工作周报

时间：2018年12月10日星期一~2018年12月17日星期一

|  |
| --- |
| 工作内容 |
| 1. 学习条件随机场的预测算法，使用的场景 2. 看了李航的统计学习方法的相关知识 |
| 问题记录 |
| 1. 预测问题转化为概率最大的路径问题，了解了理论不知道怎么去运用 2. Ontones语料处理还有点问题 |
| 待办事项 |
| 1. 结合CRF，用简单的例子表明 2. 细化理论模块，各模型的优缺点罗列 |

工作周报

时间：2018年12月03日星期一~2018年12月10日星期一

|  |
| --- |
| 工作内容 |
| 1. 学习条件随机场的数学模型 2. 学习了马尔可夫链，马尔可夫过程,HMM的相关资料 |
| 问题记录 |
| 1. 感觉前面学习的不太扎实，有些知识点MEMM的求解串不起来，关联性少 2. 数学模型理解很吃力 |
| 待办事项 |
| 1. 讲MEMM,HMM,CRF结合一下，交流 2. Onone进行了解 |

工作周报

时间：2018年11月26日星期一~2018年12月3日星期一

|  |
| --- |
| 工作内容 |
| 1. 对训练数据进行优化，建立了语义角色的model 2. 学习了HMM的模型和CRF原理，利用爬虫尝试爬数据 |
| 问题记录 |
| 1. 利用爬虫爬数据，有些数据的传递加了密是利用AJAX来进行传递，不太好获取 2. Model训练数据量比较少，跑出来结果还有些缺陷 |
| 待办事项 |
| 1. 深入学习crf 和将CRF、HMM MEMM的关系串在一起 2. Onone进行了解 3. 接触RNN和wordtovec，了解深度学习和词向量是什么 |

工作周报

时间：2018年11月19日星期一~2018年11月26日星期一

|  |
| --- |
| 工作内容 |
| 1. 对训练数据进行优化，联合CTB和CPB 2. 看了两本书nlp汉语自然语言处理和老师发的书 |
| 问题记录 |
| 1. 在提取CPB的时候，发现数据不匹配，根据CPB所给的标记提取不了里面的内容，查了资料发现要配合CTB进行对数据的预处理，然后又找到了CTB8.0对里面的相关数据进行了解和处理，配合CPB对内容进行获取 |
| 待办事项 |
| 1. 用优化的数据生成新的model 2. CRF原理研究.对训练的model进行测评 |

工作周报

时间：2018年11月12日星期一~2018年11月19日星期一

|  |
| --- |
| 工作内容 |
| 1. 训练语义测试的model 2. 学习CRFsuit |
| 问题记录 |
| 1. 前几天还在弄北大网库的语料，根据他的说明文档，对其语料有了一定的了解，知道了是以说明为区分点但是不会运用，之后老师又给了CPB的语料，又通过两天的时间对语料进行了解，CPB其中的映射内容都来自宾州树库 Treebank7.0(CTB)又花了一部分的时间对CPB进行解析出他的句子，但是最后没有成功的提取出来句法树，对模块的训练因为数据格式的问题，匹配也不成功 |
| 待办事项 |
| 1. 学习CRFsuit 2. 优化训练数据 |

工作周报

时间：2018年10月29日星期一~2018年11月5日星期一

|  |
| --- |
| 工作内容 |
| 1. 训练了crf的model，对项目进行版本控制，能够利用crf运行分词 2. 对crf的使用进行的总结 3. 调用ltp实现语义决策分析 |
| 问题记录 |
| 1. 对项目进行github版本控制，建立版本库，对项目进行上传，有些文件扫描不到，没有上传上去，还有些由于文件太大上传不了，需要下载Git Lfs插件 2. Crf跑的时候合并说是超过了List的range 3. 下载了ltp，在网上大概找了几种种方法1.安装Python的扩展包pylyp，Pip报错，Vcexit status 2。安装whl文件，又提示python版本不匹配3.直接调用编译好的ltp可执行文件，又出现乱码。 |
| 待办事项 |
| 1. 掌握ltp中srl的预料和model，了解model中以什么为区分点，尽量自己根据ltp训练出自己的model，并成功调用 2. 理解并利用解码维特比算法 |

工作周报

时间：2018年10月22日星期一~2018年10月28日星期一

|  |
| --- |
| 工作内容 |
| 1. 完成了URL提取的模块实现 2. 实现了CRF的demo运行，已成功跑成 |
| 问题记录 |
| 1. Url是利用正则表达是提取相关的网页代码，匹配的时候回出现一些不必要的代码 |
| 待办事项 |
| 1. 理解并利用crf的learn和test建立起自己模块的测试数据 2. 理解并利用解码维特比算法 |