110-2
Natural Language Processing

HW - 2

TA: Kuei-Chun Kao

Task introduction

- Information Extraction from text
- You can use many ruled based method ensemble to get better F1-score
- No ML methods
- Tips: you can Use spacy and Get NER, POS, parsing-tree…etc
- Learning Objectives: try any rule based method and statistical method, and do error analysis for each method to improve your score

Inputs

- This file contains a set of Web sentences
- Sentence id (id): the id from sentences.txt
- Argument1 (S): the first noun phrase (Subject)
- Relation (V): the relation phrase (Verb)
- Argument2 (O): the second noun phrase (Object)
- Sentence

id	S	V	O	sentence
1515	premiums	reduce	benefits	The nation 's health maintenance organizations...

Outputs

- Determine whether the given triplets is correct from the corresponding sentence.
- Output the corresponding id’s binary label
- Kaggle Link: https://www.kaggle.com/t/e25d476dd3b44c8baa8bed88851a154a
- Displayed name: <student_ID>
- Submission format: .csv file (You can also see from sample_submission.csv)
- Evaluation metric: F1 score

	baseline.csv	0.50810

Kaggle submission

- You may submit up to 5 results each day (UTC).
- Up to 2 submissions will be considered for the private leaderboard

remember to select **2** results for your final scores before the competition ends!

Requirements

- Python only
- No plagiarism!
- At the top of your Source code

```
#Author: Kuei-Chun Kao

#Student ID: 1234567

#HW ID: hw2

#Due Date: 01/30/2020
```

Submission

- Deadline
  - Submit Zip to E3 before **4/15 11:59 PM**
  - No Late Submission, thanks!
- Format
  - Source code: Hw2_<StudentID>.py (py only)
  - Report file: Hw2_<StudentID>.pdf (pdf only)
  - Make sure the .py file contains the **correct execution results and formats**.
  - If can’t compile correctly, no score for you
  - Zip file: Hw2_<StudentID>.zip (zip only)
- Any question can ask me on E3, answer your question ASAP

Grading policy

- Ranking score in Private Leaderboard (60%)
  - If your score > public baseline score , ur score in this part = leaderboard score
  - Otherwise, ur score in this part = (leaderboard score*0.8)
- Code comments and formats (10%)
- Report (30%)
  - Ur report must contains three kinds of rule-based method
- I can only see your last submission.
- Do not submit your model or dataset.
- If your code is not reasonable, your final grade will be multiplied by 0.8!
- You should NOT modify your prediction files manually.
- Do NOT share codes or prediction files with any living creatures.
- Do NOT use any approaches to submit your results more than 5 times a day.

Bonus

- If your ranking is top 3 in class, you can share implementation and ideas in class -> get 3 points bonus!