

Assignment 3 Chinese Event Extraction



- In this assignment, you will need to use sequence labeling models for Chinese event extraction.
- Event information are defined as two parts:
 - *Trigger*: the main word that most clearly expresses the occurrence of an event.
 - Argument: an entity, temporal expression or value that plays a certain role in the event.
- For example:

"因特尔在中国成立了研究中心"

- "成立" is the trigger of type Business
- "英特尔", "中国" and "研究中心" are the arguments of type Agent, Place and Org





- This task is separated as two subtasks:
 - Trigger labeling: identify the trigger word in the sentence, and classify it to the following 8 types:

```
Life|Transaction|Movement|Business|
Conflict|Contact|Personnel|Justice
```

 Argument labeling: identify all the arguments in the sentence, and classify them to 35 types (some are listed below, all types could be found in the training file):

```
Person|Place|Buyer|Seller|
Beneficiary|Price|Artifact|Origin|
Destination|Giver|Recipient|Money|
```

- You are required to use both HMM and CRF models for this task. You can use any toolkit for their implementation.
- Note that the performance of HMM can be very poor.



Input

A sequence of segmented Chinese words.

Output

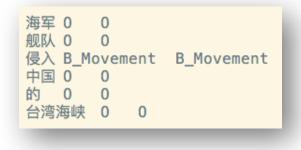
Label each word with 'T_type' (trigger), 'A_type' (argument) or 'O' (neither trigger nor argument). Save your labeling result after the real label separated with tab.

跨党大台权促6号成6号立

fg1:input



fg2: training instance



fg3: testing result



- trigger_train.txt & trigger_test.txt :
 - These two files contain 1,918 and 669 instances for training and testing, respectively.
 - Each line contains one word and its label separated by tabs.
 - Instances are separated by blank line.
- argument_train.txt & argument_test.txt :
 - These two files contain 2,131 and 997 instances for training and testing, respectively.
- Your job is to predict the sequence label for instances in test files, and write your predictions in result files. The labels in test files are only for evaluation.
- eval.py
 - This file can help you evaluate your model's recall, accuracy, precision and F1-score.



- Generate a zip file and name it as "sid_homework-3.zip".
- It should include a python file named "extraction.py", two output files named "trigger_result.txt" and "argument_result.txt", and a written report named "chinese event extraction.pdf".
- Program: codes should be written in python.
- Report: the report needs to be written in English with no more than 4 pages.



- We will mark your homework based on the four criteria:
 - Final accuracy (20%)
 - Program (30%)
 - Report (40%)
 - HMM implementation (10%)



- Submit your homework via E-learning system.
- Deadline: Mid-night at December 8th 2017
- If you have any questions about this homework, send email to TA or our course mailbox.

- TA in Charge
 - 杨依莹(zoeyangyy@163.com)