

STOCK PREDICTION USING TWITTER SENTIMENTAL ANALYSIS

ABSTRACT

Stock movement prediction is a challenging problem: the market is highly stochastic, and we make temporally-dependent predictions from chaotic data. We treat these three complexities and present a novel deep generative model jointly exploiting text and price signals for this task. Unlike the conservative methods, we are also using Twitter Sentimental Analysis to train our CNN model. Which results in an exceptional 70% accuracy

METHODOLOGY

Data is prepared by using a 30 day window and sliding it over the complete time period. For sentiment, neutral is assigned value of 0, positive 1 and negative -1. At start of each window a sentiment score is set to zero and value of each day is added to it. Score corresponding to each is added day is set as sentiment value of that day. The closing price of that day is also stored. Normal values of price and sentiment score is calculated and used. It is one instance of training data. The label for this instance is closing price of 31st day. Random 5% data is separated from training data as test data. CNN is used as a prediction model. There are two 1D convolution layers each with activation function LeakyRelu with dropout set to 0.5. There is a full connected layer with 64 neurons and a softmax layer with 2 neurons.

EXPERIMENTAL SETTINGS

The data set we used is StockNet dataset courtesy of Yumo Xu which can be obtained from Github.

The stock data is present for working days and do not contain weekend price. We predicted this price by calculating average of previous and next present day. For n days we used the formula (x + y)/2 where x is previous present day and y is next present day for n days. We have also stored only closing price. Tweet data present is raw data generated from twitter. We have gathered tweet text, favorite count and retweet count. We have used TextBlob library to calculate sentiment of tweet. Favorite count and Retweet count are added together as weight. There are three types of sentiment ie Positive, Negative, Neutral. We calculate the maximum sentiment value and assign sentiment to that day. We only used data of Apple for prediction. It contains sentiment value and closing price of stock for each day from 01/01/2014 to 30/06/2016.

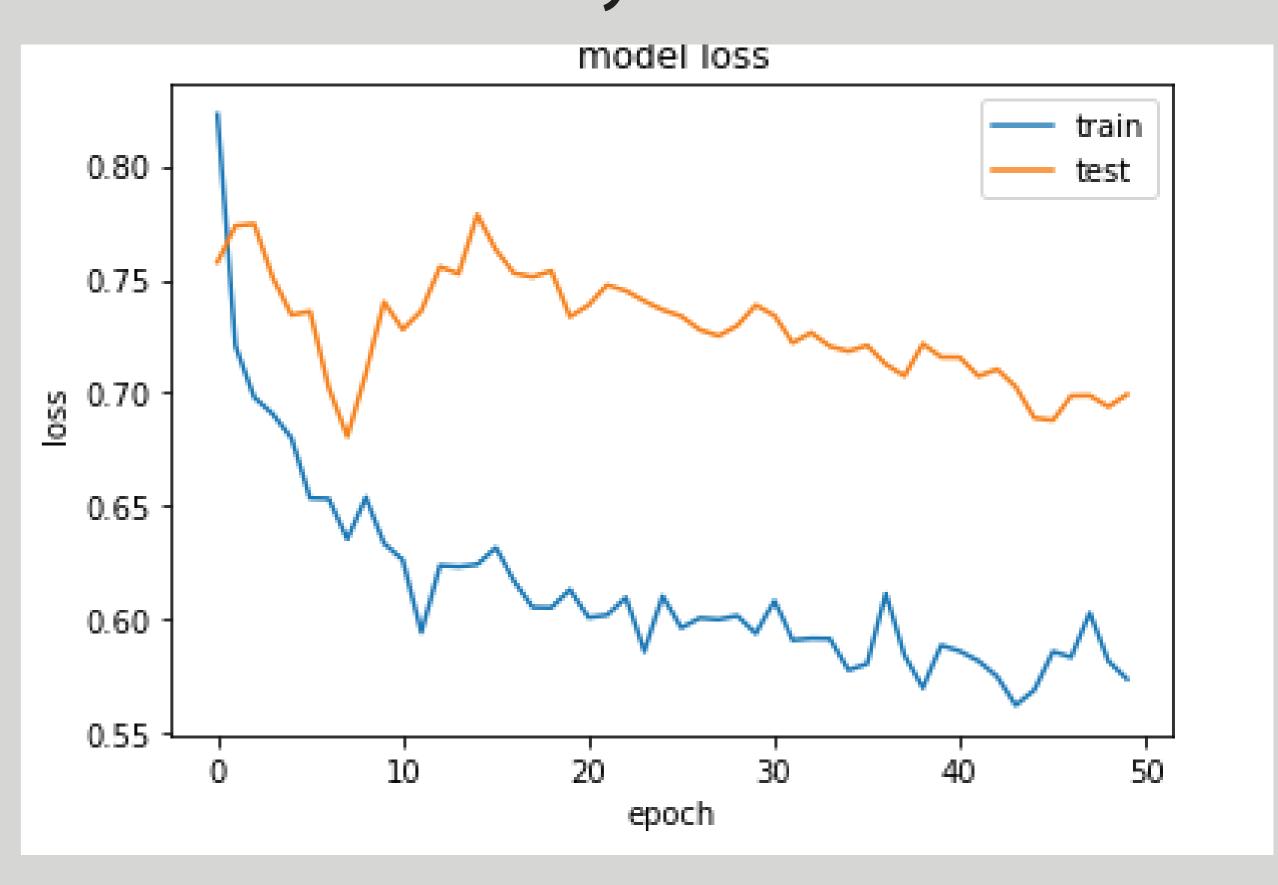


Timeseries plot of Sentiment and Stock price for January 2014

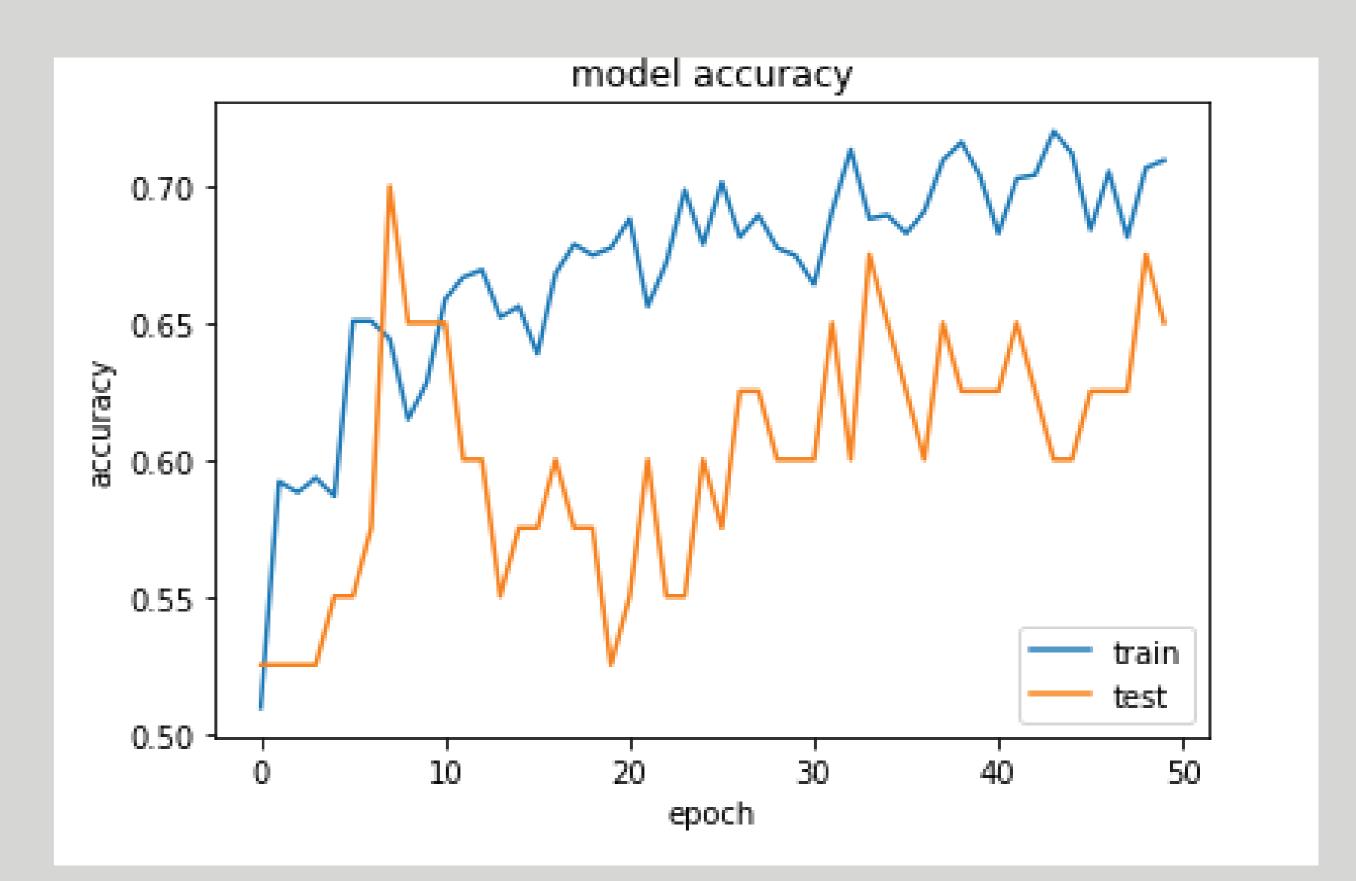
CONCLUSION & RESULTS

Maximum accuracy computed over the period of 50 epochs is 69.99%. It is obvious that it is quite better than MLP and other basic neural network techniques which gives mere

~55-65% accuracy



Train and Test loss against 50 epochs



Train and Test accuracy against 50 epochs

FUTURE WORKS

Accuracy can be improved in future works using LSTM.

Our model can be improved using news articles sentimental analysis as well. That can result in ~75-80% accuracy which is quite good and could be used for practical purposes.

M. Rizwan, Ahmed Zubair, Hamza Anjum FAST NU Lahore Tel: 111 128 128

REFERENCES