Predicting Stock Market Movements Using Global News

Joshua Chang, Callan Hoskins, Jason Kurohara

CS230: Deep Learning, Department of Computer Science, Stanford University

Stanford Computer Science

Abstract

Financial markets are inherently volatile and reflectc diverse macroeconomic and microeconomic trends. In this research project, we use natural language processing in conjunction with a deep neural network to predict the daily change in price of the Dow Jones Industrial Average. After trying a variety of models with input data (previous days' stock price changes, news, and weather), our most successful model performed significantly better (61.31%) than random guessing for predicting whether the stock market would rise or fall on a given day. While the stock market is inherently stochastic, our study offers interesting insights into how global news headlines can influence stock price movements.

Stanford University

Background

The price of a company's shares is subject to a wide range of variables, some intuitive and some not. The task of using these inputs to predict a stock's price movement is a trillion-dollar industry. Past successful efforts in this field have used tweets and event detection algorithms to predict day-to-day trends in the stock market.

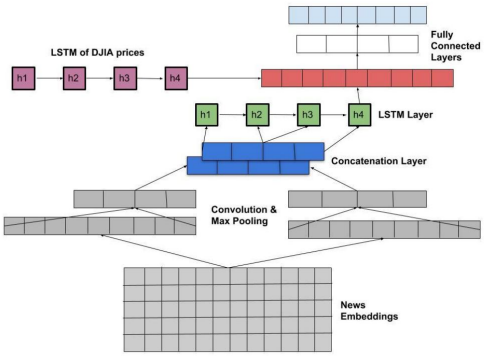
Problem

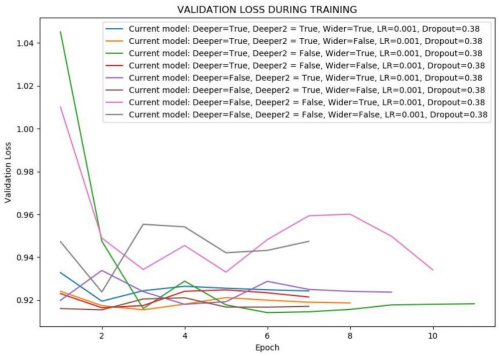
Given twenty-five of a day's most popular news headlines, predict the change in price of the Dow Jones Industrial Average (DJIA).

Dataset and Features

We use 25 news headlines for each day between August 8, 2008 and July 1, 2016. News headlines range from "The President of France says if Brexit won, so can Donald Trump" to "Spain arrests three Pakistanis accused of promoting militancy". For each day, we also include the previous five days' DJIA price changes. This is to capture the momentum of the DJIA, as previous trends in the market influence the current price.

Model





For our final model, we tested eight different architectures, adjusting the depth (1 to 4 FC and CNN layers) and width (1x or 2x hidden dimensions).

Results

Model	Accuracy	RMSE	MAE
Weather Logistic Regression	51.80%	1.0948	62.63
News with CNN & LSTM *(Baseline)*	54.52%	0.1612	169.32
News and Previous Stock Price *(1 FC layer, 1 CNN, 2x Hidden Dims)*	55.28%	1.0068	71.40
News and Previous Stock Price *(2 FC layers, 2 CNN, 1x Hidden Dims)*	58.28%	0.9776	68.27
News and Previous Stock Price *(3 FC layers, 3 CNN, 2x Hidden Dims)*	61.31%	0.9721	68.67
News and Previous Stock Price *(4 FC layers, 4 CNN, 1x Hidden Dims)*	61.31%	0.9707	69.87

Discussion

Since predicting the magnitude of the stock price change is difficult, we found that our models are best measured by a binary classification task of predicting a rise or fall in the DJIA on each day. Deeper networks proved to have the highest accuracy and lowest RMS error. We achieved an accuracy of 61.31% on our best model, which is significantly better than 50% random chance.

Future

Our model responded well to the addition of relevant data. Given more time, we would incorporate additional inputs such as presidential tweets or the prices of foreign financial indices. We would also like to refine our architecture to test deeper models with more layers, and adapt our evaluation metrics to analyze the magnitude of the stock prices rather than just their change.