Project Proposal

Submitted by: Rahul Gupta, Pan He (Gabby), Tesa Ho

The stock market, now 25 Trillion, is one pivotal player in US economy. Predicting stock returns, using either company's financials or quantitative factors, is one of the most challenging and rewarding work. In this project, we would like to attempt to predict stocks returns with cutting-edge technology in deep learning.

The application of convolutional neural networks (CNN) to time series prediction is a relatively new area. Several papers have attempted to predict returns of stocks and other financial instruments with recurrent neural networks (RNNs), in particular Long Short-Term Memory (LSTM). Persio and Honchar laid the framework for comparing different neural network models for stock prediction in their paper *Artificial Neural Networks architectures for stock price prediction: comparisons and applications*. The authors tested a Multi-Layer Perceptron (MLP), CNN and LSTM separately on the S&P 500 index prices. The results indicated that the CNN was able to model the financial time series better than the other models.

Inspired by these works, we would like to expand the stock prediction scope to individual stock and propose to use following deep learning model architectures for daily stock return prediction.

- LSTM
- CNN
- CNN + LSTM

One potential improvement area is to use multiple time series data and consider correlation among them. Previous works are usually focus on a singular time series and fail to consider the correlations between other time series. In addition, we will use reinforcement learning algorithm to construct a new portfolio based on the model output and test it against benchmark.

Our investable universe is S&P500 – the top 500 US companies based on market cap – a good representation of the US stock market. We will collect daily stock data of S&P500 constituents for the past 10 years from intrinio.com and Bloomberg terminal. Concretely, the time series data for each stock includes:

- Daily open/close, high/low price
- Trading volume
- Market cap
- Earnings
- Sector/industry classification

The output of the model could be two types: buy/sell signal for each stock; or alpha (ranking) of the stocks in the universe. In either way, we will construct a new portfolio and compare it against benchmark with following financial metrics:

- Absolute/relative returns
- Tracking errors
- Sharpe/Sortino ratios
- Maximum drawdown
- Profit and loss distribution

In addition, we will evaluate the prediction via statistical evaluation methods such as classification accuracy, area under curve, r-squared and mean squared error.