Assignment #7

Twitter US Airline

Text Sentiment Classification

Due on Jan 19, 23:59

Overview

- Sentiment classification is the automated process of identifying opinions in text and labeling them as positive, negative, or neutral, based on the emotions customers express within them.
- In this assignment, you need to train a recurrent neural network (RNN) or fine-tune a pre-trained language model (e.g., BERT) to predict the sentiment of given tweet.
- You can use pre-trained model.

Dataset

- Twitter US Airline Sentiment from kaggle
- Twitter data was scraped from February of 2015 about each major U.S. airline
- Contributors were asked to first classify positive, negative, and neutral tweets, followed by categorizing negative reasons.

- This assignment dataset link
- We resample the data and split it into three groups: train, val and test
- Replace sentiment by (positive, 2) (neutral, 1) (negative, 0)

Your task

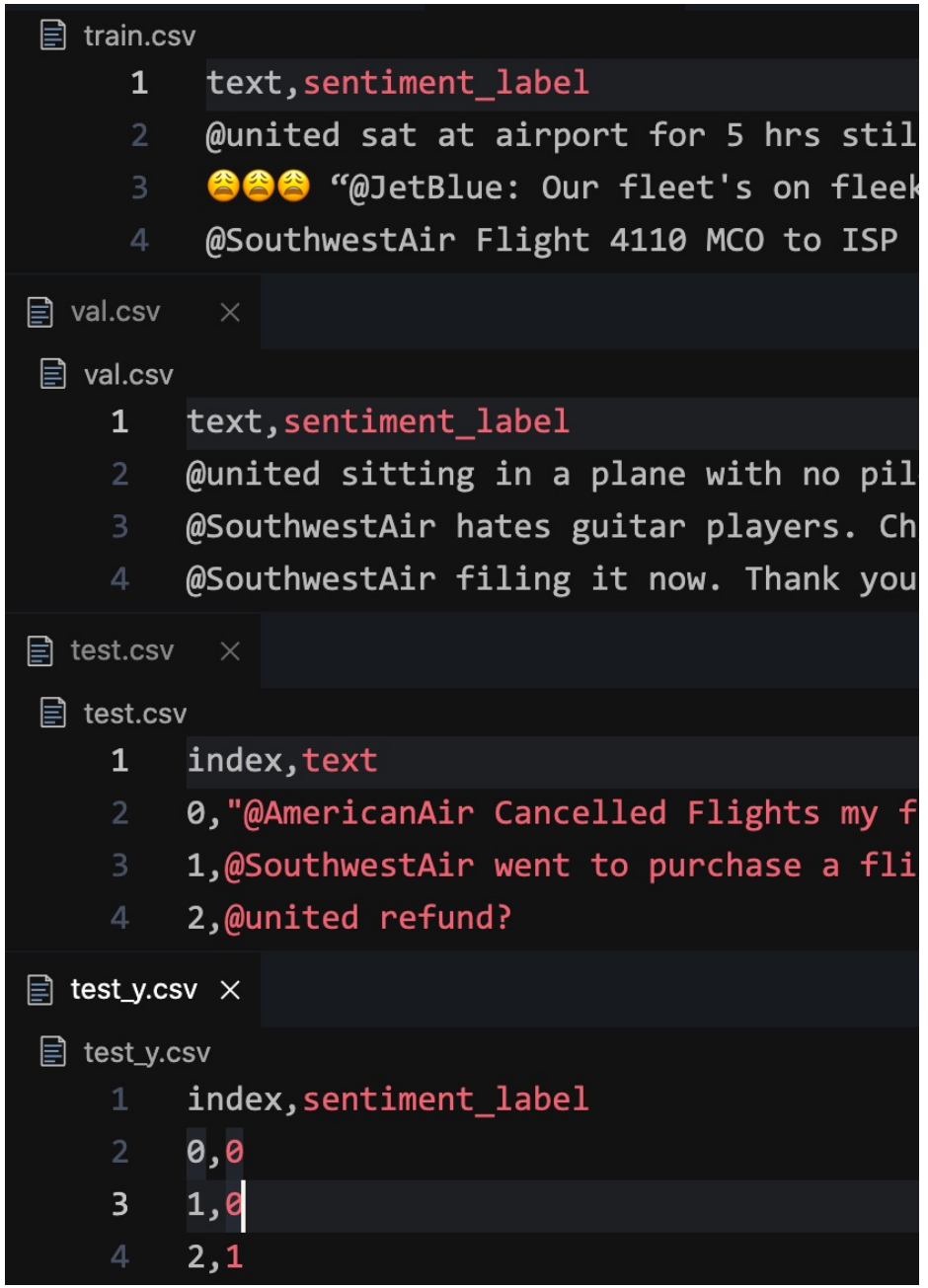
- Skeleton code: https://colab.research.google.com/drive/1i6bqF82EbMY7dnLYuPWM_o0DcF2ceuLx
- Using word embedding to represent the word
  - You can use torch.nn.Embedding to learn word embeddings
  - Example: LSTM for part-of-speech tagging
  - Or use pre-trained GloVe or fastText word embeddings for better performance
  - Example: torchtext, Deep Learning For NLP with PyTorch and Torchtext
  - Notice : You need use all text (train, val, test) to get word embeddings

- Using a pre-trained model of your choice, you are to build a deep network that predicts the sentiment of a given tweet.
  - PyTorch-transformers pre-trained models

Your task (cont.)

- Output is three sentiment polarity
  - Positive: 2
  - Neutral: 1
  - Negative: 0

- Submission format:
  - Follow the index number in test.csv



Things you cannot do

- You cannot submit results predicted by others.
- You cannot copy trained models from others.
- You cannot copy code from others, internet, GitHub …
- You cannot collect more images to train your model in order to boost performance.

Any violation will result in 0 score!

Submission

- Submit your predictions on the test tweets to Kaggle for evaluation.
- Kaggle competition

https://www.kaggle.com/t/6002ce45761a47399b40ecd8f81e3715

- Remember to change your Team Name
- Evaluate by accuracy

- Submit your code to the CU.
- File name: assignment7.ipynb

Grading

- 100 points competition
  - Bonus points to top 3 teams
  - Top 3 teams will share their approaches in class