Applied Mathematical Modelling 33 (2009) 2444–2454

Contents lists available at ScienceDirect

Applied Mathematical Modelling

journal homepage: www.elsevier.com/locate/apm

Electric load forecasting by support vector model

Wei-Chiang Hong *

Department of Information Management, Oriental Institute of Technology 58, Section 2, Sichuan Road, Panchiao, Taipei County 220, Taiwan

A R T I C L E I N F O

Article history:
Received 24 April 2008
Received in revised form 17 July 2008
Accepted 18 July 2008
Available online 5 August 2008

Keywords:
Support vector regression (SVR)
Immune algorithm (IA)
Electric load forecasting

A B S T R A C T

Accurately electric load forecasting has become the most important management goal, however, electric load often presents nonlinear data patterns. Therefore, a rigid forecasting approach with strong general nonlinear mapping capabilities is essential. Support vector regression (SVR) applies the structural risk minimization principle to minimize an upper bound of the generalization errors, rather than minimizing the training errors which are used by ANNs. The purpose of this paper is to present a SVR model with immune algorithm (IA) to forecast the electric loads, IA is applied to the parameter determine of SVR model. The empirical results indicate that the SVR model with IA (SVRIA) results in better forecasting performance than the other methods, namely SVMG, regression model, and ANN model.

1. Introduction

In the recent years, along with the power system privatized and deregulated, the issue of accurately electric load forecasting has received more attention in a regional or a national system. The error of electric load forecasting may increase the operating cost [1–3]. Therefore, overestimation of future load results in excess supply, and it is also not welcome to the international energy network. In the contrast, underestimation of load leads to a failure in providing enough reserve and implies high costs in peaking unit. Adequate electric production requires each member of the global cooperation being able to forecast its demands accurately. However, it is complex to predict the electric load, because the influencing factors include climate factors, social activities, and seasonal factors. Climate factors depend on the temperature and humidity; social factors imply human social activities including work, school and entertainment affecting the electric load; seasonal factors then include seasonal climate change and load growth year after year.

In the last few decades, there are widespread references with regard to the efforts improving the accuracy of forecasting methods. One of these methods is a weather-insensitive approach which used historical load data to infer the future electric load. It is famous known as Box-Jenkins' ARIMA [4–7], which is theoretically based on univariate time sequences. In addition, Christianse [8] and Park et al. [9] proposed exponential smoothing models by Fourier series transformation to forecast electric load. Douglas et al. [10] considered verifying the impacts of forecasting model in terms of temperature. They combined Bayesian estimation with dynamic linear model into load forecasting. The results indicate that the presented model is suitable for predicting load with imperfect weather information. The disadvantage of these methods is time consuming, particularly for the situation while the number of variables is increased. Recently, to avoid a lot of variables selection problem, Azadeh et al. [11] employ fuzzy system to provide an ideal rule base to determine which type of ARMA models should be used, the results also indicate that the integrated approach outperform those novel intelligent computing models. Wang et al. [12] propose hybrid ARMAX (auto-regressive and moving average with exogenous variables) model with particle

* Tel.: +886 2 7738 0145x327#55; fax: +886 2 7738 6310.
E-mail address: samuelhong@ieee.org

0307-904X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.apm.2008.07.010

swarm optimization to efficiently solve the problem of trapping into local minimum which is caused by exogenous variable (e.g., weather condition). Their results also reveal that the proposed approach has superior forecasting accuracy.

To achieve the accuracy of load forecasting, state space and Kalman filtering technologies, developed to reduce the difference between actual loads and prediction loads (random error), are employed in load forecasting model. This approach introduces the periodic component of load as a random process. It requires historical data more than 3–10-year to construct the periodic load variation and to estimate the dependent variables (load or temperature) of power system [13–15]. Moghram and Rahman [16] proposed a model based on this technique and verified that the proposed model outperforms another four forecasting methods (multiple linear regression, time series, exponential smoothing, and knowledge-based approach). The disadvantage of these methods is difficult to avoid the observation noise in the forecasting process especially multivariable considered. Recently, Al-Hamadi and Soliman [17] employ fuzzy rule-based logic, by utilizing a moving window of current values of weather data as well as the recent past history of load and weather data, to recursively estimate the optimal fuzzy parameters for each hour load of the day. Amjady [18] proposes hybrid model of the forecast-aided state estimator (FASE) and the multi-layer perceptron (MLP) neural network to forecast short-term bus load of power systems. The proposed hybrid model has been examined on a real power system, and the results show that the hybrid method has better prediction accuracy than the other models, such as MLP, FASE, and the periodic auto-regression (PAR) model.

Regression models construct the causal-effect relationships between electric load and independent variables. The most popular models are linear regression, proposed by Asbury [19], considering the "weather" variable into forecasting model. Papalexopoulos and Hesterberg [20] added the factors of "holiday" and "temperature" into their proposed model. The proposed model used weight least square method to obtain robust parameter estimation encountering with the heteroskedasticity. Soliman et al. [21] proposed a multivariate linear regression model in load forecasting, including temperature, wind cooling/humidity factors. The empirical results indicate that the proposed model outperforms the harmonic model as well as the hybrid model. Similarly, Mirasgedis et al. [22] also incorporate weather meteorological variables, such as relative humidity, heating, and cooling degree-days to forecast electricity demand in Greece. In contrast, Mohamed and Bodger [23] employ economic and geographic variables (such as GDP, electricity price, and population) to forecast electricity consumption in New Zealand. These models are based on linear assumption, however, these independent variables are unjustified to be used because of the terms are known to be nonlinear. Recently, Tsekouras et al. [24] introduce a nonlinear multivariable regression approach to forecast annual load, by considering correlation analysis with weighting factors to select appropriate input variables.

In the recent decade, lots of researches had tried to apply the artificial intelligent techniques to improve the accuracy of the load forecasting issue. Knowledge-based expert system (KBES) and artificial neural networks (ANNs) are the popular representatives. The KBES approaches constructed electric load forecasting by simulating the experiences of the system operators who were well-experienced in the processes of electricity generation, such as Rahman and Bhatnagar [25]. The characteristic feature of this approach is rule-based, which implied that the system transformed new rule from received information. In other word, an expert capability which is training by the existence presuming will be made much increasing accuracy of forecasting [25–27]. This approach is derivation of the rules from on-the-job training and sometimes transforming the information logic to equations could be impractical. Recently, applications of fuzzy inference system and fuzzy theory in load forecasting are also received attentions, Ying and Pan [28] introduce adaptive network fuzzy inference system (ANFIS), by looking for the mapping relation between the input and output data to determine the optimal distribution of membership functions, to forecast regional load. Pai [29] and Pandian et al. [30] all employ fuzzy approaches to get superior performance in terms of load forecasting.

Meanwhile, lots of researches also had tried to apply ANNs to improve the load forecasting accuracy. Park et al. [31] proposed a 3-layer back-propagation neural network to daily load forecasting problems. The inputs include three indices of temperature: average, peak and lowest loads. The outputs are peak loads. The proposed model outperforms the regression model and the time series model in terms of forecasting accuracy index, mean absolute percent error (MAPE). Novak [32] applied the radial basis function (RBF) neural networks to forecast electricity load. The results indicate that RBF is at least 11 times faster and more reliable than the back-propagation neural networks. Darbellay and Slama [33] applied the ANNs to predict the electricity load in Czech. The experimental results indicate that the proposed ANN model outperform the ARIMA model in terms of forecasting accuracy index, normalized mean square error (NMSE). Abdel-Aal [34] proposed an Abductive network to conduct one-hour-ahead load forecast for five years. Hourly temperature and hourly load data are considered. The results of the proposed model are very promising in terms of forecasting accuracy index, MAPE. Hsu and Chen [35] employed the ANNs model to forecast the regional electricity load in Taiwan. The empirical results indicate that proposed model is superior to traditional regression model. Recently, applications of hybrid ANNs model with statistical methods or other intelligent approaches have received a lot of attentions, such as hybrid with Bayesian inference [36,37], self-organizing map [38,39], wavelet transform [40,41], particle swarm optimization [42], and dynamic mechanism [43].

The support vector machines (SVMs) implement the structural risk minimization (SRM) principle rather than empirical risk minimization principle implemented by most of the traditional neural network models. Based on this principle, SVMs achieve an optimum networks structure. In addition, the SVMs will be equivalent to solving a linear constrained quadratic programming problem so that the solution of SVMs is always unique and globally optimal. Along with the introduction of Vapnik's ε-insensitive loss function [44], SVMs also have been extended to solve nonlinear regression estimation problems. Therefore, SVMs are successfully in time series forecasting. Cao [45] used the SVMs experts for time series forecasting. The generalized SVMs experts contained a two-stage neural network architecture. The numerical results indicated that the SVMs experts are capable to outperform the single SVMs models in terms of generalization comparison. Cao and Gu [46] proposed a dynamic SVMs model to deal with non-stationary time series problems. Experiment results showed that the DSVMs outperform standard SVMs in forecasting non-stationary time series. Meanwhile, Tay and Cao [47] used SVMs in forecasting financial time series. The numerical results indicated that the SVMs are superior to the multi-layer back-propagation neural network in financial time series forecasting. Hong and Pai [48] applied SVMs to predict engine reliability. Their experimental results indicated that SVMs outperform Duane model, ARIMA model and general regression neural networks model. Pai et al. [49] proposed a multi-factor support vector machine model to forecast Taiwanese demand for travel to Hong Kong from 1967 to 1996. They indicated that the proposed MSVM model outperforms BP model, FF model, Holt's model, MA model, Naïve model, and Multiple- regression model. For electric load forecasting, Chen et al. [50] are the pioneers for proposing a SVM model, which was the winning entry of a competition aiming at mid-term load forecasting (predicting daily maximum load of the next 31 days) organized by EUNITE network in 2001, to solve the problem. They discuss in detail how SVM, a new learning technique, is successfully applied to load forecasting. Pai and Hong [51] employed the concepts of Jordan recurrent neural networks to construct recurrent SVR model in Taiwan regional long-term load forecasting. In addition, they used genetic algorithms to determine approximate optimal parameters in the proposed RSVMG model. They concluded that RSVMG outperformed other models, such as SVMG, ANN, and regression models. Similarly, Pai and Hong [52] proposed a hybrid model of SVR and simulated annealing (SA) algorithms to forecast Taiwan long-term electric load. In which, SA is employed to select approximate optimal parameters in the proposed SVMSA model. Conclusively, they indicated that SVMSA is superior to ARIMA and GRNN models in terms of MAPE, MAD, and NRMSE.

According to the brief review of literature related to electric load forecasting approaches, this study attempts to develop a support vector regression (SVR) model for forecasting Taiwan regional electricity load. Meanwhile, the most novel optimum search algorithm, immune algorithm (IA), is applied to determine three parameters (σ, C, and ε) in a SVR model. A numerical example in the literature [35] is employed to compare the forecasting performance of the proposed model. In addition, some particular comparison of optimum search algorithms in determining three parameters of a SVR is conducted. The remainder of this paper is organized as follows. The SVR and IA (SVRIA) are introduced in Section 2. A numerical example is presented in Section 3. Conclusions are discussed in Section 4.

2. Methodology

2.1. SVR model

The support vector machines (SVMs) were proposed by Vapnik [53]. The basic concept of the SVR is to map nonlinearly the original data x into a higher dimensional feature space. Hence, given a set of data $G = \{(x_i, d_i)\}_{i=1}^{N}$ (where x_i is the input vector; d_i is the actual value, and N is the total number of data patterns), the SVR function is

$$y = f(x) = w\psi(x) + b, \tag{1}$$

where $\psi(x)$ is called feature which is nonlinear mapped from the input space $\mathbf{x}$. The w and b are coefficients which are estimated by minimizing the regularized risk function

$$R(C) = (C/N)\sum_{i=1}^{N} L_\varepsilon(d_i, y_i) + \|w\|^2/2, \tag{2}$$

where

$$L_\varepsilon(d, y) = \begin{cases} 0 & |d - y| \leqslant \varepsilon \\ |d - y| - \varepsilon & \text{otherwise}, \end{cases} \tag{3}$$

and C and ε are prescribed parameters. In Eq. (2), $L_\varepsilon(d,y)$ is called the ε-insensitive loss function. The loss equals zero if the forecasted value is within the ε-tube [54,55] (see Eq. (3)). The second term, $\|w\|^2/2$, measures the flatness of the function.

Therefore, C is considered to specify the trade-off between the empirical risk and the model flatness. Both C and ε are user-determined parameters. Two positive slack variables ζ and ζ^*, which represent the distance from actual values to the corresponding boundary values of ε-tube, are introduced. Then, Eq. (2) is transformed into the following constrained form:

Minimize

$$R(w, \zeta, \zeta^*) = \|w\|^2/2 + C\left(\sum_{i=1}^{N}(\zeta_i + \zeta_i^*)\right) \tag{4}$$

with the constraints,

$$\begin{aligned} & w\psi(x_i) + b_i - d_i \leqslant \varepsilon + \zeta_i^*, i = 1, 2, \ldots, N \\ & d_i - w\psi(x_i) - b_i \leqslant \varepsilon + \zeta_i, i = 1, 2, \ldots, N \\ & \zeta_i, \zeta_i^* \geqslant 0, i = 1, 2 \ldots, N. \end{aligned}$$

This constrained optimization problem is solved using the following primal Lagrangian form:'

$$L(w,b,\zeta,\zeta^*,\alpha_i,\alpha_i^*,\beta_i,\beta_i^*)=\frac{1}{2}\|w\|^2+C\left(\sum_{i=1}^{N}(\zeta_i+\zeta_i^*)\right)-\sum_{i=1}^{N}\beta_i[w\psi(x_i)+b-d_i+\varepsilon+\zeta_i] \\ -\sum_{i=1}^{N}\beta_i^*[d_i-w\psi(x_i)-b+\varepsilon+\zeta_i^*]-\sum_{i=1}^{N}(\alpha_i\zeta_i+\alpha_i^*\zeta_i^*). \tag{5}$$

Eq. (5) is minimized with respect to primal variables w, b, ζ and ζ^*, and maximized with respect to nonnegative Lagrangian multipliers α_i, α_i^*, β_i and β_i^*. Therefore, Eqs. (6)–(9) are obtained.

$$\frac{\partial L}{\partial w}=w-\sum_{i=1}^{N}(\beta_i-\beta_i^*)\psi(x_i)=0, \tag{6}$$

$$\frac{\partial L}{\partial b}=\sum_{i=1}^{N}(\beta_i^*-\beta_i)=0, \tag{7}$$

$$\frac{\partial L}{\partial \zeta_i}=C-\beta_i-\alpha_i=0, \tag{8}$$

$$\frac{\partial L}{\partial \zeta_i^*}=C-\beta_i^*-\alpha_i^*=0. \tag{9}$$

Finally, Karush–Kuhn–Tucker conditions are applied to the regression, and Eq. (4) thus yields the dual Lagrangian by substituting Eqs. (6)–(9) into Eq. (5). Then, the dual Lagrangian, Eq. (10), is obtained when kernel function is $K(x_i,x_j)=\psi(x_i)\psi(x_j)$,

$$\vartheta(\beta_i,\beta_i^*)=\sum_{i=1}^{N}d_i(\beta_i-\beta_i^*)-\varepsilon\sum_{i=1}^{N}(\beta_i+\beta_i^*)-\frac{1}{2}\sum_{i=1}^{N}\sum_{i=1}^{N}(\beta_i-\beta_i^*)(\beta_j-\beta_j^*)K(x_i,x_j) \tag{10}$$

subject to the constraints,

$$\sum_{i=1}^{N}(\beta_i-\beta_i^*)=0,$$
$$0\leqslant\beta_i\leqslant C,\ i=1,2,\cdots,N,$$
$$0\leqslant\beta_i^*\leqslant C,\ i=1,2,\cdots,N.$$

The Lagrange multipliers in Eq. (10) satisfy the equality $\beta_i*\beta_i^*=0$. The Lagrange multipliers β_i and β_i^*, are calculated and an optimal desired weight vector of the regression hyperplane is

$$w^*=\sum_{i=1}^{N}(\beta_i-\beta_i^*)\psi(x). \tag{11}$$

Hence, the regression function is

$$f(x,\beta,\beta^*)=\sum_{i=1}^{l}(\beta_i-\beta_i^*)K(x,x_i)+b. \tag{12}$$

Here, $K(x,x_i)$ is called the kernel function. The value of the kernel is equal to the inner product of two vectors x and x_i in the feature space $\psi(x)$ and $\psi(x_i)$, i.e., $K(x,x_i)=\psi(x)^*\psi(x_i)$. In general, there are three types of common examples of kernel function, the polynomial kernel, $K(x_i,x)=(a_1x_i^{\mathrm{T}}x+a_2)^d$ (with degree d, a_1 and a_2 represent the coefficients); the multi-layer perceptron kernel function, $K(x_i,x)=\tanh(x_i^{\mathrm{T}}x-b)$ (where b is the constant); and the Gaussian RBF kernel function, $K(x_i,x)=\exp(-\|x_i-x\|^2/2\sigma^2)$. Till now, it is hard to determine the type of kernel functions for specific data patterns [55,56]. However, any function that satisfies Mercer's condition by Vaplink [53] can be used as the Kernel function. In this work, the Gaussian function is used in the SVR. The parameters that users have to specify are the error goal ε, the constant C and the width of the radial basis function σ.

The selection of three parameters, σ, ε and C, of a SVR model is important to the accuracy of forecasting. For example, if C is too large (approximated to infinity), then the objective is to minimize the empirical risk, $L_\varepsilon(d,y)$ only, without model flatness in the optimization formulation Eq. (4). Parameter ε controls the width of the ε-insensitive loss function, which is used to fit the training data. Large ε-values result in more flat regression estimated function. Parameter σ controls the Gaussian function width, which reflects the distribution range of **x**-values of training data. Therefore, all the three parameters affect model constructing in different ways. There are lots of existing practical approaches to the selection of C and ε, such as user-defined based on priori knowledge and experience, cross-validation, and asymptotical optimization [57]. However, structural methods for efficiently and simultaneously confirming the selection of those three parameters efficiently are lacking. Therefore, immune algorithm (IA) is used in the proposed SVR model to optimize parameter selection.

2.2. Immune algorithm for parameter selection of SVR model

IA, proposed by Mori et al. [58] is used in this study, is based on the learning mechanism of natural immune systems. The natural immune system is a complex adaptive system that efficiently employs several mechanisms to recognize all cells within the body and classify those cells as self or non-self. Additionally, the non-self cells are further categorized to stimulate an appropriate type of defensive mechanism for defending against foreign invaders, such as bacteria and viruses. The lymphocyte is the main type of immune cell participating in the immune response. The lymphocyte contains two subclasses: T and B. Each subclass has its own function. When an antigen enters the bloodstream and lymphatic system, the antigen encounters B-cells, and antibodies anchored in the membrane of B-cells recognize antigens in the bacteria. T-cells, which have already received communication from macrophages about the antigen, then communicate with B-cells and stimulate their proliferation. The proliferated B-cells turn into memory cells and produce antibodies. After the antibodies enter the bloodstream via the heart, the antibodies bind to antigens and kill them with the help of macrophages and other proteins.

Analogous to the natural immune system, the IA has the ability to seek out the best solution for optimization problems. In the IA procedure, the optimization problem can be viewed as antigens. Conversely, the feasible solutions of the optimization problem are treated as antibodies (B-cells). The detail procedure is as followings (see Fig. 1).

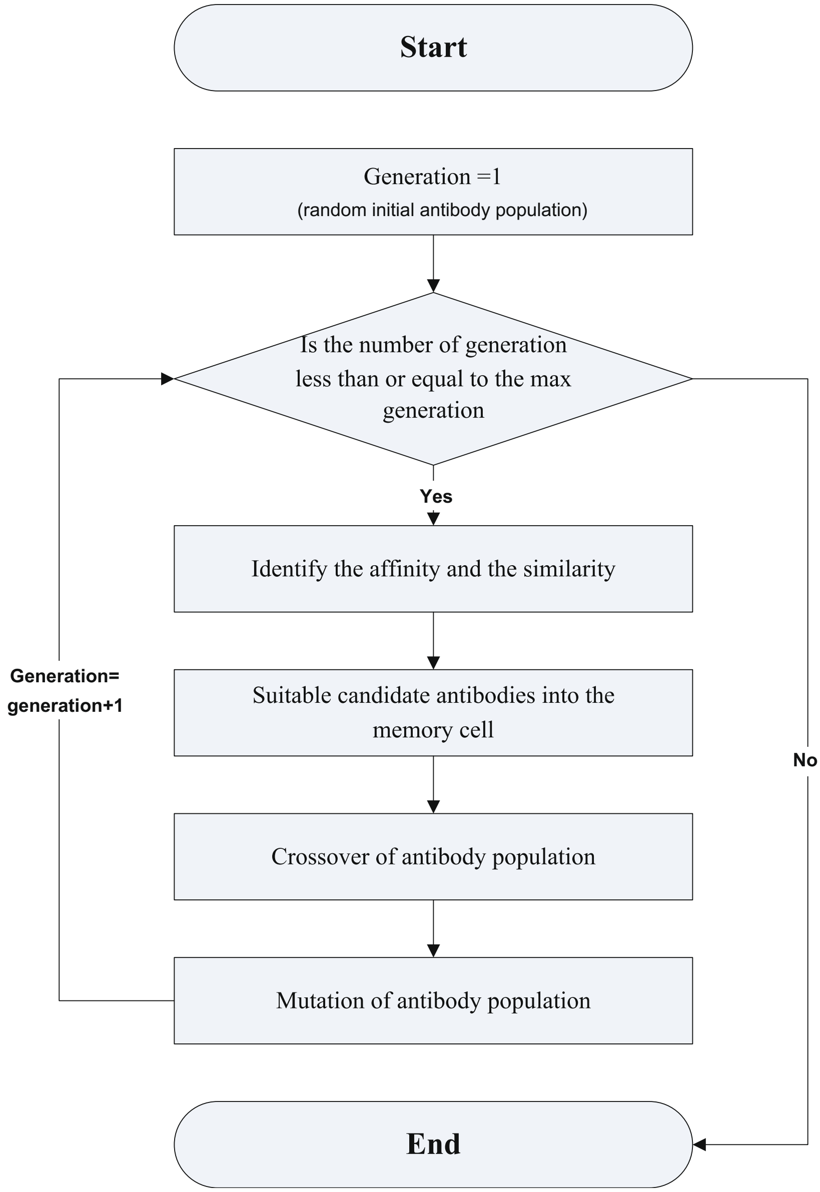

Fig. 1. The procedure of immune algorithm.

Step 1: Random initialization of antibody population

The initial antibody population represented by binary-code string, including three parameters (σ, C, and ε) of a SVR model, is generated randomly. For example, assume that an antibody contains 12 binary codes to represent three SVR parameters. Each parameter is thus expressed by four binary codes. Therefore, for example, assume the set-boundaries for parameters σ, C, and ε are 2, 10, and 0.5 respectively, then, the antibody with binary-code "1 0 0 1 0 1 0 1 0 0 1 1" implies that the real values of the three parameters σ, C, and ε are 1.125, 3.125, and 0.09375, respectively. The number of initial antibodies is the same as the size of the memory cell. The size of the memory cell is set to ten in this study.

Step 2: Identifying the affinity and the similarity

A higher affinity value implies that an antibody has a higher activation with an antigen. To maintain the diversity of the antibodies stored in the memory cells, the antibodies with lower similarity have higher probability of being included in the memory cell. Therefore, an antibody with a higher affinity value and a lower similarity value has a good likelihood of entering the memory cells. The affinity between the antibody and antigen is defined as

$$Ag_k = 1/(1 + d_k), \tag{13}$$

where d_k denotes the SVR forecasting errors obtained by the antibody k.The similarity between antibodies is expressed as

$$Ab_{ij} = 1/(1 + T_{ij}), \tag{14}$$

where T_{ij} denotes the difference between the two SVR forecasting errors obtained by the antibodies inside (existed) and outside (will be entering) the memory cell.

Step 3: Selection of antibodies in the memory cell

Antibodies with higher values of Ag_k are considered to be potential candidates for entering the memory cell. However, the potential antibody candidates with Ab_{ij} values exceeding a certain threshold are not qualified to enter the memory cell. In this investigation, the threshold value is set to 0.9.

Step 4: Crossover and mutation of antibody population

New antibodies are created via crossover and mutation operations. To perform crossovers, strings representing antibodies are paired randomly. Moreover, segments of paired strings between two determined break-points are swapped. Mutations are performed randomly by converting a "1" code into a "0" code or a "0" code in to a "1" code. The crossover and mutation rates are determined using probabilities. In this investigation, the probabilities are set to 0.5 and 0.1 for crossover and mutation, respectively.

Step 5: Stopping criteria

If the number of generations equals a given scale, then the best antibody is a solution, otherwise return to Step 2.The IA is used to seek a better combination of the three parameters in SVR. The value of the mean absolute percent error (MAPE) is used as the criterion (the smallest value of MAPE) of forecasting errors to determine the suitable parameters used in SVR model, which is given by Eq. (15).

2.3. Index of performance evaluation

In this study, the forecasting error index, namely MAPE (mean absolute percent error), is used as forecasting accuracy measures. The index is shown as follows, respectively,

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{d_i - y_i}{d_i}\right| \times 100\%, \tag{15}$$

where N is the number of forecasting periods; d_i is the actual production value at period i; and y_i is the forecasting load value of Taiwan power demand at period i.

3. Numerical example

3.1. Data set

This study uses Taiwan regional electric load data to compare the forecasting performances of SVRIA models with those of ANN and regression models proposed by Hsu and Chen [35]. In addition, due to the same application type of SVM, authors' previous proposed model, SVMG, is also involved in comparison (notice that RSVMG model had not only contained the concepts of SVM but also the applications of Jordan recurrent neural network, therefore, RSVMG model is beyond the same comparable scope with respect to this proposed SVRIA and could not be involved in comparison). Table 1 lists Taiwanese historical annual load demand data used in this example. Totally, there are 20 data (from 1981 to 2000) of Taiwan regional electricity load. It is necessary to compare the forecast performance on the same basis. Therefore, the data are divided into three data sets: the training data set (12 years, from 1981 to 1992), validation data set (4 years, from 1993 to 1996), and the testing data set (4 years, from 1997 to 2000). The data sets are listed in Table 2.

Table 1
Taiwan regional annual load demand (from 1981 to 2000)

Year	Northern regional load values	Central regional load values	Southern regional load values	Eastern regional load values
1981	3388	1663	2272	122
1982	3523	1829	2346	127
1983	3752	2157	2494	148
1984	4296	2219	2686	142
1985	4250	2190	2829	143
1986	5013	2638	3172	176
1987	5745	2812	3351	206
1988	6320	3265	3655	227
1989	6844	3376	3823	236
1990	7613	3655	4256	243
1991	7551	4043	4548	264
1992	8352	4425	4803	292
1993	8781	4594	5192	307
1994	9400	4771	5352	325
1995	10,254	4483	5797	343
1996	10,719	4935	6369	363
1997	11,222	5061	6336	358
1998	11,642	5246	6318	397
1999	11,981	5233	6259	401
2000	12,924	5633	6804	420

Unit: MW.

Table 2
Training, validation, and testing data sets of the proposed model

Data sets	SVRIA model	SVMG model (Pai and Hong [31])	ANN model (Hsu and Chen [23])
Training data	1981–1992	1981–1992	1981–1996
Validation data	1993–1996	1993–1996	
Testing data	1997–2000	1997–2000	1997–2000

3.2. Parameters determination of three models

In the training stage, the rolling-based forecasting procedure (see Fig. 2) is conducted, which dividing training data into two subsets, namely fed-in (8 load data) and fed-out (4 load data) respectively. Firstly, the primary 8 load data of fed-in subset are feeding into the SVRIA model, and the structural risk minimization principle is employed to minimize the training error, then obtain one-step ahead forecasting load, namely the 9th forecasting load. Secondly, the next 8 load data, including 7 of the fed-in subset data (from 2nd to 8th) pulsing the 9th data in the fed-out subset, are similarly again fed into the SVRIA model, the structural risk minimization principle is also employed to minimize the training error, then obtain one-step ahead forecasting load, namely the 10th forecasting load. Repeat the rolling-based forecasting procedure till the 12th forecasting load is obtained. Meanwhile, training error in this training stage is also obtained. Different regions of Taiwan electric loads in a time series are fed into the SVRIA model to forecast electric load in the next validation period.

While training errors improvement occurs, the three kernel parameters, σ, C, and ε of SVMIA model adjusted by IA are employed to calculate the validation error. Then, the adjusted parameters with minimum validation error are selected as the most appropriate parameters. Finally, a four-steps-ahead policy is used to forecast electric load in each region. Note that the testing data sets are not used for modeling but for examining the accuracy of the forecasting model. The forecasting results and the suitable parameters for the different regional SVRIA models are illustrated in Table 3.

3.3. Forecasting results and discussions

The forecasting results of various forecasting models with accuracy index, MAPE (see Eq. (15)), are illustrated in Table 4. Firstly, the proposed SVRIA model has smaller MAPE values (except the northern region presents a little inaccuracy, accuracy difference 0.23%) than the ANN and regression models. Particularly, the ANN model seems failing to capture the load decreasing trend from 1998 to 1999 both in the central and southern regions. Secondly, SVRIA model is superior to SVMG model in terms of MAPE.

For the first forecasting results regarding to the facts that the ANN model fails to capture the load decreasing trend from 1998 to 1999 both in the central and southern regions, this is the known drawback that the ANN model is time consuming to be required sufficient training data to learn more ‘‘expert rules” to approximately predict the transition trend of electric load due to other accident events, such as (1) Missile Military Maneuvers between Taiwan Strait in 1995 leads lots of businesses to decide to move out Taiwan (those businesses’ manufacturing factories are often set up in the central and southern

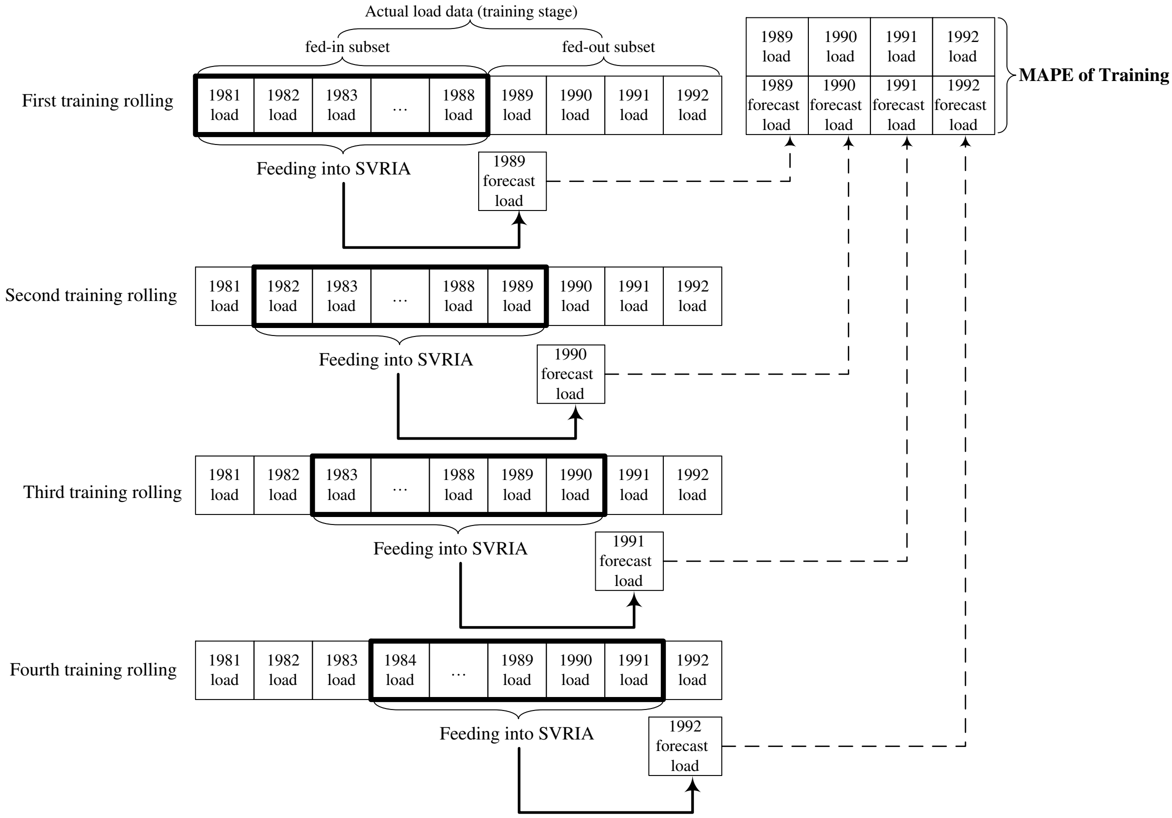

Fig. 2. The rolling-base forecasting procedure (training stage).

Table 3
Forecasting results and parameters of SVRIA models

Regions	Parameters			MAPE of testing (%)
	σ	C	ε	
Northern	0.710	2.095×10^{10}	2.32	1.29
Central	8.085	6.860×10^{10}	1.22	1.36
Southern	0.778	1.626×10^{10}	4.53	1.74
Eastern	12.920	1.894×10^{10}	1.50	2.45

regions); and (2) September 21 Earthquake in Central Taiwan in 1999 leads electricity supply out of control at least 6 months. The electric load data employed in this manuscript is only 20 data (from 1981 to 2000) which may not provide sufficient data tendency information for the training process while ANN modeling. In the meanwhile, the two accident events mentioned above did not seriously affect the northern region, thus, the actual electric load did not appear deceasing trend from 1998 to 1999, ANN model could easily learn to capture the transition tendency of electric load. For the eastern region, due to Southeast Asian Financial Crisis in 1997, the tourism industries in the region all suffered from the economic crisis, global arrival figures for 1997 reflected the difficult market conditions in the growth rate decline, thus, the actual electric load did not grow increasingly in 1997, however, ANN model optimistically forecasted the large growth and led irretrievable forecasting error. Original SVR model focuses on, as mentioned previously, empirical risk minimization (ε-insensitive loss function) rather than "expert rules" learning to approximately predict the transition trend of electric load. This advantages lead SVR model to be able to deal with any data pattern no matter data tendencies may present fluctuation or sustained increasing or decreasing types. Contributing to forecasting accuracy is that SVR model almost provides the smallest MAPE values in each regional electric load forecasting. However, expert rules learning is useful to be involved into the SVR model, SVRIA model is the empirical representative.

It is interesting to address the superiority that the SVRIA model outperforms the SVMG model. In the IA procedure, the feasible solutions of the optimization problem are treated as antibodies (B-cells) and have retained into memory cells. These

Table 4
Forecasting results of SVRIA, SVMG, ANN, and regression models

Year	Actual	SVRIA	SVMG	ANN	Regression
Northern regional					
1997	11,222	11,229	11,213	10,991	11,262
1998	11,642	11,642	11,747	11,643	12,162
1999	11,981	12,056	12,173	11,804	12,395
2000	12,924	12,349	12,543	12,834	13,122
MAPE		**1.29**	**1.40**	**1.06**	**2.45**
Southern regional					
1997	6336	6432	6265	6305	6493
1998	6318	6448	6389	6476	6868
1999	6259	6295	6346	6537	7013
2000	6804	6615	6513	6672	7481
MAPE		**1.74**	**2.02**	**2.48**	**8.29**
Central regional					
1997	5061	5017	5060	5112	5361
1998	5246	5278	5203	5301	5711
1999	5233	5245	5230	5350	5780
2000	5633	5432	5297	5572	6131
MAPE		**1.35**	**1.81**	**1.73**	**8.52**
Eastern regional					
1997	358	391	358	378	380
1998	397	392	373	403	407
1999	401	402	397	410	413
2000	420	420	408	435	440
MAPE		**2.45**	**2.65**	**3.62**	**4.1**

Unit: MW.

memory cells play the role of stocks for useful learned expert rules to avoid disappearing while crossover or mutation processes are conducting. In addition, the function of memory cells could also maintain the optimizing diversity via retaining higher-affinity-value antibody which has a higher activation with an antigen, thus, the optimizing diversity could be led to the orientation of higher-affinity, i.e., leading to approximate optimum solution, meanwhile, avoiding arbitrarily searching of genetic algorithms. This is the primary reason that the SVRIA model could provide more accurate forecasting results than the SVMG model.

4. Conclusions

From the historical data, the Taiwan regional electric load values show a strong growth trends, particularly in the northern region. This is a common electric load phenomenon in the developing countries. However, the growth rate in electric demand seems to become the rigorous mission to avoid overproduction or underproduction electricity load. In this paper, we employed a novel forecasting technique, SVRIA, to examine its potentiality in forecasting regional electric loads. Three other forecasting approaches, the SVMG, ANN, and the regression models, are used to compare the forecasting performance. Experiment results indicate that the proposed SVRIA model outperforms the other approaches in terms of forecasting accuracy, except the northern region.

This study is the first to apply the SVR model with IA to forecast electric load. Authors, so far, had finished a series of researches regarding suitable parameters determine of a SVR load forecasting model by employing novel optimization algorithms (including genetic algorithms, simulated annealing algorithm, and immune algorithm; see Pai and Hong [51,52]). Those empirical results obtained in these studies reveal that the hybrid model of SVR and proposed algorithms is a valid alternative in the electric industry while original SVR model is applied (see Pai and Hong [51,52], and this manuscript). If the data pattern is more complicate where the original SVR model could not obtain any ideal forecasting results, it is feasible to apply the recurrent SVR model (Pai and Hong [51]). Of course, recurrent SVR models are time consuming than original SVR models. On the other hand, electric load forecasting model ought to contain several social factors to increase its explanation capabilities, i.e., multivariate forecasting models, such as social activities and seasonal factors could be introduced into the SVRIA model to forecast electric load. In addition, some other advanced searching techniques to determine the suitable parameters should be combined with the SVR model to forecast electricity demand. Finally, other type of kernel functions employing should be the advanced research issue of the SVR theory.

Acknowledgement

This research was conducted with the support of National Science Council, Taiwan (NSC 97-2410-H-161-001).

References

[1] D.W. Bunn, Forecasting loads and prices in competitive power markets, Proc. IEEE 88 (2) (2000) 163–169.
[2] A.P. Douglas, A.M. Breipohl, F.N. Lee, R. Adapa, Risk due to load forecast uncertainty in short term power system planning, IEEE Trans. Power Syst. 13 (4) (1998) 1493–1499.
[3] G. Gross, F.D. Galiana, Short term load forecasting, Proc. IEEE 75 (12) (1987) 1558–1573.
[4] G.E.P. Box, G.M. Jenkins, Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco, 1970.
[5] J.F. Chen, W.M. Wang, C.M. Huang, Analysis of an adaptive time-series autoregressive moving-average (ARMA) model for short-term load forecasting, Electric Power Syst. Res. 34 (3) (1995) 187–196.
[6] S. Vemuri, D. Hill, R. Balasubramanian, Load forecasting using stochastic models, in: Proceeding of the 8th Power Industrial Computing Application Conference, 1973, pp. 31–37.
[7] H. Wang, N.N. Schulz, Using AMR data for load estimation for distribution system analysis, Electric Power Syst. Res. 76 (5) (2006) 336–342.
[8] W.R. Christianse, Short term load forecasting using general exponential smoothing, IEEE Trans. Power Apparatus Syst. PAS-90 (1971) 900–911.
[9] J.H. Park, Y.M. Park, K.Y. Lee, Composite modeling for adaptive short-term load forecasting, IEEE Trans. Power Syst. 6 (1) (1991) 450–457.
[10] A.P. Douglas, A.M. Breipohl, F.N. Lee, R. Adapa, The impact of temperature forecast uncertainty on Bayesian load forecasting, IEEE Trans. Power Syst. 13 (4) (1998) 1507–1513.
[11] A. Azadeh, M. Saberi, S.F. Ghaderi, A. Gitiforouz, V. Ebrahimipour, Improved estimation of electricity demand function by integration of fuzzy system and data mining approach, Energy Convers. Manage. 49 (8) (2008) 2165–2177.
[12] B. Wang, N.-L. Tai, H.-Q. Zhai, J. Ye, J.-D. Zhu, L.-B. Qi, A new ARMAX model based on evolutionary algorithm and particle swarm optimization for short-term load forecasting, Electric Power Syst. Res. 78 (10) (2008) 1679–1685.
[13] R.G. Brown, Introduction to Random Signal Analysis and Kalman Filtering, John Wiley & Sons, Inc., New York, 1983.
[14] A. Gelb, Applied Optimal Estimation, The MIT Press, MA, 1974.
[15] D.J. Trudnowski, W.L. McReynolds, J.M. Johnson, Real-time very short-term load prediction for power-system automatic generation control, IEEE Trans. Control Syst. Technol. 9 (2) (2001) 254–260.
[16] I. Moghram, S. Rahman, Analysis and evaluation of five short-term load forecasting techniques, IEEE Trans. Power Syst. 4 (4) (1989) 1484–1491.
[17] H.M. Al-Hamadi, S.A. Soliman, Fuzzy short-term electric load forecasting using Kalman filter, IEE Proc. C 153 (2) (2006) 217–227.
[18] N. Amjady, Short-term bus load forecasting of power systems by a new hybrid method, IEEE Trans. Power Syst. 22 (1) (2007) 333–341.
[19] C. Asbury, Weather load model for electric demand energy forecasting, IEEE Trans. Power Apparatus Syst. PAS-94 (1975) 1111–1116.
[20] A.D. Papalexopoulos, T.C. Hesterberg, A regression-based approach to short-term system load forecasting, IEEE Trans. Power Syst. 5 (4) (1990) 1535–1547.
[21] S.A. Soliman, S. Persaud, K. El-Nagar, M.E. El-Hawary, Application of least absolute value parameter estimation based on linear programming to short-term load forecasting, Int. J. Electrical Power Energy Syst. 19 (3) (1997) 209–216.
[22] S. Mirasgedis, Y. Safaridis, E. Georgopoulou, D.P. Lalas, M. Moschovits, F. Karagiannis, D. Papakonstantinou, Models for mid-term electricity demand forecasting incorporating weather influences, Energy 31 (2–3) (2006) 208–227.
[23] Z. Mohamed, P. Bodger, Forecasting electricity consumption in New Zealand using economic and demographic variables, Energy 30 (10) (2005) 1833–1843.
[24] G.J. Tsekouras, E.N. Dialynas, N.D. Hatziargyriou, S. Kavatza, A non-linear multivariable regression model for midterm energy forecasting of power systems, Electric Power Syst. Res. 77 (12) (2007) 1560–1568.
[25] S. Rahman, R. Bhatnagar, An expert system based algorithm for short-term load forecasting, IEEE Trans. Power Syst. 3 (2) (1998) 392–399.
[26] C.C. Chiu, L.J. Kao, D.F. Cook, Combining a neural network with a rule-based expert system approach for short-term power load forecasting in Taiwan, Expert Syst. Appl. 13 (4) (1997) 299–305.
[27] S. Rahman, O. Hazim, A generalized knowledge-based short-term load-forecasting technique, IEEE Trans. Power Syst. 8 (2) (1993) 508–514.
[28] L.-C. Ying, M.-C. Pan, Using adaptive network based fuzzy inference system to forecast regional electricity loads, Energy Convers. Manage. 49 (2) (2008) 205–211.
[29] P.-F. Pai, Hybrid ellipsoidal fuzzy systems in forecasting regional electricity loads, Energy Convers. Manage. 47 (15–16) (2006) 2283–2289.
[30] S.C. Pandian, K. Duraiswamy, C.C.A. Rajan, N. Kanagaraj, Fuzzy approach for short term load forecasting, Electric Power Syst. Res. 76 (6–7) (2006) 541–548.
[31] D.C. Park, M.A. El-Sharkawi, R.J. Marks II., L.E. Atlas, M.J. Damborg, Electric load forecasting using an artificial neural network, IEEE Trans. Power Syst. 6 (2) (1991) 442–449.
[32] B. Novak, Superfast autoconfiguring artificial neural networks and their application to power systems, Electric Power Syst. Res. 35 (1) (1995) 11–16.
[33] G.A. Darbellay, M. Slama, Forecasting the short-term demand for electricity – do neural networks stand a better chance?, Int J. Forecast. 16 (1) (2000) 71–83.
[34] R.E. Abdel-Aal, Short-term hourly load forecasting using abductive networks, IEEE Trans. Power Syst. 19 (1) (2004) 164–173.
[35] C.C. Hsu, C.Y. Chen, Regional load forecasting in Taiwan – application of artificial neural networks, Energy Convers. Manage. 44 (12) (2003) 1941–1949.
[36] L.M. Saini, Peak load forecasting using Bayesian regularization, Resilient and adaptive backpropagation learning based artificial neural networks, Electric Power Syst. Res. 78 (7) (2008) 1302–1310.
[37] P. Lauret, E. Fock, R.N. Randrianarivony, J.-F. Manicom-Ramsamy, Bayesian neural network approach to short time load forecasting, Energy Convers. Manage. 49 (5) (2008) 1156–1166.
[38] M.R. Amin-Naseri, A.R. Soroush, Combined use of unsupervised and supervised learning for daily peak load forecasting, Energy Convers. Manage. 49 (6) (2008) 1302–1308.
[39] O.A.S. Carpinteiro, R.C. Leme, A.C.Z. de Souza, C.A.M. Pinheiro, E.M. Moreira, Long-term load forecasting via a hierarchical neural model with time integrators, Electric Power Syst. Res. 77 (3–4) (2007) 371–378.
[40] J. Cao, X. Lin, Study of hourly and daily solar irradiation forecast using diagonal recurrent wavelet neural networks, Energy Convers. Manage. 49 (6) (2008) 1396–1406.
[41] N. Tai, J. Stenzel, H. Wu, Techniques of applying wavelet transform into combined model for short-term load forecasting, Electric Power Syst. Res. 76 (6–7) (2006) 525–533.
[42] M. El-Telbany, F. El-Karmi, Short-term forecasting of Jordanian electricity demand using particle swarm optimization, Electric Power Syst. Res. 78 (3) (2008) 425–433.
[43] M. Ghiassi, D.K. Zimbra, H. Saidane, Medium term system load forecasting with a dynamic artificial neural network model, Electric Power Syst. Res. 76 (5) (2006) 302–316.
[44] V. Vapnik, S. Golowich, A. Smola, Support vector machine for function approximation, regression estimation, and signal processing, Adv. Neural Inform. Process. Syst. 9 (1996) 281–287.
[45] L. Cao, Support vector machines experts for time series forecasting, Neurcomputing 51 (2003) 321–339.
[46] L. Cao, Q. Gu, Dynamic support vector machines for non-stationary time series forecasting, Intel. Data Anal. 6 (2002) 67–83.
[47] F.E.H. Tay, L. Cao, Application of support vector machines in financial time series forecasting, Omega 29 (4) (2001) 309–317.
[48] W.C. Hong, P.F. Pai, Predicting engine reliability by support vector machines, Int. J. Adv. Manufact. Technol. 28 (1–2) (2006) 154–161.
[49] P.F. Pai, W.C. Hong, C.S. Lin, Forecasting tourism demand using a multi-factor support vector machine model, Lecture Notes Artif. Intel. 3801 (2005) 513–521.

[50] B.J. Chen, M.W. Chang, C.J. Lin, Load forecasting using support vector machines: a study on EUNITE competition 2001, IEEE Trans. Power Syst. 19 (4) (2004) 1821–1830.
[51] P.F. Pai, W.C. Hong, Forecasting regional electric load based on recurrent support vector machines with genetic algorithms, Electric Power Syst. Res. 74 (3) (2005) 417–425.
[52] P.F. Pai, W.C. Hong, Support vector machines with simulated annealing algorithms in electricity load forecasting, Energy Convers. Manage. 46 (17) (2005) 2669–2688.
[53] V. Vapnik, The Nature of Statistic Learning Theory, Springer–Verlag, New York, 1995.
[54] H. Drucker, C.J.C. Burges, L. Kaufman, A. Smola, V.N. Vapnik, Support vector regression machines, Adv. Neural Inform. Process. Syst. 9 (1997) 155–161.
[55] K. Vojislav, Learning and Soft Computing – Support Vector Machines, Neural Networks and Fuzzy Logic Models, The MIT Press, MA, 2001.
[56] S. Amari, S. Wu, Improving support vector machine classifiers by modifying kernel functions, Neural Networks 12 (6) (1999) 783–789.
[57] V. Cherkassky, Y. Ma, Practical selection of SVM parameters and noise estimation for SVM regression, Neural Networks 17 (1) (2004) 113–126.
[58] K. Mori, M. Tsukiyama, T. Fukuda, Immune algorithm with searching diversity and its application to resource allocation problem, Trans. Inst. Electrical Eng. Jpn. 113-C (10) (1993) 872–878.