Journal on Big Data JBD, vol.1, no.3, pp.135-144, 2019

On Visualization Analysis of Stock Data

Yue Cai[1], Zeying Song[1], Guang Sun[1, *], Jing Wang[1], Ziyi Guo[1], Yi Zuo[1], Xiaoping Fan[1], Jianjun Zhang[2] and Lin Lang[1]

Abstract: Big data technology is changing with each passing day, generating massive amounts of data every day. These data have large capacity, many types, fast growth, and valuable features. The same is true for the stock investment market. The growth of the amount of stock data generated every day is difficult to predict. The price trend in the stock market is uncertain, and the valuable information hidden in the stock data is difficult to detect. For example, the price trend of stocks, profit trends, how to make a reasonable speculation on the price trend of stocks and profit trends is a major problem that needs to be solved at this stage. This article uses the Python language to visually analyze, calculate, and predict each stock. Realize the integration and calculation of stock data to help people find out the valuable information hidden in stocks. The method proposed in this paper has been tested and proved to be feasible. It can reasonably extract, analyze and calculate the stock data, and predict the stock price trend to a certain extent.

Keywords: Data visualization, stock data, data analysis.

1 Introduction

With the booming of big data technology, more and more data is waiting for us to identify, read, classify, and calculate. In view of the stock market, how to use big data technology to explore the many valuable information hidden in stocks and help shareholders to buy stocks reasonably and get the maximum profit is a difficult problem. At present, big data research in China's stock field is still not perfect.

The financial industry has been using advanced mathematics and statistics for some time. Before the 1980s, the banking and finance industries were considered "boring"; investment banks were separate from commercial banks, and the main task of the industry was to deal with "simple" (compared to today) financial functions, such as loan. The Reagan administration's reduced regulation and the application of mathematics have transformed the industry from a boring banking industry to what it is today. After that, finance became a science and became a force for the advancement and development of mathematics. For example, a major advance in mathematics is the derivation of the Black-Scholes formula. It is used for stock pricing (a contract that gives stockholders the right to buy and sell from stock issuers at a certain price). However, bad statistical models, including the Black-Scholes model, bear some of the nicknames that led to the 2008 financial crisis.

[1] Hunan University of Finance and Economics, Changsha, 410205, China.
[2] Hunan Normal University, Changsha, 410081, China.
* Corresponding Author: Guang Sun. Email: simon5115@163.com.

In recent years, computer science has joined the ranks of higher mathematics, bringing revolutionary changes to financial and securities trading (buy and sell of financial products for profit). Today's transactions are mainly done by computers: algorithms can make trading decisions at speeds that are difficult for humans to achieve (see the limitations of light speed as a bottleneck in system design). Machine learning and data mining are also being used more and more widely in the financial sector. Visually, this momentum will continue. In fact, a large part of the algorithmic transactions are high frequency transactions (HFT). Although the algorithm is faster than manual, these techniques are still very new and are used in an area that is known for its instability and high risk.

This paper investigates simple strategy analysis based on Python tools to obtain stock data. Including moving averages, how to use moving averages to develop trading strategies, how to make decisions to enter and exit the stock market, and how to use backtesting to evaluate a decision.

2 Getting and visualizing stock market data

2.1 Getting data from Yahoo Finance

The data must be obtained before analyzing the data. Stock market data can be obtained from Yahoo!, Finance, Google Finance and other places. At the same time, the pandas package provides a way to easily get data from the above websites. This article uses Yahoo Financial data. The following figure is the part of data:

	High	Low	...	Volume	Adj Close
Date			...		
2018-12-31	159.360001	156.479996	...	35003500.0	156.463837
2019-01-02	158.850006	154.229996	...	37039700.0	156.642365
2019-01-03	145.720001	142.000000	...	91244100.0	141.039642
2019-01-04	148.550003	143.800003	...	58607100.0	147.060516
2019-01-07	148.830002	145.899994	...	54777800.0	146.733185

where Open is the starting price of the day (not the closing price of the previous day); high is the highest price of the stock on the day; low is the lowest price of the stock on the day; close is the stock price of the closing time. Volume refers to the number of transactions. Adjust close is the closing price after adjustment for corporate behavior. Although the stock price is basically determined by the trader, stock splits (the stock split refers to the behavior of the listed company to split the existing stock into two, the new stock price is half of the original stock) and dividends (dividend. dividends per share) will also affect stock prices.

2.2 Stock market data visualization analysis

Let us consider visualizing it after getting the data. For example, the chart below reflects the stock price changes of Apple's 2018-01 to 2019-05:

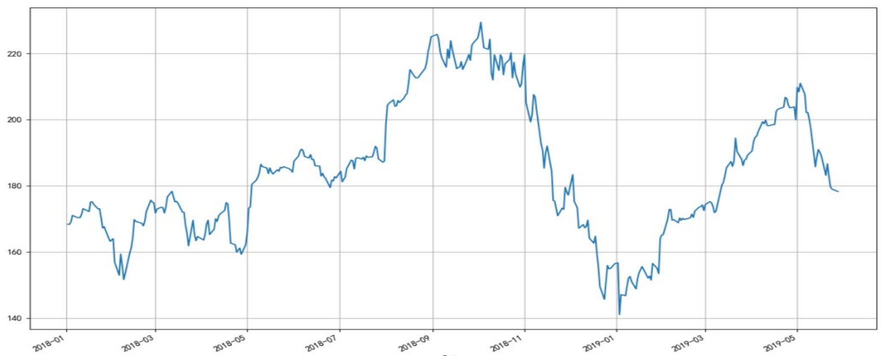

Figure 1: Apple's stock price changes from 2018-01 to 2019-05

The line graph is feasible, but the data for each day has at least four variables (opening, stock price, stock price and closing), and we hope to find a way to see the four without having to draw four different lines. A visualization method for the trend of variables. In general, we use the candlestick chart (also known as the Japanese candlestick chart) to visualize financial data. The candlestick chart was first used by Japanese rice merchants in the 18th century.

Figure 2: Changes in the stock price of Apple's 2018-01 to 2019-05

The black line in the candlestick chart represents the closing price of the trading day is higher than the opening price (profit), and the red line represents the opening price of the trading day is higher than the closing price (loss). The tick marks the highest and lowest prices of the day's trade (the hatch is used to indicate which side of the body is open and which is closed). Candlestick charts are widely used in financial and technical analysis in trading decisions, using the shape, color and position of the body.

We may want to present different financial products on a map: So we can compare different stocks, compare stocks to the market, or look at other securities, such as exchange-traded funds (ETFs). Below I show the data of stocks of different technology companies and how to adjust the data to get the data lines together.

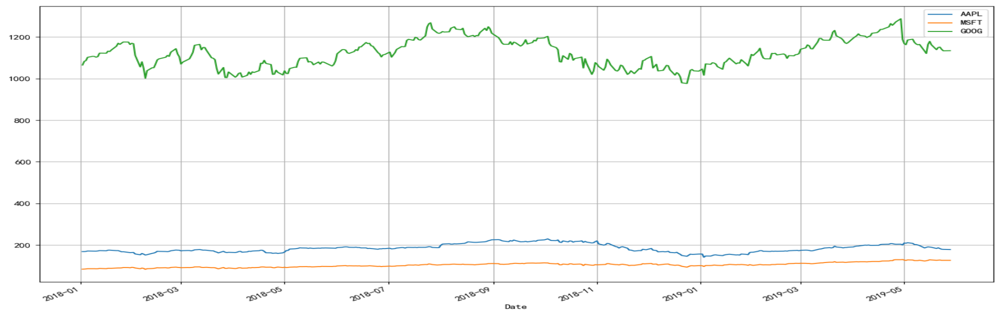

Figure 3: Apple, Microsoft, Google 2018-01 to 2019-05 stock value fluctuations

But where is the problem with this chart? Although the absolute value of the price is very important (expensive stocks are difficult to buy, this will not only affect their volatility, but also affect the difficulty of trading them), but in the transaction, we pay more attention to the price of each stock. Change instead of its price. Google's stock price is higher than Apple's. This difference makes Apple and Microsoft's stocks seem to be very volatile, and that's not the case.

One solution is to plot with two different scales. One scale is used for data from Apple and Microsoft; another scale is used to represent Google's data.

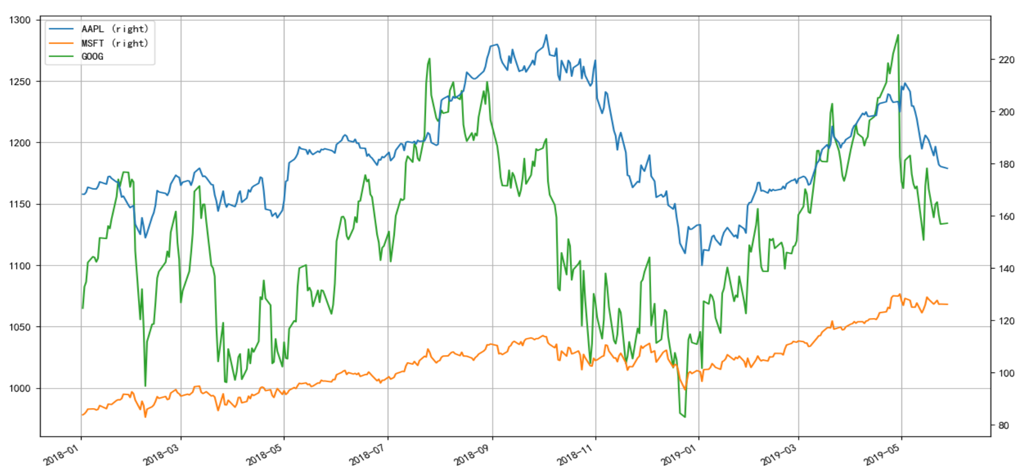

Figure 4: Apple, Microsoft, Google 2018-01 to 2019-05 stock value fluctuations chart (based on Google's scale)

This requires us to make the necessary data conversions. There are many ways to convert data. One of the conversion methods is to compare the stocks of each trading day with the stock prices starting from the time period we care about. That is:

$$return_{t,0} = \frac{price_t}{price_0} \tag{1}$$

This requires converting the data in the stock object as follows:

stock_return=stocks.apply(lambda x: x / x[0])

stock_return.head()

Result:

	AAPL	MSFT	GOOG
Date			
2018-12-31	1.000000	1.000000	1.000000
2019-01-02	1.001141	0.995570	1.009888
2019-01-03	0.901420	0.958945	0.981122
2019-01-04	0.939901	1.003544	1.033893
2019-01-07	0.937809	1.004824	1.031653

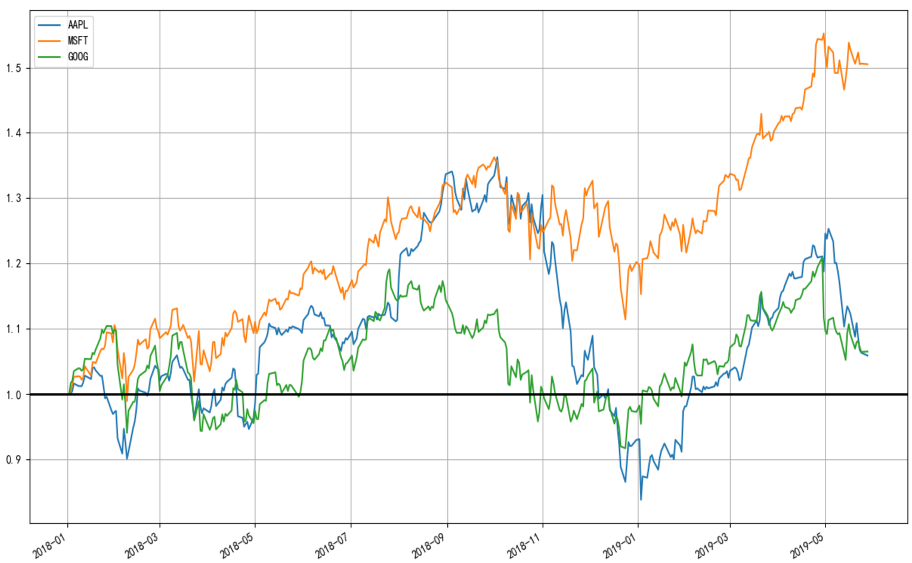

Figure 5: Apple, Microsoft, Google 2018-01 to 2019-05 stock returns

This picture is much more useful. Now we can see how high the yield of each stock is from the date we care about. And we can see that the correlation between these stocks is very high. They basically move in the same direction, which is difficult to observe in other types of charts.

We can also use the daily stock value changes to map. One possible method is the ratio of the stock value of $t+1$ and the t of the day after the day we use the share price of the day:

$$growth_t = \frac{price_{t+1} - price_t}{price_t} \tag{2}$$

We can also compare the price of the day before:

$$increase_t = \frac{price_t - price_{t-1}}{price_t} \tag{3}$$

The above formulas are not the same and may lead us to different conclusions, but we can use logarithmic differences to represent stock price changes:

$$change_t = \log(price_t) - \log(price_{t-1}) \tag{4}$$

The following code demonstrates how to calculate and visualize the logarithmic difference of stocks:

stock_change=stocks.apply(lambda x: np.log(x)-np.log(x.shift(1)))#shift moves dates back by 1.

stock_change.head()

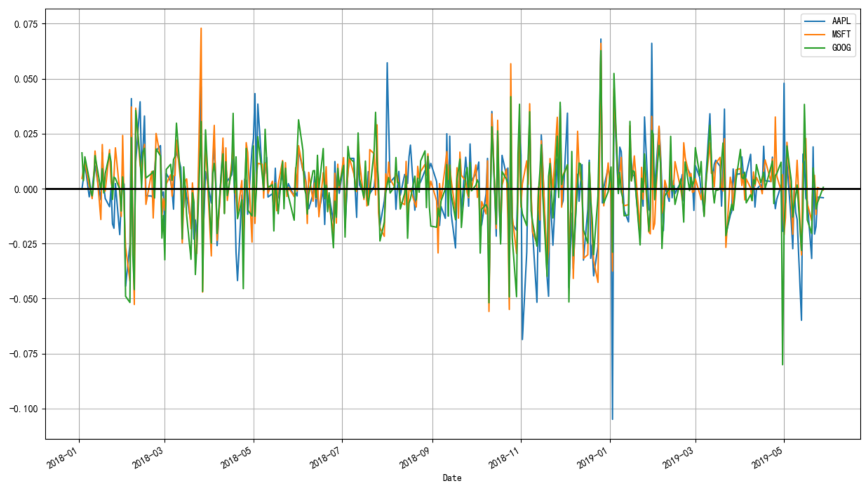

Figure 6: Apple, Microsoft, Google 2018-01 to 2019-05 stock value fluctuations

The difference in returns from the beginning of the relative time period can clearly see the overall trend of different securities. The gap between different trading days is used in more ways to predict stock market prices, and they cannot be ignored.

3 Analysis of stock market trends by moving average of stock values

3.1 Moving average in the stock market

The EXPMA (Exponential Moving Average) translation index smoothing moving average is the moving average, which is developed in response to the lack of the moving average as a backward indicator. To solve the problem, once the price has been removed from the moving average, the average cannot react immediately, moving average can reduce similar shortcomings.

In general, many investors use the KDJ indicator and the MACD indicator as important indicators for buying and selling. When the KDJ indicator and the MACD indicator of the market or individual stocks form a dead fork at a high level, they usually sell. However, because the market's main force often performs reverse operations, it often leads to "top-top" and "bottom-bottom" situations, so this indicator often fails.

According to the above phenomenon, many investors use the moving average as the main basis for buying and selling stocks, meaning that when the monthly indicators form a long position, they usually buy. Conversely, when the monthly indicators form a short position, they usually sell. However, as the market's main force often conducts anti-technical operations, it deliberately causes the stock price to break through and break through the above-mentioned moving averages. This will cause many people's chips to be dismissed.

In summary, in order to overcome the above shortcomings, this paper introduces the moving average index. The main advantages of this index are: the length of the moving average is complemented, and the KDJ and MACD indicators are "golden fork" and "dead". "Fork" and other functions. Therefore, the indicator has a high success rate and accuracy, which provides a good point for the bottom-selling and escape of individual stocks, and is a good helper for investors to adopt short-term and medium-term decision-making.

3.2 Stock market analysis based on moving average

Let's take a look at how to find the trend of stock prices.

A moving average of q days (used MA_t^q to represent) is defined as the average of the previous q days for a certain point in time t.

$$MA_t^q = \frac{1}{q}\sum_{i=0}^{q-1} x_{t-i} \tag{5}$$

Moving averages can make a series of data smoother and help us find trends. The larger the q value, the less sensitive the moving average is to short-term fluctuations. The basic purpose of moving averages is to identify trends from noise. Fast moving averages have a smaller q, which is closer to the stock price; slow moving averages have larger q values, which makes them less sensitive to fluctuations and thus more stable.

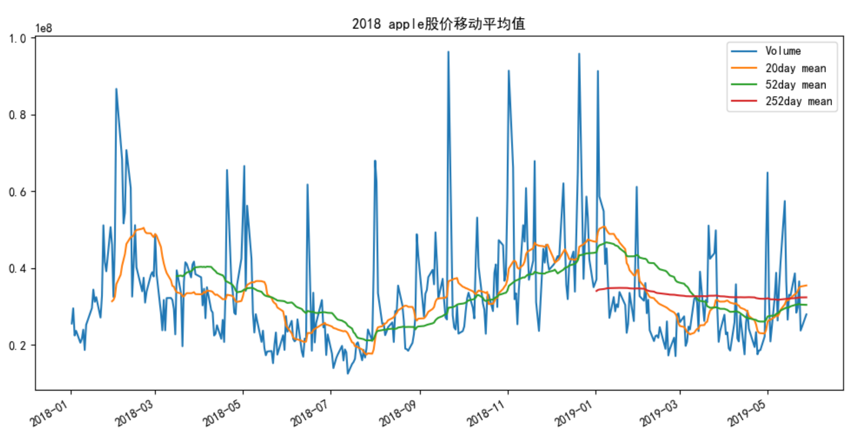

Figure 7: Apple stock price moving average change

Note that the starting point time of the average is very late. We have to wait until 20 days before we can start calculating this value. This problem is a more serious problem for moving averages over a longer period of time. Because I want I can calculate the 200-day moving average, I will expand the data we get for Apple stock.

The 20-day moving average is very sensitive to small changes, while the 252-day moving average has the least fluctuations. The 252-day moving average here shows the overall trend: the stock value remains generally stable. The information represented by the 20-day moving average is a bear market bull market alternation, followed by a bull market. The intersection of these averages is the trading information point, they represent a change in the trend of the stock price and you need to make a corresponding decision that can make a profit.

4 Conclusion

In the background of big data era, there have been preliminary applications in stock price forecasting, but how to improve forecasting accuracy is still a problem to be explored. The application research of stock forecasting in the financial field is still in the exploratory stage. Most of the applications are mainly using the powerful learning ability of deep learning, but the excessive dependence on experiments for parameter adjustment leads to the lack of generalization ability of the model, although the improved algorithm can make The parameters are obtained through self-learning, but as the complexity of the construction of the model structure of the algorithm increases, the selection of various parameters is still not guided by a clear theory. At the same time, how to make economical explanations for the post-experiment data is also lacking. It is often a good result but it is impossible to draw an accurate theory.

This paper proposes a stock profit analysis design scheme, which extracts, stores, and calculates the stock data, and obtains the market value information of the stock. Through the analysis of stock market value, the stock's development trend is reasonably predicted. In the era of big data, the use of distributed parallel computing frameworks and storage frameworks to solve practical problems has become an inevitable trend. In recent years, stock market research has been increasing, analyzing stock information, and forecasting has become a hot research direction. Stock analysis and forecasting can help investors to extract key information in stocks and accurately grasp the dynamics of the stock market in order to obtain higher profits and avoid stock traps. There are many factors that influence the stock price trend and the profit. We cannot take all the factors into consideration in the platform. Therefore, it is very difficult to achieve accurate forecasting of stock prices. Even if we can analyze the profit of each stock, there are still errors, but the analysis of the overall trend of the stock is still very helpful.

Acknowledgement: This research work is supported by Hunan Provincial Education Science 13th Five Year Plan (Grant No. XJK016BXX001), Social Science Foundation of Hunan Province (Grant No. 17YBA049), Hunan Provincial Natural Science Foundation of China (Grant No. 2017JJ2016). The work is also supported by Open foundation for University Innovation Platform from Hunan Province, China (Grant No. 18K103) and the 2011 Collaborative Innovation Center of Big Data for Financial and Economical Asset

Development and Utility in Universities of Hunan Province. We also thank the anonymous reviewers for their valuable comments and insightful suggestions.

References

Chen, M. (2015): Visualization technology research based on complex data processing. *Computer and Digital Engineering*, vol. 43, no. 11, pp. 194-197.

Chen, Y.; Lin, X. L. (2016): An associated hierarchical data visualization method based on heat map and radiation loop. *Journal of Computer-Aided Design and Graphics*, vol. 32, no. 7, pp. 12-15.

Dai, G. Z.; Chen, W.; Hang, W. X. (2013): Information visualization and visual analysis: challenges and opportunities. *Chinese Science*, vol. 43, no. 1, pp. 178-184.

Du, X. L. (2004): Overview of big data research. *Computer Application Research*, vol. 31, no. 6, pp. 1612-1616.

He, X. P.; Huang, L. (2015): The visual analysis of the evolution path of big data field, research hotspots and frontiers. *Modern Intelligence*, vol. 35, no. 4, pp. 46-51.

Ping, Q. (2013): The major trend of data operation is mainstreaming. *Internet Weekly*, vol. 13, no. 7, pp. 26-46.

Qi, C. H.; Gao, S. P. (2014): Hospital information data mining and data visualization. *China Science and Technology Information*, vol. 2014, no. 22, pp. 39-42.

Sun, G.; Li, F. H.; Jiang, W. D. (2019): Brief talk about big data graph analysis and visualization. *Journal on Big Data*, vol. 1, no. 1, pp. 25-38.

Tu, X. L. (2004): Overview of big data research. *Computer Application Research*, vol. 31, no. 6, pp. 1612-1616.

Wang, R. S. (2016): Research on spatial and temporal multidimensional data visualization in big data environment. *Computer Education*, vol. 19, no. 5, pp. 94-97.

Wu, C. R.; Zapevalova, E.; Chen, Y. W.; Li, F. (2018): Time optimization of multiple knowledge transfers in the big data environment. *Computers, Materials & Continua*, vol. 54, no. 3, pp. 269-285.

Wu, J. X.; Wang, J. H. (2010): Visual perception model based on situational awareness theory. *Modern Library and Information Technology*, no. Z, pp. 9-14.

Wu, Y. B.; Guo, F. L. (2015): Big data visualization analysis. *Computer Education*, vol. 12, no. 5, pp. 94-97.

Xiao, B.; Wang, Z.; Liu, Q.; Liu, X. D. (2018): An improved mini batch k-means algorithm based on mapreduce with big data. *Computers, Materials & Continua*, vol. 56, no. 3, pp. 365-379.

Xue, Q.; Lei, X. L. (2014): Review of big data systems and analysis techniques. *Journal of Software*, vol. 25, no. 5, pp. 1889-1908.

Xie, R. (2014): The beauty of big data visualization. *Internet Weekly*, vol. 15, no. 6, pp. 32-34.

Yang, Y. B.; Liu, B. (2014): Review of research on information visualization. *Journal of Hebei University of Science and Technology*, pp. 18-23.

Yuan, S. Y. (2014): Vision, brain and image-the role and influence of neural mechanism in the process of painting. *Research on Art Education*, no. 7, pp. 56-57.

Zen, Y. (2014): Overview of big data research. *Computer Application Research*, vol. 11, no. 6, pp. 162-1616.

Zhang, H. J.; Zhang, X. L.; Wang, Y. P. (2018): Research on complex cognitive activities in visual analysis. *Journal of Computer Aided Design and Graphics*, vol. 30, no. 1, pp. 57-67.

Zhou, S.; Liang, W. L.; Li, J. G. (2018): Improved VGG model for road traffic sign recognition. *Computers, Materials & Continua*, vol. 57, no. 1, pp. 11-24.