Stock Price Prediction Model using Candlestick Pattern Feature

Leslie C.O. Tiong, David C.L. Ngo, and Yunli Lee

Abstract—Since late 1980s, stock prediction has become a challenging task in stock market due to the uncertainty movement of stock price. In our study, we utilized candlestick pattern to test the following two AI techniques - Artificial Neural Network (ANN) and Support Vector Machine (SVM) separately in learning the various patterns and behaviours of stock price for stock prediction. The experimental results of Dell Inc stock prices and currency exchange EUR – USD demonstrated significant low rates of Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). This research tells that our proposed features extracted from candlestick pattern could offers a reliable prediction for next day stock price.

Index Terms—Stock Prediction, Candlestick Pattern, Artificial Neural Network, Support Vector Machine.

1 INTRODUCTION

In general, it is an interesting topic to determine when to buy, sell or hold that stock in financial circle. It is not an easy task due to its non-stationary data, thus, many trials and experiments have been conducted with Artificial Intelligence (AI) techniques in the area of computer science. The non-stationary characteristic is caused by uncertainty of information from the historical data in the stock market. Therefore, in order for investments to yield a significant profit, predicting the stock price has become an important and challenging task.

In the early stage of a financial circle, some researchers have begun to propose and implement technical indicators analysis methods such as Exponential Moving Average (EMA), Moving Average Convergence Divergence (MACD) and candlestick pattern for predicting the stock price [1], [2], [3]. Due to the expressed growth of computer technology, researchers from the area of computer science have applied AI techniques like Artificial Neural Network (ANN), Hidden Markov Model (HMM), Support Vector Machine (SVM) and Expert System (ES) in their experiments of stock prediction. According to their experiments’ results, they proved that the AI techniques have a better approach in learning the patterns and behaviours of stock price for stock prediction [4], [5], [6], [7].

Despite the fact that AI techniques can be beneficial to a certain degree, it also has its own limitations which require improved. For ANN and HMM, both of these algorithms act as black box in the training process. Therefore, it is unable to be explained the training process and the complexity of computation for each parameters. Thereby designing a model will be very complex and complicated, which is considered as their downside. SVM has the potential and ability to learn and optimise the behaviour, but its limitation is defining problem representation. Meanwhile, ES has a better ability in optimising and verifying with raw data but the limitation of the technique lacks of learning ability.

Through understanding the limitation of AI techniques, this study aims to use the best characteristics of AI techniques - ANN and SVM to overcome their limitation for predicting stock price. The objective of this research is to implement ANN and SVM separately by learning the candlestick patterns as proposed features for developing and improving the accuracy of stock prediction.

In our experiments, we utilised the candlestick pattern to represent the behaviour and pattern of stock data as proposed features, and then used the AI techniques - ANN and SVM as classifiers to learn and train with the patterns for predicting the stock price. The findings of the experiments will predict the candlestick position of the stock price and prematurely minimise the Mean Absolute Error (MAE) and Root Mean Absolute Error (RMAE) of the model in resulting the position of candlestick pattern.

The remaining sections in this article are categorised as following: Section 2 will brief the basic idea and concept of the stock data analysis methods (candlestick pattern and OHLC chart) and AI techniques (ANN and SVM). In section 3, structure and design of the experiments will be discussed. The result of our experiments will show in section 4. Section 5 is the conclusion and future work.

2 THE BASIC IDEA AND CONCEPT OF STOCK DATA ANALYSIS METHODS AND AI TECHNIQUES

2.1 Stock Data Analysis Methods

Candlestick pattern is one of the oldest types of chart which have been used for stock prediction in financial...
circle. It visualises the stock price pattern and provides the signal of continuation and inversion about stock trend [8], [9].

Fig. 1 shows the examples of candlestick pattern.

1. The body of candlestick is called real body which represents the price range between close and open prices.
2. The vertical line above and bottom of the real body are called upper shadow and lower shadow which represent the highest and the lowest price of the period.
3. The “black” real body illustrates that the open price is higher than close price which shows the stock trend is decreasing, and when close price is higher than open price shows the “white” real body representing that the stock trend is increasing.

In OHLC chart, it has a technique to classify the position of the OHLC bar into four types: up, down, inside and outside [11]. Fig. 3 shows the ideas to identify the classification of the OHLC chart.

2.2 AI Techniques

ANN is a field of computational science with various types of methods which tries to solve real world problems by offering strong solutions [12]. Generally, ANN is a multi-layer network which includes input layer, hidden layer and output layer shown in Fig. 4. It mirrors the basic characteristic of a human brain, and has the ability of human to learn and generate its own knowledge from its surroundings [13].
From our experiment, and experiments, SVM is also reliable features in the learning stage. As a result, those studies has the ability for recognising patterns with the selected features. According to the experiment which was conducted by Cao and Tay [22], they proved that SVM has the ability for recognising patterns with the selected features in the learning stage. As a result, those studies and experiments, SVM is also reliable with selected features in learning patterns.

### 3 The Structure and Design of Experiments

#### 3.1 Data Analysis

From our experiment, the stock price was studied, namely *Dell Inc* and currency exchange as *EUR – USD*. Those data were downloaded from Yahoo Finance [23] and MarketWatch [24].

The datasets of those companies available have the attributes shown in Table 1.

#### Table 1. Stock Data Attributes

<table>
<thead>
<tr>
<th>Date</th>
<th>Volume</th>
<th>Open</th>
<th>Close</th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>12/1/2011</td>
<td>1,000,312</td>
<td>23.39</td>
<td>23.45</td>
<td>23.49</td>
<td>23.34</td>
</tr>
</tbody>
</table>

In the data analysis part, we utilised the basic idea of candlestick pattern which have been introduced in Section II and create a few extra features: real body (RB), upper shadow (US), lower shadow (LS), real body colour (Colour) and the position of candlestick (CP).

Logically, candlestick pattern can be formalised in numeric and nominal values that should be expressed in relative unit. Real body (RB) can be converted by using the following equation:

\[
RB = [|C - O| / O] \times 100\% \tag{1}
\]

where \(C\) - close price and \(O\) - open price.

The upper shadow (US) and lower shadow (LS) of candlestick pattern can be represented with numeric values. When the close price is greater than open price in the candlestick pattern, the US and LS values can be converted as following equation:

\[
US = [(C - O) / (H - O)] \times 100\% \tag{2}
\]

and

\[
LS = [(C - O) / (C - L)] \times 100\% \tag{3}
\]

where \(C\) - close price, \(O\) - open price, \(H\) - high price and \(L\) - low price in candlestick pattern.

Unfortunately, when the open price is greater than close price in the candlestick pattern, the US and LS values can be converted into following equation:

\[
US = [(O - C) / (H - C)] \times 100\% \tag{4}
\]

and

\[
LS = [(O - C) / (O - L)] \times 100\% \tag{5}
\]

where \(C\) - close price, \(O\) - open price, \(H\) - high price and \(L\) - low price in candlestick pattern.

Above equations were adopted and modified from Stock Market Forecast [25].

To determine the colour of the real body in candlestick pattern, we used the basic concept which introduced in Section 2.

For defining the position of real body in the candlestick, we implemented the concept and trend of OHLC chart to determine the position of real body as the following:

\[\text{Fig. 4. Illustration of ANN Structure. This illustration shows the basic concept of ANN.}\]
3.2 Training and Testing Environment

In this stage, we utilised the WEKA open source library to create our own application and develop the classifiers for predicting the stock. We chose Multi-Layer Perceptron [26] as ANN classifier and Sequential Minimal Optimization [27]as SVM classifier with its default settings from WEKA. Both classifiers are used for learning the candlestick pattern based on the proposed features which were analysed from data analysis part.

Based on the proposed features, we utilised RB, US, LS, colour of RB and candlestick position to represent one day candlestick pattern. Then, we used these features for one year duration as input parameters for ANN and SVM classifiers to learn the patterns.

Fig. 6 shows the general structure design of our experiments. The experiments are based on one year data in 2011, we divided into 70% for training and 30% for testing. The duration of dataset started from January to December in 2011. We also used 100 new dataset for model validation. The duration of new dataset started from January to June in 2012. In the training dataset, we had generated data for 5 days, 11 days, 1 day, 12 hours in 2 hours interval, and 24 hours in 2 hours interval. Tables 2 and 3 showed the few experiments that used different candlestick patterns as input parameters to test and validate the model.

**TABLE 2. DETAIL OF EXPERIMENT FOR USING DATA BASED ON DIFFERENT DAYS**

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Number of Days</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5 days of Candlestick Pattern</td>
<td>Dell Inc and EUR-USD</td>
</tr>
<tr>
<td>2</td>
<td>11 Days of Candlestick Pattern</td>
<td>Dell Inc and EUR-USD</td>
</tr>
<tr>
<td>3</td>
<td>5 days of Candlestick Pattern</td>
<td>Dell Inc and Currency Exchange EUR-USD</td>
</tr>
<tr>
<td>4</td>
<td>1 day of Candlestick Pattern</td>
<td>Dell Inc and Currency Exchange EUR-USD</td>
</tr>
</tbody>
</table>

**TABLE 3. DETAIL OF EXPERIMENT FOR USING DATA BASED ON DIFFERENT HOURS**

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Number of Days</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>12 hours of data with 2 hours interval of the RB’s pattern</td>
<td>EUR-USD</td>
</tr>
<tr>
<td>6</td>
<td>24 hours of data with 2 hours interval of the RB’s pattern</td>
<td>EUR-USD</td>
</tr>
</tbody>
</table>

**Fig. 5. Rules for Identifying the Real Body Position. It determines the position of real body in Candlestick Chart.**

IF (“BLACK” AND “WHITE”)

IF (close_price>open_price_prev AND open_price== close_price_prev )

Then RB Position = “UP”;

IF (open_price<close_price_prev AND close_price== close_price_prev )

Then RB Position = “DOWN”;

IF (close_price>open_price_prev AND open_price== close_price_prev )

Then RB Position = “OUTSIDE”;

IF (close_price== open_price_prev AND open_price== close_price_prev )

Then RB Position = “INSIDE”;

IF (“BLACK” AND “BLACK”)

IF (open_price>open_price_prev AND close_price== close_price_prev )

Then RB Position = “UP”;

IF (close_price<close_price_prev AND open_price== open_price_prev )

Then RB Position = “DOWN”;

IF (open_price>open_price_prev AND close_price== close_price_prev )

Then RB Position = “OUTSIDE”;

IF (open_price== open_price_prev AND close_price== close_price_prev )

Then RB Position = “INSIDE”;

IF (“WHITE” AND “WHITE”)

IF (close_price>close_price_prev AND open_price== open_price_prev )

Then RB Position = “UP”;

IF (open_price<open_price_prev AND close_price== close_price_prev )

Then RB Position = “DOWN”;

IF (close_price>open_price_prev AND open_price== open_price_prev )

Then RB Position = “OUTSIDE”;

IF (close_price== close_price_prev AND open_price== open_price_prev )

Then RB Position = “INSIDE”;

Fig. 5: Rules for Identifying the Real Body Position. It determines the position of real body in Candlestick Chart.
In the experiments, we utilised the concept of candlestick patterns for determining the RB position. Furthermore, we also tested the OHLC chart which was introduced in Section II for determining the position of candlestick instead of using the position of RB.

In the first experiment shown in Table 2, we used 5 days of candlestick pattern as input parameters to train and predict the RB position for the next day. As for the second experiment in Table 2, we utilised 11 days of candlestick pattern as input parameters to train and predict the following day of RB's position.

In the third and fourth experiments, we used the original concept of OHLC chart. We made some changes in the proposed features by providing extra attribute values to determine that candlestick position as whether it was a ‘large up’, ‘small up’, ‘large down’, ‘small down’, ‘large inside’, ‘small inside’, ‘large outside’ or ‘small outside’. The third experiment used 5 days of candlestick patterns to train and predict the candlestick position on the next day. In the fourth experiment, it used 1 day of candlestick pattern to train and predict the candlestick position for the next day.

From the fifth experiment, we have used 12 hours of data with 2 hours interval of the RB's pattern as input parameters to train the model. Once the classifiers have trained with the patterns, they will predict the next coming hour of RB position based on the previous information of the candlestick patterns. In the sixth experiment, we utilised 24 hours of data with 2 hours interval of the RB's pattern from candlestick to train and predict the next hour of position.

**4 EXPERIMENT RESULTS**

This study emphasises in using candlestick pattern to produce the investment signal of buying and selling. Predicting the candlestick position is more reliable because it provides supported information for investment strategy.

For the result of our experiments displayed in Tables 4 and 5, we used MAE and RMAE to evaluate the performance of our model that we have implemented. In Table 4, we could see that the MAE and RMSE of first and third experiments tested with AI techniques - ANN and SVM as classifiers gave us lower rates compare to others by using different amount of daily data. In the experiments, we found that 5 days candlestick pattern which gave us more sufficient results in testing. The proposed feature was able to avoid redundant and irrelevant for ANN and SVM in learning the candlestick patterns.

Furthermore, we also used hourly data with the same concept to represent the candlestick patterns. According to Table 5, the sixth experiment gave us the most acceptable result of lower rates for MAE and RMSE. As can be seen, hourly data to represent candlestick pattern also improve the ANN and SVM in learning the candlestick patterns.

For the model validation, we have tested 100 new dataset for Dell Inc and EUR – USD. As can be seen, the first, third and sixth experiments could classify correctly the position of candlestick patterns with 60% and above using daily and hourly data.

**5 CONCLUSION AND FUTURE WORK**

In this study, we presented to use the candlestick pattern features to train with ANN and SVM separately for predicting the stock.

The evaluation results showed in Tables 6 proved that the proposed features of candlestick pattern we chose can be used as input parameters for ANN and SVM algorithms to learn and identify the behaviour and pattern of candlestick for stock prediction with lower rates in MAE and RMSE. In addition to this study, the proposed features in data analysis part are very important due to the non-stationary raw data from the stock market for AI techniques to learn and train the pattern.

In the near future, we will improve the proposed features with different type of format to train with other AI techniques such as HMM, which is quite popular in the field of machine learning researches for the stock prediction.

© 2013 International Journal of Interactive Digital Media | 62


<table>
<thead>
<tr>
<th>Experiment</th>
<th>Classifiers</th>
<th>Dell Inc MAE</th>
<th>Dell Inc RMSE</th>
<th>EUR-USD MAE</th>
<th>EUR-USD RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ANN</td>
<td>0.0591</td>
<td>0.2154</td>
<td>0.0385</td>
<td>0.1708</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>0.0721</td>
<td>0.2185</td>
<td>0.0641</td>
<td>0.1671</td>
</tr>
<tr>
<td>2</td>
<td>ANN</td>
<td>0.0615</td>
<td>0.2224</td>
<td>0.0396</td>
<td>0.1704</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>0.0792</td>
<td>0.2342</td>
<td>0.0572</td>
<td>0.1938</td>
</tr>
<tr>
<td>3</td>
<td>ANN</td>
<td>0.0615</td>
<td>0.1856</td>
<td>0.0396</td>
<td>0.1704</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>0.0911</td>
<td>0.2349</td>
<td>0.0396</td>
<td>0.1704</td>
</tr>
<tr>
<td>4</td>
<td>ANN</td>
<td>0.0938</td>
<td>0.2482</td>
<td>0.0615</td>
<td>0.1960</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>0.0918</td>
<td>0.2606</td>
<td>0.0721</td>
<td>0.1806</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Classifiers</th>
<th>EUR-USD MAE</th>
<th>EUR-USD RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>ANN</td>
<td>0.0113</td>
<td>0.0943</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>0.0175</td>
<td>0.0898</td>
</tr>
<tr>
<td>6</td>
<td>ANN</td>
<td>0.0111</td>
<td>0.0921</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>0.0173</td>
<td>0.0909</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Classifiers</th>
<th>Number of Tested Data</th>
<th>Dell Inc Correct</th>
<th>Dell Inc Incorrect</th>
<th>EUR-USD Correct</th>
<th>EUR-USD Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ANN</td>
<td>100</td>
<td>60%</td>
<td>40%</td>
<td>61%</td>
<td>39%</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td></td>
<td>61%</td>
<td>39%</td>
<td>60%</td>
<td>40%</td>
</tr>
<tr>
<td>2</td>
<td>ANN</td>
<td>100</td>
<td>40%</td>
<td>60%</td>
<td>40%</td>
<td>60%</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td></td>
<td>41%</td>
<td>53%</td>
<td>40%</td>
<td>60%</td>
</tr>
<tr>
<td>3</td>
<td>ANN</td>
<td>100</td>
<td>63%</td>
<td>37%</td>
<td>61%</td>
<td>39%</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td></td>
<td>61%</td>
<td>39%</td>
<td>60%</td>
<td>40%</td>
</tr>
<tr>
<td>4</td>
<td>ANN</td>
<td>100</td>
<td>50%</td>
<td>50%</td>
<td>49%</td>
<td>51%</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td></td>
<td>49%</td>
<td>51%</td>
<td>48%</td>
<td>52%</td>
</tr>
<tr>
<td>5</td>
<td>ANN</td>
<td>100</td>
<td>59%</td>
<td>41%</td>
<td>56%</td>
<td>44%</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td></td>
<td>58%</td>
<td>42%</td>
<td>56%</td>
<td>44%</td>
</tr>
<tr>
<td>6</td>
<td>ANN</td>
<td>100</td>
<td>62%</td>
<td>38%</td>
<td>61%</td>
<td>39%</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td></td>
<td>64%</td>
<td>39%</td>
<td>60%</td>
<td>40%</td>
</tr>
</tbody>
</table>

REFERENCES


