Predicting Financial Time Series Data Using Hybrid Model

Bashar Al-hnaity and Maysam Abbod

Abstract Prediction of Financial time series is described as one of the most challenging tasks of time series prediction, due to its characteristics and their dynamic nature. Support vector regression (SVR), Support vector machine (SVM) and back propagation neural network (BPNN) are the most popular data mining techniques in prediction financial time series. In this paper a hybrid combination model is introduced to combine the three models and to be most beneficial of them all. Quantization factor is used in this paper for the first time to improve the single SVM and SVR prediction output. And also genetic algorithm (GA) used to determine the weights of the proposed model. FTSE100, S&P 500 and Nikkei 225 daily index closing prices are used to evaluate the proposed model performance. The proposed hybrid model numerical results shows the outperform result over all other single model, traditional simple average combiner and the traditional time series model Autoregressive (AR).

1 Introduction

In modern financial time series prediction, predicting stock prices has been regarded as one of the most challenging applications. Thus, various numbers of models have been depicted to support the investors with more precise predictions. However, stock prices are influenced by different numbers of factors and the nonlinear relationships between factors existed in different periods such that predicting the value of stock prices or trends is considered as an extremely difficult task for the investors. There-

Bashar Al-hnaity

Electronic and Computer Engineering Department, Brunel University, London, Uxbridge, UB8 3PH, UK, e-mail: bashar.Al-hnaity@brunel.ac.uk

Maysam Abbod

Electronic and Computer Engineering Department, Brunel University, London, Uxbridge, UB8 3PH, UK e-mail: e-mail: maysam.abbod@brunel.ac.uk

1

fore, many methods and approaches have been introduced and employed in the prediction of stock prices and trends since the start of the stock market.

Additionally, researchers proposed numerous conventional numerical prediction models. However, traditional statistical models such ARCH, GARCH, ARMA, ARIMA and AR have failed to capture the complexity and behavior of the stock price. George Box stated in his work that "essentially, all models are wrong, but some are useful"[1]. Therefore, researchers have introduced more advanced nonlinear techniques including Support Vector Machine (SVM), Support Vector Regression (SVR), Neural Network (NN). Due to inherent characteristics such as nonlinear, non-stationary, noisy, high degree of uncertainty and hidden relationships of the financial time series, single artificial intelligence and other conventional techniques have failed to capture its non-stationary property and accurately describe its moving tendency. Therefore, researches and market participants have paid a tremendous attention in order to tackle such problem. Thus, various models architecture and new algorithm have been introduced and developed in the literature to alleviate the influence of noise.

A various number of models and theories have been implemented by researchers in order to improve the prediction performance. Different techniques has been combined with single machine learning algorithms, for example Zhang and Wu (2009) [2] have integrated a Back-Propagation neural network (BPNN) with an improved Bacterial Chemotaxis Optimization (IBCO). Anther method was proposed combining data preprocessing methods, genetic algorithms and Levenberg-Marquardt (LM) algorithm in learning BPNN by [3]. In their studies data transformation and selection of input variables were used under data preprocessing in order to improve the model's prediction performance. Moreover, the obtained result has proved that the model is capable of dealing with data fluctuations as well as yielding a better prediction performance. And also, a hybrid artificial intelligent model was proposed by [4] combining genetic algorithms with a feed forward neural network to predict stock exchange index. The latest research indicates that in short term prediction models based on artificial intelligence techniques outperform traditional statistical based models [5]. However, and also as a result of the dramatic move in stock indices in response to many complex factors, there is plenty of room for enhancing intelligent prediction models. Therefore, researchers have introduced and tried to combine and optimize, different algorithms and take the initiative to build new hybrids models in order to increase prediction speed and accuracy.

There are many example of these attempts such as; Armano et al [6]who optimized GA with ANN to predict stock indices; SVM were also combined with PSO in order to carry out the prediction of stock index by Shen and Zhang [7]. Kazem's (2013) [8] prediction model, chaotic mapping, firefly algorithm and support vector regression (SVR) were proposed to predict stock market prices. In Kazem's study SVR-CFA model was for the first time introduced and the results were compared with the SVR-GA (Genetic Algorithm), SVR-CGA (Chaotic Genetic Algorithm), SVR-FA (Firefly Algorithm) and ANN, and the ANFIS model, and the new adopted model SVR-CFA outperformed the other compared models. The seasonal Support Vector Regression (SSVR) model was developed by [9] to predict seasonal time se-

ries data. Moreover, in SSVR models in order to determine the model parameters, hybrid genetic algorithm and tabu search (GA,TS) algorithms were implemented. And also on the same data sets, Sessional Auto regressive Integrated Moving Average (SARIMA) and SVR were used for prediction, however the empirical results based on prediction accuracy indicated that the proposed model SSVR outperformed both models (SVR and SARIMA). A novel hybrid model to predict future evolution of a various stock indices was developed by integrating a genetic algorithm based on optimal time scale feature extractions with support vector machines.Neural network, pure SVMs and traditional GARCH models were used as a benchmark and prediction performances were compared and thus the proposed hybrid model's prediction performance was the best. Root mean squared error (RMSE) is one of the main utilized prediction models for performance measurement, however the reduction in this standard statistical measurement was significantly high. A novel hybrid ensemble model to predict FTSE 100 next day closing price was proposed by Alhnaity and Abbod [10]. Genetic Algorithm (GA) was used to determent the weight of the hybrid combination model, and thus, the empirical results showed that the proposed hybrid model utilizing GA has outperformed the rest utilized approaches in their study.

In this Chapter, the FTSE 100, S&P 500 and Nikkei225 daily closing prices will be predicted. Different data mining techniques will be utilized to capture the non-liner characteristics in a stock prices time series. The proposed approach is to combine the support vector machine SVM, support vector regression SVR and BPNN as the weight of these models are determined by the GA. The evaluation of the proposed model performance will be done by using it on the FTSE 100, S&P 500 and Nikkei225 daily closing price as the illustrative examples. The below examples show that the proposed model outperform all single models and with regard to the possible changes in the data structure, proposed model is shown to be more robust. The reminder of this chapter is organized as follow: in section 2 the component models SVM, SVR and BPNN are briefly introduced. Section 3 demonstrates the prediction procedures and the hybrid methodology. The experimental result exhibits in section 4. Finally conclusions and future work are in section 5.

2 Methodology

2.1 Benchmark Prediction Model

In this chapter a traditional prediction model Simple Auto-regressive model (AR) is used, in order to benchmark the performance efficiency of the utilized models. Moreover, a simple average in this thesis is used as a benchmark combination method.

2.1.1 Simple Auto-regressive Model

The autoregressive model (AR) in this study is used as a benchmark model to evaluate the prediction power between the utilized models based on the relative improvements on root mean square error. Equation 1 illustrates the AR model used.

$$y_t = a_1 y(t-1) + a_2 y(t-2) + \ldots, a_t y(t-p) \tag{1}$$

Equation 1 $y(t)$ is the predicted stock price based on the past close daily price, $y(t) - y(t-n)$ and the coefficients of AR model are $a_1 - a_n$. 5 lagged daily price is the order of which is used in the AR model was varied and found to give better prediction result. The model coefficients were determined by using the implemented regress function in the MATLAB.

2.1.2 Prediction Combination Techniques

Combining different prediction techniques has been investigated widely in the literature. In the short range prediction combining the various techniques is more useful according to [2], [11]. Timmermann [12] stated in his study that using simple average may work as well as more sophisticated approaches. In this chapter simple average is used as a benchmark combination model. Equation 2 illustrates the calculation of combination prediction method at time t [13].

$$f_{SM}^t = \left(f_{M1}^t + f_{M2}^t + f_{M3}^t \right) \div 3 \tag{2}$$

2.2 Artificial Neural Network

In recent years predicting financial time series utilizing Artificial Neural Network (ANN) applications has increased dramatically. The idea of ANN can be seen before reaching the output, where the filtration of the inputs through one or more hidden layers each of which consists of hidden units, or nodes is considered as the main idea. Thus, final output is related to the intermediate output [14].

The ability of learning from data through adaptive changing structure based on external or internal information that flows through the network during the learning phase and generates output variables based on learning is one of the most important advantages of ANN. Furthermore, the non-linear nature of ANN is also a valuable quality. ANNs are classified as non-linear data modeling tool, thus, one of the main purposes of utilizing such a model is to find the patterns in data or modeling complex relationships between inputs and outputs. Hence, an explicit model-based approach fails,but ANNs adapt irregularities and unusual of features in a time series of interest.

The application of ANNs has been popularly utilized in financial time series prediction modeling. A comprehensive review of ANNs and their application in a various finance domains is given in section 1. However, as with many other techniques, ANNs have some disadvantages such as not allowing much of understanding of the data which might be because they are not explicit models. Therefore, providing a black box for the prediction process is considered as a disadvantage. The danger of over-fitting the in-sample training data is also a major ANN methods drawback [15]. In terms of the goodness of fit the performance of in-sample data sets is good, and is what ANNs are trained on. However, in out-of-sample data it is conditional on not breaking the structure in the data sets. In accordance with Balestrassi et al [16] excessive training time and a large number of parameters that must be experimentally selected in order to generate good prediction are considered as other drawbacks facing ANN applications. In this chapter the ANN BPNN is used to predict financial data, subsection 2.2.1 demonstrates this model.

2.2.1 Back Propagation Neural Network

In modeling time series with non-linear structures, the most commonly used structure is three layers feed-forward back propagation [17]. The weights are determined in the back propagation process by building connections among the nodes based on data training, producing a least-mean-square error measure of the actual or desired and the estimated values from the output of the neural network. The initial values are assigned for the connection weights. In order to update the weights, the error between the predicted and actual output values is back propagated via the network. Minimizing of the error the desired and predicted output attempts occurs after the procedure of supervised learning [18]. The architecture of this network contains a hidden layer of neurons with non-linear transfer function and an output layer of neurons with non-linear transfer function and an output layer of neurons with linear transfer functions. Figure 1 illustrates the architecture of a back propagation network, where $x_j (j = 1, 2, ..., n)$ represent the input variables; $z_i (j = 1, 2, ..., m)$ represent the outputs of neurons in the hidden layer; and $y_t (t = 1, 2, ..., l)$ represent the outputs of the neural network [19].

In theory a neural network has the ability to simulate any kind of data pattern given a sufficient training. Training the neural network will determine the perfect weight to achieve the correct outputs. The following steps illustrate the training process of updating the weights values [20]. The first stage is hidden layers: the bellow Equation explains how the outputs of all neurons in the hidden layer are calculated:

$$net_i = \sum_{j=0}^{n} w_{ji} x_j v i = 1, 2, ... m \tag{3}$$

$$z_i = f_H (net_i) i = 1, 2, ... m \tag{4}$$

Where net_i is the activation value of the ith node, z_i is the output of the hidden layer, and f_H is called the activation Equation of a node; in this chapter a sigmoid function
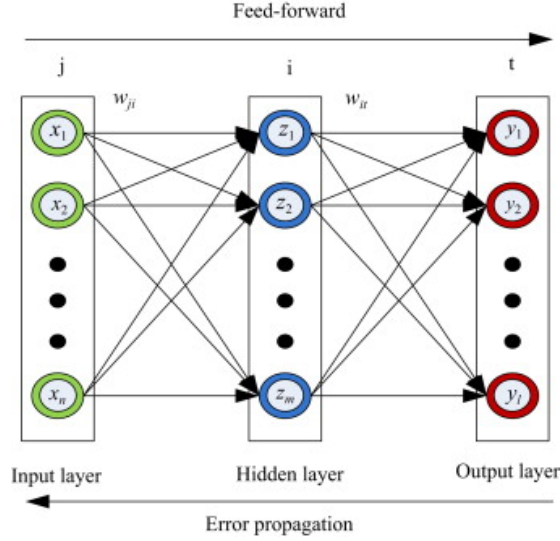
Fig. 1: Architecture of feed forward back propagation neural network [19].

is utilized. Equation 5 explains the utilized sigmoid activation equation.

$$f_H(x) = \frac{1}{1 + exp(-x)} \tag{5}$$

Second stage the output: The outputs of all neurons in the output layer are given as Equation 6 illustrated bellow:

$$y_t = f_t \left(\sum_{i=0}^{m} w_{it} z_i \right) t = 1, 2, ...l \tag{6}$$

The activation Equation is $f_t(t = 1, 2, ..., l)$, which is usually a linear equation. The weights are assigned with random values initially, and are modified by the delta rule according to the learning samples traditionally. The topology in this study is determine by a set of trial and error search conducted to choose the best number of neurons experiments. Different range of 20 to 5 neurons in hidden layer two layer feed forward back propagation network were tried. The model stopped training after reaching the pre-determined number of epochs. The ideal topology was selected based on the lowest Mean Square Error.

2.3 Support Vector Machine

The SVM theory was developed by Vladimir Vapnik in 1995. It is considered as one of the most important breakthroughs in machine learning field and can be applied in classification and regression [21]. In modeling the SVM, the main goal is to select the optimal hyperplane in high dimensional space ensuring that the upper bound of generalization error is minimal. SVM can only directly deal with linear samples but mapping the original space into a higher dimensional space can make the analysis of a non-linear sample possible [22] [23]. For example if the data point (x_i, y_i) was given randomly and independently generated from an unknown function, the approximate function formed by SVM is as follows:

$g(x) = w\emptyset(x) + b.\emptyset(x)$ is the feature and non-linear mapped from the input space x. w and b are both coefficients and can be estimated by minimizing the regularized risk Equation.

$$R(C) = C\frac{1}{N}\sum_{i=1}^{N}L(d_i, y_i)\frac{1}{2} + ||w||^2 \tag{7}$$

$$L(d, y) = \begin{cases} |d - y| - \varepsilon & |d - y| \geq \varepsilon \\ 0 & other, \end{cases} \tag{8}$$

C and ε in Equation 7 and 8 are prescribed parameters. C is called the regularization constant while ε is referred to as the regularization constant. $L(d, y)$ Is the intensive loss function and the term $C\frac{1}{N}\sum_{i=1}^{N}L(d_i, y_i)$ is the empirical error while the $\frac{1}{2} + ||w||^2$ indicates the flatness of the function. The trade-off between the empirical risk and flatness of the model is measured by C. Since introducing positive slack variables ζ and ζ^* equation 8 transformed to the following:

$$R(w, \zeta, \zeta^*) = \frac{1}{2}ww^T + C \times \left(\sum_{i=1}^{N}(\zeta, \zeta^*)\right) \tag{9}$$

Subject to:

$$w\phi(x_i) + b_i - d_i \leq \varepsilon + \zeta_{i*} \tag{10}$$

$$d_i - w\phi(x_i) - b_i \leq \varepsilon + \zeta_i \tag{11}$$

$$\zeta_i, \zeta_{i*} \geq 0 \tag{12}$$

The decision Equation (kernel function) comes up finally after the Lagrange multipliers are introduced and optimality constraints exploited. The equation 13 is the formed of kernel Equation:

$$f(x) = \sum_{i}^{l}\left(\alpha_i - \alpha_i'\right)K(x_i, x_j) + b \tag{13}$$

Where α_i' are Lagrange multipliers. The satisfy equalities $\alpha_i \times \alpha_i' = 0, \alpha_i \geq 0, \alpha_i' \geq 0$. The kernel value is the same with the inner product of two vectors x_i and x_j in

the feature space $\phi(x_i)$ and $\phi(x_j)$. The most popular kernel function is Radial Basis Function (RBF) it is form in Equation 14.

$$K(x_i, x_j) = exp\left(-\gamma ||x_i - x_j||^2\right) \tag{14}$$

Theoretical background, geometric interpretation, unique solution and mathematical tractability are the main advantages which has made SVM attract researchers and investors interest and be applied to many applications in different fields such as prediction of financial time series.

2.4 Support Vector Regression

As explained in subsection 2.3 the idea of SVM is to constructs a hyperplane or set of hyperplanes in a high or infinite dimensional space, which can be used for classification. In the regression problem the same margin concept used in SVM is used. The goal of solving a regression problem is to construct a hyperplane that is as close to as many of the data points as possible. Choosing a hyperplane with small norm is considered as a main objective, while simultaneously minimizing the sum of the distances from the data points to the hyperplane [24].

In case the of solving regression problems using SVM, SVM became known as the support vector regression (SVR) where the aim is to find a function f with parameters w and b by minimizing the following regression risk:

$$R(f) = \frac{1}{2}(x, w) + C\sum_{i=1}^{N} l(f(x_i), y_i) \tag{15}$$

C in Equation 15 is a trade-off term, the margin in SVM is the first term which is used in measuring VC-dimension [20].

$$f(x, w, b) = (w, \phi(x)) + b, \tag{16}$$

In the equation 16 $\phi(x) : x \to \Omega$ is kernel function, mapping x into in the high dimensional space, SVR and as proposed by [24]. The $-\varepsilon$ insensitive loss function is used as follows:

$$l(y, f(x)) = \begin{cases} 0, if|y - f(x) < \varepsilon \\ |y - f(x)| - \varepsilon, Otherwise \end{cases} \tag{17}$$

Equation 18 constrained minimisation problem is equivalent to previous minimisation Equation 15.

Min

$$y\left(w, b, \zeta^* = \frac{1}{2}(w, w) + C\sum_{i=1}^{N} (\zeta_i + \zeta_{i*})\right) \tag{18}$$

Subject to:

$$y_i - ((w, \phi(x_i) + b)) \le \varepsilon + \zeta_i, \tag{19}$$

$$((w, \phi(x_i)) + b) - y_i \le \varepsilon + \zeta_{i*}, \tag{20}$$

$$\zeta_i^* \ge 0 \tag{21}$$

In sample (x_i, y_i) the ζ_i and ζ_{i*} measure the up error and down error. Maximizing the dual function or in other words constructing the dual problem of this optimization problem (primal problem) by large method is a standard method to solve the above minimization problem. There are four common kernel functions; among these, this chapter will be utilizing the radial basis function (RBF). In accordance with [25], [26] and [27] the RBF kernel function is the most widely applied in SVR. Equation 22 is defining the kernel RBF, where the width of the RBF is denoted by σ. Furthermore, Cherkassky and Ma [25] suggested that the value of σ must be between 0.1 and 0.5 in order for SVR model to achieve the best performance. In this chapter σ value is determined as 0.1.

$$K(x_i, x_i) = exp\left(\frac{-||x_i - x_j||^2}{2\sigma^2}\right) \tag{22}$$

3 Prediction Procedures

Financial time series data are characteristically non-linear, non-stationary, and noisy, with high degree of uncertainty and hidden relationships. These characteristics are a reason for information unavailability, affecting the behavior of financial markets between past and future captured prices. According to Li and Ma [28] traditional linear methods and the majority of sophisticated non-linear machine learning models have failed to capture the complexity and the non-linearity that exists in financial time series, particularly during uncertainty periods such the credit crisis in 2008 [29]. Financial time series characteristics imply that the statistical distributions of time series can change over time. The cause of these changes may be economic fluctuations, or political and environmental events [30] [31]. As a result, it has been verified in the literature that no single method or model works well to capture financial time series characteristics properly and accurately, which leads to different and inaccurate financial time series prediction results [32] [33]. In order to address these issues, a new two step hybrid model is proposed to capture the non-linear characteristics in a stock price time series. The proposed approach is to combine the support vector machine SVM, support vector regression SVR and BPNN as the weight of these models are determined by the GA.

3.1 Hybrid Combination Model GA-WA

Combining different prediction techniques have been investigated widely in the literature. In the short range prediction combining the various techniques is more useful according to [2], [11]. In accordance to Timmermann's study, using the simple average may work as well as more sophisticated approaches. However, using one model can produce a more accurate prediction than any other methods. Therefore, simple averages would not be sufficient in such cases [12]. Compared with different prediction models, the hybrid prediction method is based on a certain linear combination. The assumption for the actual value in period t by model i is $f_{it}\ (i = 1, 2, ...m)$, the corresponding prediction error will be $e_{it} = y_t - f_{it}$. And also the weight vector will be $W = [w_1, w_2, ...w_m]^T$ .. Then in the hybrid model the predicted value is computed as follows [34] [35]:

$$\hat{y}_t = \sum_{i=1}^{m} w_i f_{it}\ (t = 1, 2, ...n) \tag{23}$$

$$\sum_{i=1}^{m} w_i = 1 \tag{24}$$

Equation 23 can be expressed in another from, such:

$$\hat{y} = FW$$
$$where \quad \hat{y} = [\hat{y}_1, \hat{y}_2, ...\hat{y}_n]^T, F = [f_{it}]_{n \times m} \tag{25}$$

The error for the prediction model can be formed as Equation 26 illustrates.

$$e_t = y_t - \hat{y}_t = \sum_{i=1}^{m} w_i y_t - \sum_{i=1}^{m} w_i f_{it} =$$
$$\sum_{i=1}^{m} w_1 w_i (y_t - f_{it}) = \sum_{i=1}^{m} w_i e_{it} \tag{26}$$

This research proposes a hybrid model of combining SVR, BPNN and SVM.

$$\hat{Y}_{combined_t} = \frac{w_1 \hat{Y}_{SVR} + w_2 \hat{Y}_{BPNN} + w_3 \hat{Y}_{SVM}}{(w_1 + w_2 + w_3)} \tag{27}$$

The prediction values in period t are $\hat{Y}_{combined_t}, \hat{Y}_{SVR}, \hat{Y}_{BPNN}$ and $\hat{Y}_{SVM}$ for the hybrid, SVM, BPNN and SVR models, where the assigned weights are w_1, w_2, w_3 respectively, with $\sum_{i=1}^{3} w_i = 1.0 \le w_i \le 1$.

The most important step in developing a hybrid prediction model is to determine the perfect weight for each individual model. Setting $w_1 = w_2 = w_3 = 1/3$ in Function 27 is the simplest combination method for the three prediction models. Nevertheless, in many cases equal weights cannot achieve the best prediction result. Therefore, this chapter adopts a hybrid approach utilizing GA as an optimizer to determine the optimal weight for each prediction model. Figure 2 illustrates the architecture of the GA-WA hybrid model.
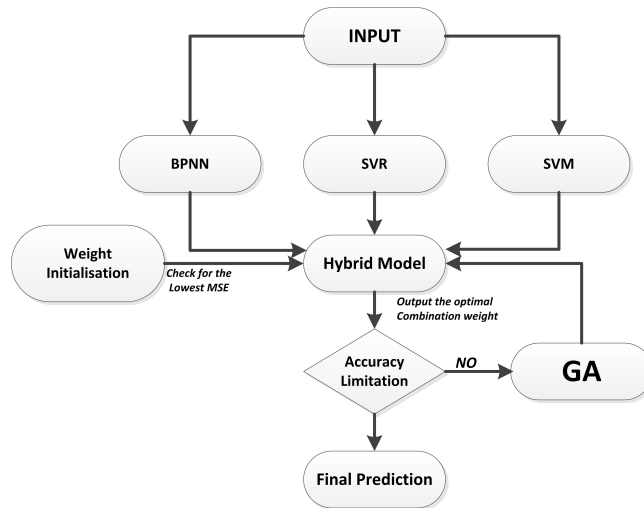
Fig. 2: The flow chart of the GA-WA hybrid model.

3.1.1 Genetic Algorithm

Genetic Algorithm GA is a well known tool in computational method modeled on Darwinian selection mechanism. GA principles were proposed by Holland [36] , and developed by Goldbreg [37] and Koza [38]. Thus, the main purpose of using such algorithm, is to solve optimization problems such determining the optimal weights for the proposed hybrid model in this research. In comparison to other conventional optimization methods, GA have many differences, which contributed in making GA more efficient in searching for the optimal solution. The following points exhibited those differences [39].

- Computing the strings in GA algorithms is done by encoding and decoding discrete points than using the original parameters values. Thus, GAs tackling problems associated to discontinuity or non-differentiability function, where, traditional calculus methods have failed to work. Therefore, due to adaptation of binary strings,such characteristic allows GAs to fit computing logic operations better.
- The prior information is not important, and thus there is no need for such information as the primary population is randomly generated. GA using a fitness function in order to evaluate the suggested solution.
- Initialization, selection and reproduction whether, crossover and mutation are what GA depends on in searching process, which involve random factors. As a results, searching process in GAs for every single execution will be stand alone even the ones who are under the identical parameter setting, which perhaps may effect the results.

3.2 Data Pre-processing

Selecting and pre-processing the data are crucial steps in any modeling effort, particularly for generalizing a new predictive model. Data sets are divided into two sets: training and testing, which is explained in the below Subsection. Thus, the preprocessing steps for financial time series data are as follows:

- Data Cleaning:
  In real world, data tend to be incomplete and inconsistent. Therefore, the purpose of data cleaning (data cleansing) is to fill in missing values and correct inconsistencies in the data. The daily closing price of the FTSE 100, S&P 500 and Nikkie 225 are used as an input, and due to public holidays in the stock market, there were missing values in the data set. Thus, treating the missing data can be done by different methods such as; ignore the tuple, fill in the missing value manually, use a global constant to fill in the missing value, use the the attribute mean to fill in the missing value, use the attribute mean for all samples belonging to the same class as the given tuple and use the most probable value to fill in the missing values. This Chapter adopted the ignore the tuple method. In other words, in order to treat the missing information in the data set the tuples were excluded from the data set [40].
- Attribute creation:
  In the presented data in section must be able to be sufficiently utilized in prediction proposed models. Therefore, and according to Fuller [41], a real valued time series is considered as a set of random variables indexed in time or it can be a set of observations ordered in time such as performance degradation data. In accordance with traditional setting of time series regression, the broad concept of the target value is represented as an unknown function for the input vector x_t of the p-lagged last value of y itself, as Equation 28 illustrated, where p is the number of back time steps which in this research is 5. Thus the input array and the output array are respectively as follows in equations 29 and 30 , where t is the number of training data.

$$y_t = f(x_t) = f(y_{t-1}, y_{t-2}, \ldots, y_{t-p}) \tag{28}$$

$$X = \begin{bmatrix} y_1 & y_2 & \cdots & y_p \\ y_2 & y_3 & \cdots & y_{p+1} \\ y_{t-p} & y_{t-p+1} & \cdots & y_{t-1} \end{bmatrix} = \begin{vmatrix} x_{p+1} \\ x_{p+2} \\ \vdots \\ x_t \end{vmatrix} \tag{29}$$

$$Y = \begin{vmatrix} y_{p+1} \\ y_{p+2} \\ \vdots \\ y_t \end{vmatrix} \tag{30}$$

 Moreover, these steps are for the training in-sample data set and testing out-of-sample data set. However, this research adopted a new approach in forming the
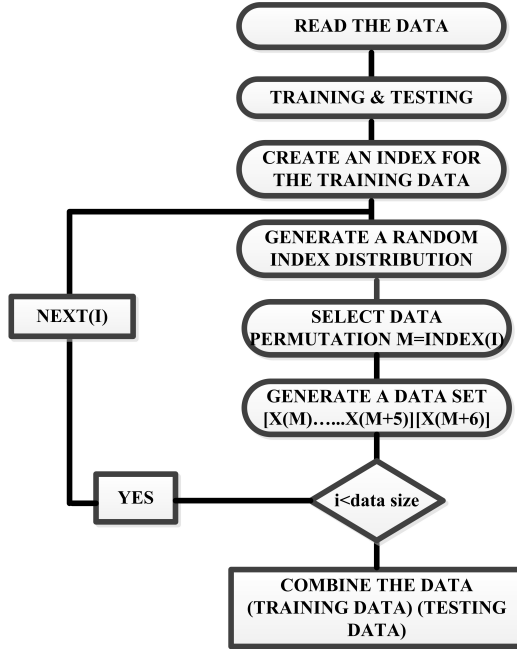
Fig. 3: The data preprocessing frame.

training in-sample data set and testing out-of-sample data set in order to achieve the best prediction results. The following forms illustrate the steps of this forming: for training data: $Tr = [x_t......x_{t+5}] [x_{t+6}]$, where m is a random number permutation $1 < m < p$, p is the data size. For testing data: $Ts = [x_t....x_{t+5}] [x_{t+6}]$, where t is $1 : p$, P is the testing data set size. Figure 3 illustrate the architecture of the data prepossessing.

3.3 Quantization factor

This chapter adopts a new approach to enhance the prediction output of SVM and SVR models. The quantization factor is for the first time introduced to SVM and SVR. As explained above in the methodology in section SVM and SVR the model input is (x_i, y_i). After adding the optimal factors which were determined by trial and error, the optimal factor was chosen from a range of factors between 10 and 100. The optimal factor was selected based on the lowest mean square error. The below steps illustrate the change in the input after introducing the quantization factor:

$$X prim = X_i \div factor \tag{31}$$

$$Y\,prim = Y_i \div factor \tag{32}$$

The model inputs become $(X\,prim_i, Y\,prim_i)$. After applying the above model SVM and SVR and to get the final prediction result the chosen factor is multiplied by the output of each model as illustrated in Equation 33 and 34:

$$X\,prim_{pred} = X\,prim \times Factor \tag{33}$$

$$Y\,prim_{pred} = Y\,prim \times Factor \tag{34}$$

This method has been proposed to enhance the performance and prediction ability of SVM and SVR techniques. To the best of our knowledge this might be the first time this approach is introduced and utilized in financial time series predictions (stocks indies predictions). Furthermore, the test was carried out on the training data set of the three stock indies (FTSE100, S&P500 and Nikkei225). first, for FTSE100, the best $Qfactor$ in SVR model is 100. Second, for S&P500 the best $Qfactor$ is 100. Third, for Nikkei225 the best $Qfactor$ is 100. Moreover, in SVM prediction model the best $Qfactor$ is 100 for FTSE100 data set, 10 is the best $Qfactor$ for S&P500 and 10 is the best $Qfactor$ for the Nikkei225 data set.

3.4 Prediction Evaluation

3.4.1 Standard Statistical Measure

The accuracy of prediction is referred to as "goodness of fit" which in other words means how well the prediction models fit a set of observations. Moreover, prediction accuracy can also be regarded as an optimists term for prediction errors as stated in [42]. Hence, prediction error represents the difference between the predicted value and the actual value. Therefore, the prediction error can be defined as: if x_n is the actual observations at time n and $\hat{x}_n$ is the predicted value for the same period, then the prediction error takes the following form:

$$E_n = x_n - \tilde{x}_n \tag{35}$$

Generally, in equation 35 $\hat{x}_n$ is one step ahead prediction so E_n is the one step ahead prediction error. The accuracy measurement can be classified under five categories according to [43]: scale-dependent measures, measures based on percentage errors, measures based on relative errors, relative measures and scaled error. However, in this research the most popular and commonly available statistical accuracy measures are used as Table 1 demonstrates:

MSE, RMSE and MAE are scale-dependent measures which are commonly utilized to compare different methods on the same data set. In addition to standard statistical measures, standard deviation is used to observe variations. Moreover, a well known cross correlation method of computing the degree of relationship ex-

Table 1: Statistical accuracy measures

Name	Category
Mean Square Error (MSE)	Scale-dependent measures.
Root Mean Square Error (RMSE)	Scale-dependent measures.
Mean Absolute Error (MAE)	Scale-dependent measures.
Cross correlation coefficient (R)	Method of computing the degree of relationship between variable.
Standard Deviation (SD)	An statistical test used to observe variations.

isting between the variables is also used in this study [**?**]. Table 2 illustrates the formulas utilized to measure prediction accuracy.

Table 2: Error metrics equation to measure prediction accuracy

Abbrev.	Formulas
MSE	$= \dfrac{1}{N} \sum_{i=1}^{N} (x_i - \hat{x}_i)^2$
RMSE	$= \sqrt{\dfrac{1}{N} \sum_{i=1}^{N} (x_i - \hat{x}_i)^2}$
MAE	$= \dfrac{1}{N} \sum_{i=t}^{N} \dfrac{|x_i - \hat{x}_i|}{|x_i|} =$
R	$= \dfrac{\sum_{i=1}^{N} (x_i - \bar{x}_i)\left(\hat{x}_i - \hat{\bar{x}}_i\right)}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x}_i)^2 \sum_{i=1}^{N} \left(\hat{x}_i - \hat{\bar{x}}_i\right)^2}}$
SD	$= \sqrt{\dfrac{\sum (x_i - \hat{x}_i)^2}{N - 1}}$

These statistical analyses after they are computed will provide the required information regarding the prediction accuracy and will strengthen the conclusions. However, after considering the pros and cons of the above-mentioned statistical ac-

curacy measures, MSE and RMSE are chosen for performance comparison on the same data set, and for comparison across data sets MAE is utilized in this research [44].

4 Experimentation Design and Results

4.1 Data Sets

This chapter proposes a hybrid model to predict the next day index closing price. In order to test the proposed model, The first data set is the daily closing price of the FTSE 100 index. It is used to demonstrate the predictability of the single approach model. The first 7788 observations (03/01/1984 - 29/10/2013) are used as the in-sample data set (training set). The last 250 observations (30/10/2013- 30/10/2014) are used as the out-sample set (testing set). The second data set is the daily closing price of the S&P 500 index which is also used to demonstrate and validate the predictability of the proposed single approach method. The first 16230 observations (03/01/1950 - 02/07/2014) are used as the in-sample data set (training set). The last 250 observations (03/07/2014 - 30/06/2015) are used as the out-sample set (testing set). The third data set is the daily closing price of the Nikkei 225 index and it is also utilized to demonstrated the predictability of the proposed single approached methods. The first 7508 observations (04/01/1984 - 04/07/2014) are used as the in-sample data set (training set). The last 250 observations (07/07/2014 - 30/06/2015) are used as the out-sample set (testing set). Thus, in order to discover a robust model it is highly suggested to have a long training duration and large sample [45]. Figure 4 illustrates the data division into in-sample and out-of-sample sets.
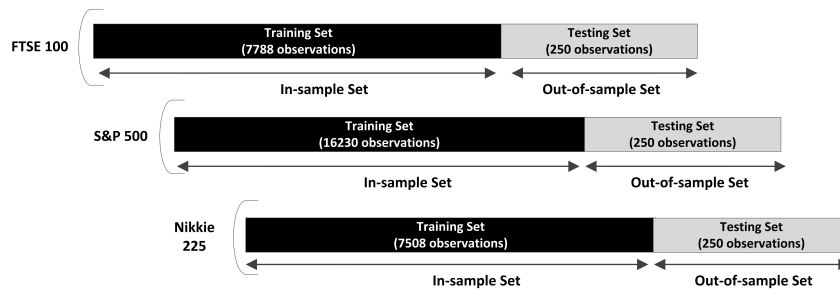

Fig. 4: In-sample and out-of-sample data sets.

4.2 BPNN, SVM and SVR Parameters Determination

This chapter proposes a hybrid method combining three different models SVM, SVR and BPNN to predict FTSE 100, S&P 500 and Nikkei 225 stocks index closing prices. First step to build the proposed method is by performing the single classical models on the data sets individually. Building a BPNN prediction model by determining the architecture as mentioned in the methodology section in BPNN part. BPNN topology has been set by trial and error techniques, two hidden layer feed forward back propagation network with only two hidden layer and a range of 20-5 is the range of neurons. The best model topology for FTSE100 data set is, [20-5] (20, 5 nodes for the two hidden layers) with learning rate of 0.3 and momentum constant of 2, the best S&P500 model topology is [10-5] (10, 5 nodes for the two hidden layers) with learning rate of 0.3 and momentum constant of 2 and for Nikkei225 it is [10-5] (10, 5 nodes for the two hidden layers) with learning rate of 0.7 and momentum constant of 10.

To build the SVR and SVM model, the method which was addressed in the methodology section and in the framework prediction model section to determine the parameters of both model and the factor is applied. In the literature there are no general rules for choosing those parameters. Thus, this research adopted the most common approach to search for the best C and γ values, which is the grid search approach [2]. The grid search approach was proposed by Lin et al [46] using a validation process in order to achieve the ability to produce good generalizations to decide parameters. The criteria of choosing the optimal C and γ parameters is by trying pairs of C and γ and the best combination of these parameters which can generate the minimum mean square error MSE is chosen to set up the prediction model.

On the grid and after identifying a better region, on this region a more specific grid search can be conducted [46]. Finally after determining these lowest cross validation prediction error parameters C and γ, these parameters are chosen to be used in the creation of the prediction models. Figure 5 illustrates the process of grid search algorithm when building the SVM, SVR prediction model.

A set of exponentially growing sequences of C and γ are used and the best parameter combination results are: for the FTSE100 prediction next day closing price is $(C = 100, \gamma = 0.0001)$, which gives the minimum mean square error MSE in the training data set and is therefore the best to set up the prediction model. For the S&P500 next day closing price prediction, the best combination parameters are $(C = 100, \gamma = 0.0001)$, which are the best combination parameters which give the minimum MSE in training data set. Furthermore, the best combination results to set up SVR prediction next day closing price model for Nikkei225 parameters are $(C = 100. \gamma = 0.0003)$ which give the minimum MSE in the training data set. Moreover, the parameters combination to set up a prediction model using SVM are: Firstly, for the FTSE100 data set the best combination of parameters are $C = 100, \gamma = 0.0003$. Secondly, for S&P500 the best combination of parameters are $C = 100, \gamma = 0.0003$. Thirdly, for Nikkei225 the best combination of parameters are $C = 90, \gamma = 0.0002$.
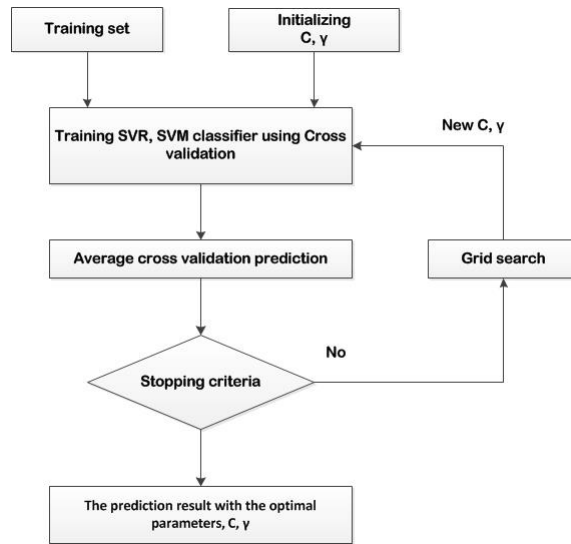
Fig. 5: Grid search algorithm for parameters selection

4.3 Prediction Results

To further explain the presented artificial intelligence techniques (SVM,SVR and BPNN) and its capability in understanding the pattern in the historical financial time series data, SVM,SVR and BPNN are applied to demonstrate their predictability of real world financial time series. In addition, this chapter presents a new hybrid model based on GA-WA method which is constructed to predict FTSE 100, S&P 500 and Nikkei 225 next day closing prices. To demonstrate the validity of the proposed methods, this section compare the obtained results of single approaches models with the results of the hybrid proposed model.

The prediction results of FTSE 100, Nikkei 225 and S&P 500 testing data sets are summarized in table 3. It can be observed from Table 3 that GA-WA hybrid model outperformed the rest of proposed methods by achieving the smallest MSE, RMSE and MAE values for FTSE 100 and S&P 500 testing data sets. Furthermore, the results of GA-WA hybrid prediction model for Nikkei 225 testing data set, is the best results in comparison with the rest of SVM, SVR, BPNN, SA and AR models. The cross correlation coefficient R of each data sets are exhibited in Table 3. It can be observed that the results of R for all methods are drawing near to 1, which implies that the predicted values and the actual are not deviating too much. Moreover, each method were run twenty times and the stander deviation was calculated. It can be observed that the results of SD for all models are relatively small, which implies hat the models are not running randomly.

Figure 6 depicts the actual testing data set of Nikkei 225 daily closing price and predicted value from the AR, SVM, BPNN, SVR, SA and GA-WA. As it can be
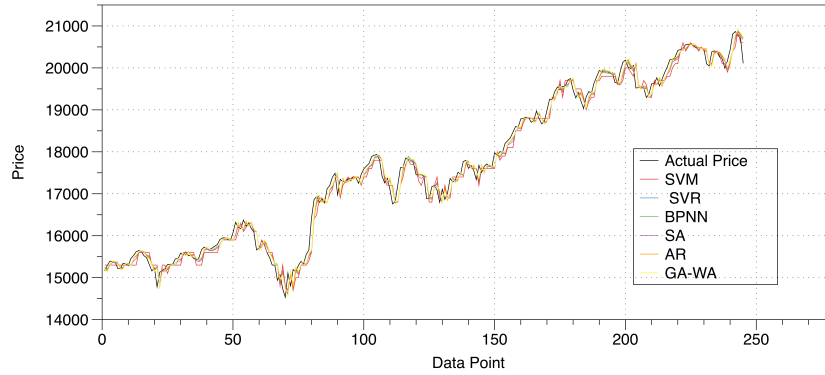
Fig. 6: The actual Nikkei 225 closing price index and its predicted value from AR, SA, SVM, SVR, BPNN and GA-WA models.
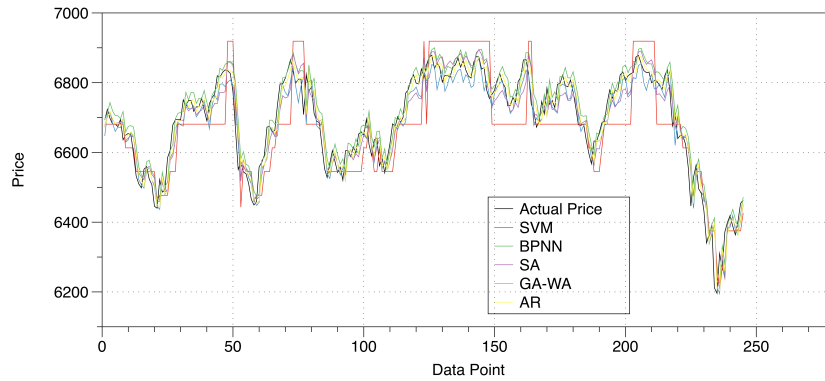

Fig. 7: The actual FTSE 100 closing price index and its predicted value from AR, SA, SVM, SVR, BPNN and GA-WA models.

observed from the Figure 6 all the utilized method have generated a good predicting results. The predicted values are very close to the actual values and to one another as well. Figure 7 presents the actual FTSE 100 closing price of the testing data set and predicted value from the AR, SVR, BPNN, SVM, SA and GA-WA. It also can be observed that the predicted obtained value from all the utilized models are very close the actual and to one another.

The predicted values of AR, SVR, SVM, BNN, SA and GA-WA for the S&P 500 closing price testing data set are illustrated in Figure 8. It can be observed from the Figure 8 that AR, SVR, SVM, SA and GA-WA predicted values are close to the actual values than BPNN model.
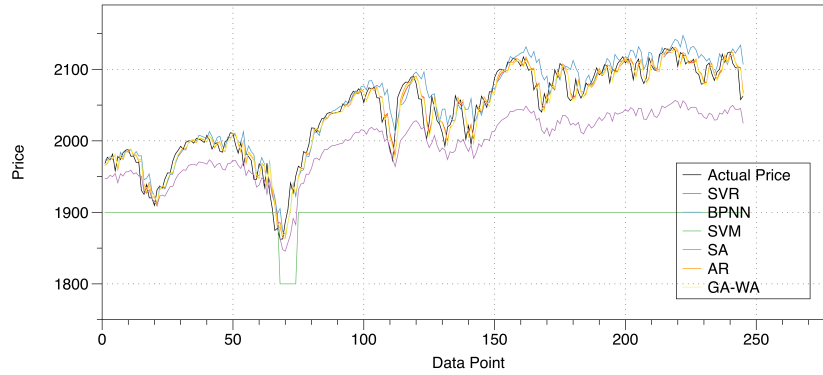
Fig. 8: The actual S&P 500 closing price index and its predicted value from AR, SA, SVM, SVR, BPNN and GA-WA models.

Table 3: The prediction result of FTSE 100, S&P 500 and Nikkei 225 using SVR,SVM, BPNN, AR, GA-WA and SA.

Index name	Models	MSE	RMSE	MAE	R	SD
FTSE100	SA	2410.96	49.10	20.45	0.99	——
	AR	2429.91	49.29	31.93	0.99	——
	SVR	2438.87	49.38	20.71	0.99	1.61
	SVM	6532.26	80.82	20.63	0.49	1.86
	BPNN	2210.75	47.02	20.41	0.99	1.74
	GA-WA	2066.65	45.46	20.15	0.99	1.64
S&P500	SA	2566.93	50.66	45.69	0.95	——
	AR	240.56	15.51	11.68	0.96	——
	SVR	244.99	15.65	11.83	0.97	1.21
	SVM	23516.58	153.35	142.20	0.36	2.09
	BPNN	602.97	22.85	18.12	0.94	3.45
	GA-WA	219.04	14.80	10.57	0.96	1.19
Nikkei225	SA	36889.34	192.06	139.98	0.99	——
	AR	34806.38	186.56	134.88	0.99	——
	SVR	34720.67	186.33	135.36	0.99	2.92
	SVM	48662.44	220.59	168.62	0.93	2.48
	BPNN	35547.91	188.50	137.028	0.99	4.00
	GA-WA	34362.89	185.37	133.49	0.99	1.79l

5 Conclusions

This paper examines the predictability of stock index time series whereas the dynamics of such a data in real situations is complex and unknown. As a result using single classical model will not produce the accurate prediction result. In this paper a hybrid combination model was introduced combining BPNN, SVM and SVR. FTSE

100 daily closing price was used to evaluate the proposed model performance. The proposed model was compared with different models and as the table 3 shows that the hybrid model outperforms all other models. From the result that were achieved in this paper, the proposed hybrid model outperform the other three single models and the benchmark models, thus future work should revolve around different hybrid combination models paradigm. Different singles models can be added and the possibility of combining them. Moreover, to test the model robustness further extension of this study can be done by testing different data sets.

References

1. Box, George EP and Draper, Norman Richard(1987) Empirical model-building and response surfaces. Wiley New York
2. Yudong Zhang and Lenan Wu(2009) Stock market prediction of s&p 500 via combination of improved bco approach and bp neural network. Expert systems with applications, 36(5):805–818 (2009)
3. Asadi, Shahrokh and Hadavandi, Esmaeil and Mehmanpazir, Farhad and Nakhostin, Mohammad Masoud(2012)Hybridization of evolutionary Levenberg–Marquardt neural networks and data pre-processing for stock market prediction. Knowledge-Based Systems, 35:245–258
4. Hadavandi, Esmaeil and Ghanbari, Arash and Abbasian-Naghneh, Salman(2010)Developing an evolutionary neural network model for stock index forecasting. Advanced Intelligent Computing Theories and Applications, Springer, Pages 407–415
5. Hill, Tim and O'Connor, Marcus and Remus, William(1996)Neural network models for time series forecasts. Management science, 42(7):1082–1092
6. Armano, Giuliano and Marchesi, Michele and Murru, Andrea(2005)A hybrid genetic-neural architecture for stock indexes forecasting. Information Sciences. Information Sciences, 170(1):3–33
7. Shen, Wei and Zhang, Yunyun and Ma, Xiaoyong(2009)Stock return forecast with LS-SVM and particle swarm optimization. International Conference on Business Intelligence and Financial Engineering, 2009. BIFE'09, IEEE , pages 143–147,
8. Kazem, Ahmad and Sharifi, Ebrahim and Hussain, Farookh Khadeer and Saberi, Morteza and Hussain, Omar Khadeer(2013)Support vector regression with chaos-based firefly algorithm for stock market price forecasting. Applied Soft Computing, 13(2):947–958
9. Pai, Ping-Feng and Lin, Kuo-Ping and Lin, Chi-Shen and Chang, Ping-Teng(2010)Time series forecasting by a seasonal support vector regression model. Expert Systems with Applications, 37(6):4261–4265
10. Al-Hnaity, Bashar and Abbod, Maysam(2015)A novel hybrid ensemble model to predict FTSE100 index by combining neural network and EEMD. Control Conference (ECC), 2015 European. IEEE, pages 3021–3028
11. Armstrong, J Scott(1989) Combining forecasts: The end of the beginning or the beginning of the end?. International Journal of Forecasting, 5(4):585–588
12. Timmermann, Allan (2006)Forecast combinations.Handbook of economic forecasting, 1:135–196
13. Xiang, C and Fu, WM (2006)Predicting the Stock Market using Multiple Models. International Conference on Control, Automation, Robotics and Vision, 2006. ICARCV'06. 9th, IEEE, pages 1–6
14. De Gooijer, Jan G and Hyndman, Rob J (2006)25 years of time series forecasting.International journal of forecasting, 22(3):443–473
15. LeBaron, Blake and Weigend, Andreas S (1998)A bootstrap evaluation of the effect of data splitting on financial time series. IEEE Transactions on Neural Networks, IEEE, 9(1):213–220

16. Balestrassi, Pedro Paulo and Popova, E and Paiva, AP de and Lima, JW Marangon (2009)De-
    sign of experiments on neural network's training for nonlinear time series forecasting. Neu-
    rocomputing, 72(4):1160–1178
17. Diaz-Robles, Luis A and Ortega, Juan C and Fu, Joshua S and Reed, Gregory D and Chow,
    Judith C and Watson, John G and Moncada-Herrera, Juan A (2008)A hybrid ARIMA and arti-
    ficial neural networks model to forecast particulate matter in urban areas: the case of Temuco,
    Chile.Atmospheric Environment, 42(35) 8331–8340
18. Kubat, Miroslav (1999) Neural networks: a comprehensive foundation by Simon Haykin,
    Macmillan. Cambridge Univ Press, ISBN 0-02-352781-7.
19. Wang, Ju-Jie and Wang, Jian-Zhou and Zhang, Zhe-George and Guo, Shu-Po (2012)Stock
    index forecasting based on a hybrid model. Omega, 40(6):758–766
20. Faruk, Durdu Ömer (2010) A hybrid neural network and ARIMA model for water quality
    time series prediction.Engineering Applications of Artificial Intelligence, 23(4):586–594
21. Cortes, Corinna and Vapnik, Vladimir (1995) Support-vector networks.Machine learning,
    Springer, 20(3): 273–297
22. Pontil, Massimiliano and Verri, Alessandro (1998)Properties of support vector ma-
    chines.Neural Computation, MIT Press, 10(4):955–974
23. Osuna, Edgar and Freund, Robert and Girosi, Federico (1997) Support vector machines:
    Training and applications.
24. Xia, Yaqing and Liu, Yulong and Chen, Zhiqian (2013)Support Vector Regression for predic-
    tion of stock trend. 2013 6th International Conference on Information Management, Innova-
    tion Management and Industrial Engineering (ICIII), IEEE, 2:123–126
25. Cherkassky, Vladimir and Ma, Yunqian (2004) Practical selection of SVM parameters and
    noise estimation for SVM regression.Neural networks, 17(1):113–126
26. Tay, Francis EH and Cao, Lijuan (2001)Application of support vector machines in financial
    time series forecasting.Omega, 29(4):309–317
27. Cao, Li-Juan and Tay, Francis EH (2003) Support vector machine with adaptive parameters in
    financial time series forecasting. IEEE Transactions on Neural Networks, 14(6):1506–1518
28. Li, Yuhong and Ma, Weihua (2010) Applications of artificial neural networks in financial eco-
    nomics: a survey. 2010 International Symposium on Computational Intelligence and Design
    (ISCID), IEEE, 1:211–214
29. Theofilatos, Konstantinos and Karathanasopoulos, Andreas and Middleton, Peter and Geor-
    gopoulos, Efstratios and Likothanassis, Spiros (2013) Modeling and Trading FTSE100 Index
    Using a Novel Sliding Window Approach Which Combines Adaptive Differential Evolution
    and Support Vector Regression.Artificial Intelligence Applications and Innovations, Springer,
    486–496
30. Hall, James W (1994) Adaptive selection of US stocks with neural nets. Trading on the edge:
    neural, genetic, and fuzzy systems for chaotic financial markets. New York: Wiley, 45–65
31. Abu-Mostafa, Yaser S and Atiya, Amir F (1996) Introduction to financial forecasting. Applied
    Intelligence, Springer, 6(3):205–213
32. Zhang, G Peter (2003) Time series forecasting using a hybrid ARIMA and neural network
    model. Neurocomputing, 50:159–175
33. Chatfield, C (1988) What is the bestmethod of forecasting. Journal of Applied Statistics,
    15(1):19–38
34. Gupta, Sunil and Wilton, Peter C (1987) Combination of forecasts: An extension. Manage-
    ment Science, 33(3):356–372
35. Christodoulos, Charisios and Michalakelis, Christos and Varoutas, Dimitris (2010) Forecast-
    ing with limited data: Combining ARIMA and diffusion models. Technological forecasting
    and social change, 77(4):558–565
36. John Henry Holland (1992) Adaptation in natural and artificial systems: an introductory anal-
    ysis with applications to biology, control, and artificial intelligence.MIT press
37. Goldberg, David E (1989) Genetic Algorithm in Search, Optimization and Machine Learning,
    Addison. W esley Publishing Company, R eading, MA, 1(98):9
38. Koza, John R(1992)Genetic programming: on the programming of computers by means of
    natural selection.MIT press, 1

39. Wu, Berlin and Chang, Chih-Li (2002) Using genetic algorithms to parameters (d, r) estimation for threshold autoregressive models. Computational Statistics & Data Analysis, 38(3):315–330
40. Han, Jiawei and Kamber, Micheline and Pei, Jian (2011) Data mining: concepts and techniques: concepts and techniques.Elsevier
41. Fuller, Wayne A (2009) Introduction to statistical time series.John Wiley & Sons, :428
42. Armstrong, Jon Scott (2001) Principles of forecasting: a handbook for researchers and practitioners.Springer Science & Business Media, 30
43. Hyndman, Rob J and Koehler, Anne B (2006) Another look at measures of forecast accuracy.International journal of forecasting, 22(4):679–688
44. Cao, Qing and Leggio, Karyl B and Schniederjans, Marc J(2005) A comparison between Fama and French's model and artificial neural networks in predicting the Chinese stock market. Computers & Operations Research, 32(10):2499–2512
45. Zhang, Dongsong and Zhou, Lina (2004) Discovering golden nuggets: data mining in financial application. Applications and Reviews, IEEE Transactions on Systems, Man, and Cybernetics, Part C, IEEE, 34(4):513–522
46. Hsu, Chih-Wei and Chang, Chih-Chung and Lin, Chih-Jen and others (2003) A practical guide to support vector classification.