JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 30, 653-667 (2014)

DOI:10.1688/JISE.2014.30.3.7

Stock Trend Prediction by Using *K*-Means and AprioriAll Algorithm for Sequential Chart Pattern Mining*

KUO-PING WU[1], YUNG-PIAO WU[1] AND HAHN-MING LEE[1,2]
[1]*Department of Computer Science and Information Engineering*
National Taiwan University of Science and Technology
Taipei, 106 Taiwan
E-mail: {wgb; M9915908; hmlee}@mail.ntust.edu.tw
[2]*Institute of Information Science*
Academia Sinica
Taipei, 115 Taiwan

In this paper we present a model to predict the stock trend based on a combination of sequential chart pattern, *K*-means and AprioriAll algorithm. The stock price sequence is truncated to charts by sliding window, then the charts are clustered by *K*-means algorithm to form chart patterns. Therefore, the chart sequences are converted to chart pattern sequences, and frequent patterns in the sequences can be extracted by AprioriAll algorithm. The existence of frequent patterns implies that some specific market behaviors often appear accompanied, thus the corresponding trend can be predicted. Experiment results show that the proposed system can produce better index return with fewer trades. Its annualized return is also better than award winning mutual funds. Therefore, the proposed method makes profits on the real market, even in a long-term usage.

Keywords: Haar wavelet, *K*-means, AprioriAll, sequential chart pattern, stock trend prediction

1. INTRODUCTION

Financial time series are very dynamic and sensitive to quick changes. It is because in part of the underlying nature of the financial domain and in part of the mix of known parameters (previous day's closing price, P/E ratio *etc.*) and unknown factors (like election results, rumors *etc.*) [1]. Financial time series are difficult to forecast because the problem is nonlinear, non-stationary and may have a lot of noises [2]. Pattern discovery, dimensionality reduction, clustering and classification, rule discovery and summarization are the main challenges of processing the time series data [3]. Stock price is one kind of time series, and using data mining techniques to predict stock trend is one of the most important issues to be investigated. Stock trend prediction aimed on developing approaches to predict the price in the future with high profits. Obviously, the prediction is difficult because the implemented models should capture the market volatility [4]. Data mining uses techniques including clustering and association rule. These techniques can forecast the future trend based on the itemsets [5]. Clustering methods group similar items while the association rule generalizes dependent variables. Representative itemsets can be obtained from trading data using these rules [6].

Received February 28, 2013; accepted June 15, 2013.
Communicated by Hung-Yu Kao, Tzung-Pei Hong, Takahira Yamaguchi, Yau-Hwang Kuo, and Vincent Shin-Mu Tseng.
* This research is supported in part by the National Science Council of Taiwan under grants number NSC 101-2218-E-011-008 and NSC 99-2221-E-011-075-MY3.

For stock market forecasting, there are two well-known analysis methods, which are technical analysis and fundamental analysis. Technical analysis accompanied with a variety of forecasting techniques such as chart analysis, cycle analysis and computerized technical trading system [7]. For example, Lo *et al.* [8] used nonparametric kernel regression to identify 10 different technical patterns, and then provided indicators that technical analysis may include. Specifically, they mentioned that even technical analysis is not necessary to generate excess trading profits, it does raise the possibility that adding value to the investment process. The challenges of stock trend prediction are:

(1) Non-linear data, non-stationary data and noise robustness [2]
(2) Large data size and necessary to update continuously [6]
(3) Time series data representation [6]
(4) Pattern discovery and clustering (similarity search), dimensionality reduction, classification, rule discovery and summarization
(5) Automatically generate trading chart patterns of different markets [9]

Finding efficient ways to summarize and visualize the stock market data is one of the most important problems in modern finance. It provides individuals or institutions useful information about the market behavior for making investment decisions. The enormous amount of data generated by the stock markets day by day has fascinated researchers to investigate this domain using different methodologies [10]. Most of the academic researches focused on modeling the observations of stock market properly and improved the accuracy of stock price forecasting. We want to study whether using the sequences of chart patterns extracted from historical financial time series data can predict stock trend effectively or not. Our aim is to find out how to use the sequential chart patterns to predict the stock price trend in the future days.

2. RELATED WORK

Stock trend prediction is an important financial research. If a market can be predicted successfully, the investors can gain improved returns. There are many factors that affect the prediction results and thus no universal model that can predict everything well for all problems or even be a single best forecasting method for all situations [11]. Therefore, many researchers have applied different analysis methods to do stock trend prediction, including association rule based approaches, chart pattern recognition, template-matching, neural networks and SVM [12, 13].

2.1 Association Rule Based Approaches

Lu *et al.* [14] presented moving prediction and *N*-dimensional inter-transaction association rules for stock prediction. Defining a maximum span along dimensional attributes can reduce the search space, which are mining parameters. It works like a sliding window in 1-dimensional cases. Only rules covered by the sliding window are concerned, thus the number of possible rules is confined. This is reasonable: in stock moving prediction, the span represents the user's interest on the predicted rise of price comparing to the current price in an expected interval of trading days. Argiddi and Apte [15] proposed that fragment based approach was promising for extracting some association rules from Indian IT Stock Market. It could be used for predicting or recommending in stock trading systems. In addition, they presented the fragment based approach to forecast association rules and recommend the customers. They also used association rules mined from stock transaction data to do future trend prediction [16].

2.2 Chart Pattern Based Approaches

For stock market technical analysis, chart patterns are the most popular and common material which becomes an important basis for numerous investors to trade accordingly. Referring to Bulkowski [17], there are up to 47 different chart patterns identified in stock price charts. Bulkowski [18] also introduced 14 chart patterns and many researchers used those common patterns to do the similar works. Lo *et al.* [8] used nonparametric kernel regression and identified 10 different technical patterns. The patterns provided indication that technical analysis could be predictive.

2.3 Template-Matching Technique Based Approaches

Many researchers have applied stock chart pattern analysis method on investment decision-making in recent years, including Leigh *et al.* [19] and Chen *et al.* [9]. All of them used template matching techniques based on pattern recognition to implement the bull or bear flags of stock chart. These studies suggest that using chart patterns can predict stock prices. Leigh *et al.* [19] used both price and volume as fitting values and attempted to detect the defined bull flag. If the cross-correlation computed fitting value is high, it represents that the stock price pattern in the selected period is close to stock chart template. Parracho *et al.* [20] used the genetic algorithm to optimize up and down pattern templates. Based on the pattern recognition algorithm developed by Lo *et al.*, Savin *et al.* [21] used experiments to confirm that head-and-shoulders price patterns is predictive for future stock returns.

2.4 Neural Network Approaches

There are many researches using artificial neural networks (ANNs). A lot of successful trials have shown that ANN can be a powerful tool for time series forecasting and modeling [22]. However, too many factors required to be tuned would affect the ANN performance. It is a challenge to design the sampling schema, choose training and testing datasets and select the effective factors for improving the prediction performance. Atsalakis *et al.* [4] surveyed articles that applied neural networks and neural fuzzy models to predict stock markets and concluded that these methods were suitable for stock market forecasting. However, it is difficult to define the structures of the models such as the hidden layers, the neurons, *etc.* Zhang *et al.* [22] presented a piecewise nonlinear model to analyzing stock market tick data. They proposed Prop NN, which can improve the predictability of stock price. They claimed that it is significantly better than the basic BPN model. Li *et al.* [23] presented a wavelet-based SOM forecasting model. Its performance was validated with the Taiwan Weighted Stock Index (TAIEX). Wu *et al.* [24]

combined back propagation neural network with data mining techniques (data sampling, data conversion and so on) and provided high accuracy for predicting Shanghai Stock Exchange Composite Index. Also, Nayak *et al.* [25] constructed neuro-genetic hybrid network and compared the performances of different models with Bombay stock exchange data. They suggested that the FLANN-GA [26] model performed well in many cases.

2.5 SVM Based Approaches

There are a lot of machine learning algorithms developed to improve testing accuracy. One of the most promising algorithms is support vector machine (SVM) which was proposed by Boser *et al.* [27]. SVM is a supervised learning algorithm that analyzes data. It uses the empirical error and a regularized term which is derived from the structural risk minimization principle [28] to construct the risk function. The decision boundary called hyperplane is determined directly by the training data, and the separating margin of decision boundary is maximized. Due to its theoretical basis of statistics foundation and excellent testing accuracy, it has been used for classification and regression analysis, and can help users make well-informed business decisions. Wang *et al.* [29] showed that the *K*-means SVM (KMSVM) algorithm can speed up the response time of classifiers by decreasing the number of support vectors while maintaining a compatible accuracy to SVM.

2.6 Related Methodology

Similarity measurement is very important to time series analysis and data mining. The most popular approach to measure the similarity between two time series is calculating the Euclidean distance on transformed representation such as DFT coefficients [30] and DWT coefficients [31]. Discrete wavelet transform (DWT) has been found to be effective in many area, including computer graphics, image processing, speech processing, and signal processing [32]. The advantage of using DWT is that it can represent the signals in multiple resolutions. DWT typically decomposes the original time sequence to different resolutions with its preceding coefficients.

The correlation coefficient is a measurement of linear dependence correlation between two variables in statistics. It gives a value between +1 and −1 inclusively, thus it is in fact a scaled covariance. It can be used as a measurement of the strength of linear dependence between two variables.

K-means clustering algorithm is one of the simplest unsupervised learning algorithms. It solves the clustering problem with the resulted centers as the clusters representation [33]. *K*-means groups samples into a specified *K* subset of clusters. The *K* cluster centers can be randomly selected from the given data set and each sample is assigned to the cluster represented by the closest cluster center. This process repeats until the difference between cluster centers in consecutive iterations converges. Liao *et al.* [10] presented using *K*-means algorithm to analysis clusters and to explore the stock clusters in order to mine stock category clusters for investment. He *et al.* [34] used *K*-means clustering algorithm and linear regression to partition stock price time series data and analysis the trend within each cluster. The results were then used for trend prediction for win-

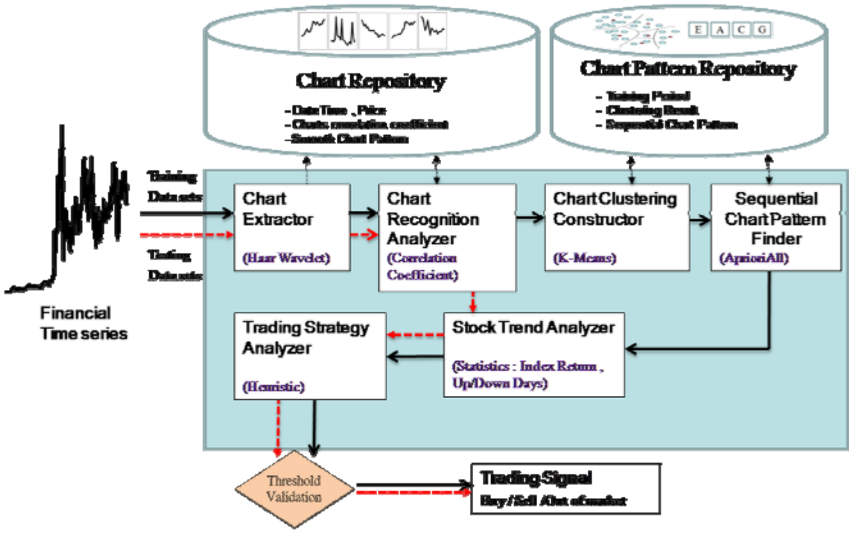


Fig. 1. System architecture of stock trend prediction by sequential chart pattern.

dowed time series data.

In addition to clustering, data mining also uses association analysis. It can be used for forecasting the time series trend [35]. Discovering association rules can be used to find out the frequent pattern in a sequence. Agrawal *et al.* [36] found that association analysis is an important data mining technique, and there are a lot of researches applied association analysis on data mining problems. The association analysis algorithms mainly determine the relationships between items or features that occur synchronously in the data. Items, subsequences or substructures which appear in the data frequently are called frequent patterns.

Apriori algorithm [37] is a typical algorithm for searching association rules. Given an itemset, the algorithm attempts to find subsets which are common according to an appearance threshold of the itemset. Agrawal and Srikant [38] mentioned the problem of mining sequential patterns, and they presented three algorithms for solving such problem. The AprioriAll algorithm was testified to perform better than the other two approaches. It is efficient and can discover temporal relationships between items in the database, and it has been used in many area, such as network security alarms, plan failure identification, analysis of Web access databases [39] and many more [40].

3. STOCK TREND PREDICTION BY SEQUENTIAL CHART PATTERN

In this section, we present a new method to predict the trend of a stock or stock market index such as TAIEX based on sequential chart pattern modeling via *K*-means and AprioriAll Algorithm. The procedure of the system is: (1) chart pattern extraction by sliding window and Haar wavelet transform (2) chart distance measurement by correlation coefficient (3) chart pattern clustering by *K*-means algorithm (4) frequent pattern extraction by AprioriAll algorithm (5) trend analysis by statistic sequential clustering index return and up/down days (6) trading strategy.

The framework of the proposed system is shown in Fig. 1. Modularized design can help users apply the proposed method to different type algorithm easily. The system includes these modules: Chart Extractor, Chart Recognition Analyzer, Chart Clustering Constructor, Sequential Chart Pattern Finder, Stock Trend Analyzer and Trading Strategy Analyzer.

(1) The Chart Extractor: We use sliding window to extract charts from the stock historical data. The sliding window covers w-day closing prices and moves in one day step. Therefore, closing prices of each continuous w days form a chart, while two successive charts correlate in time line. The raw charts are also transformed by discrete wavelet transform (DWT) and then the low frequency parts are stored in chart repository. We use DWT to remove the high frequency components of the signal, thus the transformed charts are smoothed and the dimension is reduced. We choose Haar wavelet for transformation. With Haar wavelet transform, a chart can be split to high frequency part and low frequency part. The low frequency part contains the main trend of a chart, while the high frequency part affects less. Thus we store only the low frequency part for the following-up analysis.

(2) The Chart Recognition Analyzer: We calculate the similarity between two filtered charts according to the correlation coefficient. Similarity measure is the most importance basis for time series analysis and data mining tasks. To measure the similarity between two multi-dimensional vectors, the most often used approach is the Euclidean distance. It can also be applied on transformed data such as the DFT coefficients [7] and DWT coefficients [8]. Also, the Pearson correlation coefficient is a measure of correlation (linear dependence) between two variables. The advantage of using correlation coefficient is that it does not depend on the measurement scale. With the low frequency part of charts transformed by Haar wavelet, we calculate the correlation coefficient which is between [−1, +1]. A higher value implies the two charts are more similar, and a value close to 0 implies the two charts are uncorrelated. For example, we calculate the correlation coefficients between TAIEX 2011 and 2002, 2005, 2009 three years, and plot the charts in Fig. 2. It

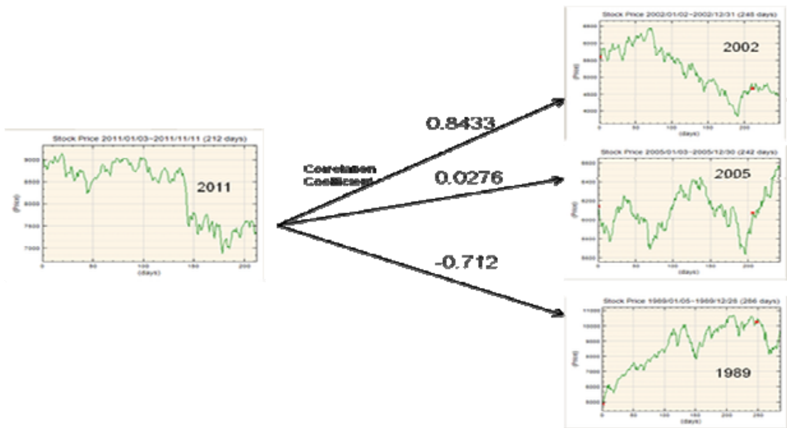


Fig. 2. Correlation coefficient can represent the similarity of charts.

can be figured out that chart 2011 looks similar to chart 2002 with a higher correlation coefficient than the others.

(3) The Chart Clustering Constructor: We use clustering algorithm to aggregate similar charts. With the charts we use correlation coefficient as the similarity measurement for clustering. Charts with high correlation coefficient will form a cluster. We use *K*-means [35] clustering algorithm. Gorunescu [41] presented that the *K*-means is well suited to generating globular cluster and faster than hierarchical clustering. The process including the following steps:

Step 1: Set the cluster correlation coefficient threshold parameter T.
Step 2: Assign all charts to 1 cluster.
Step 3: Calculate the cluster center(s).
Step 4: For each cluster, calculate the correlation coefficient. If no cluster's correlation coefficient $< T$, exit.
Step 5: Split the clusters with correlation coefficient $< T$. Then cluster all charts by *K*-means algorithm.
Step 6: Jump to step 3.

We use correlation coefficient as the measure of distance between samples. Since correlation coefficient is more related to the cluster density than K, we let K be determined automatically by given the stop criteria threshold T. After clustering, the training charts in a cluster form a chart pattern. Therefore, the number of the chart patterns are not predefined but are determined by the data distribution. The chart patterns are stored in chart pattern repository.

(4) The Sequential Chart Pattern Finder: Frequent patterns are the patterns that appear in a data set frequently. For data mining and knowledge discovery in databases, frequent pattern mining has become an important data mining task and a focused theme in data mining research [30]. For the clustered chart patterns, we use AprioriAll algorithm to find out frequent patterns which match the given minimal support and confidence with the following steps:

Step 1: Match the chart of each day to its corresponding chart pattern and form a chart pattern sequence.
Step 2: Set the number of forecasting days n.
Step 3: For every n day get the chart pattern index from the pattern sequence, by this n chart pattern subsequences are formed:
    subsequence 1: patterns of day 1, $1 + n$, $1 + 2n$,...
    subsequence 2: patterns of day 2, $2 + n$, $2 + 2n$,...
    …
    subsequence n: patterns of day n, $2n$, $3n$, ...
Step 4: Use modified AprioriAll algorithm to find out frequent patterns which match the given minimal support and confidence:
    A. L1 Sequences: fully scan the chart pattern sequence of training period.
    B. L2 Sequences: fully scan each chart pattern subsequence.

The proposed method uses the modification of AprioriAll algorithm: all repeated pattern appears in subsequence are counted.

Step 5: Get frequent patterns of the sequences.

Since a chart is the trade price sequence during the sliding window and a chart pattern represents a group of similar charts, a chart pattern sequence can represent the stock price trend for a period of time longer than the sliding window.

(5) The Stock Trend Analyzer: After we have frequent patterns of sequences, we can predict the next possible chart pattern according to the pattern the current chart belongs to. The next pattern corresponds to the charts n days after, by this we can predict the possible price trend in the future and trade accordingly. Each char pattern includes many charts. If the trends of the charts conflict, we use voting to decide the direction of the trend.

(6) The Trading Strategy Analyzer: The generalized trend prediction rules are:

- Stock trend is going up – if the index return after n days > 0 and up/down ratio > 1.
- Stock trend is going down – if the index return after n days < 0 and up/down ratio < 1.

And the trading rules are:

- Buy action – If the stock trend is going up and average index return > p% after n days.
- Sell action – If the stock trend is going down and average index return < p% after n days.

For example, we can define a buying strategy that if the index return increases 1% after 10 days, or a selling strategy that if the average index return decreases 1% after 5 days.

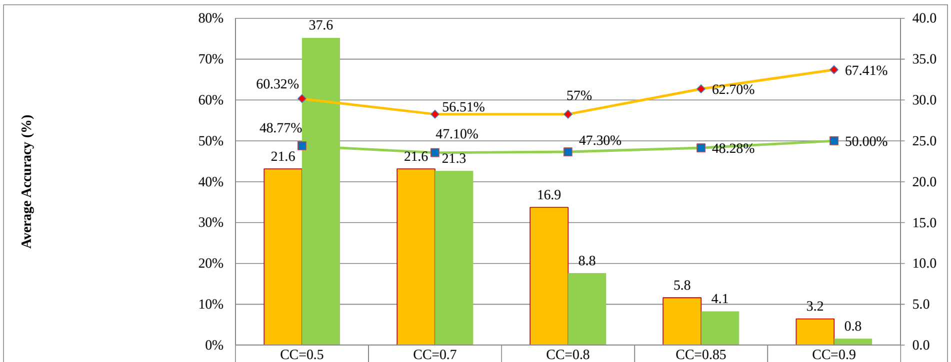
4. EXPERIMENTS AND RESULTS

The dataset we used is TAIEX dataset (1987-2011) from Taiwan Stock Exchange Corp. We tuned the system with many prior trials to determine the parameters of the proposed system, which are mentioned in Section 4.1. To make sure our method outperforms, we compare our system performance to 2 related researches (Section 4.2) and 4 price winning mutual funds (Section 4.3). The performance is evaluated according to the index return or annualized return.

4.1 Determination of Parameters

The first parameter to be determined is the correlation coefficient threshold T which affects the chart similarity measurement. $T \in \{0.5, 0.7, 0.8, 0.85, 0.9\}$ are testified. To simplify the experiments, we set the rest parameters to a moderate combination according to our experience, including sliding window size $w = 20$, AprioriAll algorithm minimal support = 3, confidence = 51%, and use the trading strategy: 1 day & 5 day average index return > 0.1% with up/down days ratio > 1 for buying, and 1 day & 5 day average index return < –0.1% with up/down days ratio < 1 for selling with 5 days buy-and-hold period. The trading strategy is chosen for easy using and comparing to the related work and can be re-design for practical usage, while the proper values of the rest parameters are testified in further experiments. As the buy-and-hold period is 5 days, the forecasting day $n = 5$. That is, we sample the chart pattern sequences to 5 interleaved subsequences, and we try to forecast the trend 5 days after.

Applying the proposed method on TAIEX from 1988 to 2006 (with 1 year testing data from 2001 to 2006 respectively and the training periods 1, 2, 3, 5, 8, 10, 13 year data backward from the testing data year, total 42 training-testing period combinations), we summarized that a larger correlation coefficient threshold T causes higher prediction accuracy (calculated as the percentage that a buying is correct after buy-and-hold period, or versa visa for selling) but fewer trades (from 21.6 trades with 56.5% accuracy for T = 0.7 to 3.2 trades with 67.4% accuracy for T = 0.9, in average). The summarized system performance when varying correlation coefficient threshold T is presented in Fig. 3, and T = 0.85 is chosen for the rest experiments.



	CC=0.5	CC=0.7	CC=0.8	CC=0.85	CC=0.9
Buy Average #N Trade per Year	21.6	21.6	16.9	5.8	3.2
Sell Average #N Trade per Year	37.6	21.3	8.8	4.1	0.8
Buy Average Accuracy(%)	60.32%	56.51%	57%	62.70%	67.41%
Sell Avearage Accuracy(%)	48.77%	47.10%	47.30%	48.28%	50.00%

Fig. 3. The effects of tuning the correlation coefficient threshold T for the proposed system.

Larger sliding window size w may result in longer charts which contain longer trends, while the charts may be harder to form compact clusters. The sliding window size $w \in \{20, 40, 60\}$ with 1 level Haar wavelet transform is testified. Within the choices, w = 20 can generate long enough smoothed charts to form chart patterns. If we simply predict the trend direction (up or down), the prediction accuracy is about 60.8%, and slightly decreases to less than 57% when w = 40 and w = 60. The summarized system performance when varying sliding window size w is shown in Fig. 4 (with T = 0.85, minimal support = 3, confidence = 51%), and w = 20 is chosen for the rest experiments.

Minimal support and confidence of AprioriAll algorithm are determined by grid search method with candidates minimal support $\in \{2, 3, 5, 8, 13, 21, 34\}$ and confidence $\in$ {5%, 10%, 15%, 20%, 30%, 40%}. The combination of minimal support = 13 and confidence = 10% are chosen thereafter because with the combinations around it the system generally has higher prediction accuracy. An example of grid search results is shown in Fig. 5.

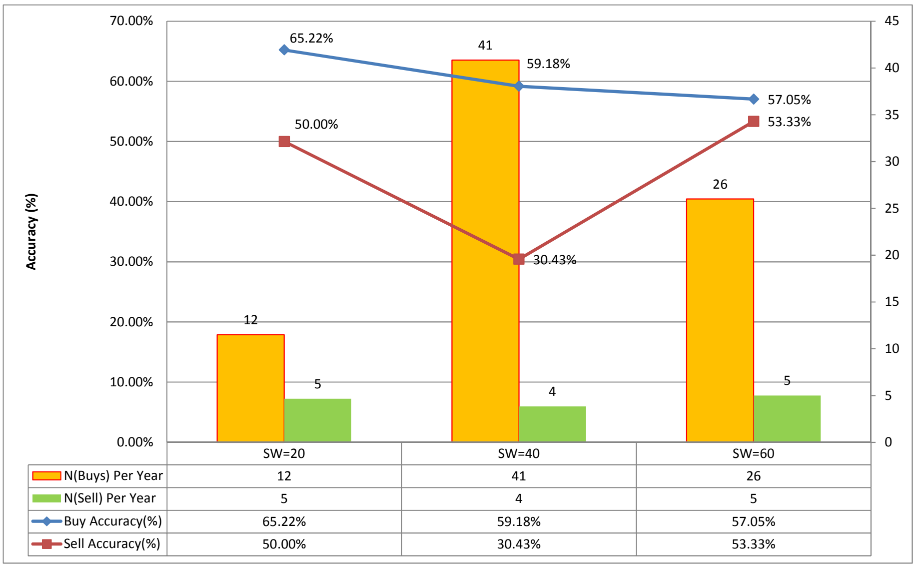


Fig. 4. The effects of tuning the sliding window size for the proposed system.

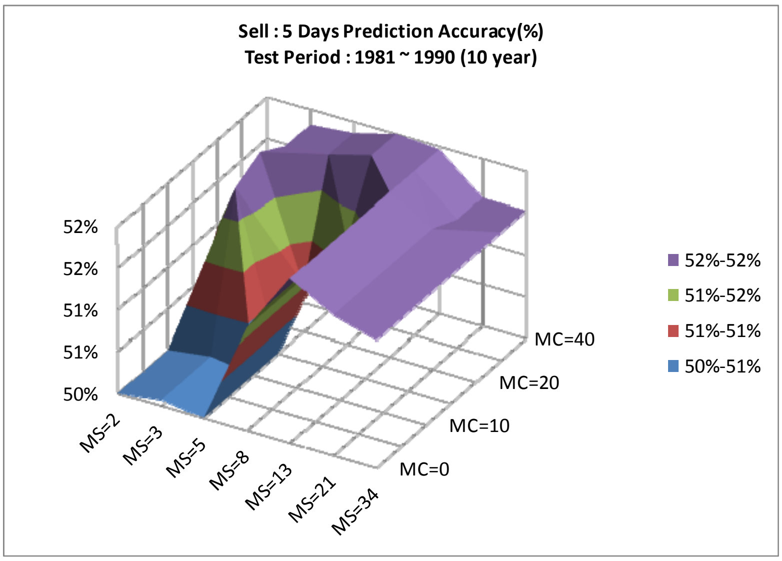


Fig. 5. Sell signal accuracy of grid search for minimal support and confidence combinations.

4.2 Comparison with Related Researches

We compare our system to Wang and Chan's method [42] and Tai-Liang Chen's method [9]. In these experiments we use the TAIXE data from (1) 1983 to 1989 for training and 1990/08/15 to 1997/02/17 for testing, and (2) 1990 to 1996 for training and 1997/02/18 to 2004/03/24 for testing. The holding period is 20 days. It can be observed in Table 1 that the average index return (%) of the proposed method is better than the other two methods. Moreover, even the proposed method results in a higher and stable average index return, we 'buy' fewer times then the other two methods. Although Chen's method outperforms in testing period (1), the system traded too many times while the index return is still low in testing period (2). When practically use a system with fewer trading signals while the index return remains may be preferred, because it reduces both the trading costs and the overtrading risk.

Table 1. Comparison with other forecasting methods.

Testing Period	Wang & Chan's method [42]		Tai-Liang Chen's method [9]		the proposed method	
	Index Return (%)	# buy	Index Return (%)	# buy	Index Return (%)	# buy
1990/08/15-1997/02/17	1.63	176	3.40	317	2.05	162
1997/02/18-2004/03/24	1.71	185	0.34	521	2.58	76
average	1.67	180.5	1.50	419	2.22	119

4.3 Comparison with Award Winning Mutual Funds

The benchmark of real market trading performance evaluation is Morningstar Fund Award Taiwan stock top 1 from 2007 to 2011 [43]. We calculate the annualized return [44] of the period of time for each year's award winning mutual fund and our method. According to Fig. 6, it is clear that the award winning mutual funds performed slightly better than the market (TAIEX) in annualized profit return, and the proposed method outperforms. This implies that the proposed method makes profits on the real market, even in a long-term period.

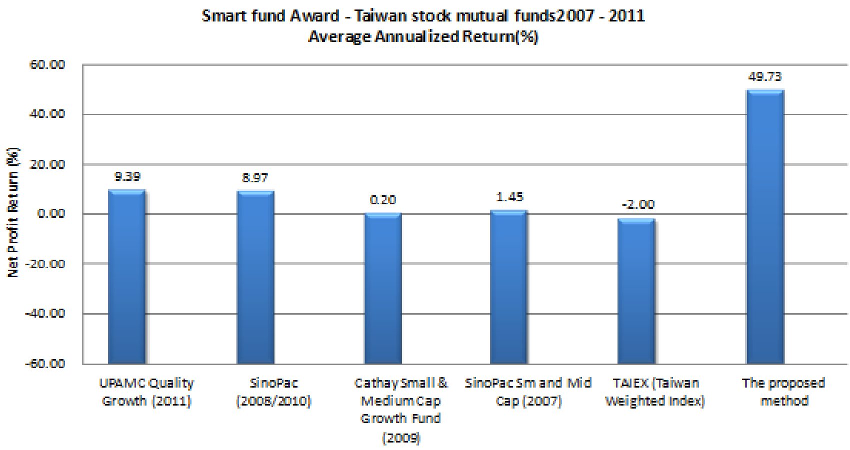


Fig. 6. Comparison with Morningstar Fund Award Taiwan stock top 1s in 2007-2011.

In addition to prediction accuracy and annualized return, a trader may concern more about how the system is applicable in real world. A trading system should have reasonable chance of winning, tolerable largest losing trade and max drawdown, *etc.* We list the performance report from 2007 to 2010 in Table 2 as a reference.

Table 2. The proposed system trading performance report from 2007 to 2010.

Experiment Period	2007	2008	2009	2010
Training Period	1999 ~ 2006	2000 ~ 2007	2001~2008	2002 ~ 2009
Total Number of Trades	3	7	4	9
Winning Trades	2	3	3	4
Percent Profitable	66.67%	42.86%	75.00%	44.44%
Total Net Profit	7.73%	9.35%	11.47%	-1.47%
Gross Profit	9.60%	14.29%	12.02%	9.32%
Loss Profit	-1.65%	-4.94%	-0.56%	-10.79%
Largest Winning Trade	5.49%	7.68%	5.81%	6.88%
Largest Losing Trade	-1.65%	-2.33%	-0.56%	-2.71%
Average Winner Trades	4.80%	4.76%	4.01%	2.33%
Average Losing Trades	-1.65%	-1.24%	-0.56%	-2.16%
Max Drawdown	-1.65%	-2.33%	-0.56%	-8.08%
Profit Factor	5.81	2.89	21.61	0.86
Initial Account Size	126,158	138,358	108,396	224,480
Return On Account	105.79%	77.53%	128.64%	-8.69%

5. CONCLUSION

We propose a novel method aimed at doing the stock trend prediction by sequential chart pattern via *K*-means and AprioriAll algorithm. With the sliding-window and *K*-means based chart pattern construction, the frequent patterns in the pattern sequences can be extracted by AprioriAll Algorithm and can be used for trend forecasting. The proposed method is able to make a financial prediction and get the excess profit value. The proposed method is effective to do the stock trend prediction. It performs better in index return and annualized return perspectives. As it is tested with the historical data for several long periods of time and trades in a lower transaction frequency, the proposed system can be used in the real market, regarded as a buying or selling signal or giving confidence to a trader's prediction of stock prices.

REFERENCES

1. E. F. Fama, "The behavior of stock-market prices," *The Journal of Business*, Vol. 38, 1965, pp. 34-105.
2. P. K. Padhiary and A. P. Mishra, "Development of improved artificial neural network model for stock market prediction," *International Journal of Engineering Science and Technology*, Vol. 3, 2011, pp. 1576-1581.
3. T.-C. Fu, "A review on time series data mining," *Engineering Applications of Artificial Intelligence*, Vol. 24, 2011, pp. 164-181.
4. G. Atsalakis and K. Valavanis, "Surveying stock market forecasting techniques part

II: Soft computing methods," *Expert Systems with Applications*, Vol. 36, 2009, pp. 5932-5941.
5. T. Brijs, G. Swinnen, K. Vanhoof, and G. Wets, "Building an association rules framework to improve product assortment decisions," *Data Mining and Knowledge Discovery*, Vol. 8, 2004, pp. 7-23.
6. D. P. Gandhmal, R. B. Parihar, and R. V. Argiddi, "An optimized approach to analyze stock market using data mining technique," in P*roceedings of International Conference on Emerging Technology Trends*, Vol. 1, 2011, pp. 38-42.
7. C.-H. Park and S. H. Irwin, "What do you know about the Return ability of technical analysis?" *Journal of Economic Surveys*, Vol. 21, 2007, pp. 786-826.
8. A. W. Lo, H. Mamaysky, and J. Wang, "Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation," *The Journal of Finance*, Vol. 55, 2000, pp. 1705-1765.
9. T. L. Chen, "Forecasting the Taiwan stock market with a stock trend recognition model based on the characteristic matrix of a bull market," *African Journal of Business Management*, Vol. 5, 2011, pp. 9947-9960.
10. S.-H. Liao, H.-H. Ho, and H.-W. Lin, "Mining stock category association and cluster on Taiwan stock market," *Expert Systems with Applications*, Vol. 35, 2008, pp. 1929.
11. N. A. Gershenfeld and A. S. Weigend, "The future of time series," in *Time Series Prediction: Forecasting the Future and Understanding the Past*, A. S. Weigend and N. A. Gershenfeld, eds., Addison-Wesley, 1993, pp. 1-70.
12. C. Slamka, B. Skiera, and M. Spann, "Prediction market performance and market liquidity: A comparison of automated market makers," *IEEE Transactions on Engineering Management*, Vol. 60, 2013, pp. 169-185.
13. A. Zhu and X. Yi, "The comparisons of four methods for financial forecast," in *Proceedings of IEEE International Conference on Automation and Logistics*, 2012, pp. 45-50.
14. H. Lu, J. Ha, and L. Feng, "Stock movement prediction and *N*-dimensional intertransaction association rules," in *Proceedings of SIGMOD Workshop Research Issues on Data Mining and Knowledge Discovery*, 1998, pp. 12:1-12:7
15. R. V. Argiddi and S. S. Apte, "Future trend prediction of Indian IT stock market using association rule mining of transaction data," *International Journal of Computer Application*, Vol. 39, 2012, pp. 30-34.
16. R. V. Argiddi and S. S. Apte, "Fragment based approach to forecast association rules from Indian IT stock transaction data," *International Journal of Computer Science and Information Technologies*, Vol. 3, 2012, pp. 3493-3497.
17. T. N. Bulkowski, *Encyclopedia of Chart Patterns*, John Wiley & Sons, Hoboken, 2000.
18. T. N. Bulkowski, *Trading Classic Chart Patterns*, John Wiley & Sons, Hoboken, 2002.
19. W. Leigh, N. Modani, R. Purvis, and T. Roberts, "Stock market trading rule discovery using technical charting heuristics," *Expert Systems with Applications*, Vol. 23, 2002, pp. 155-159.
20. P. Parracho, R. Neves, and N. Horta, "Trading with optimized uptrend and downtrend pattern templates using a genetic algorithm kernel," in *Proceedings of IEEE Congress on the Evolutionary Computation*, 2011, pp. 1895-1901.
21. N. E. E. Savin, P. A. Weller, and J. Zvingelis, "The predictive power of head-and-shoulders price patterns in the U.S. stock market," *Journal of Financial Economet-*

rics, Vol. 5, 2007, pp. 243-265.
22. G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks: The state of the art," *International Journal of Forecasting*, Vol. 14, 1998, pp. 35-62.
23. S.-T. Li and S.-C. Kuo, "Knowledge discovery in financial investment for forecasting and trading strategy through wavelet-based SOM networks," *Expert Systems with Applications*, 2008, Vol. 34, pp. 935-951.
24. M.-T. Wu and Y. Yang, "The research on stock price forecast model based on data mining of BP neural networks," in *Proceedings of the 3rd International Conference on Intelligent System Design and Engineering Applications*, 2013, pp. 1526-1529.
25. S. C. Nayak, B. B. Misra, and H. S. Behera, "Index prediction with neuro-genetic hybrid network: A comparative analysis of performance," in *Proceedings of International Conference on Computing, Communication and Applications*, 2012, pp. 1-6.
26. Y. H. Pao, S. M. Phillips, and D. J. Sobajic, "Neuralnet computing and intelligent control systems," *International Journal of Control*, Vol. 56, 1992, pp. 263-289.
27. B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, 1992, pp. 144-152.
28. K.-J. Kim, "Financial time series forecasting using support vector machines," *Neurocomputing*, Vol. 55, 2003, pp. 307-319.
29. J. Wang, X. Wu, and C. Zhang, "Support vector machines based on *K*-means clustering for real-time business intelligence systems," *International Journal of Business Intelligence and Data Mining*, Vol. 1, 2005, pp. 54-64.
30. R. Agrawal, C. Faloutsos, and A. Swami, "Efficient similarity search in sequence databases," *Foundations of Data Organization and Algorithms*, LNCS, Vol. 730, 1993, pp. 69-84.
31. K. P. Chan and A. W. C. Fu, "Efficient time series matching by wavelets," in *Proceedings of the 1st IEEE International Conference on Data Engineering*, 1999, pp. 126-133.
32. F. K.-P. Chan, A. W. C. Fu, and C. Yu, "Haar wavelets for efficient similarity search of time-series: With and without time warping," *IEEE Transactions on Knowledge and Data Engineering*, 2003, pp. 686-705.
33. J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, 1967, pp. 281-297.
34. H. He, J. Chen, H. Jin, and S. Chen, "Stock trend analysis and trading strategy," in *Proceedings of Joint Conference on Information Sciences*, 2006.
35. J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, 2nd ed., Morgan Kaufmann Publisher, San Francisco, 2006.
36. R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," in *Proceedings of ACM SIGMOID Conference on Management of Data*, 1993 pp. 207-216.
37. R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and I. Verkamo, "Fast discovery of association rules," in *Advances in Knowledge Discovery and Data Mining*, 1996, pp. 307-328.
38. R. Agrawal and R. Srikant, "Mining sequential patterns," in *Proceedings of the 11th International Conference on Data Engineering*, 1995, pp. 3-14.

39. W. Tong and H. Pi-Lian, "Web log mining by an improved AprioriAll algorithm," *World Academy of Science*, *Engineering and Technology*, Vol. 4, 2005.
40. F. Masseglia, P. Poncelet, and M. Teisseire, "Incremental mining of sequential patterns in large databases," *Data and Knowledge Engineering*, Vol. 46, 2003, pp. 97-121.
41. F. Gorunescu, *Data Mining: Concepts*, *Models and Techniques*, Springer-Verlag, Berlin, Heidelberg, 2011.
42. J.-L. Wang and S.-H. Chan, "Stock market trading rule discovery using pattern recognition and technical analysis," *Expert Systems with Applications*, Vol. 33, 2007, pp. 304-315.
43. Morningstar Fund Award – Taiwan Stock, http://smart.businessweekly.com.tw/event/2012/fundawards/01_about_3.php.
44. F. Reilly and K. Brown, *Investment Analysis and Portfolio Management*, 7th ed., South-Western Publishing, Cincinnati, 2003.

Kuo-Ping Wu (吳國賓) received the Ph.D. degree in Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan in 2008. He is now a postdoctoral researcher in Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan. His research interest includes machine learning, computer security and cloud computing.

Yung-Piao Wu (吳永標) received the master degree in Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan in 2012. He is working in Shyang-Horng Computer & CAD Consultant Ltd. since Feb. 2006 as Director Software Project Division.

Hahn-Ming Lee (李漢銘) received the Ph.D. degree in Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan in 1991. He is now a Distinguished Professor in Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, and a Research Fellow in Institute of Information Science, Academia Sinica. His research interest includes computer security, cloud computing, artificial intelligence, machine learning ad data mining.