

# 前周工作报告

九月

第三周

2017-09-11

2017-09-17

## 1. 句子关联词提取类

为了预处理阶段从句子中选出备选的关联词实现的规则正则类，经多次修改，代码可详见 [LinkCatch.py](#)

- 新建用于提取句子中关联词的新实现类
- 完成初始框架并可使用原词表提取备选关联词

## 2. Remarkable项目，负责后端实现

Remarkable项目作为后端编写逻辑

- 为前端编写数据传输与格式处理的接口
- 编写从o302格式文件转换为json格式的数据接口
- 扩展函数传参以支持o30\*格式的文件作为输入
- 完成从o301到前端的API且投入使用
- 根据罗杰学长要求，新增{sid:json\_str}类型数据接口
- 整理优化现有接口，合并清理冗余部分
- 根据需求，从后端实现了切分segment的API供调用，以减轻前端负担
- 为接口转移至cheftin/Remarkable做准备
- 迁移期间出现兼容问题，将python2代码添加对python3的适配

## 3. 为前端标注商讨确定数据结构

- 根据@shaozw的需求flatten传入数据结构中link\_position的数据维度
- 根据前端需求，数据中各部分新增用于标记位置的pos/idx字段
- 处理数据中所有多余的维数 ( Fix issue )
- 根据@luojie的需求对输入给/v1/question键值的数据类型进行转换

# Other Works

## Pytorch Research

还在调研中，尚未进行整理归纳

### 关于动态图

支持动态图的创建。现在的深度学习平台在定义模型的时候主要用两种方式：**static computation graph**(静态图模型) 和 **dynamic computation graph**(动态图模型)。绝大部分平台都采用的是**static**的定义方式，包括TensorFlow, Theano, Caffe, Keras等。静态图定义的缺陷是在处理数据前必须定义好完整的一套模型，能够处理所有的边际情况。比如在声明模型前必须知道整个数据中句子的最大长度。相反动态图模型(现有的平台比如PyTorch, Chainer, Dynet)能够非常自由的定义模型。举个例子，传统的LSTM往往处理一个句子的时候都是以**word**为单位，然后利用**word2vec**来初始化词向量。但是往往有一些很奇怪的词在**vocabulary**里是找不到的，也就是没法用**word2vec**初始化词向量。这时候你可能想用**characer-level**(字符)级别的表示来初始化那个单词，就需要借助动态图模型的定义了。简单来说动态图模型允许你在运行程序的时候动态去修正你的模型结构，来处理各种奇奇怪怪的边角输入，这在学术研究的时候的灵活性就体现出来了。

贴引用连接：

链接：<https://www.zhihu.com/question/54914188/answer/141890899>

来源：知乎

### 优势

- 可以**动态**生成computational graph，对NLP一些model来说这个特性很有用：动态结构对NLP，RNN友好
- module和weight分离，clone module、共享内存的时候就方便快速很多
- 与**numpy无缝连接**，极为简便的从numpy向torch之间切换
- 在pytorch你可以**随时打印**、**逐层打印**中间结果：过程式编程
- 图内的用variable，图外的用tensor，图内的结果要传到图外计算，就直接v.data，图外的结果要扔到图内，就Variable(t).
- 底层是C语言
- pytorch相比与torch增加了自动求导（autodiff），罗老师说过作为一个平台自动求导还是很重要的

### 劣势

- 暂时系统讲解pytorch编程的教程不多（但例程的质量尚可）
- Pytorch现在还是beta版，有一些bug以及底层优化、速度上还有待进步。

代码实例：labels\*logits

```
# torch
import torch
a = torch.Tensor([[1],[2],[3]])
b = a
print a*b

# Tensorflow
import tensorflow as tf
a = tf.reshape(tf.constant([1,2,3]),shape=[1,3])
b = a
c = b*a
sess = tf.InteractiveSession()
print a.eval()
sess.run(c)
```

## 论文阅读

《Why-Question Answering using Intra- and Inter-Sentential Causal Relations》

plenty of **templates** for verbs.

**CRF++** and **Mini-SVM**

20 answers candidates for a question.

cue phrase used for extracting a causal relation candidate a **c-marker**

## Teaching assistant

本周起于玉泉路担任助教职责

## 作业布置

第四次作业内容：

题目详见PPT “search 4.pptx” 第17页；  
请按照要求完成作业，并使用LaTeX完成电子版作业进行提交。

提示：本次作业请提交

- 1) pdf文档
- 2) 生成该pdf的LaTeX文件夹，打包成压缩包
- 3) 为了便于核查抄袭情况，请将tex文件复制一份放在压缩包外

即：本次作业提交内容为：

- 1) 201518012333333\_姓名\_04.pdf
- 2) 201518012333333\_姓名\_04.rar / 或者 .zip
- 3) 201518012333333\_姓名\_04.tex

P.S.

为方便批量处理，下划线等命名细节请务必一致  
允许使用自己的LaTeX模版，但提供的TeX与LaTeX文件夹需要可以编译出该

PDF

请使用简单一点的压缩方式，压缩包打开之后还要求下载软件的作业以后就不

改了

助教联系方式在左侧栏的“互动交流-邮件工具”中

请务必按时、按要求提交以上3份文件：

经同学反映，SEP系统的时间会提前三分钟，所以请务必尽早提交；  
按老师要求，不接受补交迟交，请大家务必准时提交作业。

## 成绩批改

详见课程网站或附件 [assignment2.xls](#) 与 [assignment3.xls](#) 文件

本次作业批改中出现以下几种问题：

1. 使用LaTeX插入手写作业照片作为答案；
2. 使用自己的LaTeX模版完成作业（不知是否允许这种情况）；
3. 作业缺题漏题；
4. 多数明显为引用的部分没有添加参考文献。

## PPT Modify

PPT for AI Course

[点此查看](#) 当前实时修改进度

## Weekly Timetable

- 2017-09-23 周六：计算所 08:30 技术交流分享会（本次为干斌学长）

## Notes

- Pytorch learning.
- Momoko for tornado learning.
- PPT for AI course