Stock Market Prediction using Candlestick Chart

Razik Batliwala[1], Muddassir Khan[2], Durvesh Bhushan[3]

[1,2]*Information Technology, PadmabhushanVasantdada Patil Pratishthan's College of Engineering, Sion Mumbai*
[3]Vaity Information Technology PadmabhushanVasantda*da Patil Pratishthan's College of Engineering,*
Sion Mumbai

Abstract—The stock market is a place where shares of publicly listed companies are traded. There shares based on brought and sold on this basis of there stock. The price of stocks and assets are an important part of the economy. There are many factors that affect share prices. However there is no specific cause for the prices to rise or fall. This makes investment subject to various risks. We proposed a novel method for prediction of stocks price on the basis of statistical data and making of candlestick chart for up to 3-4 days to find if investment are beneficial or loss of money, It is also beneficial to analyzing in future stock. In this paper we developing (LCS) Longest Common Subsequence Algorithm to retrieve numerical sequences that partially match. It also use for real time service provider to provide stock market sentiments.

Keywords—Stock price prediction; Technical analysis; Candlestick charts; Longest common subsequence algorithm for numbers; Multi numerical attributes; Nse,Bse stock average.

1. INTRODUCTION

Stock market prediction techniques play a crucial role to bring more people into market and encourage markets as a whole. Fundamental analysis and technical analysis are two popular approaches to successful stock trading.

One of the important types of technical analysis is candlestick chart patterns. The candlestick chart patterns provide short-term predictions for traders to make buy or sell decisions. While most of techniques use statistics of stock prices, the candlestick charting technique focuses on patterns among several days of candlesticks formulated by opening, high, low, and closing prices within a specific time frame, such as minute, hour, day or week. Dozens of candlestick chart patterns are identified to be signals of bullish/bearish reversals and continuations. These patterns consist of a single candlestick or a combination of multiple candlesticks.

It is also pointed out that candlestick chart pattern recognition is subjective. The candlestick chart patterns are often qualitatively described using words and illustrations. The studies adopt definitions using a series of inequalities with different parameters that specify candlestick patterns. Numerical definitions of candlestick patterns are still controversial issues.

In addition, they don't occur in time series in a strict sense because stock price fluctuation continues after intervals of several days depending on announcements of important economic indicators, economic and political news, etc. Because of these characteristics, the candlestick chart patterns are deemed to bring controversial results on predictability regarding future market trends even sort-term prediction.

2. BACKGROUND

CANDLESTICK CHART AND PATTERNS

This section introduces the formation of a candlestick. Candlestick patterns are a combination of one or more candlesticks. Samples of well-known candlestick chart patterns are shown. Because the candlestick patterns are described in natural language and illustrations, there are criticisms on their use for trend prediction by a computer.

A. Formation of Candlestick

A daily candlestick line is formed with the market's opening, high, low, and closing prices of a specific trading day. Figure 1 represents the image of a typical candlestick. The candlestick has a wide part, which is called the "real body" representing the range between the opening and closing prices of that day's trading.

If the closing price is above the opening price, then a white candlestick with black border is drawn to represent a bullish candlestick. If the opening price is above the closing price, then a filled candlestick is drawn. Normally, black color is used for filling the candle to represent a bearish candlestick. The thin lines above and below the body represent the high/low ranges. These lines and are called "shadows" and also referred to as "wicks" and "tails." The high is marked by the top of

the upper shadow and the low by the bottom of the lower shadow.

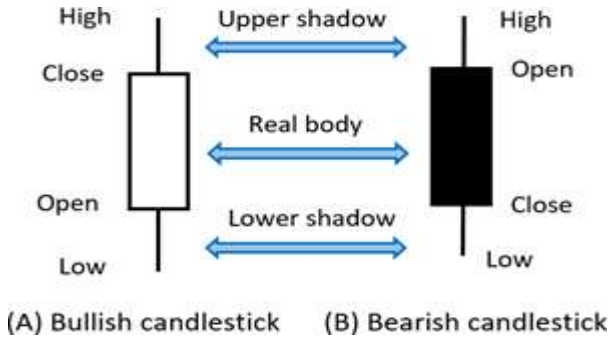


Figure 1. Candlestick formation

B. Samples of Candlestick Patterns

Dozens of candlestick patterns are identified and become popular among stock traders [2][3]. These patterns have colorful names like morning star, evening star, three white soldiers, and three black crows.

Figure 2 shows the morning star pattern which is considered as a major reversal signal when it appears in a low price zone or at a bottom. It consists of three candles, i.e., one short-bodied candle (black or white) between a preceding long black candle and a succeeding long white one. The pattern shows that the selling pressure that was there the day before is now subsiding. The third white candle overlaps with the body of the black candle showing a start of a bullish reversal. The larger the white and black candle, and the higher the white candle moves, the larger the potential reversal. The opposite version of the morning star pattern is known as the evening star pattern which is a reversal signal when it appears in a high price zone or at the end of an uptrend.

Figure 2. Morning star pattern

Figure 3 shows the three white soldiers pattern which is interpreted as a strong indication of a bullish market reversal when it appears in a low price zone. It consists of three long white candles that close progressively higher on each subsequent trading day. Each candle opens higher than the previous opening price and closes near the high price of the day, showing a steady advance of buying pressure. The opposite of the three white soldiers pattern is known as the three black crows pattern which is interpreted as a bearish signal of market trend.

Figure 3. Three white soldiers pattern

C. Criticism of Candlestick Patterns

The major criticism of the candlestick chart patterns is that the patterns are qualitatively described with words, such as "long/short candlesticks," "higher/lower trading," "strong/weak signal," supported by some illustrations [2]. What percentage of price change does "long/short" mean? Without modeling the candlestick patterns in a way that a computer can process and performing experiments comprehensively, arguments on the effectiveness of chart patterns would not come to an end.

Since it is highly possible that the existence and predictability of candlestick patterns depends on stock markets, this study focuses on the Nikkei stock average (Nikkei 225) as the first stage of study. This paper proposes a model for retrieving similar candlestick charts based on a data mining algorithm using dynamic programing technique to handle candlestick patterns including several intervals that suggest unpredictable price trends.

D. LCS Algorithm

LCS algorithm:

Let the input sequences be X[1 … m] of length m and Y[1 … n] of length n. Let D[i, j] denote the length of the longest common subsequence of X[i] and Y[j] for $0 \leqq i \leqq m$ and $0 \leqq j \leqq n$.

A) If either sequence or both sequences are empty, then the LCS is empty, i.e., D[i, 0] = 0 and D[0, j] = 0.

B) If X[i] and Y[j] match (X[i] = Y[j]), then the LCS is become longer than the previous sequences by one, i.e., D[i, j] = D[i−1, j−1] + 1.

C) If X[i] and Y[j] do not match (X[i] ≠Y[j]), then the LCS is the maximum of the previous sequences, i.e., max(D[i−1, j], D[i, j−1]).

The value of D[m, n] is the LCS of the sequences X[1 … m] and Y[1 … n]. The actual LCS sequence can be extracted by following the matrix D[i, j].

nLCS algorithm: The nLCS algorithm is derived from the LCS algorithm by replacing the match condition (X[i] =

Y[j]) with ((X[i] − Y[j]) ≦ diff) where diff is a tolerance

given by a user.

C. nLCSm: LCS for Subsequences with Multi Numerical Attributes

The idea of deriving the nLCS from the LCS can be further extend to the multi numerical attributes to obtain the nLCS for subsequences with multi numerical attributes (nLCSm).

nLCSm algorithm: Let p ($1 \leqq p$) denote the number of numerical attributes. Let Cq ($1 \leqq q \leqq p$) denote the match conditions for the qth numerical attribute. The nLCSm is derived by replacing the match condition of the nLCS, i.e., ((X[i] − Y[j]) ≦ diff), with (C1∧...∧Cq∧...∧Cp).

D. nLCSm and candlestick pattern retrieval

Given the candlestick pattern model with six parameters as depicted in Figure 4, the nLCSm algorithm can be applied to implementing the model by assigning match conditions C1 to C6 for each candlestick as follows.

C1: if a difference between closing price change of a given candlestick and that of a candidate candlestick is within the change tolerance (change_tol), then C1 is true.

C2: if a difference between body length of a given candlestick and that of a candidate candlestick is within the body tolerance (body_tol), then C2 is true.

C3: if a difference between a closing price and a 5-day moving average is within the tolerance (av5diff_tol), then C3 is true.

C4: if a difference between a closing price and a 25-day moving average is within the tolerance (av25diff_tol), then C4 is true.

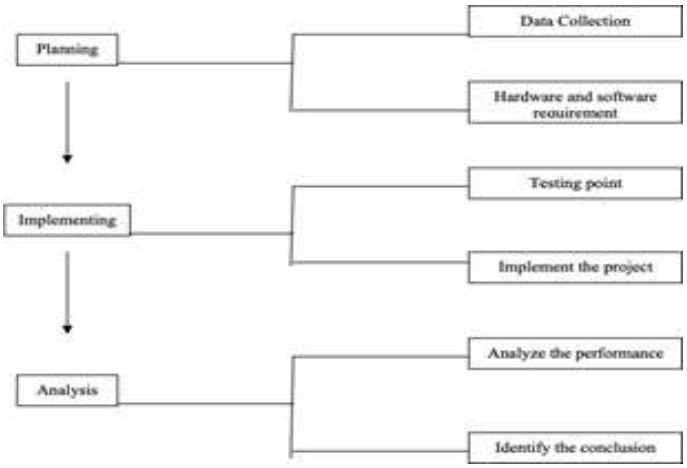
C5: if a slope of a 5-day moving average is within the given tolerance (slope5_tol), then C5 is true C6: if a slope of a 25- day moving average is within the given tolerance (slope25_tol), then C6 is true.

The 5-day moving average is calculated by the latest five days' closing prices. Because these prices are just a sample of larger population of closing prices, the sample standard deviation or Bessel's correction [2] is adopted as a measure of threshold to decide whether a given 5-day moving average is within an expected distribution.

The tolerance of 5-day moving average av5diff_tol is statistically dependent on the change tolerance change_tol. In the proposed retrieval model, av5diff_tol and av25diff_tol are calculated by the following formulas as defaults according to the definition of the sample standard deviation.

av5diff_tol = change_tol / SQRT(4)= change_tol / 2 (1) av25diff_tol = change_tol / SQRT(24)= change_tol / 4.899 (2) Thus, there are essentially four independent parameters in the proposed model, which still causes difficulties in setting parameters. Assuming that each parameter has 5 ranges of values representing, for instance, very high, high, the same level, low, and very low. The candlestick patterns of one candlestick have 5 to the power 4, i.e., 5^4= 625 cases of parameters. The patterns composed of two candlesticks have 5^(4*2)= 625*625= 390,625 cases. The patterns of tree candlesticks have 244,140,625 cases. These cases mean very wide varieties of candlestick charts leading difficulties even in setting parameters for retrieving a specific candlestick chart pattern.

3. METHODOLOGY



Planing

To identify all the information and requirement such as hardware and software, planning must be done in the proper manner. The planning phase have two main elements namely data collection and the requirements of hardware and software

Implementing

After fetching data, python will generate candle stick chart from which each candle will describe different pattern from that pattern prediction part is implemented

Analysis

After prediction it is check with the live data for analysis, for accuracy purpose.

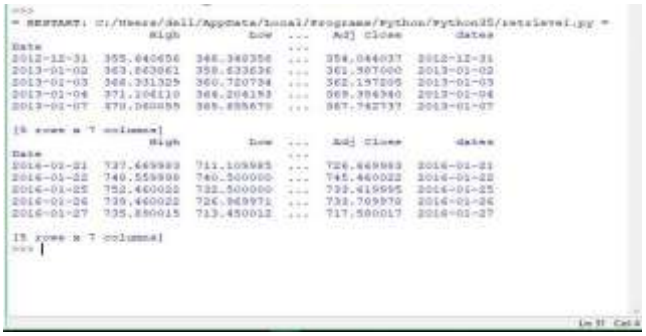


Fig. Data fetching

Above figure shows the fetching of historical data such as open, high, close and low of a particular company.

Fig. CandleStick Chart

The above diagram shows the candle stick chart which is plotted with the help of historical data.

Fig. CandleStick Chart with Moving Average

Above diagram has candle stick with moving average line. Green color denotes moving average line.

4. RELATED WORK

As the existing system only predict the stock price of any company on the basis of stock high, low, close and open price, which leads to an accuracy rate of 35-40%. Previous system predicts future price on the basis of company balance sheet, sentiment, dividend pay, dividend yield, etc.

One drawback in this system is in can only give future growth and potential of the stock in the company.

To overcome this, we use candle stick chart. In candle stick chart candles are generated with the low high close and open price. This data can form several types of candle which can be used to predict further stock price. Each and every candle have their own feature, such as doii candle, hammer candle, morning star, evening star, falling star, etc.

a) Other studies conclude that applying certain candlestick patterns is profitable at least for short-term trading [8]-[11]. Chootong and Sornil [8] propose a trading strategy combining price movement patterns, candlestick chart patterns, and trading indicators. A neural network is employed to determine buy and sell signals. Experimental

b) results using stock data in the Stock Exchange of Thailand show that the proposed strategy generally outperforms the use of traditional trading methods based on indicators. Zhu, Atri, and Yegen [9] examine the effectiveness of five different candlestick reversal patterns in predicting short- term stock movements using two Chinese stock data. The results of statistical analysis suggest that the patterns perform well in predicting price trend reversals.

c) Lu, Chen, and Hsu [10] apply candlestick trading strategies to the U.S. market data with several trend definitions. They find three-day reversal patterns are profitable when the transaction cost is set at 0.5%.

d) One of the obstacles of candlestick charting is the highly subjective nature of candlestick pattern [2] since the candlestick patterns are defined using words and illustrations. Tsai and Quan [11] propose an image processing technique to analyze the similarities of the candlestick charts for stock prediction instead of using numerical inequality formulas. The experimental results using the Dow Jones Industrial Average (DJIA) show that visual extraction of contents and similarity matching of candlestick charts are suitable for predicting stock movements.

5. CONCLUSION AND FUTURE WORK

This system proposes a model for retrieving candlestick charts. A numerical sequence version of the Longest Common Substring (LCS) algorithm is devised to implement the proposed model. This system can be helpful for a accurate short term prediction of stock price, which leads to a correct company share price.

The system will be able to predict the future stock price from current stock price candlestick chart strategy can be use to predict, as this strategy is most accurate withaccuracy rate of 80-85%. It will predict time duration for investment.

6. FURTHER RESARCH

As part of this project it is planned to conduct further research into the impact of unusually high volume of trading and the direction of the stock price.

7. REFERENCES

[1] V. Drakopoulou, "A Review of Fundamental and Technical Stock Analysis Techniques," Journal of Stock & Forex Trading vol. 5, pp. 1–8, Nov. 2015.

[2] R. Yamini Nivetha and Dr. C. Dhaya "Developing a Prediction Model for Stock Analysis" 2017,

[3] C. F. Tasi and Y. C. Hsiao, "Combining multiple feature selection methods for stock Prediction: Union, intersection, and multi- intersection approaches," Decision Support Systems, vol. 50, no. 1, pp. 258-269, 2010.

[4] C.-F. Tsai and Z.-Y. Quan, "Stock Prediction by Searching for Similarities in Candlestick Charts," Journal ACM Transactions on Management Information Systems (TMIS), vol. 5, Article No. 9, July 2014.

[5] W. Lan and G. Zhao, "Stocks network of coal and power sectors in china stock markets," Information Computing and Applications, pp. 201–208, 2017.

[6] W.-Q. Huang, X.-T. Zhuang, and S. Yao, "A network analysis of the Chinese stock market," Physica A: Statistical Mechanics and its Applications, vol. 388, no. 14, pp. 2956– 2964