Dialogue Processing on Twitter

JAFARI Somaye, OLUOKUN Adedayo
University of Lorraine
June 6, 2017

Introduction

In order to understand dialogues, the ability to model and automatically understand discourse structure is essential.
To achieve this, the identification of dialogue acts (DAs) is a useful first step. In this work, the identification of dialogue acts was done by annotating a corpus of tweets based on a dialogue act set defined.

The goal of this project is to analyse and annotate Twitter dialogues and build a neural network model that automatically classifies the Twitter dialogues into corresponding dialogue acts .

Work done

Dialogue act annotation

1039 English tweets were annotated by two annotators using 10 dialogue acts. Each dialogue contains at least 5 tweets. The inter annotator agreement was 85.12%.

Dialogue act modelling

The dialogues were modelled using a Bi-directional LSTM with dialogue history taken into consideration and also a LSTM without dialogue history taken into consideration.
The original tag set containing the 10 dialogue acts were merged into 4 categories so as to have more balanced corpus.

Results

4 tag set

Accuracy of 0.53 was gotten with the LSTM modelled without dialogue history taken into consideration while 0.60 was gotten with the Bi-directional LSTM with dialogue history taken into consideration.

10 tag set

Accuracy of 0.46 was gotten with the LSTM modelled without dialogue history taken.

Conclusion

Different experiments on the model showed that having a more balanced corpus with respect to the dialogue acts improved the accuracy of the model. The best accuracy was achieved when the dialogue history was taken into consideration. An increase in the training data from 162 tweets to 1039 tweets increased the accuracy of the model.
These three important points can be taken into consideration to improve the model for future work.