

前周工作报告

九月

第一周

2017-08-28

2017-09-03

1. 解决线上运行中产生的问题

出错时无法准确定位错误位置，为了报错时可以立刻定位到错误所在，从而便于快速debug，做出以下几处修改作为解决方案：

- 增加异常抛出
Add `raise` to determine error type.
- 增加可用于debug的中间变量输出
Modify `mid_output` for debugging.
- 增加判断异常文本直接跳过的逻辑
Add judgement for whether skip current document or not.
- 在python逻辑层面增强系统稳定性
Advanced stability

2. 更新规则库

在阅读部分输出结果的过程中，发现几处先前未能发现的连接词与逻辑漏洞进行修改

- Modify several rules.
- Update `link_word` list.
- Add reference features for each `sentence_dict`.

3. 为因果标注编写后端逻辑

对于形同“由于”、“...所致”等问题难以通过规则制作正确的训练数据，故需要标注系统进行辅助

- 预留标注数据字段
Add placeholder for later marking functions.

- 实现标注类（类名：LabelMark）基础框架
[Basic version Alpha for LabelMark module.](#)
- 为标注类制作input接口，可以把o301\o302\o401文件提取所需信息并转化成便于分析的形式
[Basic pre_processing for sentence_dict type.](#)
- 为数据处理实现了一些低时间复杂度的工具函数
[Several new functions as data_tools.](#)
- 与前端的前辈商讨后，最终确定了基本方案，并给出API接口
[Finished API for ppt of 因果标注相关.pptx](#)

4. 为前端标注商讨确定数据结构

- 初期决定数据结构，并完成第一版的 [因果关系标注.pptx](#)
[Determines the basic data structure in trans.](#)
- 后续经罗老师提醒，需要顾及到多种情况的因果标注，进行了改进
[advanced data structure for updated 因果标注相关.pptx](#)
[Modified data structure and describe for label marking.](#)

5. 一些隐含问题的处理

线上运行时，发现当两个及以上相同segment相继出现的时候会造成position字段错误，进行修正，如“A/B/B/C/D”中的两个B字段均判定为 `position=1` 类似型的错误

- [Fix segment position calculation in causality.](#)
- [Fix issues for possible recurring location](#)

Other Works

- ICT_Auto_Login

<http://blog.csdn.net/okcd00/article/details/77677326>

与李宏伟学长一起为实验室的自动登陆进行了基于PhantomJS模拟点击（我）与python爬虫（宏伟学长）两种方法的尝试，成功为组里解决了新的登陆机制下的自动登陆问题

- PPT Modify

Weekly News

- 国科大中关村校区自 2017-08-28 至 2017-09-28 停止热水供应
 - 影响：该周四处奔波于各处借用浴室，其中两天尝试使用学校浴室排队时间也相对较长
 - 后续：曹荣禹回来之后，下班回去可以和曹荣禹一同前往科一招借用浴室，可以省下不少时间。