

Assignment #4 Text Sentiment Classification

Announcement

- Kaggle Deadline: 2018/06/06 11:59 P.M. (GMT+8)
 - Github Deadline: 2018/06/07 11:59 P.M. (GMT+8)
-

Link

- 投影片連結 <https://docs.google.com/presentation/d/1DRY1qawZ1FKTQgKimdpLgG2Vs0IAfDw2Xpv4fD5ceVE/edit#slide=id.p22>
 - Kaggle 連結 <https://www.kaggle.com/c/ml-2018spring-hw5/overview>
 - 遲交表單
 - 小老師申請表單
 - report template
-

Sample code

Report

P1: 請說明你實作的 **RNN model**，其模型架構、訓練過程和準確率為何？ (1%)

- 作業要求: 請將你的 model 架構(summery)、訓練細節(ex. epoch、optimizer、loss fuction...)與準確率附上。

P2: 請說明你實作的 BOW model, 其模型架構、訓練過程和準確率為何? (1%)

- 作業要求: 同上。

P3: 請比較 bag of word 與 RNN 兩種不同 model 對於"today is a good day, but it is hot"與"today is hot, but it is a good day"這兩句的情緒分數, 並討論造成差異的原因。 (1%)

- 作業要求: 請附上兩個 model 對這兩句的 prediction(不是 label), 並討論原因。

P4: 請比較"有無"包含標點符號兩種不同 tokenize 的方式, 並討論兩者對準確率的影響。 (1%)

- 作業要求: 請附上在 RNN model 上, 兩種不同 tokenize 方式的準確率, 並討論原因。

P5: 請描述在你的 semi-supervised 方法是如何標記 label, 並比較有無 semi-supervised training 對準確率的影響。 (1%)

- 作業要求: 請說明如何標記 unlabeled data 的 label, 附上兩者在 RNN model 上的準確率, 並討論原因。

Rules

- Basic : No extra dataset.
- Only toolkits below are allowed to use :
 - Python 3.5+
 - tensorflow 1.4.0
 - keras 2.0.8
 - pytorch 0.3.0
 - gensim, GloVe
 - Numpy, Pandas 0.20+, Python Standard Lib...

Posted by: [ntumlta](#)

Contact information: ntumlta2018@gmail.com.

Course information: [Machine Learning \(2018, Spring\) @ National Taiwan University](#).