ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

**Международный институт экономики и финансов**

Жевакин Максим Павлович

**распознавание закономерностей для предсказания доходностей**

**(Pattern Recognition For Return Prediction)**

Выпускная квалификационная работа

БАКАЛАВРСКАЯ РАБОТА

по направлению подготовки 38.03.01 «Экономика»

образовательная программа «**Программа двух дипломов по экономике НИУ ВШЭ и Лондонского университета**»

|  |  |
| --- | --- |
| Рецензент  ученая степень, звание  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  И.О. Фамилия  Москва 2020 | Научный руководитель  ученая степень, звание  Д. И. Малахов |

**Contents:**

1. **Introduction**
2. **Financial Markets Analysis**
   1. **Candlestick charts**
   2. **Technical Analysis**
   3. **Technical Indicators**
   4. **Chart Pattern Analysis**
3. **Model framework**
4. **ANN (Artificial neural network)**
5. **CNN (Convolutional neural network)**

**2.1. Dataset**

**2.2. Pattern labeling and extraction (CNN)**

**2.3. Data augmentation**

**2.4. Data processing**

**2.5. Time-series imaging**

1. **RNN (Recurrent Neural Networks)**
2. **LSTM (Long short-term memory)**
   1. **Data preprocessing and feature creation**

**4.2. Feature scaling**

**5. Neural Network Hyperparameters tuning**

1. **Model Evaluation**
2. **LSTM**
3. **Pattern Signal Backtesting**
4. **Conclusion**

**Abstract:**

As time passes debates over efficiency of technical analysis of markets in price prediction are still ongoing and nowhere to end. People are still polarized when the discussion is brought up with proponents believing in predictive power of technical analysis and proponents of efficient-market hypothesis (EMH) saying that technical analysis has nothing to do with real-world financial market analysis. This thesis applies latest neuro network architectures and data extraction methods to pattern recognition analysis and proposes methodology for encoding of trading patterns of CNN model into three field matrices GASF GADF MTF in order to achieve more accurate classification. Empirical part consists of comparing baseline LSTM model’s prediction sequence with and without pattern and evaluating backtesting pattern prediction model on S&P 500 1-minute high frequency data.

1. **Introduction**

As time passes debates over efficiency of technical analysis of markets in price prediction are still ongoing and nowhere to end. People are still polarized when the discussion is brought up with proponents believing in predictive power of technical analysis and proponents of efficient-market hypothesis (EMH) saying that technical analysis has nothing to do with real-world financial market analysis. Inspired by the innovations in the classification field, specifically time-series classification and new neural network architectures achieving high accuracy and sometimes overperforming human capabilities in classification, I try to apply the newest methods to make an educated experiment whether technical pattern recognition may be successfully implemented to returns prediction.

Taking into consideration experience in time-series classification of previous research studies [2], this paper proposes encoding of trading patterns into three field matrices GASF GADF MTF for CNN model to classify patterns more accurately and then test resulting pattern prediction for predictive power by feeding predictions into LSTM model that also controls for previous returns dynamics and signals from technical indicators. The testing processes are considered in two stages: comparison of baseline LSTM model with and without pattern prediction sequence and backtesting pattern prediction model on S&P 500 1-minute high frequency data to understand whether there is a possibility to make abnormal profits from classified chart patterns.

1. **Financial Markets Analysis**
   1. **Candlestick charts**

In order to observe price dynamics of desired stock throughout time, traders and investors usually speculate Japanese candlesticks charts of a stock. Candlestick charting technique was developed by Japanese rice trader in 18th century to picture price fluctuations of rice. Basically, this chart concisely pictures interperiod dynamics into each candle by showing four main figures of an inter-period dynamic (OHLC – Opening price, Highest price, Lowest price, Closing price). This ease of representation difficult price fluctuation brought analysts, investors, traders to make their observations referring to this charting scheme. Candlestick chart is basing element in pattern analysis and technical analysis.

* 1. **Technical Analysis**

Technical analysis is an analysis methodology for forecasting direction of market data based on past values or patterns. Principles of technical analysis are derived from years of observation of market fluctuations, but the most famous work on this topic is Dow Theory which was derived from Dow Jones collection of writings, also other prominent pioneers who played part in forming classical technical analysis: Nelson Elliott, William Delbert Gann and Richard Wyckoff. However, modern technical analysis has transformed dramatically with the rise in researches in other scientific fields such as statistics, quantitative analysis and economics.

Although, the whole area of technical analysis is heavily disputed by proponents of efficient-market hypothesis (all of past price movements already incorporated in the current price), there are a lot of researches and studies that prove that technical analysis and pattern prediction has predictive power and even show abnormal expected conditional returns [1].

* 1. **Technical Indicators**

Technical indicators can be mainly split into overlays and oscillators.

Overlays are technical indicators that use the same scale as prices and they are plotted on top of price charts include moving averages, other extensions of moving averages (exponential, weighted moving averages) and Bollinger Bands.

Oscillators are technical indicators that fluctuate between a local minimum and maximum, usually plotted separately from price chart. Examples of oscillators include the stochastic oscillator, MACD or RSI.

MACD (Moving Average Convergence/Divergence)

An Indicator used for analyzing strength and direction of price trends, by measuring consistency of short-term and long-term tendencies of price fluctuations. Basing blocks of this indicator are exponential moving averages of prices. Conventional method of trading with this indicator uses EMA’s of periods 12, 26 and 9. EMA of period 12 (representing short-term trend) has to be subtracted by EMA of period 26 (long-term trend) this forms quick moving MACD line which would be positive in uptrend movement of price and negative in downtrend. EMA of period 9 from quick moving MACD line forms signaling line. This all renders to two lines which sometimes overlay and form “trading signals”.

RSI (Relative Strength Index)

An Indicator used for measuring strength of a trend and upcoming shift of aa current trend. Very easy instrument to interpret which is why it is one of the most famous. The Indicator bases on the upward and downward changes in closing prices smoothed by some period of exponential moving average. Convergence or divergence of an instrument line with market price trend is usually a sign of a trend shift.

Stochastic Oscillator

An instrument that measures percentage of market being in oversold or overbought region. The indicator is created by scaling closing price to highest and lowest trading price over some smoothing period. Traders choose a point of an entrance when indicator nears 100% or 0% (for 100% market is signaling that it is overbought and for 0% market is signaling that is oversold).

* 1. **Chart Pattern Analysis**

Pattern analysis (or chart analysis) is a subset of technical analysis which is more concerned with figures and patterns that are formed by candlesticks. These patterns were observed throughout history of financial markets to have signaling power of price movement. There are myriad of patterns, but in this study most famous ones would be tested for prediction power (H&S, IH&S, BBOT and BTOP). One of the given explanation for these patterns to have predictive abilities is self-fulfilling prophecy theory [6] which claims that patterns can predict prices because active market agents believe in its predictive power and act upon this belief.

All these patterns can be incorporated into concise formulas, where sequence are local extrema.

H&S (Head and Shoulders)

The pattern usually makes a signal to sell an asset

Изображение выглядит как текст

Автоматически созданное описание

Figure Head and Shoulders Pattern

IH&S (Inverted Head and Shoulders)

The pattern usually makes a signal to buy an asset

Figure 2 Inverted Head and Shoulder Pattern

BBOT (Broadening Bottoms)

The pattern usually makes a signal to buy an asset

Изображение выглядит как текст

Автоматически созданное описание

Figure 3 Broadening Bottoms Pattern

BTOP (Broadening Tops)

The pattern usually makes a signal to sell an asset

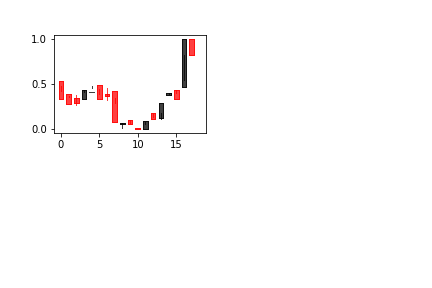


Figure Broadening Tops Pattern

1. **Model framework**

As this study revolves around predictive power of chart patterns and technical indicators, the choice was made to use CNN coupled with LSTM (CNN concentrating on pattern classification and LSTM predicting price with a set of technical indicators). This section follows all the details of the data preprocessing, data extraction, hyperparameters tuning and two used architectures which are LSTM and CNN.

1. **ANN (Artificial neural network)**

A neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates. In this sense, neural networks refer to systems of neurons, either organic or artificial in nature. Neural network structures have been used since 1990’s in financial markets due to its possibility to approximate complex functions.

1. **CNN (Convolutional neural network)**

CNN is a special architecture of artificial neural networks suggested by Yann LeCun in 1988 which by use of filter convolution may be able to extract underlying abstractions of any image (edges and features of an image) which are shared among images of one class. CNN connectivity between layers resembles structure of animal visual cortex. CNN was inspired by Neocognition which is based on two main elements downsampling (or Max-Pool layer) layer and convolution layer.

Convolutional layer

Layer that uses filter matrices which have some values which are applied to a raw image for feature extraction. Filter simultaneously with other filters train their values during training process in order to capture representative features of learned class. First convolutional layers usually detect edges of an image subsequent layers might detect a representative part of an image like an eye if CNN is trained for human face recognition.

Max-pooling layer

This layer is much simple and less computationally expensive. Max-pooling layer does a reduction of dimensionality of an input data by passing through input a square window (n x n) and taking the biggest value out of it. This layer’s main function is prevention of overfitting as some learned features from convolution maybe specific not to the whole class, but only to this one unique image.

Dropout

Regularization operation that is used in learning stage process. Dropout goes through a given network and with some probability get rid of neurons. This puts more weight on other neurons which in turn reduces chances of overfitting as it makes network more open to generalization over data.

Coordinate convolution layer

New type of convolutional layer proposed by Uber AI Labs in 2018 in their paper [5], An Intriguing Failing of Convolutional Neural Networks and the CoordConv Solution (\*reference\*). As convolutional layers are used as main building blocks almost in all state-of-the-art vision networks, it is expected to perform infallibly or with little error rate at any classification task. CNN’s advantage in image classification is that network is very accurate at classifying an object no matter the disposition of an object on a presented image, this is possible because of filters that can detect features and parts of an object all around the image. However, it is not the case for a lot of task, as it was found at the scientific study by Uber’s research center.

This new layer showed little advantage at simple classification of objects as they are usually not attached to certain points of an image. Although, it is expected crucial for network to show little generalization at trading pattern classification as it was mentioned above patterns form sequentially, thus extrema have to be still sequentially placed.

CoordConv makes it possible by adding extension of coordinates into separate channels to simple convolutional layer. This extension makes each layer possible to recognize their position in Cartesian coordinates space.

LeNet-5

An architecture that is one of the simplest and most famous structures in image classification proposed by Yann LeCun in 1998 [4]. This architecture is referenced and modified to produce a proper model for classifying trading patterns.

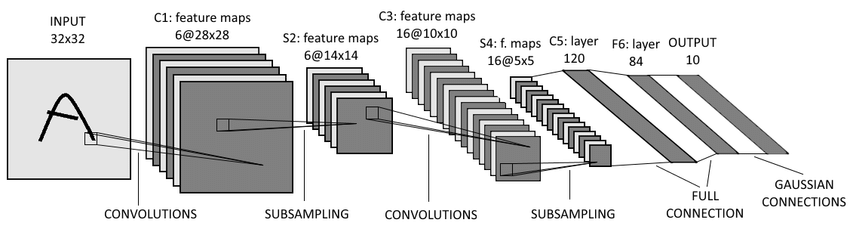
The Architecture of LeNet-5 consists of two convolutional layers, one max-pooling layer and two fully connected dense layers.

Figure 5 Architecture of LeNet-5 (Source: Yann Lecun, 1997, 7 page)

The presumed proper fit model for the given classification task has two CoordConv layers and one fully connected layer. In order to get the best performing model, hyperparameter tuning has to be carried out, the process is described later in the paper.

* 1. **Dataset**

Dataset for further analysis consisted from csv files with daily dynamics of opening, closing, high, low prices and volume. Scraped daily OHLC data of the most liquid assets in the S&P 500 from yahoo finance, the choice of liquidity is essential as less liquid market are more susceptible to big price fluctuations and less possible for robust technical analysis prediction.

Extracted financial data was cleaned from omitted information and time intervals when prices were not recorded in a form of candlestick.

Training set is split into 5 classes: 4 classes being trading patterns baring a trading signal and 5th class being “alien” class that bears no information for us.

* 1. **Pattern labeling and extraction (CNN)**

In order to correctly extract patterns (H&S, IH&S, BBOT, BTOP), kernel regression was used on a dataset of very liquid assets. A method that approximates non-linear function of a given dataset, denoising financial stochastic data in this case. Smoothed out function then used on a set of rules to find recognizable trading pattern.

For the right fit of a function a parameter is decided beforehand, is being bandwidth of each point placing more or less weight on farther price movements which would either overfit the function or underfit it. To find the best fit cross-validation technique was used with simple loss function representing how far the true price is far from kernel estimated price. The best function corresponds to the one with parameter h that has the lowest loss value.

“Alien” class extraction is done manually by observing no aforementioned pattern on an extracted image and with intent to capture random candlestick movements for the model not to mistakenly treat random pattern as one of the four chosen trading patterns.

Nadaraya–Watson kernel regression:

* 1. **Data augmentation**

The process of Fmainly picking patterns is heavily time consuming, but for more robust model larger data set is crucial. Model trained on small number of samples is susceptible to overfitting and leads to poor performance. Training set can be expanded by rotation, inversion and synthetic generation of time-series.

Extracted patterns do not have same length due to patterns in the market chart being of random different sizes which creates more difficulties for the size of training set because CNN has to be fed samples holding equal properties, same number of candlesticks or being of the same time interval. Although, an easy method could be used to generalize patterns of different size to a desired number of timestamps. Synthetic reduction and synthetic expanding of each sample until a sample is reduced or stretched to the needed time interval. Reduction may be proceeded in a following algorithm: finding subsequent candlesticks moving in the same direction (descending or ascending candlesticks) and combining their OHLC data together in proper manner (for ascending – combination of lowest and opening prices of first candle with highest and closing values of second candle, for descending – combination of highest and opening prices of first candle with lowest and closing values of second candle). Extension method could be carried out simply by splitting random candlestick in a randomly chosen value between opening and closing price into separate instances.

As H&S and IH&S patterns are symmetric data set of these classes can be successfully expanded by inversion of their candlestick sequence. However, for such patterns as BBOT and BTOP this is not a case as they do not hold this property.

To prevent class imbalance, another method is used which takes a training sample and slightly changes each candlestick of a sample by randomly picking distortion values from normal distribution.

* 1. **Data processing**

After manually picking suited patterns, an algorithm extracted OHLC data of each training sample for further processing.

* 1. **Time-series imaging**

A big problem arise with raw time-series classification because of time-series sequence linear dependence convolution filters would fail at extracting raw features without temporal dependence context. The study “Imaging Time-Series to Improve Classification and Imputation” [2] has shown that a NN better recognizes temporal dependency through transformation of time-series into Gramian Transformation Field or Markov Transformation Field.

Gramian Angular Field

Before encoding opening, closing, highest, lowest price into GAF, time-series have to be normalized as GAF could capture absolute values in its calculations. After that they have to be transformed into polar coordinates.

Markov Transition Field

First of all, data has to be properly scaled before building Markov transition matrix. Transition matrix is a (Q x Q) matrix which counts the probability of transitioning from one value of a given time-series to another value throughout all given series.

is given by the frequency with which a point in quantile is followed by a

point in quantile

Figure 6 MTF, GADF and GASF representation of closing time-series of H&S

**

Throughout all the process of input preparation each training sample pattern should produce 4 images (opening, closing, low, high time-series) for each field method, totaling 12 pre-processed images per each raw sample. Each image has a scale timestamp x timestamp. CNN takes as input three representations (MTF, GASF, GADF) of a sample pattern.

1. **RNN (Recurrent Neural Networks)**

A type of neural networks where connections between elements create direct sequence. This possibility allows a network to work with sequential data and have some memory of previous inputs helping network to have more context to work with. However, due to vanishing gradient problem this simple neural network may not be able to capture long-term relationships of sequential data, but only short-term.

1. **LSTM (Long short-term memory)**

A network of RNN type which can capture short-term as well as long-term relationships between sequenced data.

Recurrent Dropout

Recurrent dropout has similar effects as simple dropout on a network performance, though it is much more important to use it instead of simple dropout as recurrent network is based on recurrent connections which can also by overfitting hurt performance of a model.

Online learning

Online learning is a training method which updates weights of a network with each data point this actually benefits a network, makes it more flexible, but in return for more time training. Online learning requires model’s training batch size to be set to 1. In addition, model evaluation becomes much complex because of instability of output during one-batch learning.

* 1. **Data preprocessing and feature creation**

For LSTM model to be able to predict returns direction of the next day based on a set of technical indicators and past returns which have to be extracted for each timestamp of time-series.

A set of 22 technical indicators including MACD, RSI, Momentum, Stochastic Oscillator, 27 values of past returns ranging from previous day log returns to 80-week logarithmic returns (Appendix A) and volume value are used for binary prediction of next day closing price being greater than today’s closing price.

* 1. **Feature scaling**

Since features can widely deviate, features have to be scaled to a normal distribution. If features not scaled than they may be problem that certain features may have greater effect because of greater absolute value undermining effect of much smaller values that also contribute greatly to prediction.

1. **Neural Network Hyperparameters tuning**

Each neural network architecture has a number of parameters that have to be preliminary indicated before training. These parameters play main role in network succeeding in its task.

Hyperparameter is carried out by grid search where all presumed range of parameters form all possible architectures and the one with the least loss is chosen. To evaluate correctly the best fit grid search is coupled with cross-validation.

For LSTM: LSTM units in each layer, recurrent dropout value.

For CNN: number of convolution filters and size of filters in convolution layers, dropout value.

For both: batch size (online learning), number of epochs, loss function and optimizer algorithm.

CNN results of hyperparameter search: filter size (3,3) with padding, dropout value 0.2, 32 batch size, first conv layer 64 filters and second conv layer 64 layers.

LSTM results of hyperparameter search: first layer LSTM units 80, second layer LSTM units 30, batch size 1 (as we have a case of online learning) and recurrent dropout 0.2.

1. **Model Evaluation**
2. **LSTM**

Evaluation of LSTM model for stock price movement prediction is very unstable, due to stochastic nature of asset’s prices. Proper evaluation of a model involves cross-validation in sequence also known as forward chaining imitating a process of real time prediction where forecast is made on the past data for future values of a time-series. For this experiment 10 splits were chosen.

Baseline LSTM model’s features simply contain returns of opening, low, close, high prices and volume with five day lagged values for each feature trying to capture short-term return dependency on nearest previous dynamics.

Baseline LSTM model:

|  |  |  |
| --- | --- | --- |
| Forward-chaining split | Loss | Accuracy |
| 1 | 0.7418955564498901 | 55.147% |
| 2 | 0.7407013773918152 | 52.2059% |
| 3 | 0.7284354567527771 | 47.0588% |
| 4 | 0.6913175582885742 | 55.8824% |
| 5 | 0.6867634057998657 | 55.1471% |
| 6 | 0.706450343132019 | 51.4706% |
| 7 | 0.705305814743042 | 47.7941% |
| 8 | 0.6745472550392151 | 61.0294% |
| 9 | 0.7322924733161926 | 40.4412% |
| 10 | 0.741936981678009 | 52.2059% |
| **Average:** | **0.71496462225914** | **51.8382%** |

Results are positive on average, as predictive power of model is better than random guessing, although model is not reliable, because some testing windows show very poor performance of 40% prediction rate.

Baseline model + Technical indicators:

Second model is more complex as it also includes technical indicator signals (Appendix A).

|  |  |  |
| --- | --- | --- |
| Forward-chaining split | Loss | Accuracy |
| 1 | 0.6903282403945923 | 55.147% |
| 2 | 0.7137702703475952 | 50.3224% |
| 3 | 0.7246361374855042 | 47.7941% |
| 4 | 0.6713104844093323 | 56.6176% |
| 5 | 0.6789352297782898 | 59.5588% |
| 6 | 0.697299599647522 | 52.9412% |
| 7 | 0.7226393818855286 | 49.1177% |
| 8 | 0.6932458281517029 | 51.3235% |
| 9 | 0.7346256375312805 | 47.6471% |
| 10 | 0.7017077207565308 | 56.6176% |
| **Average:** | **0.7028498530387879** | **52.7088%** |

There is an overall performance increase for each test window. Overall accuracy tends to rarely predict worse than with 50% accuracy.

Baseline model + Technical indicators + Pattern recognition:

Third model includes CNN pattern prediction output where each pattern found correspond to 1 and 0 means no pattern found on a given timestamp.

|  |  |  |
| --- | --- | --- |
| Forward-chaining split | Loss | Accuracy |
| 1 | 0.7233170866966248 | 55.147% |
| 2 | 0.7105453014373779 | 49.2647% |
| 3 | 0.7175766730308533 | 47.7941% |
| 4 | 0.6831021308898926 | 59.5588% |
| 5 | 0.6833411455154419 | 58.08823% |
| 6 | 0.6870191097259521 | 52.9412% |
| 7 | 0.7003855514526367 | 49.2647% |
| 8 | 0.6752483248710632 | 63.2353% |
| 9 | 0.7176839113235474 | 48.5294% |
| 10 | 0.7119864821434021 | 58.08823% |
| **Average:** | **0.7010205717086792** | **54.1912%** |

Model shows little increase in performance which is positive and shows that patterns have predictive power in them, although increase in performance is small and signifies that predictions are not robust.

Possible fallbacks of LSTM model for returns prediction

One of the most probable fallbacks of proposed LSTM model is standard loss function which doesn’t take in to account strength of price movement and treat returns that move in a same direction as being equal, although incrementally there may be substantial difference in returns

1. **Pattern Signal Backtesting**

To put pattern classified figures to a real test, a trading strategy is proposed: recognized patterns are used as entry signals by convention (H&S – sell, IH&S – buy, BBOT – buy and BTOP – sell) with exit signal triggered by the use of a trailing stop.

Trailing stop moves only in one direction when a stock price gets to another price peak, it moves percentagewise only in favorable direction limiting losses and locking in additional profits.

Parameters for the strategy: starting percentage value of stop loss and cut-off prediction percent to get rid of patterns that are not quite resemble ours. 0.5% is a conservative value for stop-loss parameter which is taken from a closing price on a signaled day and 85% cut-off value for recognition of patterns as being decent.

For backtesting purposes dataset of high-frequency data of S&P 500 was used (only a subset from 2010-11-14 to 2018-12-31 of 1-minute data was found in open source [7]).

Results show 533.36% return over the period which beats simple strategy to buy and hold which amounts to only 209.93%. However, this only holds in perfect market condition without transaction fees which would completely take away all the profit margin.

1. **Conclusion**

Returns prediction is a very hard task and implementing simply past returns dynamics can only result in a performance that is not significantly better than random guessing. Neural networks are around for more than two decades which means that all abnormal profits from simple neural architectures are already taken by experienced traders in the major markets. Although, results are still in favorable range when considering pattern recognition on multiple timestamp intervals, consideration of real-world limitations as transaction fees undermine any profit possibility. There are still great deal of studying could be done in this area, for example, taking into consideration more patterns and covering more timestamp interval could boost the performance of the model in real-world conditions.

**Appendix A. Technical indicators**

1. Max\_last\_52\_weeks\_signal

Maximum сlose price for the previous 52 weeks signal (Tuesday-Tuesday).

1. Mean\_week\_50\_signal, Mean\_week\_200\_signal

Weekly mean of MA of 50 (200) days signal.

1. s4\_l36, s4\_l48, s8\_l36, s8\_l48, s12\_l36, s12\_l48

MA rule signal.

1. s4\_l36OBV, s4\_l48OBV, s8\_l36OBV, s8\_l48OBV, s12\_l36OBV, s12\_l48OBV

OBV rule signal.

.

1. s\_K

Stohastic Oscillator signal.

1. S\_m\_36, S\_m\_48

Momentum rule signal.

1. s.rsi.14, s.rsi.25

RSI signal.

1. *s.macd*

*MACD signal.*

*where -*

1. *s.adx*

*ADX signal.*

**References**

1. ANDREW W. LO, HARRY MAMAYSKY, AND JIANG WANG; Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation, 2000
2. Zhiguang Wang and Tim Oates; Imaging Time-Series to Improve Classification and Imputation, 2015
3. Zhichao Zou and Zihao Qu;Using LSTM in Stock prediction and Quantitative Trading, 2020
4. Yann LeCun, Leon Bottou, Yoshua Bengio, and Patrick Haffner; Gradient-Based Learning Applied to Document Recognition, 1998
5. Rosanne Liu, Joel Lehman, Piero Molino, Felipe Petroski Such, Eric Frank, Alex Sergeev, Jason Yosinski; An Intriguing Failing of Convolutional Neural Networks and the CoordConv Solution
6. M. P. Taylor and H. Allen, “The use of technical analysis in the foreign exchange market,” Journal of International Money and Finance, vol. 11, no. 3, pp. 304–314, 1992
7. FutureSharks, https://github.com/FutureSharks/financial-data