

第十周周记

周一	
完成内容	阅读了 2017 年计算技术与自动化期刊的论文《基于深度学习的医疗命名实体识别》
内容描述	首先的神经网络至少包含三层，第一层是输入层，第二层是隐含层，第三层是输出层。该架构主要思路是将实体识别看作一个分类问题。其输入是词向量表达与上下文词汇的词向量。这些词向量替代了传统机器学习方法人工定义的特征，将这些词向量输入到神经网络，然后通过隐含层将这些词向量转换为另外向量，再通过逻辑回归层进行分类，得到每个词的实体名概率，从而完成此实体识别工作。
未解决问题	无

周二	
完成内容	阅读 2017 年华南理工大学学报(自然科学版)期刊的论文《基于双向长短时记忆模型的中文分词方法》
内容描述	首先基于深度学习中的双向长短时记忆(BLSTM)神经网络模型对中文分词进行了研究。首先从大规模语料中学习中文文字的语义向量，再将字向量应用于 BLSTM 模型实现分词，并在简体中文数据集(PKU、MSRA、CTB)和繁体中文数据集(HKCityU)等数据集上进行了实验。实验表明，在不依赖特征工程的情况下，基于 BLSTM 的中文分词方法仍可取得很好的效果。
未解决问题	无

周三	
完成内容	阅读 2017 年计算机应用研究期刊的论文《基于 LSTM 网络的中文地址分词法的设计与实现》
内容描述	首先尝试将 LSTM 网络应用于中文地址分词。主要研究了： a) 对比了 LSTM 网络与 Bi-LSTM 网络在中文地址分词上的应用，两者都能得到较好性能，且差异不大； b) 尝试了几种 dropout 策略，验证了 dropout 策略在中文地址分词任务上能有效减少过拟合； c) 使用了大量未标记数据集参与词向量预训练过程，发现能提高模型的性能。
未解决问题	

周四	
完成内容	阅读 2017 年南京大学郁振庭的硕士论文《基于神经网络的词法分析研究》
内容描述	首先针对传统基于字标注的分词模型基于窗口抽取局部特征，无法捕获长距离依赖的问题，我们提出使用双向长短期记忆网络代替原有特征抽取模块，该网络既可 W 保存长距离信息也简化了特征抽取工作。其次，我们设计了基于双向长短期记忆网络的贪心模型和结构化模型。最后我们针对通用的词嵌入与具体任务不契合的问题，我们分别设计了分词和词性标注任务相关的词

	嵌入模型。
未解决问题	无

周五	
完成内容	学习 python 的基础知识
内容描述	继续学习了 python 中用 dump 保存，用 load 恢复等有关存储的知识
未解决问题	无

周末	
完成内容	完成毕业设计开题报告
内容描述	
未解决问题	无

工程汇总	
完成任务	阅读了几篇论文，完成开题报告
任务描述	
代码量	
未解决问题	

论文汇总	
论文列表	[1] 《基于深度学习的医疗命名实体识别》 [2] 《基于双向长短时记忆模型的中文分词方法》 [3] 《基于 LSTM 网络的中文地址分词法的设计与实现》 [4] 《基于神经网络的词法分析研究》
论文摘要	
未解决问题	无

下周任务	
工作	继续深入研究自己的课题
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	了解更多与我的课题相关的知识

日期:2018/03/05 -

2018/03/10