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. 详细学习最大熵模型原理和算法理论 |
| 问题记录 |
| 1. 对拉格朗日中值定理的理解不够 |
| 待办事项 |
| 1. 对算法的渗透理解 2. 把算法函数运用到代码里 |

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HMM，维特比，最大熵的使用原理和算法思想。 2. 摘要了相关的文档和算法实现，制作了相关的文档。 |
| 问题记录 |
| 1. 理解了相关的算法实现原理，但是对他们的具体实现不懂。 2. 对知识的理解没有想要的那么深，需要自己的知识的学习进行加强。 |
| 待办事项 |
| 1. 这些算法的具体使用需要学习 2. 对这些算法的更加深刻理解。 |

时间：2018年11月26日星期一~2018年12月03日星期一

|  |
| --- |
| 工作内容 |
| 1. 学习flask的用法和相关知识 2. 进一步学习的crf，HMM和MEMM的算法知识和相关原理知识。 |
| 问题记录 |
| 1. MEMM的算法理解不够深度，搜到的先关资料不是很多。 2. 对于实训的目标进行了深刻的反思和理解，端正自己学习状态。 |
| 待办事项 |
| 1. 深度学习相关的其他算法例如维特比算法。 2. 学习python下的环境的使用。 |

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. 搜索到了一些命名实体识别的相关语料，但是训练模板就很少。 2. 粗略学习了老师说的相关的Linux的简单知识。 |
| 问题记录 |
| 1. 没有找到适合学习Linux环境中配置CRF的方法。 2. 对Linux的知识掌握太少，有效的学习路径没有找到。 |
| 待办事项 |
| 1. Linux上进行安装运行crf的条件环境。 2. 需要将python下的运行结果可视化。 |

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. 网上搜索命名实体识别的相关语料和训练模板模型。 2. 用中文分词和词性标注完成命名实体识别。 3. 学习crfsuite的相关知识。 |
| 问题记录 |
| 1. 网上的命名实体识别的语料特别少，可能是我没有搜到的原因。 2. 自己的模块在使用曾辉祥的模块可以进行简单的命名实体识别。 3. 网上的命名实体识别都是在中文分词和词性标注的基础上，然后进行识别。 4. Crfsuite的环境配置没有完成，没有找到window上的使用环境。 |
| 待办事项 |
| 1. 配置crfsuite环境 2. 找到更多的相关的语料 |

时间：2018年11月05日星期一~2018年11月12日星期一

|  |
| --- |
| 工作内容 |
| 1. 网上搜索命名实体识别的方法，学习命名实体识别的操作流程，训练学长给的相关语料。 2. 学习了解命名实体的训练格式和测试的格式。 3. 运用网上的资料进行命名实体识别任务。 4. 学习相关的template的知识 |
| 问题记录 |
| 1. Template的算法简单的只能读清楚，不会写，不能掌握该方面的知识。 2. 自己的模块没有运用到团队的模块的合作，模块完成的时候比较复杂。 3. 在尝试运用词性标注完成命名实体识别的时候识别失败。 4. 自己的命名实体识别模块的训练语料不够准确，成品效果也不是很准确 |
| 待办事项 |
| 1. 是命名实体模块的识别内容更准确 2. 使用词性标注的一些功能完成或者优化命名实体识别 3. 学习template的相关知识和内容。 |

时间：2018年10月29日星期一~2018年11月05日星期一

|  |
| --- |
| 工作内容 |
| 1. 网上搜索学习摘要关键词的方法，学习摘要关键词的操作流程，跑了一些语料。 2. 按照学长的给的例子学习摘要关键词，学习别人的方法和代码。 3. 尝试用自己小组的中文分词基础上进行关键词的摘要。 4. 学习python |
| 问题记录 |
| 1. Python语言还是的还是不够，只了解到了一些python基础，许多语法结构都不会。 2. 看不懂TextRank算法的原理和使用方法。 3. Tensorflow自动文摘也没有看懂。 |
| 待办事项 |
| 1. 摘要关键词模块需要尽快加班完成 2. crf++方面的知识还需要学习 3. 需要结合团队的模块进行调用他们的方法实现功能 |