Knowledge Discovery with SOM Networks in Financial Investment Strategy

Shu-Ching Kuo[1]
Institute of Information Management, Diwan College of Management, Taiwan, R.O.C.
Email: shu-ching@mail.dwu.edu.tw

Sheng-Tun Li[1], **Yi-Chung Cheng**[1], **Men-Hsieu Ho**[2]
[1]Department of Industrial and Information Management, National Cheng Kung University, Taiwan, R.O.C.
[2]College of Business Administration, National Chiayi University, Taiwan, R.O.C.

Abstract

Recently, the recession of the global economy induces the coming of a new era of low interest-rates, which resulted in the stock market as an alternative investment channel for investors. The diversity and complication of domain knowledge existing in the stock market enhance its importance for developing a decision support system which can gather real-time pricing information for supporting decision-making in financial investment. In this study, we tackle these challenges by proposing an integrated solution on the basis of K-chart analysis and the over-whelming self-organizing map neural networks. We not only endeavor to improve the accuracy of uncovering trading signals, but also to maximize the profits of trading. The resulting decision model can help investment decision-makers of national stable funds make the most profitable decisions. In addition, financial experts can benefit from the ability of verifying or refining their tacit investment knowledge offered by the uncovered knowledge.

Keywords: *knowledge discovery, self-organizing map, trajectory analysis, decision support systems, financial investment.*

1. Introduction

Recently, due to the recession of the global economy, the US Federal Bank decreased the interest rates more than ten times, followed by the Central Bank of Taiwan (CBT). The fixed deposit interest rate of CBT dropped drastically below 1%. With the low interest-rates, the fixed deposits as an investment tool were not feasible anymore and this drove investors to search for other investment channels, finding the stock market with the lowest barrier. The diversity and complication of domain knowledge existing in the stock market makes it very difficult for investors to make correct investment decisions, since transaction speeds have become much faster nowadays. Therefore, there is a great necessity for developing a decision support system which can gather real-time pricing information for supporting decision-making in financial investment [3, 15].

Financial investment is a knowledge-intensive industry. In the past years, with the electronic transaction technology advances, vast amount of transaction data has been collected and the emergence of knowledge discovery technology sheds light toward building up a financial investment decision support system [8]. Data of financial markets are essentially time-series which bring more challenges than the traditional discrete data for uncovering the hidden knowledge [2]. Looking at an enormous time series, the prediction patterns for stock values is basically divided into two types: linear models include the regression model, AR, MA, ARIMA, GARCH, ARCH-M model, and the reflexive model that differs in its non-linear relations but still exhibit variance and covariance [9]. The fluctuations in the stock market are high-level non-linear models; therefore, with its limitations, linear models for stock values might not make predictions of the fluctuations of stock values. As for non-linear stock value prediction models, the key is that they include artificial intelligence, neural network, fuzzy system, genetic algorithm, and so on [10, 12]. With machine learning, artificial neural network models the nonlinear characteristics of time series and allows appropriate learning, expression, and presenting for decision-making purposes. By reviewing the appropriate literature, theory and own empirical results, the self-organizing map (SOM) network has shown to be an outstanding clustering model via visualization in diverse application domains [1, 11].

Data from financial markets are essentially time-series, which bring more challenges than the traditional discrete data for uncovering the hidden knowledge. In this research, K-chart patterns as the tool for technical analysis. The SOM network can be used as an extremely fine clustering tool, the trading signals are classified by performing pattern-matching with K-Chart patterns in a trained SOM and the sliding-window data. Finally, the closing price in the next day is predicted based on the patterns of the first 31 days. This study is different from relevant research, we provide investors to judge Primary Bull Markets or Primary Bear Markets by trajectory analysis. The resulting intelligence investment decision support system can help fund managers and investment decision-makers of national stable funds make profitable decisions. In addition, financial experts can benefit from the ability of verifying or refining their tacit investment knowledge offered by the uncovered knowledge.

2 Underlying Technologies

It has been one of the greatest challenges to predict the stock market. Since stock prices vary dramatically, it is important to determine when to buy and sell in order to get high returns from stock investment. In technical analysis, we used candlestick to present daily market's open, high, low, and close prices and looked at the change in body color of the K-chart to interpret the day-to-day sentiment. Another issue is the selection of a proper technology in identifying homogeneous strategy using clustering, which involves tackling the problem of high dimensionality inherent in temporal data. In the machine learning literature, the SOM network has shown to be an effective clustering technology in isolating clusters in a high dimensional space [6]. We will discuss these two major technologies in this section.

2.1 K-chart patterns

A K-chart analysis was used to elicit technical knowledge in this study. This charting technique has become very popular among traders. One of the major reasons is that K-chart is a useful tool to visualize the stock prices so that investors can catch up patterns promptly which can be used to predict future stock price movements [7]. As illustrated in Figure 1, a candlestick is composed of a rectangle and two shadow lines. The former, so-called ‘‘real body’’, indicates the difference between the opening and closing prices of a stock. If the real body of a candlestick shows that the opening prices are higher than the closing prices, the candlestick is called a ‘‘block candlestick’’, otherwise it is named ‘‘white candlestick.’’ The white candlestick implies a rising signal of a stock price and the black candlestick implies a falling signal. Investors thus can interpret the day-to-day sentiment from simply looking at the change in body color of the K-chart. The stock price patterns represented by the candlestick shapes provide important clues to predict future stock price movements. In this study, we adopt K-charts as the feature representation of stock price movements.

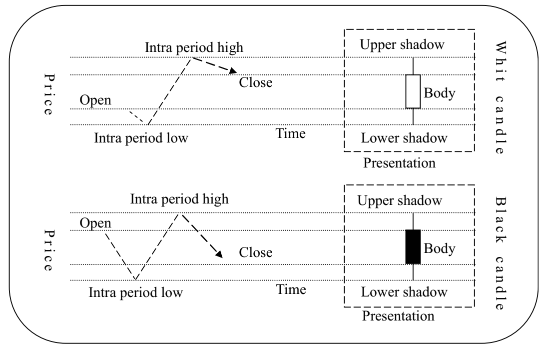

Figure 1. K-chart terms and interpretation

2.2 Self-organizing map neural networks

The SOM neural network is one of the most popular unsupervised neural network models, which quantizes the data space and simultaneously performs a topology-preserving projection from the data space onto a regular two-dimensional grid [6]. The SOM network can be used for clustering, classification, visualization and modeling. In particular, it has visualization capabilities in providing informative pictures of the data space and in exploring data vectors or whole data sets. The versatile properties of the SOM network make it a valuable tool in data mining and knowledge discovery [13, 14]. SOM have been successfully applied in various areas such as image analysis, financial investment, and traveling salesman problem [4, 5].

A basic SOM network is composed of an input layer and a Kohonen layer . The input layer contains neurons for each element in the input vector. The Kohonen layer is formed by neurons which are located on a regular, usually two-dimensional grid and are fully connected with those at the input layer. Each neuron i in the map is represented by a n-dimensional weight or reference vector, where n is equal to the number of neurons in the input layer. The neurons in the map are connected to adjacent ones by a neighborhood relation dictating the topological structure of the neurons. When an input vector $x \in R^n$ is presented to the network, the neurons in the map compete with each other to be the winner (or the best-matching unit, BMU), which is the closest to the input vector in terms of some kind of dissimilarity measure. Therefore, similar input vectors are grouped into a single neuron or neighboring ones in the map when learning is accomplished.

3. Research Methodology

In this section we present the process model used in mining Taiwan Weighted Stock Index. Figure 2 shows the self-explanatory model and the associated steps, which will be discussed in detail in this section.

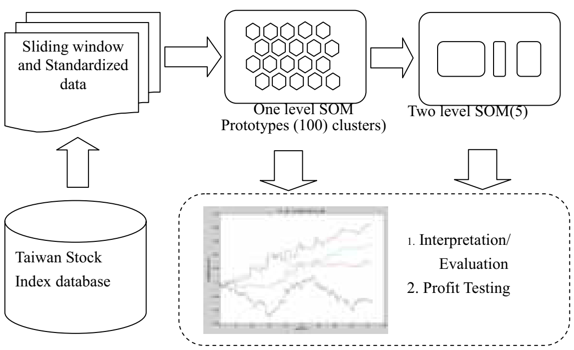

Figure 2. The proposed mining process model

3.1 Source of the data

The experimental data of the Taiwan Weighted Stock Index was gathered from the Taiwan Economic Journal Data Base. The dates ranged from January 3, 1991 to December 31, 2002, with a total of 3,306 exchange-days. This research extracted information from the opens, highs, lows, and closes prices, and these numbers are the basic elements of what the K-chart consist of. To simulate a prospective use of the neural network, the data was divided into training and test sets. The training sets were selected from January 3, 1991 through December 31, 2000. The testing set was from January 2, 2001 to December 31, 2002, with 492 data sets.

3.2 Research framework

This study uses quantitative methods and programmed methods for profitable decision-making. A K-chart analysis is used in order to elicit technical knowledge. The K-chart combines opening, highest, lowest, and closing prices, with utilize sliding windows (32 days) as a segmentation algorithm. By proceeding with standardized processes, we extract features from the time-series data and data mining. For clustering, two-level self-organizing maps are used as extraction of the whole cluster prototypes of the time series patterns [10, 13]. The first level given 100 clusters, the trading signals are classified by performing pattern-matching with K-Chart patterns in the trained SOM and the sliding-window data. Then, the closing price in the next day is predicted based on the patterns of the first 31 days. So far, first-stage daily prediction and profit testing is developed. Subsequently, the two-level SOM clusters patterns numbers are decided by the agglomeration coefficients. The second step uses the trained two-level SOM and conducts a corresponding one-level and two-level trajectory analysis, which develop trend predictions for the Primary Bull Markets and the Primary Bear Markets.

3.3 Sliding window and trajectory analysis

This study use SOM algorithms analysis the Taiwan Weighted Stock Index trading signals. In selection of the data, usage of the sliding window produced pieces of time-sequence data [10]. The sliding window technique is used to divide the time series into small windows, each moving window that is extracted from the time sequence data. The length of the sliding window was set to thirty-two. The variables p_t^o, p_t^h, p_t^l, and p_t^c were set, respectively, as the open prices, the high prices, the low prices, and the close prices of period *t*. W represents the length of the sliding window sizes; therefore, the sliding window frame of period *t*'s open prices is $x_t^o = [p_t^o, p_{t-1}^o, p_{t-2}^o, ..., p_{t-w+1}^o]$. With the same formula applied to the high prices, the low prices, and the close prices, $x_t^h = [p_t^h, p_{t-1}^h, p_{t-2}^h, ..., p_{t-w+1}^h]$, $x_t^l = [p_t^l, p_{t-1}^l, p_{t-2}^l, ..., p_{t-w+1}^l]$, and $x_t^c = [p_t^c, p_{t-1}^c, p_{t-2}^c, ..., p_{t-w+1}^c]$ are obtained, respectively. Thus, the sliding window for the whole period *t* is $x_t = [x_t^o, x_t^h, x_t^l, x_t^c]$. With the sliding window's movement on the time-axis, pieces of pattern can be formed, like the sequence database. Each of these patterns contain $[x_t^o, x_t^h, x_t^l, x_t^c]$ open prices, high prices, low prices, and close prices; and the dimensions for each pattern is 128 (32 ×4).

In the trajectory analysis of the SOM network, the test pattern series uses each clustering pattern to contrast with the minimum mean square error, in order to search for the most similar pattern and to record the pattern serial numbers according to their sequence. Figure 3 is the trajectory analysis concept, where test pattern T_1 to test pattern T_n can contrast with the clustering patterns of the first level, which with the passing of time, also goes from pattern series number c_k to pattern number c_m ($k, m \in 1, 2, ..., 100$). The clustering pattern of the one level SOM can map to the two level SOM c_1' to c_N'. Hence, when a test pattern is in progress, we can find the closest clustering pattern in the one level, reflect it onto the two level, and obtain its serial number.

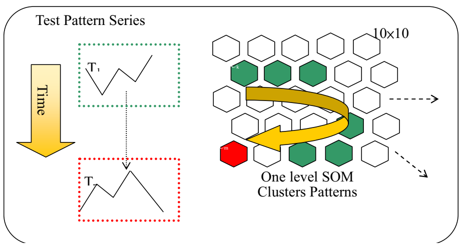

Figure 3. Trajectory analysis concept

3.4 Model of prediction by days

The model only predicts the rising or falling of next day's price with time serious moving. We fit these testing patterns using a 31 trading day window by one level SOM pattern matching. The signal of buying or selling is as follows:

$$Sig(t) = \begin{cases} Buy, if & P(t+1)\text{-}P(t)>0 \\ Sell, if & Otherwise \end{cases}$$

On the other hand, the trading strategy includes three parts: buying strategy, selling strategy, and buying and selling strategy. As follows:

- Buying strategy: It means to buy and hold until the signal show to sell. In this strategy we don't consider short sale strategy.

$$Strategy\_buy(t) = \begin{cases} Buy, if & Sig(t-1) = Sell \quad and\ Sig(t) = Buy \\ Hold, if & Sig(t-1) = Buy \quad and \quad Sig(t) = Buy \\ Sell, if & Sig(t-1) = Buy \quad and \quad Sig(t) = Sell \end{cases}$$

- Selling strategy: It means to sell (short sale) until the signal show to buy. In this strategy we don't consider long trading strategy.

$$Strategy\_sell(t) = \begin{cases} Buy, if & Sig(t-1) = Sell \quad and\ Sig(t) = Buy \\ Hold, if & Sig(t-1) = Sell \quad and \quad Sig(t) = Sell \\ Sell, if & Sig(t-1) = Buy \quad and \quad Sig(t) = Sell \end{cases}$$

- Buying and selling strategy: It includes both long and short trading strategy.

$$Strategy\_sell(t) = \begin{cases} Short \to long\ ,\ if\ Sig(t\text{-}1) = Sell\ and\ Sig(t) = Buy \\ Long \to short,\ if\ Sig(t\text{-}1) = Buy\ and\ Sig(t) = Sell \\ long\ trading,\ if\ Sig(t\text{-}1) = Buy\ and\ Sig(t) = Buy \\ short\ trading,\ if\ Sig(t\text{-}1) = Sell\ and\ Sig(t) = Sell \end{cases}$$

3.5 Model of trend prediction

We use trajectory analysis of two level SOM to divide the market trend in two parts: primary bull market and primary bear market. By this, we can observe the long-term trend of price in Taiwan stock market. The rule is as follows:

$$Trajectory\_Strategy(t) = \begin{cases} Buy,\ if\ \ PatternNumber\ (t) < 4 \\ Sell,\ if\ \ PatternNumber\ (t) \geq 4 \end{cases}$$

4. Experimental Results

This study emphasizes using short trend to produce trading signal of buying or selling. Predicting the trend pattern is more effective than predicting actual closing price because of predicting price can not provide a continuous trend prediction. However predicting the trend pattern can provide a complete support investment strategy.

We extract 10×10 SOM from the experiment data. Figure 4 shows the distribution of ten years time serious data using K-chart in Taiwan stock price weighted index. Figure 5 labels the buying or selling signal of each cluster and according to the cluster pattern, if the price of the thirty-second day is more than that of the thirty-first day, it shows buy (using red "○"); or else it shows sell (using green "×"). The testing trading strategy is as in Figure 6, according to the above-mentioned strategy rule to predict the price trend. Figure 7 is one level SOM profit chart and it shows the comparison between the three strategies and buy and hold strategy in Taiwan stock price weighted index. We can see that the profit of the research model is better than the strategy which buys and holds. From Figure 8 we can find the suitable cluster numbers using hierarchical cluster analysis and agglomeration coefficients. The paper sets five clusters in the two level SOM, as Figure 9 shows. And according to trajectory analysis of the two level SOM patterns, it identifies two parts: primary bull market and primary bear market. Results show that the first and second clusters are both primary bull market, so we set those as buying signals on trend prediction. The other clusters belong to primary bear market, so we set those as selling signals on trend prediction. By this way we can observe the long-term trend. In SOM trajectory analysis, combining time serious and one level SOM trajectory analysis allows us to understand the progress of the pattern. Figure 10 is one level SOM trajectory analysis and we can find the difference of the primary bull market and the primary bear market in it. If the trend belongs to the primary bull market or the primary bear market, the pattern trend only changes a little. But if the pattern trend is severely modified, it means the market trend will change. Figure 11 indicates two -level SOM trajectory analysis and trading strategy. Table 1 reveals all the details of the system performance. Though the win rate of the two level SOM is only fifty percent and it is worse than that of the one level SOM(62.86%). But in other indicators such as average profit, profit factor and PRR, the performance of the two level SOM is better than that of the one level SOM. And the performance of two strategies is better than that of buying and holding strategy.

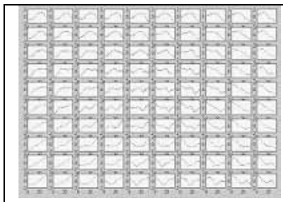

Figure 4. One level SOM K-chart map

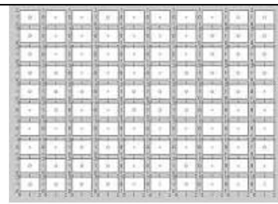

Figure 5. One level SOM clusters buy or sell label

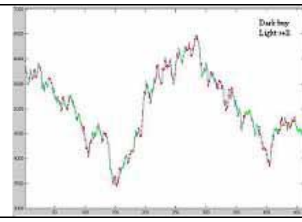

Figure 6. One level SOM trading strategy

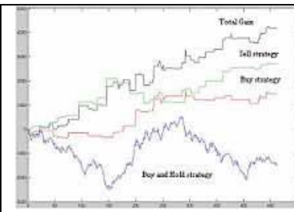

Figure 7. One level SOM profit chart

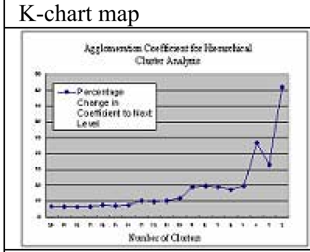

Figure 8. Hierarchical cluster analysis

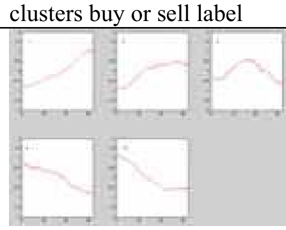

Figure 9. Two level SOM cluster map（K-chart）

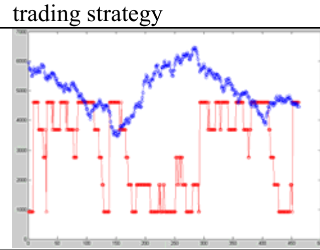

Figure 10. SOM trajectory analysis

Figure 11. Two level SOM trading strategy

Table 1. One level vs. Two level SOM trading strategy results

Trading Strategy	One level Buying Strategy	Two level Buying strategy	One level Selling trategy	Two level Selling trategy	One level Total Gain	Two level Total Gain
Win times	19	2	25	6	44	8
Loss times	16	6	10	2	26	8
Total trades	35	8	35	8	70	16
Win rate	54.29	25.00	71.43	75.00	62.86	50.00
Total profit	1466.90	941.17	2964.45	2438.68	4431.35	3379.85
Avg. profit	41.91	117.65	84.70	304.84	63.31	211.24
Gross gain	3355.30	2239.10	4889.45	2771.08	8244.75	5010.18
Gross loss	-1888.40	-1297.93	-1925.00	-332.40	-3813.40	-1630.33
Avg. gain	176.59	1119.55	195.58	461.85	187.38	626.27
Avg. loss	-118.03	-216.32	-192.50	-166.20	-146.67	-203.79
Avg. G/L	1.50	5.18	1.02	2.78	1.28	3.07
Max. Gain	665.70	2194.10	804.80	1141.90	804.80	2194.10
Max. Loss	-267.80	-497.90	-473.30	-233.70	-473.30	-497.90
Profit factor	1.78	1.73	2.54	8.34	2.16	3.07
PRR	1.83	0.85	2.97	16.84	2.28	3.07

5. Conclusions

Financial investment is a knowledge-intensive industry. Data of financial markets is essentially time-series which brings more challenges than the traditional discrete data for uncovering the hidden knowledge. In this study we explore the knowledge discovery in Taifex Index in Taiwan Futures Exchange in order to identify the important movement trends of stocks. With the outstanding abilities of clustering and visualization provided by the SOM network, a thorough trajectory analysis on the K-chart patterns provides satisfied accuracy of classification of trading signals and the profits of trading as well. In addition, the resulting visualization decision tool can help financial experts verify or refine their tacit investment knowledge offered by the uncovered knowledge.

References

[1] Flexer, A. (2001), "On the ues of self-oraganizing maps for clustering and visualization," *Intelligent Data Analysis*, Vol.5, pp. 373-384.

[2] Irma, B. F., Stelios, H. Z. and Steven, W. (2002), "Knowledge discovery techniques for predicting county investment risk," *Computers & Industrial Engineering*, Vol. 43, pp.787-800.

[3] Jan, M. F., Andrew, S., and Whinston, B. (1999), "A Web-Based Financial Trading System," *IEEE*, pp. 64-70.

[4] Jin, H., Leung, K. S. and Wong, M. L. (2001), " An integrated self-organizing map for the traveling salesman problem," *Advances in Neural Networks and Applications*, pp. 235.240.

[5] JIN, H., Leung, K. S. and Wong, M. L. (2002), "An efficient self-organizing map designed by genetic algorithms for the traveling salesman problem," *Accepted by IEEE Trans*. SMC Part B.

[6] Kohonen, T. (1997), "Self-Organizing Maps," *Springer-Verlag*, New York.

[7] Lee, K.H. and Jo, G.S. (1999), "Expert system for predicting stock market timing using a candlestick chart," *Expert Systems with Applications*, Vol. 16, pp. 357-364.

[8] Leigh, W., Modani, N., Purvis, R. and Roberts, T. (2002), "Stock market trading rule discovery using technical charting heuristics," *Expert Systems with Applications*, Vol. 23, pp.155-159.

[9] Mamaysky, L. A. and Wang, H. (2000), "Foundations of technical analysis: computational algorithms, statistical inference, and empirical implementation," *Journal of Finance*, Vol. 55, pp. 1705–1765.

[10] Shin, T. and Han, I. (2000), "Optimal signal multi-resolution by genetic algorithms to support artificial neural networks for exchange-rate forecasting," *Expert Systems with Applications*, Vol. 18, pp. 257-269.

[11] Smith, K. and Gupta, J. (2000), "Neural networks in business: techniques and applications for the operation researcher," *Computers and Operations Research*, Vol. 27, pp. 1023–1044.

[12] Van den Berg, J., Kaymak, U. and van den Bergh, W. (2004), "Financial markets analysis by using a probabilistic fuzzy modeling approach," *International Journal of Approximate Reasoning*, Vol. 35, pp. 291-305.

[13] Vesanto, J. (2000). "Using SOM in Data Mining. Licentiate's thesis," *Helsinki University of Technology*.

[14] Vesanto, J., Alhoniemi, E. (2000). "Clustering of the self-organizing map," *IEEE Transactions on Neural Networks*, vol. 11 (3), pp. 586-600.

[15] Zopounidis, C., Doumpos, C. M. and Matsatsinis, N. F. (1997), "On the use of Knowledge-based Decision Support Systems in Financial Management: A Survey," *Decision Support Systems*, Vol. 20, pp. 259-277.