Sentiment Analysis of Twitter Data for Prediction of Stock Movements

Óttar Guðmundsson <ottarg@kth.se> Xin Ren <xinr@kth.se>

Problem description:

Twitter is a social network where people post short, 140-character, status messages called tweets. Because tweets are sent out continuously, Twitter is a great way to figure out how people feel about current events. For example, how people feel about the presidential candidates, how people feel about a movie to predict box office. In this project, we will create a program that enables us to find out how people feel about a firm thus to predict its stock movements.

We will apply sentiment analysis on tweets. The total positive, negative and neutral emotions in tweets in a 3 day period are calculated successively to predict the movement of next day stock price. Then we will compare the prediction with data from Yahoo finance to evaluate the accuracy of the program.

P.S. We may reduce the window size (3 days) given the tight schedule of the project, and this may have an impact on the accuracy. However, it won't affect what we can learn from the project.

Tools:

Apache Spark, Stanford CoreNLP, Scala

Data:

Real time stream from Twitter Streaming API, Yahoo finance daily and intraday data

Methodology and algorithm:

Sentiment Analysis:
Sentiment analysis or opinion mining refers to the use of natural language processing and text analysis to identify and extract subjective information in source materials. Normally speaking, sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual of a document(s).