# ZUTNLP-Text 第10周周报

2018-2019-2 学期， 2018/11/12 – 2018/11/17

|  |  |  |
| --- | --- | --- |
|  |  | |
| 指导讨论记录： | 2018-11-12 13:20- 13:50 刘小明   1. 评审上周工作完成情况； 2. 布置本周工作内容；   2018-11-12 刘凤华 | |
| 上周工作总结 | 团队工作总结： | |
| 1. 使用crf++训练了命名实体识别的model，测试数据  2. 使用crf++对依存句法分析进行训练和测试并编写python脚本能直接在pycharm中调用  3. 已经陆续的在java项目中搭建crf实体  4. 学习HMM模型+维特比算法  5. 学习Python相关的知识点 | |
| 团队成员工作总结： | |
| 曾辉祥 | ①对template有一个更深的认识  ②使用crf在java项目中实现词性标注  ③将crf训练了命名实体识别的model |
| 卢思童 | ①使用HMM模型+维特比算法完成中文分词。 ②使用Pycharm展示CRF依存句法分析的效果。 ③深度学习特征模板template. ④学习双端LSTM实现序列标注（分词）对LSTM的结构和训练做一个简单的介绍，写文档 |
| 赵向阳 | ①网上搜索命名实体语料，用crf++ 进行训练，完成了命名实体识别。  ②简单学习了关于template的算法  ③又学习了一些python语言知识 |
| 陈金新 | ①掌握ltp利用CRF训练分词cws模型,词性标注pos模型  ②找到了语义决策免费的语料  ③学习Python对语料进行的处理 |
| 房晔 | ①粗略了解crf++各文件构造含义，可以测试训练出一个model；  ②学习Python基础知识。 |
| 张欣欣 | ①CRF理论学习，CRF++对命名实体识别实现的Python学习 |
| 支友盟 |  |
| 本周存在问题 | 团队共性问题 | |
| 1. 对文本的词性标注、命名实体识别分析还不是很智能  2. 算法太复杂，看不懂  3. 语义决策这一块对于crf的相关资料太少，没有找到相关的语料 | |
| 团队成员问题 | |
| 曾辉祥 | ①数据还是不能全部的区分开来  ②词云展示方法没有学  ③在java中适应crf可以测试出结果，但是对于数据存储出现了问题 |
| 卢思童 | ①对于维特比算法在依存句法中的使用还有待了解。 |
| 赵向阳 | ①命名实体的语料存在一些问题，一些实体的识别出现错误。  ②自己的模块没有和团队的模块进行合作运用。。 |
| 陈金新 | ①刚开始借用ltp的srl，以为可以利用直接训练出来的cws分词，pos词性标注再加上依存的model可以在srl的基础上跑出来，但是由于自己的语料和训练模板的不一样，所以在训练完依存的model，接入到srl的工程中不能成功的调用。  ②网上关于语义决策的资料比较少，大部分的训练集都是不免费的，找了几天找到一个训练集，也跟平常的不太一样。 |
| 房晔 | ①Crf++中的template模板不知道具体适应哪些模块 |
| 张欣欣 | ①wording.txt为空，只能通过命令行显示 |
| 支友盟 |  |
| 本周工作计划 | 团队总体计划 | |
| 1. 每个人跑出自己的模块  2. 开始使用crfsuit工具  3. 收集各模块的资源  4. 学习HMM隐马尔科夫模型  5. 学习维特比  6. 将各个模块部署rest服务接口 | |
| 成员个人计划 | |
| 曾辉祥 | ①使用crfsuit工具  ②部署crf相关模块的rest服务接口  ③解决数据存储问题 |
| 卢思童 | ①理解CRF的template生成的CRF特征函数采用双数组Trie树存储（树的构造与双数组的构造）。  ②学习维特比算法解码。  ③用Java实现CRF序列标注的中文依存句法分析。 |
| 赵向阳 | ①对自己的模块进行细节优化  ②学习crfsuite的使用  ③试着用crfsuite进行命名实体识别 |
| 支友盟 |
| 陈金新 | ① 通过拥有的训练集和生成的template训练出自己的语义决策model，并完成数据的测试，  ② 学习CRFsuit并且利用CRFsuit继续训练 |
| 房晔 | ①更深入地学习crf++的具体使用。  ②在pycharm中能使用三个模块的功能实现。  ③了解学习前端。 |
| 张欣欣 | ① 根据java项目加深学习  ②向卢思童学习依存句法分析 |