CS583 – Research Project

Demo date: the week of Nov. 29, 2021

Task: Classify the sentiment or opinion expressed in tweets into one of the three classes: positive (1), negative (-1), or neutral (0) (which means no opinion).

You can use any techniques or tools that you can find on the Web to do the project. Clearly, you can also design and implement your own algorithms, which is encouraged.

Training data. The training datasets are in the attached .rar file, which has a zipped Excel file containing two training datasets: tweets about Obama and tweets about Romney, which are stored in two separate sheets of the file. The tweets were downloaded from Twitter during a presidential debate between Obama and Romney some years ago. The training data has 4 classes: positive (1), negative (-1), neutral (0), and mixed (2). The mixed class (2) means that a tweet expresses both positive and negative opinions (or sentiments). In testing, we will not use class label 2. So, you can ignore class 2 in training.

Test data. Test data without class labels will be given to you at the demo time. The test data contains only positive, negative and neutral tweets about the two presidential candidates in two separate sheets of an excel file, just like the training datasets.

In training, you can build two separate models, one for Obama tweets and one for Romney tweets, or single model that can be used to classify tweets about either of the candidates.

Evaluation metrics:
1. Accuracy
2. Precision, recall and F1 score of the positive class
3. Precision, recall and F1 score of the negative class

Project team: This is a team project. Each team/group can have no more than 2 students.

What to turn in:
(1) Given the test data, your model(s) should assign a class label to each test tweet. We will compute the above evaluation metrics. **Due date**: at the demo time.
(2) A written report of no more than 4 pages describing the techniques that you have tried and their results and the lessons learned. **Due date**: a few days after the demo.

Grading: You will be graded based on the results of the above evaluation metrics and your project report. In some sense, this is a contest. You are competing with your fellow students.