Project Proposal
Dior Xiong, Jee Su Byun, Shantanu Mandal

Stock Prices Forecast with Stacked Autoencoders and LSTM

1. Introduction

During the past decades, machine learning models, such as convolutional neural networks, deep belief networks and stacked autoencoders, have been widely used to predict financial time series and gain high predictive accuracy. In our class project, we will implement a model based on the stacked autoencoders to predict the stock market.

2. Input Data

Six stock indices will be used to train and test our model. These include both developed and developing markets (such as Indian, Hong Kong and New York). Data will be collected from available APIs such as Investing.com. Three types of variables will be used as model inputs:
1. Historical data: Open, High, Low and Close price.
2. Technical indicators: Bollinger Bands, MACD and RSI.
3. Macroeconomic variables: GDP, Inflation, Exchange Rates, and Interest Rates.

3. Model

Stacked Autoencoders (SAEs) will be used to learn the deep features of financial time series in an unsupervised manner. It is a neural network consisting of multiple single layer autoencoders in which the output feature of each layer is wired to the inputs of the successive layer. The unsupervised training is done one AE at a time by minimizing the error.

Additionally, wavelet transforms (WT) and long-short term memory (LSTM) will be incorporated to increase accuracy. LSTM will be used to solve the problem of a vanishing gradient by having the memory unit retain the time related information. WT will be used to denoise the input financial time series before fed into the SAE.

4. Evaluation

We will implement a conventional RNN as a performance benchmark. We will evaluate performance from two dimensions - predictive accuracy and profitability.
1. Predictive Accuracy: Mean absolute percentage error and correlation coefficient (R).
2. Profitability: Buy-and-Sell trading strategy and Buy-and-Hold strategy for each index.

5. Expectation

Our model is expected to outperform the conventional RNN model not only in predictability but also in profitability.

Reference

Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. Plos One, 12(7). doi: 10.1371/journal.pone.0180944