

Project proposal: Relation between the DOW Jones performance and the news

Jeroen Le Maire

November 28, 2016

Introduction

Because of my interest in global economics and the stock market, I would like to do a project about this topic. One of the branches in making investment decisions is called market sentiment analysis. In this kind of analysis, the news is the main input for predicting the performance of an index/stock.

Required data

A teacher has put a dataset on Kaggle that contains the 25 biggest news topics from reddit for every day of the last eight years. Furthermore, the data is binary labeled. When the dow Jones performed well that day, the label is 1 (otherwise 0). I think i will replace the binary label with the (normalised) percentage that the Dow Jones changed. In that way I can search for news with the most impact.

Architecture and plan

First of all I would use the bag of words methodology to put the data into the right format for feeding a neural network. Then I would do some basic analysis as: which topics were in the news when the market movement was the biggest, which words occur the most on "good" days,... After the analysis it is time to build and train the neural network. The first 6 years of data will be used for training, the other 2 year for testing. I will use a fully connected network with a few layers. Probably I will start with 50 intermediary nodes and see how the network performs. If the network performs really bad, I will try to increase it or use different techniques (activation, dropout,...).