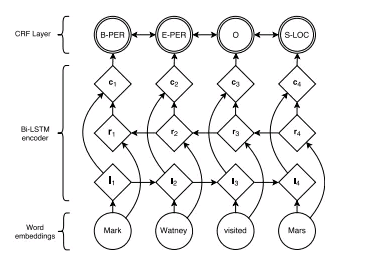


图2 BiLSTM+CRF模型图

上述图片出自[http://www.aclweb.org/anthology/N16-1030](https://link.jianshu.com?t=http%3A%2F%2Fwww.aclweb.org%2Fanthology%2FN16-1030)。

模型通过下述公式计算最优标注序列，A矩阵是标签转移概率，P矩阵是BiLSTM的预测结果。

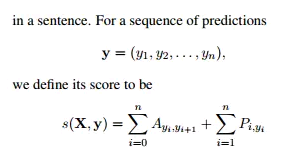


图3  CRF最优序列计算

模型训练的时候，对于每个序列 y 优化对数损失函数，调整矩阵A的值。

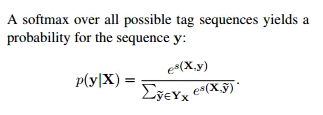


图4 条件概率计算

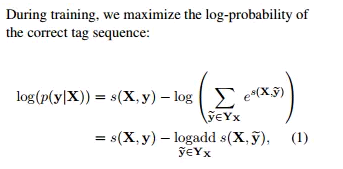


图5 CRF训练优化函数

当模型训练完成，模型预测的时候，按如下公式寻找最优路径：

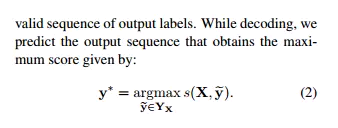


图6 BiLSTM-CRF模型预测

Y\_x表示所有可能的序列集合，y\*表示集合中使得Score函数最大的序列。

（以上为论文的核心部分，其它细节请参阅原文）

        至此，我们已经大致了解BiLSTM-CRF的原理。对于分词任务，当前词的标签基本上只与前几个和和几个词有关联。BiLSTM在学习较长句子时，可能因为模型容量问题丢弃一些重要信息，因此我在模型中加了一个CNN层，用于提取当前词的局部特征。CNN用于文本分类的模型如下。

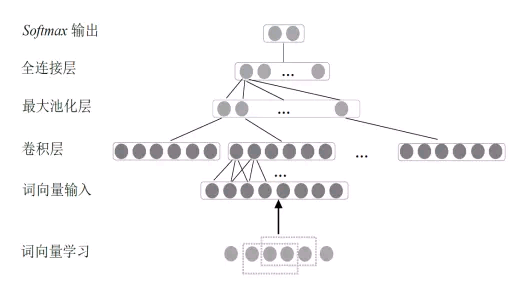


图7 CNN用于文本分类

        设句子输入维度为（120,100），经过等长卷积后得到T2（120，50），其中50为卷积核个数。对于当前词对应的50维向量中，包含了其局部上下文信息。我们将T1与T2拼接，得到T3（120,178），T3通过全连接层得到T4（120，4）,T4输入至CRF层，计算最终最优序列。最终模型BiLSTM-CNN-CRF如下。

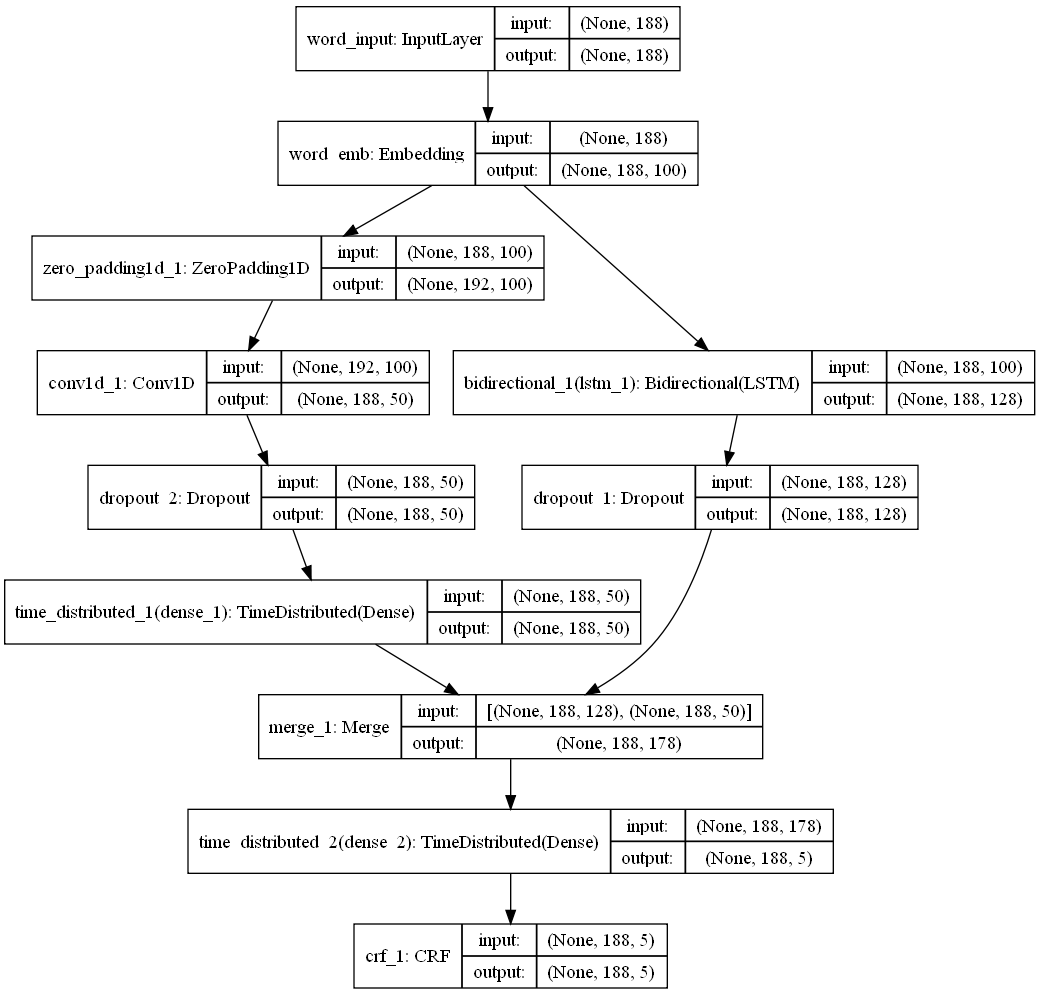


图8 BiLSTM-CNN-CRF模型图

参考资料：

<https://www.jianshu.com/p/5fea8f42caa9>

<https://www.jianshu.com/p/8f406c861cec>

<https://www.jianshu.com/p/373ce87e6f32>