

第六周周记

周一	
完成内容	阅读了 2017 年信息网络安全期刊《基于深度神经网络的命名实体识别方法研究》
内容描述	1. 介绍了几种命名实体识别的分类器，如决策树等 2. 了解了注意力机制 3. 了解了基于线性链式条件随机场的命名实体标注
未解决问题	无

周二	
完成内容	阅读了 2016 年电子科技大学硕士郑宗荣的论文《基于实例的中文分词系统设计与实现》
内容描述	1. 了解了 N 最短路径方法 2. 了解了实例精确匹配算法 3. 了解了实例类推分词算法
未解决问题	无

周三	
完成内容	阅读了 2015 年东北大学硕士王威的论文《基于统计学习的中文分词方法的研究》
内容描述	1. 了解了基于 L-BFGS 的模型训练算法 2. 了解了基于 Viterbi 的模型预测算法
未解决问题	无

周四	
完成内容	阅读了 2017 年华中师范大学的硕士周寅的论文《融合深度学习特征与浅层机器学习特征的中文分词关键技术研究》
内容描述	1. 学习了基于 BP 的中文分词方法 2. 学习了基于 RNN 的中文分词方法 3. 学习了基于 LSTM 的中文分词方法
未解决问题	无

周五	
完成内容	阅读了 2015 年北京工业大学硕士王国昱的论文《基于深度学习的中文命名实体识别研究》
内容描述	1. 了解了神经网络理论基础和反向传播算法 2. 了解了神经网络里 DNN 模型的构建
未解决问题	无

周末	
完成内容	看了莫烦 python 的视频《什么是 LSTM RNN 循环神经网络》和《什么是循环神经网络 RNN》

内容描述	对 LSTM RNN 循环神经网络和 RNN 循环神经网络有了更深入的了解
未解决问题	无

工程汇总	
完成任务	如上所述
任务描述	如上所述
代码量	无
未解决问题	无

论文汇总	
论文列表	[1] 《基于深度神经网络的命名实体识别方法研究》 [2] 《基于实例的中文分词系统设计与实现》 [3] 《基于统计学习的中文分词方法的研究》 [4] 《融合深度学习特征与浅层机器学习特征的中文分词关键技术研究》 [5] 《基于深度学习的中文命名实体识别研究》
论文摘要	<p>目前深度学习在图像处理领域被广泛运用，也取得了可喜的成绩，同时在自然语言领域也有不错的效果。本文将机器学习和深度学习的方法运用到中文分词领域，对传统的基于机器学习中文分词方法进行改进，提升分词效果。本文将标记语料按字向量化后，注入 L S T M 将语言中上下文关系添加到向量中，为接下来的条件随机场分词提供了充足的上下文信息，从而提升分词的准确率。L S T M 相比于卷积神经网络优点在于能够保留上下文的依赖信息，相比于普通循环神经网络的优点在于不易产生梯度弥散和梯度爆炸保留长距离依赖信息，从而更好的支撑分词效果提升。</p> <p>本文在北京语言大学提供的语料库上对提出的模型进行实验验证，并对传统模型在同一数据集上实验，对比分词效果。实验表明，融合深度学习特征与浅层机器学习特征的中文分词相比于传统的机器学习分词、概率模型分词、字标注分词和词典分词效果有了一定程度的提升。在北京大学语料库的分词中，我们的实验结果取得了 92.80% 的 F 值，比常规的神经网络分词方法提升了 1% 的效果。</p>
未解决问题	无

下周任务	
工作	继续查找相关的论文和继续阅读相关文献。
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:2018/02/05 -

2018/02/10