



A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting

Cheng-Lung Huang*, Cheng-Yi Tsai

Department of Information Management, National Kaohsiung First University of Science and Technology, 2, Juoyue Road, Nantz District, Kaohsiung 811, Taiwan, ROC

Abstract

Stock market price index prediction is regarded as a challenging task of the financial time series prediction process. Support vector regression (SVR) has successfully solved prediction problems in many domains, including the stock market. This paper hybridizes SVR with the self-organizing feature map (SOFM) technique and a filter-based feature selection to reduce the cost of training time and to improve prediction accuracies. The hybrid system conducts the following processes: filter-based feature selection to choose important input attributes; SOFM algorithm to cluster the training samples; and SVR to predict the stock market price index. The proposed model was demonstrated using a real future dataset – Taiwan index futures (FITX) to predict the next day's price index. The experiment results show that the proposed SOFM-SVR is an improvement over the traditional single SVR in average prediction accuracy and training time. © 2007 Elsevier Ltd. All rights reserved.

Keywords: Support vector regression; Self-organizing feature map; Feature selection; Prediction

1. Introduction

Data mining is a rapidly-growing technology in the information processing industry. It has been applied to various disciplines, such as engineering, business, management, science, as well as the military. In the financial domain, data mining can be used to assist with the prediction of stock prices, credit scores, and even bankruptcy potential. Stock market price prediction is regarded as a challenging task of the financial time series prediction process since the financial market is a complex, evolutionary, and non-linear dynamic system (Abu-Mostafa & Atiya, 1996). In the last decade many studies have been conducted in mining financial time series data, including traditional statistical approaches and data mining techniques. In the area of financial stock market forecasting, many studies focused on application of Artificial Neural Networks (ANNs); the ANNs have been reported as good methods to predict

financial stock market levels (Kim & Han, 2000; Kimoto, Asakawa, Yoda, & Takeoka, 1990; Quah & Srinivasan, 1999; Roman & Jameel, 1996; Tsaih, Hsu, & Lai, 1998).

Recently, the support vector machine (SVM) method, which was first suggested by Vapnik (1995), has recently been used in a range of applications, including financial stock market prediction (Cao, 2003; Huang, Nakamori, & Wang, 2005; Kim, 2003; Tay & Cao, 2001). The SVM technique, in general, is widely regarded as the state of art classifier and previous researches indicated that SVM prediction approaches are superior to neural networks approaches (Chen & Shih, 2006; Huang, Chen, & Wang, 2007; Huang, Chen, Hsu, Chen, & Wu, 2004; Lee, 2007). Initially developed for solving classification problems, SVM techniques can be successfully applied in regression (i.e., for a functional approximation problems) (Drucker, Burges, Kaufman, Smola, & Vapnik, 1997; Vapnik, Golowich, & Smola, 1997). Unlike pattern recognition problems where the desired outputs are discrete values (e.g., Boolean), the support vector regression (SVR) deals with 'real valued' functions. The SVR is derived from the structural risk minimization principle to estimate a function by

* Corresponding author. Tel.: +886 7 6011000x4127; fax: +886 7 6011042.

E-mail address: clhuang@ccms.nkfust.edu.tw (C.-L. Huang).

minimizing an upper bound of the generalization error (Huang et al., 2005; Schölkopf, Burges, & Vapnik, 1995). Previous researches reported the SVR model has successfully solved prediction problem in many domains (Kim, 2003; Pai & Lin, 2005; Thissen, Van Brakel, De Weijer, Melssen, & Buydens, 2003). However, improving the prediction accuracy still remains a prime issue of concern in the area of prediction. Especially for stock market prediction, even a slight improvement on prediction accuracy could have a positive impact on the profit of investment.

It was reported that a hybrid system in prediction and classification achieved a higher performance level against the traditional system (Brabazon & Keenan, 2004; Chang & Chang, 2006; Kim & Shin, 2007; Lekakos & Giaglis, 2007; Min, Lee, & Han, 2006). In order to improve the prediction accuracy of the traditional SVR method and to reduce its long training time, this study proposes a hybrid SOFM-SVR which combines SVR with self-organizing feature map (SOFM). Combining SOFM with classification or prediction algorithms has successfully solved many classification and prediction problems such as with ANN (Li, Zhang, & Wang, 2006; Wang & Yan, 2004; Wong, Wong, Gedeon, & Fung, 2003), SVM (Wu, Liu, Xu, Peng, & Rudy, 2004), SVR (Cao, 2003; Martín-Merino & Román, 2006; Tay & Cao, 2001), and with other methods (Chang & Liao, 2006; Chen, Grant-Muller, Mussone, & Montgomery, 2001). According to the research of Tay and Cao (2001), the approach combining SOFM with SVR can successfully predict stock market activity. In this study, the SOFM-SVR is combined with the filter-based feature selection to predict Taiwan index futures (FITX) that are collected from the Taiwan futures exchanges. By combining the SOFM-SVR with filter-based feature selection, an improvement in training time, prediction accuracy, and the ability to select a better feature subset is achieved. To train a SOFM-SVR model, a two-stage process was performed. In the first stage, the SOFM algorithm clusters the training data into several disjointed clusters; the individual SVR model for each cluster is then constructed in the second stage. The multiple SVR models provide various patterns in predicting the stock market, so the SOFM-SVR is able to cope with the fluctuation of stock market values. Using the filter-based feature selection method, the important feature subset can be selected. With a proper feature subset, the data complexity can be reduced and prediction accuracy can be improved.

This paper is organized as follows: Section 2 describes the basic concepts of SVR and SOFM. Section 3 describes the proposed SOFM-SVR with a filter-based feature selection model. Section 4 presents the experimental results of predicting the FITX. Section 5 provides remarks and conclusions.

2. Brief introduction to SVR and SOFM

2.1. Support vector regression

In this section the basic SVR concepts are briefly described which can also be found in Schölkopf and Smola

(2000), Cristianini and Taylor (2000) and Kecman (2001). Given a training set (\mathbf{x}_i, y_i) , $i = 1, 2, \dots, m$, where the input variable $\mathbf{x}_i \in R^n$ is the n -dimensional vector; and the response variable $y_i \in R$ is the continuous value. SVR builds the linear regression function as the following form:

$$f(\mathbf{x}, \mathbf{w}) = \mathbf{w}^T \mathbf{x} + b \quad (1)$$

Based on the Vapnik's linear ε -Insensitivity loss (error) function [see Eq. (2)], the linear regression $f(\mathbf{x}, \mathbf{w})$ is estimated by simultaneously minimizing $\|\mathbf{w}\|^2$ and the sum of the linear ε -Insensitivity losses [as shown in Eq. (4)] for the SOFM value of parameter C . The constant C , which influences a trade-off between an approximation error and the weights vector norm $\|\mathbf{w}\|$, is a design parameter chosen by the user.

$$|y - f(\mathbf{x}, \mathbf{w})|_\varepsilon = \begin{cases} 0, & \text{if } |y - f(\mathbf{x}, \mathbf{w})| \leq \varepsilon \\ |y - f(\mathbf{x}, \mathbf{w})| - \varepsilon, & \text{otherwise} \end{cases} \quad (2)$$

$$R = \frac{1}{2} \|\mathbf{w}\|^2 + C \left(\sum_{i=1}^m |y_i - f(\mathbf{x}_i, \mathbf{w})|_\varepsilon \right) \quad (3)$$

Minimizing the risk R is equivalent to minimizing the following risk:

$$R_{\mathbf{w}, \zeta, \zeta^*} = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m (\zeta + \zeta^*) \quad (4)$$

under constraints:

$$(\mathbf{w}^T \mathbf{x}_i + b) - y_i \leq \varepsilon + \zeta_i \quad (5)$$

$$y_i - (\mathbf{w}^T \mathbf{x}_i + b) \leq \varepsilon + \zeta_i^* \quad (6)$$

$$\zeta_i, \zeta_i^* \geq 0, \quad i = 1, 2, \dots, m \quad (7)$$

where ζ_i and ζ_i^* are slack variables, one for exceeding the target value by more than ε , and the other for being more than ε below the target. As with procedures applied to SVM classifiers (Cristianini & Taylor, 2000; Kecman, 2001; Schölkopf & Smola, 2000), this constrained optimization problem is solved by applying the Lagrangian theory and the Karush Kuhn-Tucker condition to obtain the optimal desired weights vector of the regression function.

The non-linear SVR maps the training samples from the input vectors into a higher-dimensional feature space via a mapping function Φ . By performing such a mapping method, in the feature space the learning algorithm will be able to obtain a linear regression function by applying the linear regression SVR formulation. In the final expression for a predictor function, training data only appears in the form of scalar products $\mathbf{x}_i^T \mathbf{x}_j$ which are replaced by scalar products $\Phi^T(\mathbf{x}_i)\Phi(\mathbf{x}_j)$. The scalar product $\Phi^T(\mathbf{x}_i)\Phi(\mathbf{x}_j)$ is calculated directly by computing a kernel function, $K(\mathbf{x}_i, \mathbf{x}_j)$, that is, $\Phi^T(\mathbf{x}_i)\Phi(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j)$, to avoid having to perform a mapping $\Phi(\mathbf{x})$. The most popular kernel function is Radial Basis Function (RBF), as shown in Eq. (8).

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \quad (8)$$

2.2. Self-organizing feature map

The SOFM proposed by Kohonen (Kohonen, 1989, 1997; Kohonen, Raivio, Simula, Venta, & Henriksson, 1990) is an unsupervised and competitive learning algorithm. The learning algorithm spatially forms the topological maps in which the relations of similarity among the input samples (i.e., high-dimensional space) are converted into spatial relationships among the neurons (i.e., two-dimensional grid) that respond to these inputs. Kohonen's learning algorithm updates not only the weights of the winning neuron, but also those of the spatially-close neurons, based on the activation zone for each neuron. The training steps, including competitive and weight adaptation processes, are repeated until the stopping criteria are met as follows:

1. Introduce an input vector \mathbf{x} into the net.
2. In the competitive process, compute Eq. (9) (the distance between input vector and all neurons of output layer), and find the winning neuron.

$$\|\mathbf{x} - \mathbf{w}_i\| \quad i = 1, 2, \dots, \ell \quad (9)$$

where ℓ is the number of output neurons; \mathbf{w} is the weight vector.

3. In the weight adaptation process, update the new weights at time $t + 1$ of the winning neuron and its neighborhood neurons defined by the activation zone.

$$\mathbf{w}_i(t + 1) = \mathbf{w}_i(t) + \eta(t)\pi_{i,i^*}(t)(\mathbf{x}(t) - \mathbf{w}_i(t)) \quad (10)$$

where $\eta(t)$ is the learning rate and $\pi_{i,i^*}(t)$ is the topological neighborhood depending on lateral distance between the neuron i and the winner neuron i^* . A Gaussian-like neighborhood function centered at winning neuron i^* is defined by Eq. (11):

$$\pi_{i,i^*}(t) = \exp\left(\frac{-\|\mathbf{p}_i - \mathbf{p}_{i^*}\|}{g(t)}\right) \quad (11)$$

where the \mathbf{p}_i is the position vector of the neighborhood neuron, the \mathbf{p}_{i^*} is the winning neuron, and $g(t)$ is a parameter which is gradually decreased.

3. The proposed system framework

3.1. Overview of the SOFM-SVR model

In this study, a two-stage hybrid method was implemented which combines the SVR prediction capability with the SOFM clustering algorithm, as illustrated in Fig. 1. In the first stage, SOFM clusters the training data into several disjointed clusters. Each cluster contains similar objects. In the second stage, an individual SVR model for each cluster is constructed. In order to construct an accurate SVR model, the kernel function and loss function must be selected. Properly setting the parameters can improve the

SVR prediction accuracy. The parameter gamma (γ) of the RBF kernel, the parameter ε of the loss function, and the soft margin constant C (i.e., penalty parameter) must be optimized to improve the prediction accuracy.

This study used the RBF kernel because it can analyze high-dimensional data and requires only two parameters, a soft margin constant (C) and a gamma (γ) (Hsu, Chang, & Lin, 2003). The Grid algorithm (Hsu et al., 2003) is adapted to find the best parameters of C , γ , and ε . The detailed steps to train and test the two-stage SOFM-SVR model, as shown in Fig. 2, are described in the next two sections.

3.2. Training the hybrid SOFM-SVR model

In this study, SVR is combined with SOFM to construct the predicted model. In order to guarantee valid results for making predictions regarding new data, the data set was randomly partitioned into a training set and an independent test set. The training set adjusts the model prediction parameters, while the test set evaluates the prediction accuracy of the trained model. The SOFM-SVR model is established according to the following steps:

1. Scaling the training set: All input variables are scaled during the data preprocessing stage. The two main advantages of scaling are to avoid attributes in greater numeric ranges from dominating those in smaller numeric ranges, and to prevent numerical difficulties during the calculation (Hsu et al., 2003). Feature value scaling can help increase SVR accuracy according to experimental results of this study. Generally, each feature can be linearly scaled to the $[0, 1]$ range using the following formula [see Eq. (12)], where x is original value, x' is the scaled value, \max_a is the maximum value of feature a , and \min_a is the minimum value of feature a .

$$x' = \frac{x - \min_a}{\max_a - \min_a} \quad (12)$$

2. Clustering the training dataset (first stage): Cluster property includes objects within a cluster that have high similarity and objects in other clusters that are very dissimilar (Han & Kamber, 2001). This study used the unsupervised neural network SOFM to cluster the training data. An advantage of SOFM over other clustering algorithms is its ability to visualize high-dimensional data using a two-dimensional grid while preserving the similarity between data points as much as possible (Roussinov & Chen, 1999). Another positive aspect is that the SOFM clustering algorithm is able to determine the number of clusters under a pre-defined two-dimensional grid. The clustering performed in this study was based on the scaled training data. After the training data was clustered, the SVR model was constructed for each cluster in the next step.

21, 1998, which is based on the Taiwan stock exchange (TAIEX) value-weighted index.

The FITX data set covers a six-year period, from January 4, 2000 to February 20, 2006. A total of 1,540 pairs of daily observations were made, as shown in Fig. 5. The data set was split into five subsets by slicing the time period. Each subset contains a training period around five years (i.e., 1,340 cases) and a testing period around two months (i.e., 40 cases), as shown in Table 1. A prediction model was built for each of the five data sets, and the average performance for the five data sets was then used to evaluate the reliability of the proposed approach.

4.2. Input variables consideration

Thirteen technical indicators are considered as input variables to forecast the daily price in the stock price index. Previous researches indicated SOFM variables are relevant for predicting and classifying the stock price index including the relative strength index (RSI), the moving average convergence and divergence (MACD), moving average (MA), William's oscillator percent R (WR), psychological line (PSY), stochastic K index, and stochastic D index (Achelis, 1995; Kim, 2003; Kim & Lee, 2004; Klassen, 2005). Other SOFM input variables suggested by the domain experts were adapted in this study including directional indicator up (+DI), directional indicator down (−DI), BIAS, volume ratio (VR), A ratio (AR) and B ratio (BR). The descriptions and definitions of these input variables are presented in Appendix A.

4.3. Performance measures

To demonstrate the effectiveness of the proposed SOFM-SVR model, this study used the statistical metrics of the mean square error (MSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). The definitions of MSE, MAE and MAPE are provided in Eqs. (14)–(16), respectively.

$$MSE = \frac{\sum_{i=1}^n (A_i - F_i)^2}{n} \quad (14)$$

$$MAE = \frac{\sum_{i=1}^n |A_i - F_i|}{n} \quad (15)$$

$$MAPE = \frac{\sum_{i=1}^n |A_i - F_i| / A_i}{n} \times 100 \quad (16)$$

Table 1

Dividing the FITX data set into five experimental sub-datasets

Dataset	Training period	Test period
1	2000/01/05–2005/05/04	2005/05/05–2005/06/29
2	2000/03/04–2005/06/29	2005/06/30–2005/08/26
3	2000/04/27–2005/08/26	2005/08/29–2005/10/25
4	2000/06/19–2005/10/25	2005/10/26–2005/12/20
5	2000/08/09–2005/12/20	2005/12/21–2006/02/22

where A_i is the actual value of sample i , F_i is a predicted value of sample i , and n is the number of samples.

4.4. Experimental results

The implementation platform was carried out via Matlab 7.0, a mathematical development environment, by extending the Libsvm version 2.82, which was originally designed by Chang and Lin (2001). The empirical evaluation was performed on a Pentium IV CPU running at 2.8 GHz with 256 MB RAM. The experimental results are shown in the sections that follow.

4.4.1. Determining the number of clusters

Models based on the maximum numbers of 3, 4, and 5 clusters in the SOFM are constructed to investigate the impacts of the number of clusters on the performances. Fig. 6 shows the predicted value of the SOFM-SVR with various numbers of clusters in data set number 1. The average prediction errors, shown in Table 2, are computed by averaging the prediction errors of the five data sets. The average MAPE errors for the three models are 3.59002%, 3.90044%, and 4.3148%, respectively. It was determined that the model of three clusters is the best according to the criteria of MAPE as well as to the MSE and MAE. To precisely compare the prediction error of the test set, a nonparametric Wilcoxon signed rank test (i.e., a test for the two-related samples) was performed. Since the population distributions of the performance measures are unknown, a nonparametric test is suggested for the performance comparison of the two models (Conover, 1980). Additionally, the experiments used identical training and test sets for each “treatment” (i.e., model), thus the test set is not independent (Salzberg, 1997). Therefore, the nonparametric with dependent samples was adapted to compare the performances. As shown in Table 3, the model

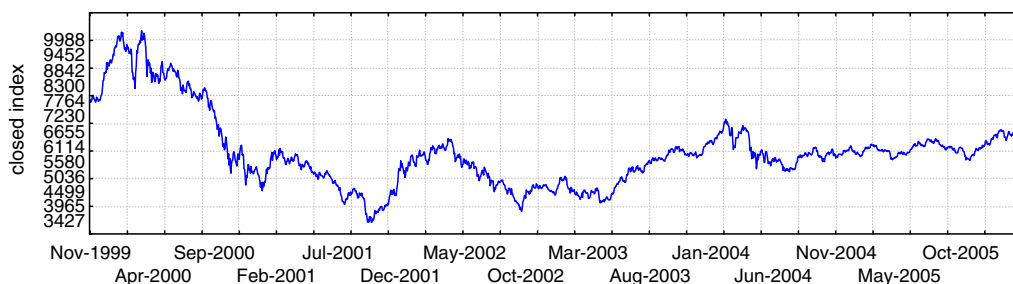


Fig. 5. The FITX from November, 1999 to February, 2006.

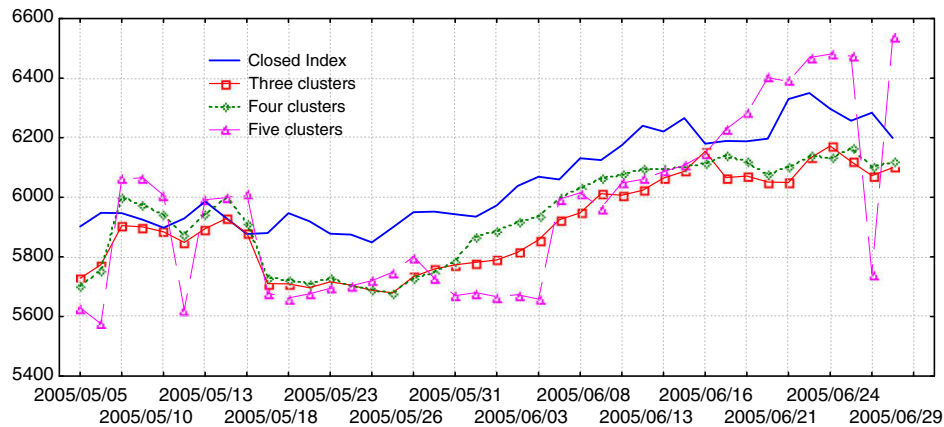


Fig. 6. The predicted value of the SOFM-SVR with various numbers of clusters in dataset #1.

Table 2

The accuracy measures with various numbers of clusters

Number of clusters	MSE	MAE	MAPE (%)	Time (min)
3 Clusters	74260.73 \pm 64047.00	223.5331 \pm 115.96	3.59002 \pm 1.85	419.9717 \pm 26.60
4 Clusters	91781.47 \pm 78616.04	243.6887 \pm 131.12	3.90044 \pm 2.07	446.2202 \pm 13.46
5 Clusters	99859.31 \pm 68100.27	268.6646 \pm 107.51	4.3148 \pm 1.70	338.4418 \pm 28.67

with three clusters significantly outperforms the others (with 0.005 significance), in MSE, MAE and MAPE, based on the nonparametric Wilcoxon signed rank tests with two-related (dependent) samples. That is, for the cases of the FITX data set, the SOFM-SVR model with three clusters is more appropriate than other models with four or five clusters, based on the criteria of minimizing the prediction errors. The detail prediction errors of the model with three clusters are illustrated in Table 4.

4.4.2. SOFM-SVR with filter-based feature selection

The models of SOFM-SVR with filter-based feature selection for each data set were built based on three clusters determined in the previous section. The prediction errors for each model are shown in Table 5. It was determined the model with feature selection is more accurate than the model without feature selection based on the average MAPE errors for the two models: 1.7726% from Table 5 and 3.59002% from Table 4, respectively. This is also true from the results of the Wilcoxon signed rank test on the prediction errors, as shown Table 6. Each dataset has a different selected feature subset. Fig. 7 illustrates the features' relative importance (i.e., frequency) for the selected fea-

Table 3

Wilcoxon sign rank test on the prediction errors for the SOFM-SVR with various numbers of clusters

Accuracy criteria	3 Clusters vs. 4 clusters	3 Clusters vs. 5 clusters	4 Clusters vs. 5 clusters
MSE	0.345	0.043 ^a	0.5
MAE	0.345	0.043 ^a	0.138
MAPE	0.345	0.043 ^a	0.138

^a Achieving 0.05 significance (2-tailed).

Table 4

Results of SOFM-SVR using three clusters

Dataset	MSE	MAE	MAPE (%)	Time (min)
1	26705	148.27	2.4447	422.91
2	189930	428.34	6.8011	451.95
3	95206	266.62	4.4282	413.56
4	14489	99.40	1.6289	373.63
5	44974	175.03	2.6472	437.80
Average	74260.73	223.53	3.59002	419.97
Deviation	64047.00	115.96	1.847173	26.60

tures for each data set. For the case of the FITX data set, MA10, MACD9, PSY10, D10, K10, +DI10, and -DI10 are important input attributes.

4.4.3. Comparison with the traditional single SVR

This study compared the prediction accuracy of the SOFM-SVR method with that of the traditional single SVR. The difference between SOFM-SVR and the traditional single SVR is that the former built a SVR model for each cluster, and the latter built only the single SVR model for the whole training set. More specifically, the same RBF kernel, ϵ -insensitive loss function, and grid

Table 5

Results of SOFM-SVR with selected features

Dataset	MSE	MAE	MAPE (%)	Time (min)
1	22418	118.52	1.9338	428.99
2	41091	164.32	2.6168	384.38
3	9289	78.96	1.3238	389.62
4	10432	79.67	1.3143	383.24
5	15492	110.95	1.6743	324.46
Average	19744.40	110.48	1.7726	382.14
Deviation	11635.67	31.33	0.481619	33.42

Table 6
Wilcoxon sign rank test on the prediction errors: SOFM-SVR with selected features vs. SOFM-SVR with all features

Prediction error	MSE	MAE	MAPE
SOFM-SVR with selected features vs. SOFM-SVR with all features	0.043115 ^a	0.043115 ^a	0.043115 ^a

^a Achieving 0.05 significance (2-tailed).

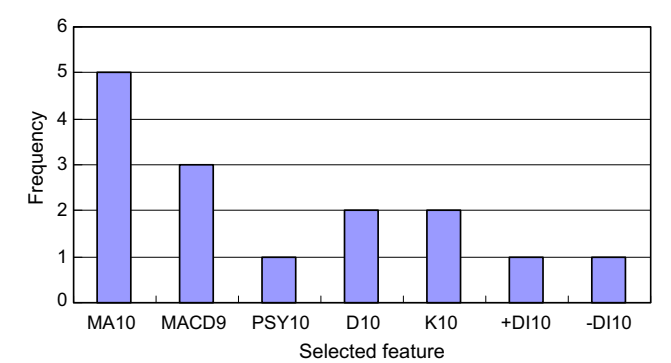


Fig. 7. Relative importance of the selected features.

Table 7
Wilcoxon sign rank test: SOFM-SVR vs. single SVR

	MSE	MAE	MAPE	Time
SOFM-SVR with selected features vs. single SVR with selected features	0.224917	0.224917	0.224917	0.043115 ^a
SOFM-SVR with selected features vs. single SVR with all features	0.500185	0.685831	0.685831	0.043115 ^a

^a Achieving 0.05 significance.

search approach were used in the single SVR. In addition, the experiment for the single SVR is conducted on the five data sets that are the same as the data sets of the SOFM-SVR in order to have a fair basis of comparison.

The prediction errors of the single SVR with selected features and that with all features are shown in Table 8 and 9, respectively. The average MAPE errors for the two models are 3.47252% from Table 8 and 3.5424% from Table 9, respectively. The SOFM-SVR with feature selection (i.e., MAPE of 1.7726% from Table 5) is more accurate than the single SVR with feature selection and that without feature selection. Although the Wilcoxon signed

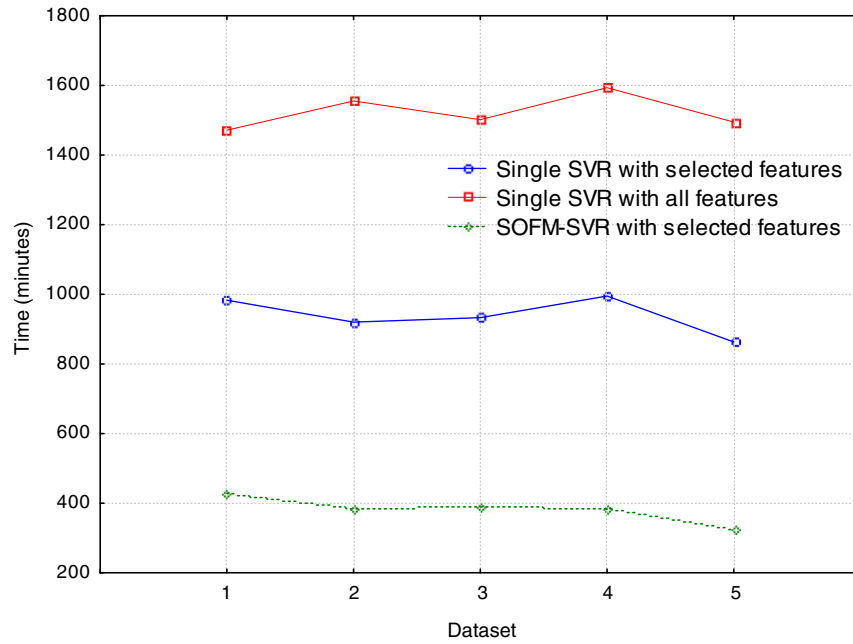


Fig. 9. Training time comparisons: SOFM-SVR vs. single SVRs.

rank test on the prediction errors are not significant enough, as shown in Table 7, the average MSE, MAE, and MAPE errors of the SOFM-SVR are superior to that of the single SVR with and without feature selection. Fig. 8 shows the MAPEs for the three models. For the stock market prediction, even a slight improvement on prediction accuracy could have a positive impact on the profit of investment. This study showed the SOFM-SVR improved the prediction accuracy against the traditional single SVR.

The average training time for the SOFM-SVR with selected features, single SVR with selected features, and single SVR with all features are 382.14 (see Table 5), 938.77 (see Table 8), and 1523.41 min (see Table 9), respectively. The training times shown in Tables 5 and 8 are collected by constructing the SOFM-SVR model and single SVR model with the selected feature subset. Fig. 9 shows the training time for the three models. The SOFM-SVR significantly outperforms the single SVR according to the Wilcoxon signed rank test on the training times, as shown in Table 7. One disadvantage of SVM is that the training time scales somewhere between quadratic and cubic with respect to the number of training samples. Thus, a large amount of computation time will be involved when SVM is applied for solving large problems (Cao & Tay, 2003; Deng, Jin, & Zhong, 2005). By dividing the samples into smaller clusters, the training time for each cluster can be reduced.

5. Conclusion

This study introduced a hybrid SOFM-SVR with filter-based feature selection to improve the prediction accuracy and to reduce the training time for the financial daily stock index prediction. The SOFM algorithm clusters the train-

ing data into several disjointed clusters. The individual SVR model for each cluster is then constructed.

According to the empirical experiments in forecasting daily FITX, the SOFM-SVR improves the average prediction accuracy, and thus may lead to a capital gain. Additionally, the SOFM-SVR model reduces the training time. Because the training time of the SVR is determined by the size of training samples, dividing the training samples into clusters of small sample sizes reduces the training time at each cluster, and thus saves a large amount of training time.

The proposed approach with multiple SVR models provides various patterns in predicting the daily FITX, so it was able to cope with the fluctuation of stock market values. It also yielded good prediction accuracy. In addition, with fewer and more appropriate input variables generated by a filter-based feature selection approach, the SOFM-SVR model can maintain its accuracy and shorten the training time. Further research directions are using optimization algorithms (e.g., genetic algorithms) to optimize the SVR parameters and performing feature selection using a wrapper-based approach that combines SVR with other optimization tools (e.g., genetic algorithms and ant colony optimization).

Appendix A. The descriptions and definition of input attributes

1. MA10: 10-day moving average.

$$MA(n)_t = \frac{1}{n} \times \sum_{i=t-n+1}^t C_i \quad (1)$$

where n is the length of time period, C_i is the close price at time i , and t is current day.

2. RSI5: 5-day relative strength index.

$$RSI(n)_t = \left(\frac{UPC_n / UD_n}{UPC_n / UD_n + DPC_n / DD_n} \right) \times 100 \quad (2)$$

where UD_n is the number of upward days (during previous n days period), DD_n is the number of downward days, UPC_n is the summation of close prices at upward days, and DPC_n is the summation of close prices at downward days.

3. K9: 9-day stochastic index K

$$K(n)_t = \frac{2}{3} \times K(n)_{t-1} + \frac{1}{3} \times \frac{C_t - HP_n}{HP_n - LP_n} \quad (3)$$

where HP_n is the highest high price (in previous n days period), and LP_n is the lowest low price.

4. D9: 9-day stochastic index D

$$D(n)_t = \frac{2}{3} \times D(n)_{t-1} + \frac{1}{3} \times K(n)_t \quad (4)$$

5. MACD9: 9-day moving average convergence/divergence.

$$MACD(n)_t = MACD(n)_{t-1} + \frac{2}{n+1} \times (DIFF_t - MACD(n)_{t-1}) \quad (5)$$

where $DIFF_t = EMA(12)_t - EMA(26)_t$ is represented as the difference between a 26-day and 12-day exponential moving average. $EMA(k)_t = EMA(k)_{t-1} + \alpha \times (C_t - EMA(k)_{t-1})$, where $\alpha = 2/(k+1)$, k is time period of the k -day exponential moving average.

6. WR10: 10-day Larry Williams' oscillator.

$$WR(n)_t = \left(\frac{HP_n - C_t}{HP_n - LP_n} \right) \times 100 \quad (6)$$

7. PSY10: 10-day psychological line.

$$PSY(n)_t = \left(\frac{UD_n}{n} \right) \times 100 \quad (7)$$

where UD_n is the number of upward days during previous n days.

8. +DI10: directional indicator up.

$$\left(\sum_{i=t-n+1}^t (+DM_i) / n \right) / \left(\sum_{i=t-n+1}^t TR_i / n \right) \times 100 \quad (8)$$

where $+DM_i = H_i - H_{i-1}$, H_i is high price at time i , $TR_i = \max\{H_i - L_i, H_i - C_{i-1}, L_i - C_{i-1}\}$, and L_i is low price at time i .

9. -DI10: directional indicator down.

$$-DI(n)_t = \left(\sum_{i=t-n+1}^t (-DM_i) / n \right) / \left(\sum_{i=t-n+1}^t TR_i / n \right) \times 100 \quad (9)$$

where $-DM_i = L_i - L_{i-1}$.

10. BIAS5: 5-day BIAS. It is a momentum indicator which measures deviation level between close price and moving average.

$$BIAS(n)_t = \frac{C_t - MA_n}{MA_n} \times 100 \quad (10)$$

11. VR10: 10-day volume ratio.

$$VR(n)_t = (UV_n - TV_n/2) / (DV_n - TV_n/2) \quad (11)$$

where TV_n is the volume summation at previous n days, UV_n is the volume summation at upward days, and DV_n is the volume summation at downward days.

12. AR20: 20-day A ratio.

$$AR(n)_t = \sum_{i=t-n+1}^t (H_i - O_i) / \sum_{i=t-n+1}^t (O_i - L_i) \quad (12)$$

where O_i is open price at time i .

13. BR20: 20-day B ratio.

$$BR(n)_t = \sum_{i=t-n+1}^t (H_i - C_{i-1}) / \sum_{i=t-n+1}^t (C_{i-1} - L_i) \quad (13)$$

References

- Abu-Mostafa, Y. S., & Atiya, A. F. (1996). Introduction to financial forecasting. *Applied Intelligence*, 6(3), 205–213.
- Achelis, S. B. (1995). *Technical analysis from A to Z*. Chicago: Probus Publishing.
- Brabazon, A., & Keenan, P. (2004). A hybrid genetic model for the prediction of corporate failure. *Computational Management Science*, 1(3–4), 293–310.
- Cao, L. (2003). Support vector machines experts for time series forecasting. *Neurocomputing*, 51, 321–339.
- Cao, L. J., & Tay, F. E. H. (2003). Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Transactions on Neural Networks*, 14(6), 1506–1518.
- Chang, C. C., & Lin, C. J. (2001). LIBSVM: A library for support vector machines. Software available at: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Chang, F. J., & Chang, Y. T. (2006). Adaptive neuro-fuzzy inference system for prediction of water level in reservoir. *Advances in Water Resources*, 29(1), 1–10.
- Chang, P. C., & Liao, T. W. (2006). Combining SOM and fuzzy rule base for flow time prediction in semiconductor manufacturing factory. *Applied Soft Computing*, 6(2), 198–206.
- Chen, H., Grant-Muller, S., Mussone, L., & Montgomery, F. (2001). A study of hybrid neural network approaches and the effects of missing data on traffic forecasting. *Neural Computing and Applications*, 10(3), 277–286.
- Chen, W. H., & Shih, J. Y. (2006). A study of Taiwan's issuer credit rating systems using support vector machines. *Expert Systems with Applications*, 30(3), 427–435.
- Conover, W. (1980). *Practical nonparametric statistics* (2nd ed.). New York: Wiley & Sons.
- Cristianini, N., & Taylor, J. S. (2000). *An introduction to support vector machines*. Cambridge UK: Cambridge University Press.
- Deng, Y. F., Jin, X., & Zhong, Y. X. (2005). Ensemble SVR for prediction of time series. In *Proceedings of the Fourth International Conference on Machine Learning and Cybernetics* (pp. 18–21). Guangzhou, China.
- Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., & Vapnik, V. (1997). *Support vector regression machines*. *Advances in Neural Information Processing Systems* (Vol. 9). Cambridge MA: MIT press.
- Guyon, I., & Elisseeff, A. (2003). An introduction to variable feature selection. *Journal of Machine Learning Research*, 3(7–8), 1157–1182.
- Han, J., & Kamber, M. (2001). *Data mining: Concepts and techniques*. CA: Morgan Kaufmann Publishers San Francisco, USA.

- Hsu, C. W., Chang, C. C., & Lin, C. J. (2003). A practical guide to support vector classification. Available at: <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>.
- Huang, Z., Chen, H., Hsu, C. J., Chen, W. H., & Wu, S. (2004). Credit rating analysis with support vector machines and neural networks: A market comparative study. *Decision Support Systems*, 37(4), 543–558.
- Huang, C. L., Chen, M. C., & Wang, C. J. (2007). Credit card scoring with a data mining approach based on support vector machine. *Expert Systems with Applications*, 33(4), 847–856.
- Huang, W., Nakamori, Y., & Wang, S. Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers and Operations Research*, 32(10), 2513–2522.
- John, G., Kohavi, R., & Peger, K. (1994). Irrelevant features and the subset selection problem. In *Proceedings of the eleventh international conference on machine learning* (pp. 121–129). San Mateo, USA.
- Kecman, V. (2001). *Learning and soft computing*. Cambridge, MA: The MIT Press.
- Kim, K. J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, 55(1–2), 307–319.
- Kim, K. J., & Han, I. (2000). Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Expert Systems with Applications*, 19(2), 125–132.
- Kim, K. J., & Lee, W. B. (2004). Stock market prediction using artificial neural networks with optimal feature transformation. *Neural Computing and Applications*, 13(3), 255–260.
- Kimoto, T., Asakawa, K., Yoda, M., & Takeoka, M. (1990). Stock market prediction system with modular neural network. In *Proceedings of the international joint conference on neural networks* (Vol. 1, pp. 1–6). San Diego, USA.
- Kim, H. J., & Shin, K. S. (2007). A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets. *Applied Soft Computing*, 7(2), 569–576.
- Klassen, M. (2005). Investigation of some technical indexes in stock forecasting using neural networks. In *Proceedings of the third world enformatika conference* (pp. 75–79). Istanbul, Turkey.
- Kohavi, R., & John, G. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97(1–2), 273–324.
- Kohonen, T. (1989). *Self organization and associative memory* (3rd ed.). Berlin: Springer-Verlag.
- Kohonen, T. (1997). *Self-organized maps*. New York: Springer.
- Kohonen, T., Raivio, K., Simula, O., Venta, O., & Henriksson, J. (1990). Combining linear equalization and self-organizing adaptation in dynamic discrete-signal detection. In *Proceedings of the international joint conference on neural networks* (Vol. 1, pp. 223–228). San Diego, USA.
- Lee, Y. L. (2007). Application of support vector machines to corporate credit rating prediction. *Expert Systems with Applications*, 33(1), 67–74.
- Lekakos, G., & Giaglis, G. (2007). A hybrid approach for improving predictive accuracy of collaborative filtering algorithms. *User Modeling and User Adapted Interaction*, 17(5), 5–40.
- Li, C. F., Zhang, J. B., Wang Z. Y., & Wang, S. T. (2006). A fully automated pattern classification method of combining self-organizing map with generalization regression neural network. In *Proceedings of the 13th international conference on neural information processing* (pp. 132–139). Hong Kong, China.
- Li, J., Manry, M. T., Narasimha, P. L., & Yu, C. (2006). Feature selection using a piecewise linear network. *IEEE Transactions on Neural Networks*, 17(5), 1101–1115.
- Liu, H., & Motoda, H. (1998). *Feature selection for knowledge discovery and data mining*. Norwell, MA: Kluwer Academic.
- Martín-Merino, M., & Román, J. (2006). A new SOM algorithm for electricity load forecasting. In *Proceedings of the 13th international conference on neural information processing* (pp. 995–1003). Hong Kong, China.
- Min, S. H., Lee, J., & Han, I. (2006). Hybrid genetic algorithms and support vector machines for bankruptcy prediction. *Expert Systems with Applications*, 31(3), 652–660.
- Pai, P. F., & Lin, C. S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega*, 33(6), 497–505.
- Quah, T. S., & Srinivasan, B. (1999). Improving returns on stock investment through neural network selection. *Expert Systems with Applications*, 17(4), 295–301.
- Roman, J., & Jameel, A. (1996). Backpropagation and recurrent neural networks in financial analysis of multiple stock market returns. In *IEEE system sciences proceedings of the twenty-ninth Hawaii international conference* (Vol. 2, pp. 454–460). Hawaii, USA.
- Roussinov, D. G., & Chen, H. (1999). Document clustering for electronic meetings: An experimental comparison of two techniques. *Decision Support Systems*, 27(1–2), 67–79.
- Salzberg, S. L. (1997). On comparing classifiers: Pitfalls to avoid and a recommended approach. *Data Mining and Knowledge Discovery*, 1(3), 317–327.
- Schölkopf, B., Burges, C., & Vapnik, V. (1995). Extracting support data for a given task. In *Proceedings of the first international conference on knowledge discovery and data mining* (pp. 252–257). Montréal, Canada.
- Schölkopf, B., & Smola, A. J. (2000). *Statistical learning and kernel methods*. Cambridge, MA, USA: MIT Press.
- Tay, F. E. H., & Cao, L. J. (2001). Application of support vector machines in financial time series forecasting. *Omega*, 9(4), 309–317.
- Tay, F. E. H., & Cao, L. J. (2001). Improved financial time series forecasting by combining support vector machines with self-organizing feature map. *Intelligent Data Analysis*, 5(4), 339–354.
- Thissen, U., Van Brakel, R., De Weijer, A. P., Melssen, W. J., & Buydens, L. M. C. (2003). Using support vector machines for time series prediction. *Chemometrics and Intelligent Laboratory System*, 69(1), 35–49.
- Tsaih, R., Hsu, Y., & Lai, C. C. (1998). Forecasting S&P 500 stock index futures with a hybrid AI system. *Decision Support Systems*, 23(2), 161–174.
- Vapnik, V. (1995). *The nature of statistical learning theory*. New York: Springer-Verlag.
- Vapnik, V., Golowich, S., & Smola, A. (1997). Support vector method for function approximation, regression estimation, and signal processing. *Advances in Neural Information Processing Systems* (Vol. 9, pp. 281–287). Cambridge, MA: MIT press.
- Wang, J., & Yan, D. (2004). A high precision prediction method by using combination of ELMAN and SOM neural networks. In *Proceedings of the first international symposium on neural networks* (pp. 943–949). Dalian, China.
- Wong, K. W., Wong, P. M., Gedeon, T. D., & Fung, C. C. (2003). Rainfall prediction model using soft computing technique. *Soft Computing*, 7(6), 434–438.
- Wu, W., Liu, X., Xu, M., Peng, J., & Rudy, S. (2004). A hybrid SOM-SVM method for analyzing zebra fish gene expression. In *Proceedings of the 17th international conference on pattern recognition* (pp. 323–326). Cambridge, England.