Stock Prices Forecast with Stacked Autoencoders and LSTM

1. Data Collection

We initially select 11 stock indices to test the prediction ability of the proposed model. Those indices include:

- Korean: KOSPI (Korea Composite Stock Price Index)
- Vietnam: VN30 (VN30 Equal Weight Index)
- Bangladesh: DS30 (Dhaka Stock Exchange 30)
- Hong Kong: SHI (The Hang Seng Index)
- London: FTSE 100 Index (The Financial Times Stock Exchange 100 Index)
- USA: DJIA (Dow Jones Industrial Average)
- China: SSEC (Shanghai Stock Exchange Composite)
- India: BSESN (S&P Bombay Stock Exchange Sensitive Index)
- Switzerland: SMI (Swiss Market Index)
- Brazil: IBOVESPA

Beside, we also choose technology indices for extra model testing:

◊ Korean: KOSDAQ (acronym of Korean Securities Dealers Automated Quotations)
◊ USA: NASDAQ (Nasdaq Composite)
◊ Indian: Nifty IT (NIFTYIT)
◊ China: CQQQ (Invesco China Technology ETF)
◊ London: FTSE AIM All-Share - Technology

As we presented here, for example, the KOSPI represents Korean stock market, and DS30 presents Bangladesh stock market, etc., etc.

At this moment, we collected daily data from www.investing.com. There are seven data attributes we adopted:

- Date: The date which we collected the price
- Price: The corresponding stock price
- Open: The price when the market open
- High: The highest price the market reach
- Low: The lowest price the market reach
- Vol: The total number of trades occurs
- Change: The percentages fluctuated (compared with the price on last day)

We also considered to collect data from other sources: (not free)

- Quantopian: https://www.quantopian.com
- Quandle: https://quandl.com
- NASDAQ: https://www.nasdaq.com/market-activity/quotes/historical
- Eoddata: http://www.eoddata.com/download.aspx
- TAMU May Business School: https://mays.tamu.edu/innovation-research-center/data-sets-in-mays-business-school/
- First Trade Data: http://firstratedata.com/

However, we still have several difficulties when collecting the data:

a) We wish to narrow down the industry sector such as oil, gold and technologies. But we didn't find all the corresponding index which reflect them individually. This is something we will refine later on.
b) We also wish to find the hourly data which can help us better analyze the short term trend pattern. But the data normally is not free. We will continue working on that later.
c) Data-set which heavily affected the markets such as political events, festivals, trader war, etc. We need to collect that data too

d) Other indicators such as 52 weeks high points, buy and sell data respectively, etc.

2. Wavelet Transform

A function $\psi \in L^2(\mathbb{R})$ is called an orthonormal wavelet if it can be used to define a Hilbert basis, that is a complete orthonormal system, for the Hilbert space $L^2(\mathbb{R})$ of square integrable functions.

The Hilbert basis is constructed as the family of functions $\{\psi_{jk} : j, k \in \mathbb{Z}\}$ by means of dyadic translations and dilations of ψ,

$$\psi_{jk}(x) = 2^{\frac{j}{2}} \psi\left(2^j x - k\right) \quad \text{for integers } j, k \in \mathbb{Z}.$$

The continuous wavelet transform of a function $x(t)$ at a scale (a>0) $a \in \mathbb{R}^{+*}$ and translational value $b \in \mathbb{R}$ is expressed by the following integral

$$X_w(a, b) = \frac{1}{|a|^{1/2}} \int_{-\infty}^{\infty} x(t) \overline{\psi}\left(\frac{t-b}{a}\right) dt$$

where $\psi(t)$ is a continuous function in both the time domain and the frequency domain called the mother wavelet and the overline represents operation of complex conjugate. The main purpose of the mother wavelet is to provide a source function to generate the daughter wavelets which are simply the translated and scaled versions of the mother wavelet. To recover the original signal $x(t)$, the first inverse continuous wavelet transform can be exploited.

Since wavelet transform has the ability to decompose complex information and patterns into elementary forms, it is commonly used in acoustics processing and pattern recognition, but it has also been proposed as an instantaneous frequency estimator. Moreover, wavelet transforms can be applied to the following scientific research areas:

edge and corner detection, partial differential equation solving, transient detection, filter design, electrocardiogram (ECG) analysis, texture analysis, business information analysis and gait analysis.

Other than that, wavelet transform is applied for data denoising in this study since it has the ability to handle the non-stationary financial time series data. The key property of wavelet transform is that it can analyze the frequency components of financial time series with time simultaneously compared with the Fourier transform. Consequently, wavelet is useful in handling highly irregular financial time series. This study applies the Haar function as the wavelet basis function because it can not only decompose the financial time series into time and frequency domain but also reduce the pro- cessing time significantly. The wavelet transform with the Haar function as a basis has a time complexity of O(n) with n denoting the size of the time series.

3. Stacked Autoencoder

In our project, autoencoders will be applied for layer-wise training for the OHLC variables and technical indicators. Single layer AE is a three-layer neural network, the first layer being the input layer and the third layer being the reconstruction layer, respectively. The second layer is the hidden layer, designed to generate the deep feature for this single layer AE. The aim of training the single layer AE is to minimize the error between the input vector and the reconstruction vector. The first step of the forward propagation of single layer AE is mapping the input vector to the hidden layer, while the second step is to reconstruct the input vector by mapping the hidden vector to the reconstruction layer.

The activate function can have many alternatives such as sigmoid function, rectified linear unit (ReLU) and hyperbolic tangent. In this project, we set this to be a sigmoid function. The model learns a hidden feature from input by reconstructing it on the output layer. Stacked autoencoders is constructed by stacking a sequence of single-layer AEs layer by layer. The single-layer autoencoder maps the input daily variables into the first hidden vector. After training the first single-layer autoencoder, the hidden layer is reserved as the input layer of the second single-layer autoencoder. Therefore, the input layer of the subsequent AE is the hidden layer of the previous AE.

For training, the gradient descent algorithm will be used for solving the optimization problem in SAEs, and completing parameter optimization. Each layer is trained using the same gradient descent algorithm as a single-layer AE by solving the optimization function and feeds the hidden vector into the subsequent AE. The weights and bias of the reconstruction layer after finishing training each single-layer AE will be cast away. Depth plays an important role in SAE because it determines qualities like invariance and abstraction of the extracted feature. In this project, the depth of the SAE will be set to 5.

4. LSTM

LSTM networks consists of an input layer, several hidden layers and an output layer. Number of input layer is same as the number of features. At the beginning we plan to use the time series data as sequence to train the LSTM. We choose LSTM because it doesn’t have the problem of vanishing gradients. LSTM is a gated version of recurrent neural network. The main drawback of recurrent neural network is it can not use previously seen data due to vanishing gradient problem. It is commonly known that in stock data pattern is the key factor to move the market in technical analysis. Hand coded model such as clustering or feature extraction is sometime difficult in time series data such as stock value. So the main purpose of the LSTM is to capture pattern throughout the time series data. We have plan to show different performance benchmark comparison between LSTM and other deep neural network. With our intuition we are confident that LSTM will work better that other feed forward network.

5. Future Plan:

The future objective of our projects can be divided into many folds.

One, we want to show different deep neural network model comparison on stock analysis and find out the best fitted model for our case.

Two, different benchmark comparison with market alpha.

Third, make a framework that is easily extendable with different prediction model.

6. Conclusion:

In this report we discussed two different types of model stacked auto encoder and LSTM. We want to both feature extract and building our prediction model via different neural network topologies. Main goal of this project is to find out the optimized model through different analysis and cross validation of the data. We'll also show detail result on different stock benchmarks.

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

Ramsey JB. The contribution of wavelets to the analysis of economic and financial data. Philosophical Transactions of the Royal Society B Biological Sciences. 1999; 357(357):2593–606.

Popoola A, Ahmad K, editors. Testing the Suitability of Wavelet Preprocessing for TSK Fuzzy Models. IEEE International Conference on Fuzzy Systems; 2006.

Abramovich F, Besbeas P, Sapatinas T. Empirical Bayes approach to block wavelet function estima- tion. Computational Statistics & Data Analysis. 2002; 39(4):435–51.

Hsieh TJ, Hsiao HF, Yeh WC. Forecasting stock markets using wavelet transforms and recurrent neural networks: An integrated system based on artificial bee colony algorithm. Applied Soft Computing. 2011; 11(2):2510–25.