Identifying and Categorizing Offensive Language in Social Media

2019 NLP Project 2

Task Description

This task requires the annotators to give their judgements on whether a tweet is offensive or not.

1. The annotators mark the tweet as being offensive or not offensive.
2. If the tweet is offensive then the annotators need to tell if the offense is targeted towards somebody or something or it is not targeted.
3. If the offense is targeted then the annotators also need to tell who it is targeted against.

Data

Download: [link]

Example:

1. Sub-task A: Offensive language identification
   a. Hey @LIRR , you are disgusting. - **Offensive**
   b. A true American literary icon. #PhilipRoth will be missed. - **Not offensive**
2. Sub-task B: Automatic categorization of offense types
   a. I mean I'm dating to get fucking attention - **Offensive Untargeted**
   b. Hey @LIRR , you are disgusting. - **Offensive, Targeted Insult**
3. Sub-task C: Offense target identification
   a. @BreFields1 @jonesebonee18 fuck you lol - **Offensive, Targeted Insult, Individual**
   b. @Top_Sergeant Assuming liberals are unarmed would be a grave mistake by the deplorables.
      - **Offensive, Targeted Insult, Group**

Data (conti.)

tsv format:

```
1   id   tweet      subtask_a    subtask_b    subtask_c
2   86426    @USER She should ask a few native Americans what their take on this is. OFF UNT NULL
3   90194    @USER @USER Go home you're drunk!!! @USER #MAGA #Trump2020 🇺🇸us🇺🇸 URL   OFF TIN IND
4   16820    Amazon is investigating Chinese employees who are selling internal data to third-party seller
    #TCOT  NOT NULL    NULL
5   62688    "@USER Someone should'veTaken"" this piece of shit to a volcano. 😂""" OFF UNT NULL
6   43605    @USER @USER Obama wanted liberals &amp; illegals to move into red states    NOT NULL    NULL
```

Evaluation

macro-averaged F1-score

- **sklearn.metrics.f1_score()**

Report

- report.pdf (no more than 6 pages; font size 12pt; 中英皆可)
  - Name and ID
  - Division of Work
  - Methods
  - Experiments (ex. acuracy on the given test set)
  - Discussion
  - Conclusion
  - Agree to share your report with your classmates? (YES/NO)
    - Will appear in 作業觀摩
- Code
  - Write the proper comment for each part and function
  - readme.txt (如何執行)
- Grading Policy
  - report: 70%; performance: 30%
  - Bonus: 1st~ 3rd +10%; 4th~6th +5%

Submission (Upload to CEIBA)

- project2_team_<team number>.zip
  - report_team_<team number>.pdf
  - code_team_<team number>
    - Readme
    - *script 1*
    - *script 2*
    - *...*

- testing 準確率寫在 report 裡就好, 不用繳交結果。

- project2_team_0.zip
  - report_team_0.pdf
  - code_team_0
    - readme.txt
    - function.py
    - main.py

Important Dates

- 5/24 release project 2
- 6/20 期末考
- 6/23 report and code submission due
- 分組名單 [link]

Resource

1. OffensEval 2019 (SemEval 2019 - Task 6) [link]
2. SemEval-2019 Task 6: Identifying and Categorizing Offensive Language in Social Media (OffensEval) [link]
3. An Exploration of State-of-the-art Methods for Offensive Language Detection [link]
4. Predicting the Type and Target of Offensive Posts in Social Media [link]
5. Inducing a Lexicon of Abusive Words -- A Feature-Based Approach [link]
6. The International Workshop on Semantic Evaluation [link] 搜尋 "task 6"