



# A hybrid procedure for stock price prediction by integrating self-organizing map and genetic programming

Chih-Ming Hsu \*

Department of Business Administration, Minghsin University of Science and Technology, 1 Hsin-Hsing Road, Hsin-Fong, Hsinchu 304, Taiwan, ROC

## ARTICLE INFO

### Keywords:

Stock price prediction  
Self-organizing map  
Genetic programming

## ABSTRACT

Stock price prediction is a very important financial topic, and is considered a challenging task and worthy of the considerable attention received from both researchers and practitioners. Stock price series have properties of high volatility, complexity, dynamics and turbulence, thus the implicit relationship between the stock price and predictors is quite dynamic. Hence, it is difficult to tackle the stock price prediction problems effectively by using only single soft computing technique. This study hybridizes a self-organizing map (SOM) neural network and genetic programming (GP) to develop an integrated procedure, namely, the SOM-GP procedure, in order to resolve problems inherent in stock price predictions. The SOM neural network is utilized to divide the sample data into several clusters, in such a manner that the objects within each cluster possess similar properties to each other, but differ from the objects in other clusters. The GP technique is applied to construct a mathematical prediction model that describes the functional relationship between technical indicators and the closing price of each cluster formed in the SOM neural network. The feasibility and effectiveness of the proposed hybrid SOM-GP prediction procedure are demonstrated through experiments aimed at predicting the finance and insurance sub-index of TAIEX (Taiwan stock exchange capitalization weighted stock index). Experimental results show that the proposed SOM-GP prediction procedure can be considered a feasible and effective tool for stock price predictions, as based on the overall prediction performance indices. Furthermore, it is found that the frequent and alternating rise and fall, as well as the range of daily closing prices during the period, significantly increase the difficulties of predicting.

© 2011 Elsevier Ltd. All rights reserved.

## 1. Introduction

Stock price prediction is an important financial subject, which has received considerable attention from researchers in recent years. Stock price prediction is considered a challenging task in consideration of its high volatility, complexity, dynamics, and turbulence. In the past, many attempts have been made to predict stock prices using various methodologies, which can be broadly classified into three categories, namely, fundamental analysis, technical analysis, and traditional time series forecasting. Fundamental analysis examines the basic financial information of a corporation in order to forecast profits, supply, demand, industry strength, management abilities, and other intrinsic matters affecting the market value and growth potential of a stock (Thomsett, 1998). In fundamental analysis, investors believe that the fundamentals include a corporation's financial statements, interim reports, historical financial trends, and any forecasts concerning future growth, sales, profits, etc., should rule the processes of the selection of stocks and timing of sales (Thomsett, 1999). However,

technical analysis studies the stock prices and related issues, including analysis of recent and historical price trends, cycles and factors beyond the stock price, such as dividend payments, trading volume, index trends, industry group trends and popularity, and volatility of a stock (Thomsett, 1999). Technical analysis, rather than relying solely upon historical financial information, analysts will surmise upon recent trends in stock price changes, prices and earnings relationships, the activity volume of a particular stock or industry, and other similar indicators in order to determine changes in stocks, and in the market itself (Thomsett, 1999). In addition, traditional time series forecasting techniques, such as autoregressive integrated moving average (ARIMA) (Box & Jenkins, 1970), generalized autoregressive conditional heteroskedasticity (GARCH) (Bollerslev, 1986), and multivariate regression have been applied to the prediction of stock price movements. In recent years, data mining/computational intelligence techniques have become another important approach to predict stock prices. For example, Kim and Han (2000) utilized genetic algorithms (GAs) to discretize features and determine the connection weights of artificial neural networks (ANNs), thus, predicting the stock price index. Experiments conducted on the daily Korea stock price index (KOSPI) showed that, their proposed approach outperformed the linear

\* Tel.: +886 3 5593142x3581; fax: +886 3 5593142x3573.

E-mail address: [cmhsu@must.edu.tw](mailto:cmhsu@must.edu.tw)

transformation functions of both a backpropagation neural network (BPLT) and a linear transformation with ANN, as trained by GA (GALT). Kim (2003) applied a support vector machine (SVM) to predict the stock price index, and the feasibility of applying SVM to financial forecasting was examined through comparisons with a backpropagation neural network (BPNN) and case-based reasoning (CBR). The experimental results of the daily Korea stock price index (KOSPI) investigation showed that, SVM provides a promising alternative for financial time series forecasting; moreover, it outperforms both BPNN and CBR approaches. Pai and Lin (2005) proposed a hybrid methodology through exploitation of the strengths of the autoregressive integrated moving average (ARIMA) and support vector machine (SVM) in order to forecast stock prices. The performance of the proposed model is evaluated by testing real data sets of ten stocks, and adequate results are obtained. Tsang et al. (2007) presented a stock buying/selling alert system using a feed-forward backpropagation neural network, called NN5. The system is tested with data from The Hong Kong and Shanghai Banking Corporation (HSBC) Holdings stock, located in Hong Kong, and achieved an overall hit rate of over 70%. Chang and Liu (2008) presented a Takagi–Sugeno–Kang (TSK) type fuzzy rule based system by applying a linear combination consequence of the significant technical index in order to predict stock prices. Their proposed approach was tested on the Taiwan Stock Exchange (TSE) and MediaTek Inc., and the experimental results outperformed other methodologies, such as a back-propagation neural network and multiple regression analysis. Ince and Trafalis (2008) assumed that the future value of a stock price depends on its financial indicators, although there is no existing parametric model able to explain the relationship coming from the technical analysis. Hence, they proposed two nonparametric data driven models, a support vector regression (SVR) and a multi-layer perceptron (MLP), for short term stock price predictions based on technical indicators. The experiments were conducted on the daily stock prices of ten companies traded on the NASDAQ, and comparison results indicated that the SVR approach outperformed the MLP networks in short term predictions, in terms of the mean square error. Huang and Tsai (2009) proposed a hybrid procedure using support vector regression (SVR), self-organizing feature map (SOFM), and filter-based feature selection in order to predict the stock market price index. Their proposed model was demonstrated through a case study of predictions of the next day's price index for Taiwan index futures (FITX), and the experiment results showed that the proposed approach can improve prediction accuracy and reduce the training time over the traditional single SVR model. Lai, Fan, Huang, and Chang (2009) proposed a decision-making system that integrates a data clustering technique, a fuzzy decision tree, and genetic algorithms in order to forecast stock price tendencies. Three particular stocks in the Taiwan Stock Exchange Corporation (TSEC) were selected to test the effectiveness of their proposed system, which yielded the best performance of an 82% average hit rate, in comparison with other approaches. Liang, Zhang, Xiao, and Chen (2009) presented a nonparametric methodology based on neural networks (NNs) and support vector regression (SVR) to forecast option prices. In their study, the improved conventional option pricing methods were modified to forecast the option prices, and then, the NN and SVR were further employed to decrease the forecasting errors of the parametric methods. The proposed approach was demonstrated by experimental studies upon data taken from the Hong Kong options market, which results showed that the NN and SVR approaches can significantly shrink the average forecast errors, thus, improving forecasting accuracy. Lee (2009) developed a model based on a support vector machine (SVM) with a hybrid feature selection, namely, *F*-score and supported sequential forward search (F\_SSFS), to predict the trends of stock markets. The experiments of predicting the NASDAQ index

direction were used to illustrate their proposed method, and suitable results were obtained. In addition, comparisons with information gain, symmetrical uncertainty, and correlation-based feature selection methods all indicated that their proposed model could yield the highest levels of accurate and generalized performances. Yu, Chen, Wang, and Lai (2009) presented an evolving least squares support vector machine (LSSVM) learning paradigm, with a mixed kernel based on genetic algorithms (GAs), in order to predict the trends of stock markets. The GAs were used to select the input features and optimize parameters of LSSVM. The LSSVM approach was illustrated through testing the S&P 500 index, the Dow Jones Industrial Average (DJIA) index, and New York Stock Exchange (NYSE) index, and experimental results revealed that their proposed learning paradigm was more efficient than other parameter optimization methods, and outperformed all other forecasting models in terms of the hit ratio. Zhang, An, Tang, and Hong (2009) proposed a type-2 fuzzy rule based expert system that applied technical and fundamental indices as the input variables for the analysis of stock prices. Their proposed model was tested on the stock price predictions of an automotive manufactory in Asia, and successful results were obtained.

In this study, an integrated approach based on a self-organizing map (SOM) neural network and genetic programming (GP), namely, the SOM-GP procedure, is proposed for predicting stock prices. The remainder of this paper is organized as follows: In Section 2, SOM and GP are discussed. The proposed integrated approach is presented in Section 3. Section 4 evaluates the feasibility and effectiveness of the proposed approach by a case study of predicting the finance and insurance sub-index of TAIEX. Finally, Section 5 concludes the paper.

## 2. Self-organizing map and genetic programming

### 2.1. Self-organizing map

The self-organizing map (SOM) was first introduced by Kohonen (1989), as an unsupervised and competitive learning neural network able to map a high-dimensional input data space into a lower-dimensional (typically one- or two-dimensional) space. The end-product is called a feature map able to preserve the most important topological relationships of the input data. The typical SOM consists of two layers, as shown in Fig. 1, where the input

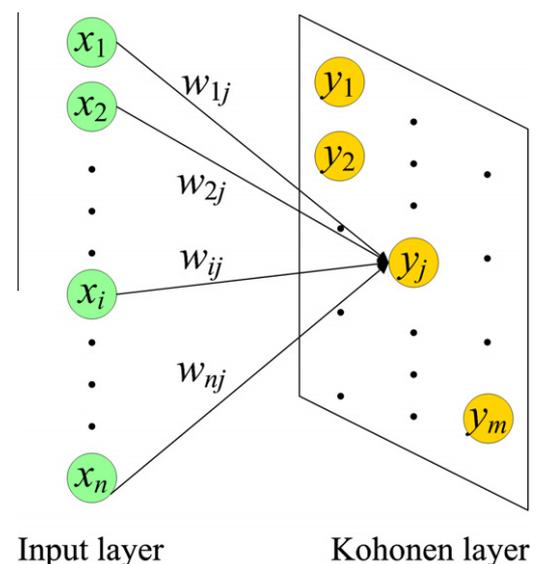


Fig. 1. Typical SOM topology consists of two layers.

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات