An Application of Decision Tree for Stock Trading Rules: A Case of the Stock Exchange of Thailand

Suchira Chaigusin*

To increase the amount of money and overcome the interest rates and the inflation rates, stock investors attempt to use various approaches for their trading strategies. Technical analysis is one of popular tools that help develop trading approaches. This study is to apply the concept of the J48 decision tree with a chosen stock of a bank company listed in the Stock Exchange of Thailand. By creating the decision tree with technical indicators using the historical data, the tree can generate rules for a trading strategy. These rules support the decisions making in buying and selling the stock, and are considered being better than ad hoc strategies.

Name of the track: Finance

Keywords: Decision Tree, Stock trading rules, the Stock Exchange of Thailand, Technical indicators

1. Introduction

Investment plays a vital role for each person in every daily life. Traditionally safe way to grow up money is to deposit the money in the bank. Saving money in the bank, however, will not make high profits and overcome the inflation rates in the long term. Therefore, changing from depositing money in the bank to own good stocks is a kind of smart ways growing money in the long run. When considering investing in stocks, there are two important things to concern: the first is selecting good stocks and the last is when to buy or sell those stocks. In terms of selecting the good stocks to invest, there are some considerations such as choosing industries that the investors have a good understanding, selecting good companies by investigating in the companies’ information and compare them to find the companies that have a great margin of safety. Beside the strategies in selecting the companies to invest, suitable times to invest are also important. The times for buying and selling stocks may consider as buying and selling signals. However, there are no concrete rules for investing in stocks; one might consider using only the selecting stocks of the companies with the high margin of safety, which would be called using fundamental analysis; one might use only buying and selling signals, which would be called technical analysis, in his/her investment; one might use both fundamental and technical analyses in his/her investment.

In terms of fundamental analysis, the ways to analyze the fundamental of a company could be top-down or bottom-up approaches or both (Thailand Securities Institute (TSI) 2014). For top-down approach, the economic analysis such as economic cycle should be firstly considered and the industry life cycle is then analyzed, and finally the analysis is done by deeply investigating into companies’ information (Thailand Securities Institute (TSI) 2014). For the bottom-up approach, the steps for analysis are from examination into companies and then analysis the industry and the economic respectively.

*Dr.Suchira Chaigusin, Faculty of Business Administration, Rajamangala University of Technology Phra Nakorn, Thailand. Email: suchira.c@rmutp.ac.th
For the technical analysis, this analysis is based on concept of prices of stocks are reflected all known factors related the stocks at all times (Fong and Tai, 2009). Moreover, the philosophy of technical analysis also leans on using the past data to predict the future, which is grounded from inductive statistics that infers generalizations, predictions, or extrapolations from that data (Murphy, 1999). In the case that if someone critic the grounds of prediction in technical analysis, they should seriously doubt the forecasting based on historical data such as weather forecasting or economic forecasting.

Many approaches use to trade stocks, some might emphasize in fundamental analysis, some might focus in technical analysis, and however, some of them might overlap of both fundamental and technical analyses, since the main goal of investment is to grow up the money. A legendary investor, Warren Buffett, who becomes aspiration of many investors and his investment strategy seem to be a magnet for investors to find value stocks which are related to his strategy. Beside that there are some of technical investors who focus in use the technical analysis to invest such as Jesse Livermore, Richard Donchian, and Edward Arthur Seykota who are well-known for making great amount of money from their investment trading systems. As economic circumstances keep changing, the investment strategies should be timely dynamic, the investors are continuing to find a new strategic approaches for their investment.

2. Technical Analysis and Decision Tree

Some technical analysts may start with candlestick chart, candlestick patterns, line chart, chart patterns, or the Dow Theory. However, some of them maybe firstly do a smoothing data, which are prices, to see trends. There are kinds of smoothing prices such as simple moving average (SMA), weighted moving average (WMA), and exponential moving average (EMA). The SMA is simply calculated from the average of the close prices of the previous n days when n is a specific number of days. The WMA is calculated by given a weight to each close price, in this manner, generally, the weight are ranked from two continuums (high to low) according to the most recent days to the less recent days. The EMA calculation concept is similar WMA. It is computed by deduct the current close price by the previous EMA then multiply the result with the multiplier, after that the result for the multiplication is added with the previous EMA. The multiplier is normally two divided by the number of days for the selected time period plus one. There might be questions of how many days in a time period should be used or which one is the best moving average should be used. There is no concrete rule for using these. It depend on stocks or the price patterns of stocks and normally people are using the one that can help them making investment profit. Besides smoothing the data price for line charts, there are some popular technical indicators such as, Moving Average Convergence Divergence (MACD), Relative Strength Index (RSI), Average Directional Index (ADX), Stochastic oscillator, volume, On-balance volume.

In the area of financial intelligence, many techniques have been applied for stock trading strategies, such as artificial neural network, support vector machine (SVM), genetic algorithm and decision tree. In addition, the technical indicators are widely used especially in system trading. The indicators can also help create trading strategies that can be implemented to a computer programs to be used as robot traders, or trading systems. Fong and Tai (2005) use the Trend-following (TF) strategies for evaluation trends using static and adaptive rules on the Hang Seng future indices. Their model could provide positive trading profits even when the stock market was in a declining state. In addition,

For investment strategy, rules for making decision on buying and selling are significant for making profits. Besides using decision tree to making decisions for credit risks, customer relationships, medical decisions, it can use for construction of trading rules for the investment strategies. Sorensen, Miller and Ooi (2000) also use a decision tree approach to select stocks. Their result was a tree structure that provides the probability of stocks to classify as a outperformance or underperformance stock. In addition, Chencen (2011) developed an investment model by using decision tree with ID3 algorithm to provide decision support for future investment in ShenZhen stock market. This decision support included the analyses of company’s financial situations and stock market situations. There are a number of researches in this area focusing on many stock markets, however, the number of studies are very limited on an emerging stock market like the Stock Exchange of Thailand (SET). This study applies decision tree with J48 algorithm to construct decision rules for investment strategy and focuses on the SET.

3. The Methodology

3.1 Technical indicators

Technical indicators are widely used in stock trading. There are many technical indicators, however, in the use of these indicators, there are no rigorous rules for apply these indicators to trading strategies. Stock investors will longer imply their selected indicators as far as they have gained profit. The selected indicators used in this study are the EMA, RSI and Stochastic Oscillator.

The Exponential Moving Average (EMA) is a widely use moving average that assigns high weight to the recent situations as the StockCharts (2013a) suggests the calculation of the EMA as following:

\[
EMA(t) = EMA(t-1) + \beta \times (c_{t,price} - EMA(t-1))
\]

\[
\text{where } \beta = \left(\frac{2}{n+1}\right), n = \text{a number of days for the selected time period}
\]

In order to reflect the short term to middle term of situations, the selected time periods in this study are five days and 25 days. Therefore, in this case n would be 5 and 25.

The Relative Strength Index (RSI) is a useful indicator showing the recent gain compared to the recent loss according to a selected time period. Its value ranges from 0 to 100. According to the StockCharts (2013b), a 14 day period is suggested and the calculation of the RSI is the following.

\[
RSI = 100 - \left(\frac{100}{RS + 1}\right)
\]

When \(RS = \frac{\text{Average Gain of 14 days}}{\text{Average Loss of 14 days}}\), in the case that the average loss of 14 days is 0 then RSI is 100.
The Stochastic oscillator uses to indicate the position of the current price or close price relative to the difference of highest price and lowest price in a selected time period. The StockCharts (2013c) provides the following computation of %K and %D,

\[
\%K = 100 \times \left( \frac{cl\_price - lst\_price}{hst\_price - lst\_price} \right)
\]

\[
\%D = \frac{1}{3} \sum_{i=1}^{3} (\%K_i)
\]

Where cl_price is the lowest price at a selected time period and hst_price is the highest price at a selected time period.

### 3.2 Decision tree

Decision trees are popular tool to make decision and generate rules understood by human being. For classifying, predicting and making decision, decision tree can be efficiently used. The decision trees use to group data with similar attributes together and also separate data with different attributes to different classes. The decision tree is formed in a hierarchy tree structure, comprising nodes and branches. Nodes can be considered into three kinds of nodes, root node, internal node or decision node and leaf or terminal node. Root node is the first decision that should be made. The branches are the possible choices or the outcome from testing criteria at the node, which can split to at least two branches from node. The root node has no incoming branches but outgoing braches. The internal nodes have both incoming and outgoing branches. The leaf nodes, having incoming branches but no outgoing branches, are the final solutions or values. Decision trees can be composed of sub-tree. The basic concept of testing attribute of data at the root or internal nodes is for splitting action to choices which can go along the branches to other nodes until get to the leaf node which is the conclusion or solution. Data use for decision tree need to explain in term of fix collection of attributes, the target outputs are predefined with discrete values. In addition, the data use to train, validate and test the decision trees must be sufficient.

In order to make a criterion to select attributes to be nodes of tree to split data, there are functions or measurements can be used. Weka, a free tool for decision tree, uses the information theory, which has the gain criterion in the algorithms called ID3 and C4.5, to construct the tree. The Gain criterion calculates from information theory and entropy (Kijssirikul, 2003). The entropy uses to measure the disorder or the mix up of the values or attributes of a given set of data. Entropy can be considered as the average number of bits that can encode the values or attributes of the training data (Kowalczyk, 2009). If we have a set with n different values in it, we can calculate the entropy as follows:

\[
I(M) = -\sum_{i} P(m_i) \log_2 P(m_i)
\]

Where P(m_i) is the probability of getting m_i from a data set, \{m_1, m_2, m_3, ..., m_n\}. When moving down to a node of decision tree, information gain is used to determine the change of entropy (Kowalczyk, 2009). This way is for measuring the average number of bits saved, when consider encoding Y, the information gain can calculate from reduce entropy of X by entropy to encode Y. ID3 is an algorithm uses the information gain to construct a tree. However, the ID3 seems to have a bias of perforation of the attribute with many possible values. Additionally, in personal practical experiments of using ID3 for generate the rules of trading; the decision trees from ID3 are sometimes very huge. To deal with the bias, the Gain Ratio Criteria is used, which calculated from the Gain divide by the value of Split Information. This decision tree is called C4.5. A simple C4.5 decision tree is called J48 decision tree, implemented in Weka.
It widely uses in many data mining works and this algorithm can be freely used by Weka, which is a free software tool for data mining.

### 3.3 Data preparation

The data used in this study are a selected banking company listed in the Stock Exchange of Thailand. The historical data from 3/1/2007 to 29/11/2013, which is about 1690 data sets. In order to customize trading rules to fit with a specific strategy of trading, the historical candlestick chart has been reviewed and the dates that decided of buying and selling have been marked. These specific dates for buying or selling signals are about two or three consecutive days. The buying and selling become values in a target attribute. Only the dates that should buy or sell are used in this study, which use short term trading strategy. The results of this showed the profit more than the buy-and-hold strategy about 240%. The buying and selling data are divided into two data sets: training data set (124 instances) and testing data set (53 instances). This study uses eight selected attributes as shows in Table1.

#### Table1: All attributes

<table>
<thead>
<tr>
<th>Target Attribute</th>
<th>Buy or Sell</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input Attributes</strong></td>
<td></td>
</tr>
<tr>
<td>X1: b/s</td>
<td>Buy or Sell</td>
</tr>
<tr>
<td>X2: close_t&gt;closetna1</td>
<td>Close price(t) is more than close price(t-1)</td>
</tr>
<tr>
<td>X3: close &gt;= ema5</td>
<td>Close price is equal or more than 5-day exponential moving average</td>
</tr>
<tr>
<td>X4: m_ema5na25 pos/neg</td>
<td>The difference of 5-day and 25-day exponential moving averages change in positive, negative, or n/a</td>
</tr>
<tr>
<td>X5: rsi14&gt;q1</td>
<td>RSI14 is more than the 14-day simple moving average of RSI, called q1</td>
</tr>
<tr>
<td>X6: rsi14&gt;q1pos/neg</td>
<td>Crossing between RSI14 and q1 be positive (RSI14 crossing up), negative (RSI14 crossing down), or n/a</td>
</tr>
<tr>
<td>X7: K&gt;D</td>
<td>%K more than %D</td>
</tr>
<tr>
<td>X8: KxD pos/neg</td>
<td>Crossing between %K and %D be positive (%K crossing up), negative (%K crossing down) or n/a</td>
</tr>
</tbody>
</table>

### 4. The Results and Findings

The results from the J48 decision tree with using 10-fold cross-validation show correctly classified instances of 91.13 % and the confusion matrix as below shows the raw numbers of instances that are correctly classified and incorrectly classified which are 113 and 11 respectively.

#### Table 2: Confusion Matrix

<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
<th>&lt;- classified as</th>
</tr>
</thead>
<tbody>
<tr>
<td>61</td>
<td>6</td>
<td>a = buy</td>
</tr>
<tr>
<td>5</td>
<td>52</td>
<td>b = sell</td>
</tr>
</tbody>
</table>

By using 7 attributes excluded the target attribute, only 4 attributes are used to create the decision tree as shown in figure 1. The stochastic oscillator which is the attribute X7 and X8 and the difference between 5-day and 25-day exponential moving averages are not used to construct the decision tree.
The 6 rules generated from the decision tree are the following
1. If RSI 14 crosses up q1 then buy
2. If RSI14 crosses down q1 then sell
3. If RSI14 does not cross q1 and RSI 14 is more than q1 then sell
4. When the crossing between RSI14 and q1 is not applicable, and RSI 14 is not more than q1, and close price is less than 5-day EMA then buy
5. When the crossing between RSI14 and q1 is not applicable, and RSI 14 is not more than q1, and close price is equal or more than 5-day EMA, and close price is increased then buy
6. When the crossing between RSI14 and q1 is not applicable, and RSI 14 is not more than q1, and close price is equal or more than 5-day EMA, and close price is not increased then sell

When testing with the testing data set, the results showed the correctly classified instances of 90.57%. The false signals resulted from Rule 3 and 4. Moreover, the consideration for using Rule 3 is that they do not move long a trend. To be able to make profits when waiting for selling signal, others rules or techniques should be taken to substitute. In the use of Rule 4, when waiting for buying signal with an uptrend, it might generate a false buying signal since the price may be close to the turning point and it may cause unprofitable or even loss. Moreover, in the trading strategy, the saying, which is considered to be the rule of thumb, is that “cut loss and let profit run”. The Rule 2 seems to work well for the cutting loss. However, it depends on traders who use mix rules to their profited trading strategies.

In the using of these rules, key considerations taken into account are firstly, the rules are made upon the strategy of trading in a short term, secondly, none of waiting signal is used in this study, so it might make more profitable by using these rules with other technical indicators or techniques, lastly, these rules are developed with a selected stock which may have a specific range of swinging prices which is able to make trading profit.

To generalize the rules of trading, the rules should be tested with many stocks and in many different ranges of times. Moreover, each indicator may customize differently for fitting different stocks. The number of days for RSI calculation in this study is 14 days, but for other stocks may depends on the characteristic of the stocks.
5. Conclusions

This study supports the work of Fong and Tai (2005) in the use of Trend-following (TF) strategies on the Hang Seng future indices, which resulted in positive trading profits, and supports the study of Kantavat and Kijsirikul (2008) in using SVM with technical indicators in their trading experiment which showed trading profitability. The J48 decision tree constructed with historical trading signals can generate 6 trading rules. The trading rules showed high correct rate for classified instances in testing data. In generalization the trading rules, the repeatedly testing with different historical stock data and many different periods of times are recommended before uses in the real trading or for further studies.

References


