NLP Unit 4 Evaluation

Nov 15, 2016, 1:30 pm to 6:30 pm

Objective

In this exercise we will build an analyzer to process the performance of banks with respect to the recent government decisions. Given a tweet corpus, this model computes the performance of each bank and assigns a score.

Steps

(a) Dataset

i. You are provided a dataset in csv form that has tweet versus label. The labels are Displeasure, Compliment and Miscellaneous

ii. You are required to use the @xxx at the beginning of the tweets in order to identify the target bank to which the tweet is addressed.

iii. Create a test dataset that contains randomly selected 500 tweets per bank. This will be used later to compute the bank's performance

(b) Building Classifier

i. You are required to build the bank performance analyzer using 3 classifiers: (a) MaxEnt (b) Deep learning classifier (c) Hybrid

ii. Train the system with the labelled corpus. Write suitable feature functions for the MaxEnt, use word2vec for the deep learning and combine these features to train a hybrid model (You should think through the design yourself, I am not providing step by step instructions so that you can think through with more freedom)

iii. Use the held out dataset and compute a performance index for each bank. Rank order these and identify best and worst performers

(c) As the tweets may have spelling errors, perform your study using both word2vec as well as lexicalized word2vec as you did during the U3 evaluation

(d) Before training the system, you may perform necessary pre processing such as tweets cleaning etc.

Deliverables

Submit the following by 6:30 pm, 15th Nov 2016:

1. Source code of classifiers, application

2. Test results: Who is the topper among banks?

Do the following by 10 pm tonight:

1. Post your results and analysis on the Facebook. Regardless of the accuracy you got, explain in the document what went right and where you could have taken a better approach to getting better results.
2. Accuracy will be given a credit of 1 mark out of 10
3. Optionally you can include any graphics, visualization etc

Best wishes from the faculty, enjoy NLP development!

Anantharaman Palacode Narayana Iyer, Chitra GM