To Stock or not to stock

Patrick Du, Kiseok Yang, Chuanyao Lin

Problem Description

- The goal of this project is to predict the specific price of a stock at any given point based on previous existing data and real time data provided by news sites. \
- One major challenge will be addressing when the market is closed and how to accomodate for it.
- We plan to categorize stocks into 3 groups, stable stocks, new stocks and volatile stocks for prediction.
- Sentiment Analysis on news headlines (Bloomberg) + bigram/trigram dictionary
- SVM for Feature Selection to optimized features
- LTSM to make prediction on the strock

Related Works

Taewook Kim, Ha Young Kim.(2019) "Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data", PLOS ONE Journals,

Mehtab, S., Sen, J., Dutta, A. (2020). "Stock Price Prediction Using Machine Learning and LSTM-Based Deep Learning Models". Conference: Second Symposium on Machine Learning and Metaheuristic Algorithms and Applications (SOMMA'20) at: Vellore Institute of Technology, Chennai, INDIA.

Nelson, D.M.Q.; Pereira, A.C.M.; de Oliveira, R.A. Stock market's price movement prediction with LSTM neural networks. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017.

B. Li, Q. Wang, and J. Hu, "Feature subset selection: A correlation based SVM filter approach," IEEJ Trans. on Electrical and Electronic Engineering, vol. 6, no. 2, pp. 173–179, March 2011.

- This paper talks about the feature selection. By applying the SVM-based feature ranking module, it helps to eliminate the features with low influence quantities and rank them. By applying the correlation-based clustering (affinity propagation), it groups the features into some clusters which has high correlation in the same cluster, but low correlation with the other clusters. Thus, we could find out a good subset by selecting one feature from each cluster as the delegate.

Shen, S., Jiang, H., & Zhang, T. *Stock Market Forecasting Using Machine Learning Algorithms*. http://cs229.stanford.edu/proj2012/ShenJiangZhang-StockMarketForecastingusingMachineLearningAlgorithms.pdf.

- This paper talks about how Sentiment Analysis can produce bias as well as introduce a way to handle the US market when it is closed. One major challenge was finding a way to accomodate for the US market while it was closed. Major success was their high prediction rate.

Kalyanaraman, V., Kazi, S., Tondulkar, R., & Oswal, S. (2014). *Sentiment Analysis on News Articles for Stocks*. Sentiment Analysis on News Articles for Stocks - IEEE Conference Publication. https://ieeexplore.ieee.org/document/7079267.

- This paper focuses on using Sentiment Analysis as a means to predict the stock market. One major issue was that a sentiment dictionary for financial terms does not exist so they had to create their own. A major success was that there was a notable increase in prediction accuracy using Sentiment Analysis.

Algorithms

Sentiment Analysis algorithm -> Being fed sentiment dictionaries to predict sentiment of newlines related to our stocks
Feature Selection via SVM -> finding which features increases accuracy of prediction
Reliably predicting the price of a stock with bidirectional LTSM -> Predicting stock price reliably.

Team Roles

Patrick

Coordinator and working on Sentiment Analysis

Kiseok

Working on LTSM

Chuanyao

Working on Feature selection with SVM

Everyone

- Organize the dataset
- Manually give sentiment values to headlines and bigram/trigrams.
- Learning necessary tech stack/libraries

Evaluation Methods

First and foremost we will be scrapping news headlines off Bloomberg to apply Sentiment Analysis on. This will then be used to predict the sentiments of the newsline related to our groups. Afterwards this will be added into our Stock objects as a feature.
We will then use Feature Selection via SVM to improve the accuracy of our prediction.
Finally we will apply LTSM on the whole dataset to find whether this stock should be invested in.
LTSM vs LTSM+Sentiment Analysis+Feature Selection

We plan to compare the results of our LTSM using Sentiment Analysis as well as Feature Selection to those results seen in LTSM.

Timeline

- Scrapping Bloomberg and creating sentiment dictionaries (Done by 11/10)
- Sentiment Analysis (Done by 11/17)
- Feature Selection (Done by 11/24)
- LTSM (Done by 12/01)
- Visualization (Done by 12/08)
- Preparing for final report and presentation (Done by due date)