

第十五周周记

周一	
完成内容	阅读博客园里的《神经网络结构在命名实体识别（NER）中的应用》
内容描述	在传统机器学习中，条件随机场（Conditional Random Field，CRF）是 NER 目前的主流模型。它的目标函数不仅考虑输入的状态特征函数，而且还包含了标签转移特征函数。在训练时可以使用 SGD 学习模型参数。在已知模型时，给输入序列求预测输出序列即求使目标函数最大化的最优序列，是一个动态规划问题，可以使用维特比算法进行解码。
未解决问题	无

周二	
完成内容	阅读 CSDN 里的一篇博客《基于深度学习的命名实体识别详解》
内容描述	命名实体识别（NER）是在自然语言处理中的一个经典问题，其应用也极为广泛。比如从一句话中识别出人名、地名，从电商的搜索中识别出产品的名字，识别药物名称等等。传统的公认比较好的处理算法是条件随机场（CRF），它是一种判别式概率模型，是随机场的一种，常用于标注或分析序列资料，如自然语言文字或是生物序列。简单是说在 NER 中应用是，给定一系列的特征去预测每个词的标签。
未解决问题	无

周三	
完成内容	阅读 CSDN 里的博客《LSTM+CRF 介绍》
内容描述	LSTM 单元最终输出的向量即可以看成是输入数据的一种表示形式，最终在打标签阶段，一般都采用 softmax 进行处理，不过这种方法在处理输出标签直接有强烈关系的数据时，效果还是有限的。特别是在实际的序列标注任务时，由于神经网络结构对数据的依赖很大，数据量的大小和质量也会严重影响模型训练的效果，故而出现了将现有的线性统计模型与神经网络结构相结合的方法，效果较好的有 LSTM 与 CRF 的结合。简单来说就是在输出端将 softmax 与 CRF 结合起来，使用 LSTM 解决提取序列特征的问题，使用 CRF 有效利用了句子级别的标记信息。
未解决问题	无

周四	
完成内容	阅读 CSDN 的博客《基于深度学习的命名实体识别 bi-lstm+crf》
内容描述	递归神经网络（RNN）已被用于在包括语言模型（Mikolov et al., 2010; Mikolov 等人, 2011）和语音识别（Graves 等人, 2005）在内的各种任务中产生有前景的结果。RNN 基于历史信息维护存储器，使得模型能够预测以长距离特征为条件的当前输出。图 1 显示了具有输入层 x ，隐藏层 h 和输出层 y 的 RNN 结构（Elman, 1990）。在命名实体标记上下文中， x 表示输入特征， y 表示标记。图 1 显示了一个命名实体识别系统，其中每个单词用其他（O）或四种实体类型（人员（PER），位置（LOC），组织（ORG）和杂项（MISC））中的一个标记。例如：欧盟拒绝德国呼吁抵制英国羊肉。被标记为：B-ORG O B-MISC O O B-MISC O，其中 B、I-标签指示实体的开始和中间位置。

未解决问题	无
-------	---

周五	
完成内容	把 LSTM-CRF 的程序导入 linux 虚拟机里面的 pycharm 编辑器中
内容描述	
未解决问题	无

周末	
完成内容	调试程序里的 bug。
内容描述	如 printf 函数由于 python 版本问题没加括号导致错误
未解决问题	无

工程汇总	
完成任务	阅读了几篇论文
任务描述	初步学习了 python 的一些基本知识
代码量	无
未解决问题	无

论文汇总	
论文列表	[1] 《神经网络结构在命名实体识别（NER）中的应用》 [2] 《基于深度学习的命名实体识别详解》 [3] 《LSTM+CRF 介绍》 [4] 《基于深度学习的命名实体识别 bi-lstm+crf》
论文摘要	
未解决问题	无

下周任务	
工作	运行程序并编写 Co-training 算法的代码
论文	继续寻找与中文分词和命名实体识别相关的论文
其他	无
汇总	下周开始准备写中期报告

日期:2018/04/09 -

2018/04/14