Integrating Traditional Stock Trading Systems: A Data Mining Approach

Ekarin Varutbangkul

Abstract—This paper proposes a new stock trading approach which integrates all trading signals from six traditional trading systems: Moving Average Convergence/Divergence, Relative Strength Index, Stochastic Oscillator, Moving Average Envelopes, Bollinger Bands, and Commodity Channel Index. The combined signals are filtered by decision trees in order to improve the quality of the trading signals. Precision of the trading signals and annual rate of return are used as metrics to evaluate the proposed trading system. From the experimental results, precision of trading signals can be improved by decision tree learning. To evaluate performance of the proposed system, the system is compared to each individual of the six traditional trading systems and the buy-and-hold strategy. From the comparison, the purposed system can provides the highest annual rate of return. Thus, the proposed approach is very promising for stock trading.

Keywords—Decision Tree, Financial Computing, Stock Trading Signal, Stock Trading System

I. INTRODUCTION

FINANCIAL COMPUTING has become a very popular research area in the last decades. Several researchers have studied to predict financial price trends because successful predictions not only help to make more profit but also help to reduce risk.

Predicting stock price trends is a high dimensional problem because stock price movements can be affected by several factors such as political and economic situations as well as their business performance. Therefore, the problem is very difficult and challenging for both investors and researchers.

Two main approaches are commonly used to solve the problem. The first approach, fundamental analysis, involves analyzing their financial statements, their management, and their competitive advantages in order to estimate the intrinsic value of the securities. The latter approach, technical analysis, uses historical prices to predict future price trends.

In this paper, we propose a new technical analysis approach which integrates traditional trading systems such as Moving Average Convergence/Divergence (MACD), Relative Strength Index (RSI), Stochastic Oscillator (SO), Moving Average Envelopes (MAE), Bollinger Bands (BB), and Commodity Channel Index (CCI). In addition, we improve the quality of combined trading signals by using decision tree learning.

II. RELATED WORK

Several research studies applied data mining techniques for predicting stock trading signals in the last decades. For example, Teixeira et al. [1] used k-nearest neighbor (k-NN) algorithm to predict trading signals of 15 securities from Sao Paulo Stock Exchange, Brazil.

Chang et al. [2] used piecewise linear representation (PLR) and artificial neural networks (ANN) to predict trading signals of three American securities and three Taiwan securities. PLR was used to identify turning points and then ANN were used to model the relationship between turning points and technical indicators.

Luo and Chen [3] used PLR and weighted support vector machine (WSVM) to predict stock trading signals of 20 securities from Shanghai Stock Exchange, China. PLR was used to identify turning points and then WSVM was used instead of ANN in [4]. They also reported that PLR-WSVM made higher profit than PLR-ANN.

Son et al. [4] forecasted trends of Korea composite stock price index (KOSPI 200). In their study, they compared several data mining techniques such as linear regression, logistic regression, artificial neural networks (ANN), and support vector machines (SVM). They extracted several technical indicators from both KOSPI 200 time series and KOSPI 200 futures time series. With dimensionality reduction, the accuracies of all classifiers are about 0.62.

Qin et al. [5] predicted direction of Shanghai Composite Index by using regression model and ANN. Features used in this study are moving average, oscillator %K and %D, volatility, price change, and volume change. They found that ANN performs better than regression model with the directional accuracy about 0.55 to 0.65.

The main difference between the related work so far and this study is the usage of traditional trading systems’ signals. Although indicators from the trading systems (MACD, RSI, SO, BB, CCI, and MAE) might be used as input features in above researches, the trading signals from these trading systems were ignored. In this study, both the technical indicators and their trading signals are used to generate the integrated trading signals.

To the best of our knowledge, integrating traditional trading systems have not been investigated in the computer science community. The only related work we found is a financial research done by Lento [6]. Lento proposed a combined signal
approach (CSA) on the individual trading rules. A buy/sell signal is taken if \( t' \) or more of 9 trading systems suggest the buy/sell actions. The approach was experimented on the 59 year period of S&P 500 Index. The result showed that CSA (2/9) performed better than all individual trading rules.

The differences between [6] and our study are as follows. First, Lento just experimented whether the combined signal approach increase profitability without building any predictive model. Second, the best results were obtained when at least two from nine trading systems agreed on the buy/sell actions, while all trading signals from the six trading systems are used in this study.

The remainder of this paper is organized as follows. In the next section, technical background is introduced. Section IV describes our methodology. Section V presents the results and discussion. Conclusion and recommendations for future work are presented in section VI.

III. TECHNICAL BACKGROUND

In this section, the six traditional trading systems and decision tree learning are introduced.

A. Moving Average Convergence/Divergence

Moving Average Convergence/Divergence (MACD) [7] is a technical indicator created by Gerald Appel in the late 1970s. MACD line can be calculated by subtracting the 26-period exponential moving average (EMA) from the 12-period EMA of closing prices as presented in (1) and (2):

\[
MACD_t = EMA_{t-12} - EMA_{t-26}
\]

(1)  

\[
EMA_t = EMA_{t-1} + 2(C_t - EMA_{t-1}) / (n + 1)
\]

(2)  

where \( C \) is a time series of closing prices, \( n \) is a number of time period, and \( t \) indicates time period \( t \).

A 9-period EMA of the MACD line is called "signal line". The MACD trading system generates buy signals when the MACD line crosses above the signal line, and generates sell signals when the MACD line crosses below the signal line.

B. Relative Strength Index

Relative Strength Index (RSI) [8] is a momentum oscillator that measures the speed and change of price movements. It was developed by J. Welles Wilder. RSI oscillates between 0 and 100. Although RSI is most typically used on a 14 periods, it is developed from MAE. BB consists of three bands: middle band, upper band, and lower band. The middle band (MB) is a moving average of closing price. Each envelope is set with the same percentage above or below the moving average. Either SMA or EMA can be used for the moving average. In this study, 10-period EMA is used and the percentage is set as 4. The calculation of upper envelope (UE) and lower envelope (LE) are presented in (10) and (11):

\[
UE_t = SMA_t \times (100 + \text{percentage}) / 100
\]

(10)  

\[
LE_t = SMA_t \times (100 - \text{percentage}) / 100
\]

(11)  

MAE trading system generate buy signals when the closing price line crosses above the lower envelope and generate sell signals when the closing price line crosses below the upper envelope.

E. Bollinger Bands

Bollinger Bands (BB) [10] is a trading system created by John Bollinger in the early 1980s. Bollinger observed that volatility was dynamic, not static. Therefore, BB was developed from MAE. BB consists of three bands: middle band, upper band, and lower band. The middle band (MB) is a moving average of closing prices. In this study, 20-period SMA of the closing prices is used as presented in (12). The upper band (UB) and the lower band (LB) can be calculated by (13) and (14):

\[
MB_t = SMA_t / (1 + \text{percentage}) / 100
\]

(12)  

\[
UB_t = SMA_t \times (100 + \text{percentage}) / 100
\]

(13)  

\[
LB_t = SMA_t \times (100 - \text{percentage}) / 100
\]

(14)  

SO trading system generates buy signals when the slow %K line crosses above the slow %D line and \( Sk_j \) is lower than 20, and generate sell signals when the slow %K line crosses below the slow %D line and \( Sk_j \) is higher than 80.
\[ UB_n(k, n, k) = MB_n(n) + k \times SD(C, n) \]  \hspace{1cm} (13)

\[ LB_n(k, n, k) = MB_n(n) - k \times SD(C, n) \]  \hspace{1cm} (14)

where \( k \) is a constant which has default value 2, and \( SD \) is standard deviation of closing prices over \( n \) periods.

BB trading system generates buy signals when the closing price line crosses above the lower band and generate sell signals when the closing price line crosses below the upper band.

**F. Commodity Chanel Index**

Commodity Chanel Index (CCI) [11] is an oscillator that measures the current price level relative to the average price over a period of time. It was created by Donald Lambert in 1980. CCI can be calculated by using (15) and (16):

\[ CCI_T = \frac{(TP_i - SMA(TP, n))}{(0.015 \times MD(TP, n))} \]  \hspace{1cm} (15)

\[ TP_i = (H_i + L_i + C_i)/3 \]  \hspace{1cm} (16)

where \( TP \) is typical price, \( H \) is high price, \( L \) is low price, \( MD \) is mean deviation of the typical prices over \( n \) periods. In this study, \( n \) is set as 20.

CCI trading system generates buy signals when the CCI line crosses above -100 and generate sell signals when the CCI line crosses below the 100.

**G. Decision tree Learning**

Decision tree learning is a data mining technique which uses a decision tree as a predictive model. A decision tree can be learned by splitting data instances into subsets by testing the value of an attribute which has the highest discriminative power for each splitting. This process is recursively repeated on each derived subset until a stopping criteria is met. Each internal node in the tree is a test of a selected attribute and branches from the node correspond to the possible values of the attribute. After the tree stops growing, each leaf node represents a class of target attribute by using majority voting of instances at the leaf node. Each path from root node to each leaf node represents a rule which is a conjunction of attributes that lead to classification. Thus, a decision tree is a set of classification rules.

The decision tree learning algorithm used in this study is C4.5 algorithm [12]. The stopping criterion of C4.5 algorithm is the minimum objects per each leaf node which is set as 20 in this study. In C4.5 algorithm, a metric used to measure discriminative power of each attribute in each splitting is gain ratio which can be calculated by (17) – (19):

\[ GainRatio(S, A) = \frac{InformationGain(S, A)}{\frac{k}{i=1} \log \frac{n_i}{n}} \]  \hspace{1cm} (17)

\[ InformationGain(S, A) = \frac{\sum_{i=1}^{k} n_i \log \frac{n_i}{n}} {\sum_{i=1}^{k} n_i} \]  \hspace{1cm} (18)

\[ Entropy(S) = -\frac{\sum_{i=1}^{k} n_i \log \frac{n_i}{n}} {n} \]  \hspace{1cm} (19)

where \( n \) is number of instances in subset \( S \) at a parent node before splitting into \( k \) subsets, \( k \) is number of possible values of an attribute \( A \), \( S_i \) represents each subset of \( S \) from \( k \) subsets, \( n_i \) is number of instances belonging to each class from \( c \) classes of target attribute.

**IV. METHODOLOGY**

The methodology of this study can be organized into three phases: data preprocessing, training, and testing, which are described in subsection A, B, and C respectively. Subsection D presents details of input features in this study.

**A. Data Preprocessing Phase**

A diagram of processes in data preprocessing phase is presented in Fig. 1.

![Fig. 1 Data Preprocessing Phase](image)

Data used in this study consists of daily trading information of 5 selected securities from 2001 to the first half of 2013. The five selected securities belong to the top 5 biggest banks in Thailand: Bangkok Bank (BBL), Krungthai Bank (KTB), Siam Commercial Bank (SCB), Kasikorn Bank (KBANK), and Bank of Ayudhya (BAY).

In the original dataset, there are 6 attributes in the daily trading information: date, opening price, high price, low price, closing price, and volume. First, we extract input features from the original dataset. Details of the extracted input features are presented in subsection D. From the input features, we generate buy and sell signals of the six trading systems as explained in section III.
We also generate target feature from the original dataset with our buy and sell conditions:

**Buy Condition**: buy today, if the average of next 10 days’ closing price is higher than today’s closing price at least 1%. Otherwise, do not buy.

**Sell Condition**: sell today, if the average of next 10 days’ closing price is lower than today’s closing price at least 1%. Otherwise, do not sell.

Then we combine all the input features, the buy signals, the sell signals, the buy target feature, and the sell target feature into a combined dataset. Finally, two datasets were generated from the combined dataset: buy-signal dataset and sell-signal dataset. In buy-signal dataset, we keep only instances which at least one trading system provide “buy signal” on those days. In the same way, we keep only instances which at least one trading system provide “sell signal” on those days for the sell-signal dataset. Therefore, the last two dataset are generated from the union of all six trading systems’ signals.

### B. Training Phase

To make predictive model more generalized we combine the buy-signal dataset of all 5 securities into only one dataset. Only instances within years 2001 to 2010 are used in training set. Another training set for sell-signal is created in the same manner. For each training set, an important subset of features is selected by using wrapper method. Then C4.5 algorithm is applied to learn a decision tree as presented in Fig. 2. At the end of training phase, two decision trees are obtained: a buy decision tree, and a sell decision tree.

![Fig. 2 Training Phase](image)

### C. Test Phase

In test phase, buy-signal dataset and sell-signal dataset of each individual security from data preprocessing phase are used to create a buy test set and a sell test set. Only instances within years 2011 to the first half of 2013 are used in the test sets.

Signals from both the buy test set and the sell test set are filtered by the buy decision tree and the sell decision tree learned in training phase. Then the filtered buy and sell signals along with the opening price from original data are used for trading simulation.

In trading simulation, the filtered buy and sell signals are combined. A buy or sell action is taken at the opening of the next day after a filtered signal appears. If a security has been already bought, it will not be bought again although another buy signal appears. For each buy transaction, a fixed budget of (one million Thai Baht, THB) is used. There is a transaction cost including commission, tax, and fee approximately 0.17 percent of the transaction for each buy and each sell transaction.

The metrics used to evaluate our approach are precision of the filtered models and annual rate of return of the approach. A diagram representing the test phase is presented in Fig. 3.

![Fig. 3 Test Phase](image)

### D. Input Features

There are 18 input features used in this study as presented in Table I.

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>INPUT FEATURES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value Indicators</td>
<td>Value Change Indicators</td>
</tr>
<tr>
<td>BB</td>
<td>BBVC</td>
</tr>
<tr>
<td>CCI</td>
<td>CCIVC</td>
</tr>
<tr>
<td>MACD</td>
<td>MACDVC</td>
</tr>
<tr>
<td>MAE</td>
<td>MAEVC</td>
</tr>
<tr>
<td>RSI</td>
<td>RSIVC</td>
</tr>
<tr>
<td>SO</td>
<td>SOVC</td>
</tr>
</tbody>
</table>

The 18 input features can be categorized into 3 groups: value indicators, value change indicators, and graphical change indicators.

For the first group, value indicators, most indicators except BB and MAE can be computed directly using equations in section III. CCI can be computed using (15). MACD can be computed using (1). RSI can be computed using (3) and SO can be computed using (8). The indicator used in the BB trading system is normally the closing price line. Therefore, we normalized the BB indicator using the lower band and the upper band as presented in (20):
The second group, value change indicators, can be computed by subtracting indicators in the first group from their values in the previous period. For example, $BBVC_t$ can be calculated by subtracting $BB_{t-1}$ from $BB_t$.

The last group, graphical change indicators, focuses on the change from previous period in the same manner as the second group. However the change of values is not considered directly. We originally create this group of indicators by using chartists’ point of view. Chartists have been interested in the crossover of the indicator line and specific constants (or specific bands or other indicator line). Most trading system used in this study except MACD has upper threshold (or upper band) and lower threshold (or lower band). Therefore, we categorize the value of each indicator into 4 locations: 1) lower than the lower threshold, 2) between the lower threshold and the middle point between lower and upper threshold, 3) between the middle point and upper threshold, and 4) higher than upper threshold. Each location is encoded with a character (i.e. A, B, C, and D). To present the change from previous period, we concatenate two characters: the first character from previous period, and the second character from current period. For example, AB represents the change from location A in previous period to location B in current period. While the indicators in the first two groups are continuous features, the indicators in the third group are nominal features.

V. RESULTS

As mentioned in section IV, the metrics used to evaluate our approach are precision of the filtered models and annual rate of return of the approach. First, we compare precision of two approaches. The first approach is called union approach (UA). In UA, the union of the sets of buy/sell signals from all 6 traditional trading systems is used directly without filtering. The second approach is called integrating approach (IA). In IA, the union set of buy/sell signals from UA is filtered by a decision tree before simulating trading activities. We compare the precision of both approaches based on our target feature and the results are presented in Table II. For buy signals, IA performs better than UA for all 5 securities. For sell signals, IA performs better than UA for most securities (except KTB).

If we consider the average precision, IA performs better than UA for both buy and sell signals. From the results in Table II, we conclude that our filtering models can improve the quality of buy and sell signals.

Then, we use the filtered buy and sell signals to simulate trading activities and then compare the annual rate of return (ARR) using IA to the ARR using other trading systems as presented in Table III.

First, we compare our approach to each individual of the six traditional trading systems (MACD, RSI, SO, MAE, BB, and CCI). From the 5 securities, IA provides the highest ARR for 2 securities (BB and SCB) and a high ARR for BB (23.32%) which is comparable to the highest one (25.56%) from SO. The ARR of KTB using IA is quite low compared to the six individual traditional trading systems which can be explained by the low precision of its sell signals in Table II. However, if we consider the average ARR of the 5 securities, IA still performs better than all individual traditional trading systems.

Next, we compare the average ARR of the 5 securities using IA and UA and the result shows that IA outperforms UA. The result reassures that our filtering models can improve the quality of buy and sell signals. While IA outperforms both UA and the six individual traditional trading systems, UA still provides lower ARR than BB. This comparison shows that using union of the buy/sell signals from the six traditional trading systems directly without filtering is not good enough to overcome the six individual traditional trading systems.

Finally, we compare IA to B&H. Since Thai stock market has grown dramatically in the last few years (the test period). Results from Table III shows that the average ARR of the five securities using B&H strategy is higher than 15 percent and none of the six individual traditional trading systems can overcome the B&H strategy. However by integrating the buy/sell signals of these six traditional trading systems and filtering the buy/sell signals to improve the quality of the signals, IA can provide higher ARR than the B&H strategy. Therefore, IA is a very promising approach for stock trading.
Beside of annual rate of return, we are also interested in number of trades, and profitable trade ratio. The profitable trade ratio of the six traditional trading systems, UA, and IA are presented in Table IV. For example, the result of using MACD on BAY is 7/24 which represented 7 profitable trades from all 24 trades in the test period (2.5 years), and the percentage is shown in the parenthesis below. MACD has high number of trades compared to others, but its profitable trade ratio is very low. While RSI, MAE, and BB have very low number of trades (less than 10 trades in 2.5 years) but their profitable trade ratio are quite high. For our approach, we want to have more number of trades to have more chances of making profit. However, the profitable trade ratio should also be acceptable. For UA, the number of trades is quite high (about 30 trades) since UA uses all buy and sell signals from all 24 trades in the test period (2.5 years), and the percentage is shown in the parenthesis below. MACD has high number of trades compared to others, but its profitable trade ratio is very low. While RSI, MAE, and BB have very low number of trades (less than 10 trades in 2.5 years) but their profitable trade ratio are quite high. For our approach, we want to have more number of trades to have more chances of making profit. However, the profitable trade ratio should also be acceptable. For UA, the number of trades is quite high (about 30 trades) since UA uses all buy and sell signals from the six traditional trading systems without filtering, but its profitable trade ratio is quite low. With filtering in IA which improves the quality of buy and sell signals, the profitable trade ratio is also increased. Although the number of trades is slightly lower because some buy and sell signals are filtering out.

VI. CONCLUSION

In this paper, we propose a new stock trading approach which combines trading signals from six traditional trading systems: MACD, RSI, SO, MAE, BB, and CCI. Then, decision trees are used as filters in order to improve the quality of the combined trading signals.

From our experimental results, we found that filtering with decision trees can improve the quality of the combined trading signals and our approach can overcome all individual trading systems, B&H strategy, and UA. Thus, the proposed approach is very promising for stock trading.

We also found that precision of trading signals has very high impact on the trading system’s performance. Since all input features used in this study are extracted from the six traditional trading systems only, the precision of the filtered trading signals are quite low. Therefore, an interesting further investigation is to employ more technical indicators as input features. Another potential work is to incorporate more trading systems and then select only trading systems with high performance to be integrated in the system.

TABLE IV
COMPARISON OF PROFITABLE TRADE RATIO

<table>
<thead>
<tr>
<th>Securities</th>
<th>MACD</th>
<th>RSI</th>
<th>SO</th>
<th>MAE</th>
<th>BB</th>
<th>CCI</th>
<th>UA</th>
<th>IA</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAY</td>
<td>7/24</td>
<td>7/8</td>
<td>14/17</td>
<td>6/7</td>
<td>6/7</td>
<td>8/11</td>
<td>21/34</td>
<td>13/16</td>
</tr>
<tr>
<td></td>
<td>(29.17%)</td>
<td>(87.50%)</td>
<td>(82.35%)</td>
<td>(85.71%)</td>
<td>(85.71%)</td>
<td>(72.73%)</td>
<td>(61.76%)</td>
<td>(81.25%)</td>
</tr>
<tr>
<td>BBL</td>
<td>7/25</td>
<td>3/5</td>
<td>12/17</td>
<td>2/3</td>
<td>4/6</td>
<td>7/11</td>
<td>16/30</td>
<td>10/17</td>
</tr>
<tr>
<td></td>
<td>(28.00%)</td>
<td>(60.00%)</td>
<td>(70.59%)</td>
<td>(66.67%)</td>
<td>(66.67%)</td>
<td>(63.64%)</td>
<td>(53.33%)</td>
<td>(58.82%)</td>
</tr>
<tr>
<td>KBANK</td>
<td>11/21</td>
<td>4/5</td>
<td>9/16</td>
<td>1/3</td>
<td>5/5</td>
<td>9/12</td>
<td>14/26</td>
<td>9/16</td>
</tr>
<tr>
<td></td>
<td>(52.38%)</td>
<td>(80.00%)</td>
<td>(56.25%)</td>
<td>(33.33%)</td>
<td>(100.00%)</td>
<td>(75.00%)</td>
<td>(53.85%)</td>
<td>(56.25%)</td>
</tr>
<tr>
<td></td>
<td>(42.86%)</td>
<td>(60.00%)</td>
<td>(60.00%)</td>
<td>(83.33%)</td>
<td>(60.00%)</td>
<td>(62.50%)</td>
<td>(54.84%)</td>
<td>(40.00%)</td>
</tr>
<tr>
<td>SCB</td>
<td>9/22</td>
<td>7/7</td>
<td>10/15</td>
<td>3/4</td>
<td>6/6</td>
<td>10/13</td>
<td>15/23</td>
<td>12/16</td>
</tr>
<tr>
<td></td>
<td>(40.91%)</td>
<td>(81.82%)</td>
<td>(60.00%)</td>
<td>(75.00%)</td>
<td>(100.00%)</td>
<td>(76.92%)</td>
<td>(65.22%)</td>
<td>(75.00%)</td>
</tr>
<tr>
<td>Total</td>
<td>43/113</td>
<td>24/30</td>
<td>54/80</td>
<td>17/23</td>
<td>24/29</td>
<td>39/55</td>
<td>83/144</td>
<td>50/80</td>
</tr>
<tr>
<td></td>
<td>(38.05%)</td>
<td>(80.00%)</td>
<td>(67.50%)</td>
<td>(73.91%)</td>
<td>(82.76%)</td>
<td>(70.91%)</td>
<td>(57.64%)</td>
<td>(62.50%)</td>
</tr>
</tbody>
</table>

REFERENCES


