Empirical Methods in Natural Language Processing

Peking University, 2022

Homework 3

Due on Sunday, July 3 at 11:59 p.m.

(For Seniors, Due on Sunday, June 19 at 11:59 p.m.)

Instructions

Please read these instructions to ensure you receive full credits on your homework.

- Submit your homework as a **zip** file through **Course**, which should include one report in PDF, your source code and one shell script.
- Any coding language is acceptable, but your code should be **your own**. Do NOT submit Jupyter or other notebooks, but the original source code only.
- You should write your report in **English** and submit it in PDF. We recommend using the official ACL style template for your report (https://github.com/acl-org/acl-style-files).
- Your code should be paired with a shell script for starting your program, and a README file describing dependencies, code structures, etc.
- There is no need to submit the data you have used. Your grade will be based on the contents of one PDF file, the original source code and the script. Additional files will be ignored.

Late submission policy

- Late homework will have 5% deducted from the final grade for each day late. The number of points deducted will be rounded to the nearest integer.
- **NO** submission will be accepted a week after the due date. It is non-negotiable.
- Your submission time will be based on the time of your last submission to Course. Therefore, do NOT resubmit after midnight on the due date unless you are confident that the new submission is significantly better to overcompensate for the points lost.
- You can resubmit as much as you like, but each time you resubmit please be sure to upload all files you want graded!

Problem Description

In Homework 2, we focus on locating the event triggers in the sentences, which is the first step of the event extraction task. Based on your last homework, you will construct a complete event extraction system in Homework 3.

This homework involves two parts.

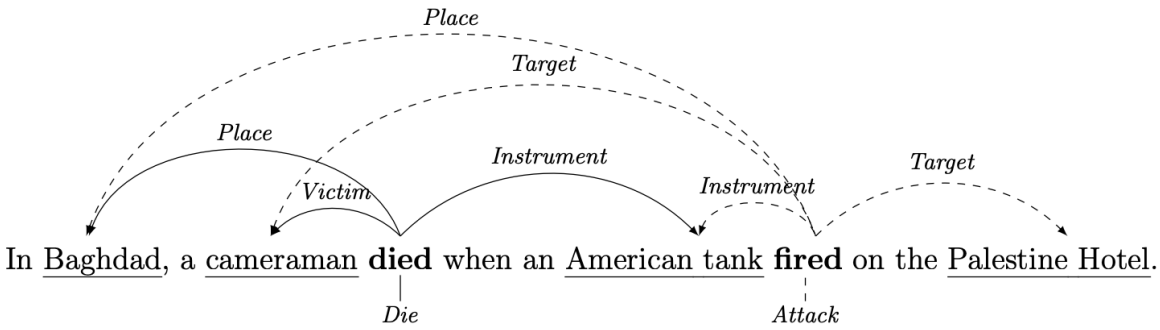
For non-seniors: You are required to complete both Part 1 and Part 2.

For seniors: Since your deadline is earlier, you are only required to complete Part 1. That will, at most, give you 80% credits of this homework. If you stick for both, then you could get 100% credits at most.

Part 1: Event Trigger Classification

In this part, you are required to extend your model in Homework 2 to support event trigger classification. For example, your model for Homework 2 is able to detect two triggers, *died* and *fired*, 2in the following sentence. For this homework,

your model should further classify them into correct event types, i.e., *Die* and *Attack*, respectively.



Part 2: Event Argument Extraction

After identifying the event triggers, an event extraction system will extract the event arguments for each trigger. An event argument is an entity that serves as a participant or attribute with a specific role in an event. In the above example, the event *Die* triggered by the word *died* has three arguments, i.e., *Baghdad*, *cameraman*, and *American tank*. These argument entities play the role of *Place*, *Victim* and *Instrument*, respectively.

In this part, you are going to implement a model for event argument extraction. Your tasks are:

1. For the triggers detected by your model in Homework 2, extract their arguments from the sentences.
2. For the extracted arguments, classify them into correct roles/types.
3. (Bonus) Jointly optimize all the modules of your event extraction system. Describe your method and the performance after joint optimization.

As the evaluation metrics in Part 1 and Part 2, you can use precision, recall, and F1. Refer to Section 4.1 in the following paper for the detailed evaluation criteria.

Li, Qi, Heng Ji, and Liang Huang. "Joint event extraction via structured prediction with global features." ACL 2013. https://aclanthology.org/P13-1008/

Report

The experiment report should include the following parts:

- Description of your models.
- Implementation details of your models.
- Results on the dataset and analysis of the results.

We recommend using the official ACL style templates for you report. (https://github.com/acl-org/acl-style-files).

Description of Shell Script

To make sure we can run your code successfully, you need to submit a shell script, named ***run.sh***. We will run this script in terminal to start your program. If we cannot start your program from the script, you can only obtain 50% of the total score at most. You can assume that this script and your source code are in the same directory and the data files are in the folder named "data". Please use **Relative Path** rather than **Absolute Path** in both your script and your source code.

Contact

If you have any question about this homework, please email TA via zhangch@pku.edu.cn