



A Novel hybrid genetic algorithm for kernel function and parameter optimization in support vector regression

Chih-Hung Wu^{a,*}, Gwo-Hshiung Tzeng^{b,c}, Rong-Ho Lin^d

^a Department of Digital Content and Technology, National Taichung University No. 140, Ming-Shen Road, Taichung 40306, Taiwan

^b Department of Business Administration, Kainan University, No. 1, Kainan Road, Luchn, Taoyuan 338, Taiwan

^c Institute of Management of Technology, National Chiao Tung University, 100, Ta-Hsueh Road, Hsinchu 300, Taiwan

^d Department of Industrial Engineering & Management, National Taipei University of Technology, No. 1, Section 3, Chung-Hsiao East Road, Taipei 106, Taiwan, ROC

ARTICLE INFO

Keywords:

Support vector regression (SVR)
Hybrid genetic algorithm (HGA)
Parameter optimization
Kernel function optimization
Electrical load forecasting
Forecasting accuracy

ABSTRACT

This study developed a novel model, HGA-SVR, for type of kernel function and kernel parameter value optimization in support vector regression (SVR), which is then applied to forecast the maximum electrical daily load. A novel hybrid genetic algorithm (HGA) was adapted to search for the optimal type of kernel function and kernel parameter values of SVR to increase the accuracy of SVR. The proposed model was tested at an electricity load forecasting competition announced on the EUNITE network. The results showed that the new HGA-SVR model outperforms the previous models. Specifically, the new HGA-SVR model can successfully identify the optimal type of kernel function and all the optimal values of the parameters of SVR with the lowest prediction error values in electricity load forecasting.

Crown Copyright © 2008 Published by Elsevier Ltd. All rights reserved.

1. Introduction

Support vector machines (SVMs) have been successfully applied to a number of applications such as including handwriting recognition, particle identification (e.g., muons), digital images identification (e.g., face identification), text categorization, bioinformatics (e.g., gene expression), function approximation and regression, and database marketing, and so on. Although SVMs have become more widely employed to forecast time-series data (Tay & Cao, 2001; Cao, 2003; Kim, 2003) and to reconstruct dynamically chaotic systems (Müller et al., 1997; Mukherjee, Osuna, & Girosi, 1997; Mattera & Haykin, 1999; Kulkarni, Jayaraman, & Kulkarni, 2003), a highly effective model can only be built after the parameters of SVMs are carefully determined (Duan, Keerthi, & Poo, 2003).

Min and Lee (2005) stated that the optimal parameter search on SVM plays a crucial role in building a prediction model with high prediction accuracy and stability. The kernel-parameters are the few tunable parameters in SVMs controlling the complexity of the resulting hypothesis (Cristianini, Campell, & Taylor, 1999). Shawkat and Kate (2007) pointed out that selecting the optimal degree of a polynomial kernel is critical to ensure good generalization of the resulting support vector machine model. They proposed an automatic selection for determining the optimal degree of polynomial kernel in SVM by Bayesian and Laplace approximation method estimation and a rule based meta-learning approach. In

addition, to construct an efficient SVM model with RBF kernel, two extra parameters: (a) sigma squared and (b) gamma, have to be carefully predetermined. However, few studies have been devoted to optimizing the parameter values of SVMs. Evolutionary algorithms often have to solve optimization problems in the presence of a wide range of problems (Dastidar, Chakrabarti, & Ray, 2005; Shin, Lee, Kim, & Zhang, 2005; Yaochu & Branke, 2005; Zhang, Sun, & Tsang, 2005). In these algorithms, genetic algorithms (GAs) have been widely and successfully applied to various types of optimization problems in recent years (Goldberg, 1989; Fogel, 1994; Cao, 2003; Alba & Dorronsoro, 2005; Aurnhammer & Tonnies, 2005; Venkatraman & Yen, 2005; Hokey, Hyun, & Chang, 2006; Cao & Wu, 1999; McCall, 2005). Therefore, this paper proposes a hybrid genetic-based SVR model, HGA-SVR, which can automatically optimize the SVR parameters integrating the real-valued genetic algorithm (RGA) and integer genetic algorithm, for increasing the predictive accuracy and capability of generalization compared with traditional machine learning models.

In addition, a wide range of approaches including time-varying splines (Harvey & Koopman, 1993), multiple regression models (Ramanathan, Engle, Granger, Vahid-Araghi, & Brace, 1997), judgmental forecasts, artificial neural networks (Hippert & Pedreira, 2001) and SVMs (Chen, Chang, & Lin, 2004; Tian & Noore, 2004) have been employed to forecast electricity load. One of the most crucial demands for the operation activities of power systems is short-term hourly load forecasting and the extension to several days in the future. Improving the accuracy of short-term load forecasting (STLF) is becoming even more significant than before due to the changing structure of the power utility industry (Tian &

* Corresponding author. Tel.: +886 939013100; fax: +886 422183270.

E-mail addresses: chwu@ntcu.edu.tw (C.-H. Wu), ghtzeng@cc.nctu.edu.tw, ghtzeng@mail.knu.edu.tw (G.-H. Tzeng).

Noore, 2004). SVMs have been applied to STLF and performed well. Unfortunately, there is still no consensus as to the perfect approach to electricity demand forecasting (Taylor & Buizza, 2003).

Several studies have proposed optimization methods which used a genetic algorithm for optimizing the SVR parameter values. To overcome the problem of SVR parameters, a GA-SVR has been proposed in a earlier paper (Hsu, Wu, Chen, & Peng, 2006) to take advantage of the GAs optimization technique. However, few studies have focused on concurrently optimizing the type of SVR kernel function and the parameters of SVR kernel function. The present study proposed a novel and specialized hybrid genetic algorithm for optimizing all the SVR parameters simultaneously. Our proposed method was applied to predicting maximum electrical daily load and its performance was analyzed. An actual case of forecasting maximum electrical daily load is illustrated to show the improvement in predictive accuracy and capability of generalization achieved by our proposed HGA-SVR model.

The remainder of this paper is organized as follows. The research gap for obtaining optimal parameters in SVR is reviewed and discussed in Section 2. Section 3 details the proposed HGA-SVR, ideas and procedures. In Section 4 an experimental example for predicting the electricity load is described to demonstrate the proposed method. Discussions are presented in Section 5 and conclusions are drawn in the final Section.

2. Basic ideas of methods for obtaining optimal parameters in SVR

SVR is a promising technique for data classification and regression (Vapnik, 1998). We briefly introduce the basic idea of SVR in the Section 2.1. To design an effective model, the values of the essential parameters in SVR must be chosen carefully in advance (Duan et al., 2003). Thus, various approaches to determine these values are discussed in Section 2.2. Although many optimization methods have been proposed, GAs is well suited to the concurrent manipulation of models with varying resolutions and structures since they can search non-linear solution spaces without requiring gradient information or a priori knowledge of model characteristics (McCall & Petrovski, 1999). The genetic algorithm employed in this study to search for the optimal values of the SVR parameter is illustrated in Section 2.3.

2.1. Support vector regression (SVR)

This subsection briefly introduces support vector regression (SVR), which can be used for time-series forecasting. Given training data $(x_1, y_1), \dots, (x_l, y_l)$, where x_i are the input vectors and y_i are the associated output values of x_i , the support vector regression is an optimization problem:

$$\min_{\omega, b, \xi_i, \xi_i^*} \frac{1}{2} \omega^T \omega + C \sum_{i=1}^l (\xi_i + \xi_i^*), \quad (1)$$

$$\text{Subject to } y_i - (\omega^T \phi(x_i) + b) \leq \varepsilon + \xi_i, \quad (2)$$

$$(\omega^T \phi(x_i) + b) - y_i \leq \varepsilon + \xi_i^*, \quad (3)$$

$$\xi_i, \xi_i^* \geq 0, i = 1, \dots, l, \quad (4)$$

where l denotes the number of samples, x_i vector of i -sample is dataset mapped to a higher dimensional space by the kernel function ϕ , vector, ξ_i represents the upper training error, and ξ_i^* is the lower training error subject to ε -insensitive tube $|y - (\omega^T \phi(x) + b)| \leq \varepsilon$. Three parameters determine the SVR quality: error cost C , width of tube, and mapping function (also called kernel function). The basic idea in SVR is to map the dataset x_i into a high-dimensional feature space via non-linear mapping. Kernel functions perform non-linear mapping between the input space

and a feature space. The approximating feature map for the Mercer kernel performs non-linear mapping. In machine learning theories, the popular kernel functions are

$$\text{Gaussian(RBF) kernel : } k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right). \quad (5)$$

$$\text{Polynomial kernel : } k(x_i, x_j) = (1 + x_i \bullet x_j)^d. \quad (6)$$

$$\text{Linear kernel : } k(x_i, x_j) = x_i^T x_j. \quad (7)$$

In Eq. (5), x_i and x_j are input vector spaces; and V denotes the variance-covariance matrix of the Gaussian kernel.

2.2. Parameter optimization

As mentioned earlier, when designing an effective model, values of the two essential parameters in SVR have to be chosen carefully in advance (Duan et al., 2003). These parameters include (1) regularization parameter C , which determines the tradeoff cost between minimizing the training error and minimizing model complexity; and (2) parameter sigma (or d) of the kernel function, which defines the non-linear mapping from the input space to some high-dimensional feature space. This investigation considers only the Gaussian kernel, namely sigma square (V), which is the variance-covariance matrix of the kernel function. Generally speaking, model selection by SVM is still performed in the standard way: by learning different SVMs and testing them on a validation set to determine the optimal value of the kernel parameters. Therefore, (Cristianini et al., 1999) proposed the Kernel-Adatron Algorithm, which can automatically perform model selection without being tested on a validation. Unfortunately, this algorithm is ineffective if the data have a flat ellipsoid distribution (Campbell, 2002). Therefore, one possible way is to consider the data distribution.

2.3. Genetic algorithms (GAs)

Evolutionary algorithms often have to solve optimization problems in the presence of a wide range of uncertainties (Yaochu & Branke, 2005). Genetic algorithms (GAs) are well suited for searching global optimal values in complex search space (multi-modal, multi-objective, non-linear, discontinuous, and highly constrained space), coupled with the fact that they work with raw objectives only when compared with conventional techniques (Holland, 1975; Goldberg, 1989; Waters & Sheble, 1993). For example, (Venkatraman & Yen, 2005) proposed a generic, two-phase framework for solving constrained optimization problems using GAs. Although many optimization methods have been proposed (e.g. Nelder-Mead simplex method), GAs are well suited to the concurrent manipulation of models with varying resolutions and structures since they can search non-linear solution spaces without requiring gradient information or a priori knowledge of model characteristics (Darwen & Xin, 1997; McCall & Petrovski, 1999). Based on fitness sharing, the learning system of GAs outperforms the tit-for-tat strategy against unseen test opponents. They learn using a "black box" simulation, with minimal prior knowledge of the learning task (Darwen & Xin, 1997).

In addition, the problem in binary coding lies in the fact that a long string always occupies the computer memory even though only a few bits are actually involved in the crossover and mutation operations. This is especially the case when a lot of parameters have to be adjusted in the same problem and a higher precision is required for the final result. This is also the main problem when initialing values of parameters of SVM in advance. To overcome this inefficient use of computer memory, the underlying real-valued crossover and mutation algorithm are employed (Huang & Huang, 1997). Contrary to the binary genetic algorithm (BGA), the real-valued genetic algorithm (RGA) uses real value as a

parameter of the chromosomes in the population without the coding and encoding process prior to calculating the fitness value (Haupt & Haupt, 1998). Consequently, the RGA is more straightforward, faster, and more efficient than the BGA. Recently, a hybrid GA (HGA) has been proposed by (Li & Aggarwal, 2000) to take advantage of both GAs and the local search techniques for speeding up the search effectiveness and to overcome the premature convergence problem. (Li & Aggarwal, 2000) proposed a relaxed hybrid genetic algorithm (RHGA) to economically allocate power generation in a fast, accurate, and relaxed manner.

3. Design of the hybrid genetic-based SVR (HGA-SVR) model for improving predictive accuracy

In this section, we describe the design of our proposed novel HGA-SVR model. The optimization process of HGA-SVR is intro-

duced in the first section. The basic idea of non-linear SVR model is described in the next section. The design of chromosome representations, fitness function and genetic operators in our novel HGA-SVR are discussed in the final sections.

3.1. Our proposed novel HGA-SVR model

In our proposed novel HGA-SVR model, the type of kernel and the parameter value of SVR are dynamically optimized by implementing the evolutionary process, and the SVR model then performs the prediction task using these optimal values. Our approach simultaneously determines the appropriate type of kernel function and optimal kernel parameter values for optimizing the SVR model to fit various datasets. The overall process of our proposed approach is illustrated in Fig. 1. The types of kernel function and optimal values of the SVR's parameters are determined by our

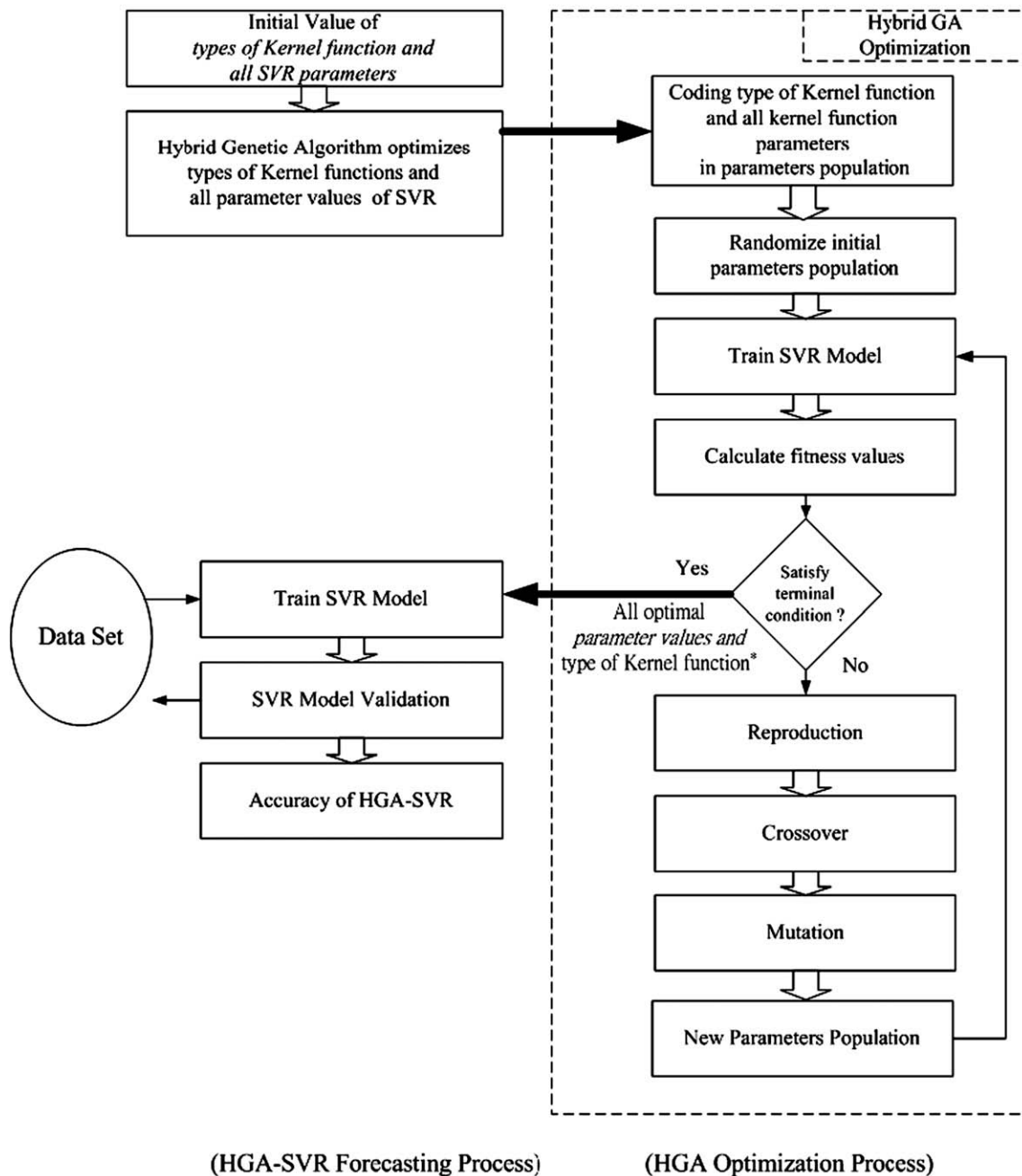


Fig. 1. The optimization process of HGA-SVR.

proposed novel HGAs with a randomly generated initial population of chromosomes. The types of kernel function (Gaussian (RBF) kernel, polynomial kernel, and linear kernel) and all the values of the parameters are directly coded into the chromosomes with integers and real-valued numbers, respectively. The proposed model can implement either the roulette-wheel method or the tournament method for selecting chromosomes. Adewuya's crossover method and boundary mutation method were used to modify the chromosome. Only the one best chromosome in each generation survives to move on to the succeeding generation.

Christiani and Shawe-Taylor (2000) proposed the Kernel-Adatron Algorithm, which can automatically select models without them being tested on a validation data. Unfortunately, this algorithm is ineffective if the data have a flat ellipsoid distribution (Campbell, 2002). Unfortunately, this may happen often in the real world. Therefore, rather than applying the Kernel-Adatron Algorithm, a new method named HGA-SVR was developed in this study to optimize all the parameters of SVR simultaneously. The major SVR training and validation tool used in this study has been previously developed (Pelckmans et al., 2002; Suykens, Van Gestel, De Brabanter, De Moor, & Vandewalle, 2002). The proposed model was developed and implemented in the MATLAB 7.1. The main tool used, LIBSVM, for training and validating the SVR was developed by Pelckmans et al. (2002). By using this tool, Comak et al. (2007) integrated the fuzzy weight pre-processing for the medical decision making system and obtained the highest classification accuracy in their dataset. Thus, we believe our proposed HGA-SVR model is able to handle huge data sets and can easily and efficiently be combined with the integer genetic algorithm and real-valued genetic algorithm for developing the hybrid genetic algorithm.

3.2. The non-linear SVR model

The SVR model can be represented as follows. The non-linear objective function maximizes

$$\text{Max } W(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{j=1}^l \alpha_j y_j y_i (k(\mathbf{x}_i, \mathbf{x}_j)) \tag{8}$$

$$\text{Subject to } 0 \leq \alpha_i \leq C, \quad i = 1, \dots, l, \tag{9}$$

$$\sum_{i=1}^l \alpha_i y_i = 0. \tag{10}$$

The optimal weight w^* and bias are determined by solving the quadratic programming problem.

$$w^* = \sum_{i=1}^l \alpha_i^* y_i \mathbf{x}_i, \tag{11}$$

$$b^* = y_i - w^{*T} \mathbf{x}_i. \tag{12}$$

The optimal decision function is as follows:

$$f(\mathbf{x}) = \text{sign} \left(\sum_{i=1}^l y_i \alpha_i^* k(\mathbf{x}, \mathbf{x}_i) + b^* \right). \tag{13}$$

3.3. The proposed HGA

The proposed HGA was revised and combined with the integer genetic algorithm and real-valued genetic algorithm in order to obtain a higher precise value under various ranges of parameter values. The HGA is designed as follows.

3.3.1. Chromosome representations

Unlike applying traditional GAs, when using a HGA for optimization problems, all of the corresponding parameters and types of kernel function can be coded directly to form a chromosome. Hence, the representation of the chromosome is straightforward in a HGA. All the parameters of SVR were directly coded to form the chromosome in the present approach. Consequently, chromosome \mathbf{X} was represented as $\mathbf{X} = \{KT_i, P_1, P_2\}$, where P_1 and P_2 denote the type of kernel function, and the first and second parameter values, respectively. The gene structure of our proposed HGA is shown as Fig. 2.

KT_i denotes the types of kernel function which includes three types of kernel function as follows.

$$\text{Linear kernel : } k(x_i, x_j) = x_i^T x_j \tag{14}$$

$$\text{Polynomial kernel : } k(x_i, x_j) = (x_i^T x_j + t)^d \tag{15}$$

where t is the intercept and d the degree of the polynomial.

$$\text{Gaussian(RBF)kernel : } k(x_i, x_j) = \exp \left(-\frac{\|x_i - x_j\|^2}{2\sigma^2} \right) \tag{16}$$

with σ^2 the variance of the Gaussian kernel.

The values zero, one, and two denote that the system will choose 'Linear kernel', 'Polynomail kernel', and 'Gaussian (RBF) kernel', respectively. The first part of the HGA will be implemented in the integer value type GA.

$P1i$: optimal parameter 1; $P2i$: optimal parameter 2.

The various types of SVM kernel function and sufficient kernel function parameters that need to be optimized are summarized in Table 1. The definition and type of essential parameters in SVR is based on the definition of LSSVM tool.

Parameter C is the penalty (cost) parameter of the training error in the RBF kernel function. Parameter d denotes the degree of polynomial kernel function, t denotes the constant term of the polynomial kernel function, and ϵ denotes the epsilon-insensitive value in epsilon-SVR. In the LIB-SVM tool, we don't need the ϵ parameters for using SVR.

Table 1
Types of various kernel function and sufficient kernel function parameters

KT_i		P_{1i} (parameter 1)	P_{2i} (parameter 1)
0	Linear kernel	gamma	-
1	Poly kernel	d	t
2	RBF kernel	C	σ

Notes: - denotes no parameter needed; and gamma, d , t , C , σ denote various types of kernel function parameters.

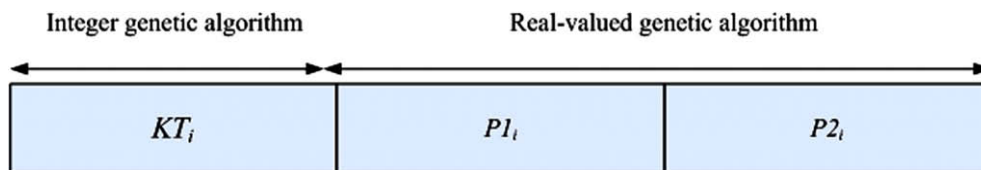


Fig. 2. Gene structure of our proposed HGA (population i).

3.3.2. Genetic operators

The real-valued genetic algorithm uses selection, crossover, and mutation operators to generate the offspring of the existing population. The proposed HGA-SVR model incorporates two well-known selection methods: roulette-wheel method and tournament method. The tournament selection method is adopted here to decide whether or not a chromosome can survive into the next generation. The chromosomes that survive into the next generation are then placed in a mating pool for the crossover and mutation operations. Once a pair of chromosomes has been selected for crossover, one or more randomly selected positions are assigned into the to-be-crossed chromosomes. The newly-crossed chromosomes then combine with the rest of the chromosomes to generate a new population. However, the problem of frequent overloading occurs when the RGA is used to optimize values. In this study we used the method proposed by (Adewuya, 1996), a genetic algorithm with real-valued chromosomes in order to avoid a post-crossover overload problem. The mutation operation follows the crossover to determine whether or not a chromosome should mutate to the next generation. In this study, uniform mutation was designed in the presented model.

Uniform mutation

$$X^{\text{old}} = \{x_1, x_2, \dots, x_n\}, \quad (17)$$

$$X_k^{\text{new}} = LB_k + r \times (UB_k - LB_k), \quad (18)$$

$$X^{\text{new}} = \{x_1, x_2, \dots, x_k^{\text{new}}, \dots, x_n\} \quad (19)$$

where n denotes the number of parameters, r represents a random number range (0, 1), and k is the mutation location. LB and UB are the low and upper bounds of the parameter, respectively. LB_k and UB_k denote the low and upper bounds in location k , respectively. X^{old} represents the population before the mutation operation; and X^{new} represents the new population after the mutation operation.

However, the major problem for optimizing all parameters of SVR is that various kernel function parameters have a different range of parameter values. Therefore, we proposed that the new GA operators in our proposed HGA deal with the range of SVM parameter values. The new GA operators are shown in Fig. 3.

Our proposed HGA adopts different GA operators in the integer GA the real-valued GA. As shown in Fig. 3, the HGA is divided into two parts—the integer GA and the real-valued GA. Our method selects the same GA reproduction operator and crossover operators. However, in this study we designed a different GA mutation operator (i.e. method1 and method2 in Fig. 3) for limiting the range of the parameter value. The revised mutation operator in KTi (new method1) is designed by MOD function calculation (remainder) and ROUND function calculation (by converting the real-value into the integer value) to limit the range of the value. The revised mutation operator in KTi (new method 2) is first calculated via uniform mutation operators and then converts the real-value into the integer value (The KTi value must be an integer value to map the coding design). Finally, we believe that the boundary mutation which adopts the upper bound and the lower bound does not need to be redesigned. The revised parts are shown in red in Fig. 3.

3.3.3. The fitness function

A fitness function assessing the performance for each chromosome must be designed before searching for the optimal values of the SVR parameters. Several measurement indicators have been proposed and employed to evaluate the prediction accuracy of models such as MAPE, RMSE, and the maximum error in time-series prediction problems. To compare the results achieved by the present model with those of the EUNITE competition, this study employed MAPE, which is the same fitness function used in the above-mentioned competition.

4. Experimental example for predicting electricity load

In this section, the effectiveness of the proposed HGA-SVR model was demonstrated by forecasting the daily electricity loading problem as announced on the 'Worldwide Competition within the EUNITE Network¹'. The set problem was to predict the maximum daily electricity load for January 1999 using daily half-an-hour electricity load values, average daily temperatures, and a list of public holidays for the period from 1997 to 1999. There is no consensus as to the best approach to forecast electricity load (Taylor & Buizza, 2003). The winning model, SVM, demonstrated a superior predictive accuracy compared with the traditional neural network models that were employed in the EUNITE competition (e.g. functional network², Back-propagation ANN³, adaptive logic networks⁴). In view of the above, we used our proposed HGA-SVR model to predict the maximum daily values of electricity load and compared its prediction performance with that of other models employed in the previous EUNITE competition.

4.1. Descriptions of competition data and structure

The competition data files include Load1997.xls, Load1998.xls, Temperature 1997.xls, Temperature 1998.xls, and Holidays.xls, which were downloaded from the EUNITE network. The file, Load1997and 8.xls, contains all half-hour electricity load values for 1997 and 1998. Temperature199X.xls comprises the average daily temperatures for the same two years. Holiday.xls describes the occurrence of holidays in the period 1997 to 1999. Furthermore, the prediction file, Load1999.xls, comprises the maximum electricity load values and half-hour loads in January of 1999. All data formats are listed in Table 2.

4.2. Data analysis

Variable selection plays a critical role in building a SVR model as well as traditional time-series prediction models. Therefore, this study first analyzed the data to ensure that all essential variables were included in the GA-SVR model. Only when all essential variables are included can the model yield a satisfactory prediction performance.

4.2.1. Temperature influence

As mentioned in most data mining research, the data sets must be analyzed and cleaned before the proposed model is applied to them. The maximum electrical loads were strongly influenced by the temperature factor, with a negative correlation existing between the two, as shown in Fig. 4. Specifically, people require a higher electricity load to keep warm in cold weather. Despite the change in the daily temperature, the data of the maximum loads, as shown in Fig. 5, also showed a seasonal pattern. There was a recurrent high peak of electricity demand during the winter and a lower peak during the summer. According to previous studies, the distribution of temperature shows Gaussian characteristics (The indexes for the Gaussian curve are: $a = 20.85$, $b = 196.04$, $c = 64.85$, respectively⁵).

¹ European Network on Intelligent Technologies for Smart Adaptive Systems (EUNITE) network organized a competition on the short-term prediction problem in 2001 (<http://neuron.tuke.sk/competition/index.php>).

² <http://neuron.tuke.sk/competition/reports/BerthaGuijaro.pdf>

³ <http://neuron.tuke.sk/competition/reports/DaliborZivcak.pdf>

⁴ <http://neuron.tuke.sk/competition/reports/DavidEsp.pdf>

⁵ <http://neuron.tuke.sk/competition/reports/DaliborZivcak.pdf>

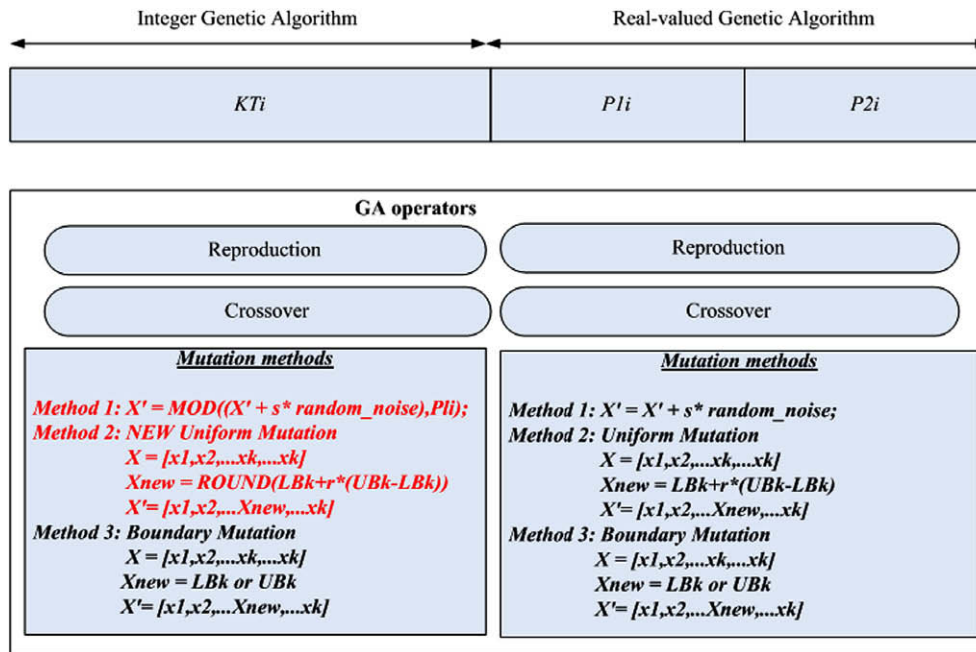


Fig. 3. The new GA operators in our proposed HGA.

Table 2
Given data formats

Data files	Content and format description							
(Training)	Date			Half-hour loads			(etc.)	Max. Loads
	Year	Month	Day	00:30	01:00	01:30..		
Load 1997.xls	1997	1	1	797	794	784	.. (etc.)	797
Load 1998.xls	1997	1	2	704	697	704	.. (etc.)	777
 (etc.)	...
	1998	12	31	716	703	690	.. (etc.)	733
	1999	1	1	751	735	714	.. (etc.)	751
(Predicting) (etc.)	...
Load 1999.xls	1999	1	31	712	720	694	.. (etc.)	743
				Date				Temperature [°C]
(Training)				01/01/97				-7.6
Temperature 1997.xls				02/01/97				-6.3
Temperature 1998.xls			
				12/31/98				-8.7
(Predicting)				01/01/99				-10.7
Temperature 1999.xls			
				01/31/99				-6.0
				<u>(Training)</u>				<u>(Predicting)</u>
				<u>Holiday-1997</u>				<u>Holiday-1999</u>
Holidays.xls				1997/01/01				1999/01/01
				1997/01/06				1999/01/06
				1997/03/28				1999/04/02
			
				1997/12/31				1999/12/31

4.2.2. Maximum load and the holiday effect

Fig. 6 displays a non-linear pattern of the maximum electricity loads during 1997 and 1998. The descriptive statistical information of the maximum loads is summarized in Table 3. The descriptive statistical information revealed that the lowest peak of electricity demand during 1997 and 1998 was 464 and the highest peak of electricity demand was 876. Moreover, the average demand was 670.8 with high volatility. The data sets also offered holiday information to help predict the maximum electricity loads, because earlier work in this area noted that holidays will influence the maximum load demand. According to public holiday information,

the electricity load is generally lower during the holidays and varies with the type of holiday.

4.3. Modeling

Kernel and variable selection are an important step for SVR modeling. Since the electricity load is a non-linear function of the weather variables (Taylor & Buizza, 2003) and since some variables (see Fig. 6) seemed to be more properly used here than others for fitting the electricity load data, this study chose three major kernel function types of SVR (linear, poly, and RBF) for the data mapping

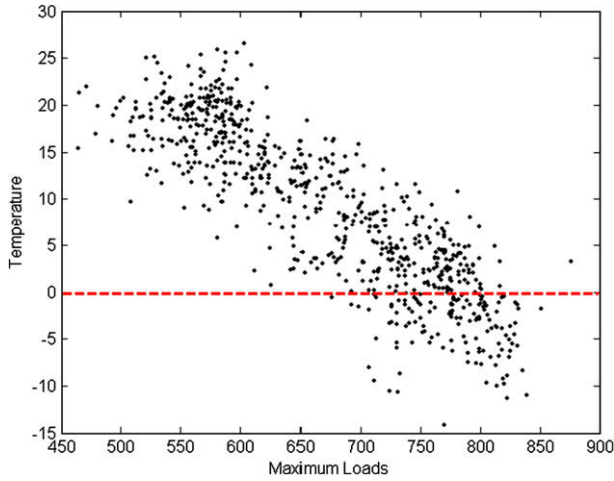


Fig. 4. Weather influence.

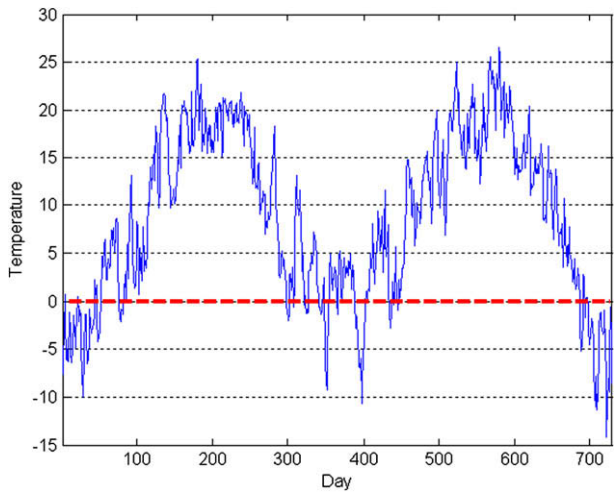


Fig. 5. Seasonal pattern in temperature.

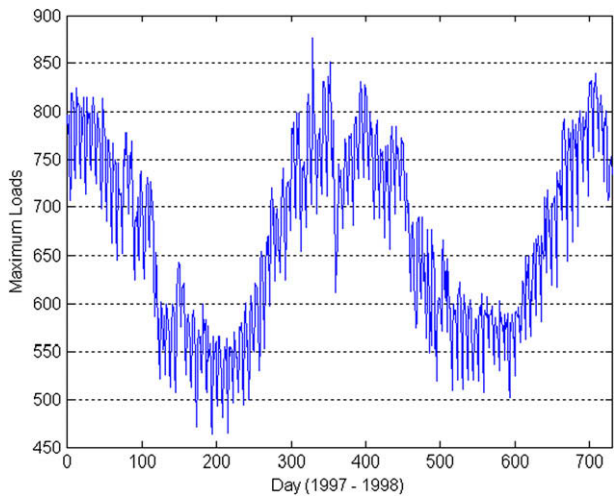


Fig. 6. Maximum loads from 1997 to 1998.

function and obtained the HGA-SVR parameters by HGA evolution. The daily electricity loads in the training data were adopted as the target value y_i , and the daily temperature values and public holiday

Table 3
Descriptive statistics on maximum loads

Statistics	Value
Minimum	464
Maximum	876
Mean	670.8
Std.	93.54
Range	412
Skewness	-.043
Kurtosis	-1.235

information were adopted as the input variables x_i in our model. For the holiday variable, a code of one or zero was used to indicate whether or not a day was a holiday. In addition, lagged demands, such as day-head inputs, which might be useful in short-term demand forecasting were not included in the input variables of this short-term forecasting problem. Extra variable information was not used for modeling. In other words, this work adopted the same variables that were selected by previous competitors in the EUNITE competition for modeling.

4.4. Results evaluation

To provide a comparison with the prior prediction ability of SVR models in the ‘Worldwide Competition within the EUNITE Network’, this work evaluated the HGA-SVR model according to the same criteria employed in the above mentioned competition.

1. Magnitude of MAPE error

$$MAPE = 100 * \frac{\sum_{i=1}^n \frac{|L_{R_i} - L_{P_i}|}{L_{R_i}}}{n} \tag{20}$$

L_{R_i} denotes the real value of the maximum daily electrical load on day “i” of 1999, and L_{P_i} represents the predicted maximum daily electrical load on the “ith” day of 1999, hence n is the number of days in January of 1999, hence $n = 31$.

2. Magnitude of Maximum Error

$$M = \max(|L_{R_i} - L_{P_i}|) \tag{21}$$

i represents the day in January of 1999, where $i = 1, 2, \dots, 31$

4.5. Design of parameters and fitness function

Some parameters have to be determined in advance before using HGA-SVR to forecast the electricity loads. Table 4 summarizes all HGA-SVR training parameters. The values of individual parameters and the value of the fitness function depend on the prior experiences of HGA-SVR training and problem type. Moreover, the fitness function is designed using the formula of the first

Table 4
HGA-SVR training parameters

Parameter	Value
Population size	20
Generations	50–100
Gamma range	0–1000
Sigma range	0–1000
Selection method	tournament
Mutation method	uniform
Snoise	100
Elite	yes
Mutation rate	0.5
Problem type	minimum

criterion (Eq. (14)), MAPE, and its value is taken as the fitness value in this HGA-SVR.

From Table 4, a uniform mutation method with high mutation ratio was selected to avoid the local optimum and pre-maturity problems. The present study activated the elite mechanism to ensure that the MAPE was efficiently minimized and that it remained in a convergent state during the early generation evolution. Consequently, both the RMSE and maximum error fluctuated sharply with the generation evolution. Meanwhile, the population size and the generations were increased to ensure that the global optimum values of all the parameters could be found. Fig. 7 illustrates the whole optimization process of MAPE in the proposed HGA-SVR.

The focus of the issue here was to predict the real maximum electricity loads in January 1999. Fig. 8 shows the results of the HGA-SVR conducted. Although the real values fluctuated sharply during January 1999, our prediction values (dashed line) were still very close to the real values (solid line).

In the proposed model, the best MAPE was 0.76, RMSE=7.73 and the maximum error (MW) was 20.88. The optimal type of kernel function is the Poly kernel function, and the optimal values of parameters 1 and 2 of SVR were 4.42 and 184.98, respectively. Comparing the results obtained by HGA-SVR with the previous results revealed that the best MAPE generated by our previous work, GA-SVR in the EUNITE dataset was 0.8501 (Hsu et al., 2006). Table 5 lists the results of our previously proposed GA-SVR during various generations. The new HGA-SVR model outperformed the previous

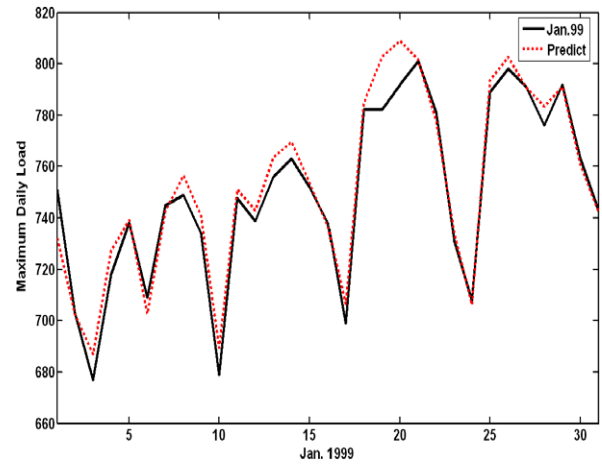


Fig. 8a. Prediction for January 1999 (generations = 50) (MAPE: 0.76, RMSE = 7.73; Max. error = 20.88) (polynomial kernel with optimal $d = 4.42^*$; optimal $t = 184.98^*$).

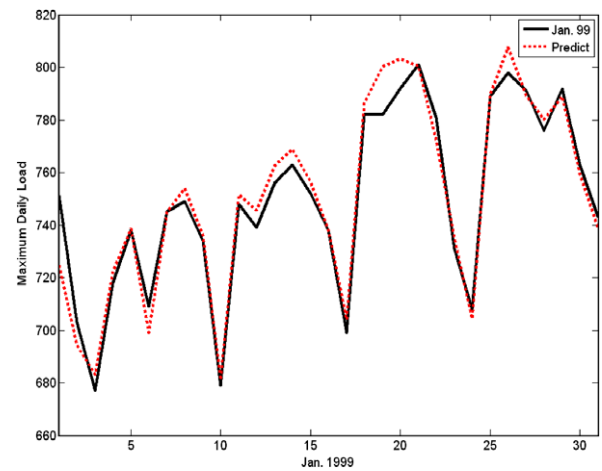


Fig. 8b. Prediction for January 1999 (generations = 100) (MAPE: 0.75, RMSE = 7.77; Max. error = 26.34) (polynomial kernel with optimal $d = 4.0^*$; optimal $t = 186.34^*$).

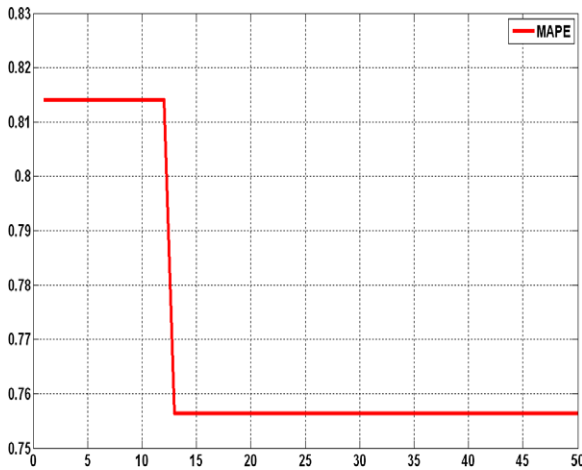


Fig. 7a. Optimization process of MAPE in HGA-SVR (50 generations).

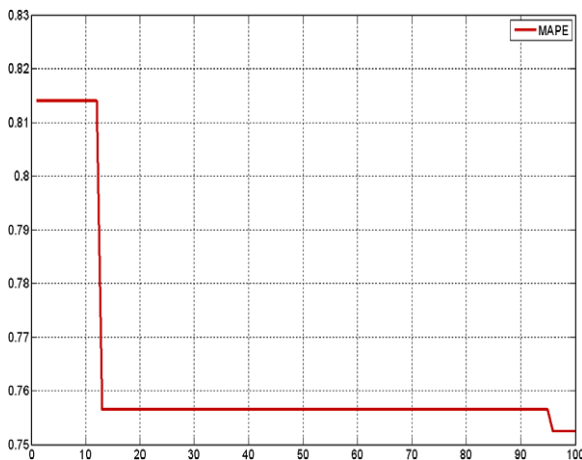


Fig. 7b. Optimization process of MAPE in HGA-SVR (100 generations).

Table 5

Results in various generations of GA-SVR

	Generations			
	50	100	200	500
RMSE	9.68	9.70	9.60	9.46
MAPE	0.8551	0.8540	0.8519	0.8501
Max. error	38.47	38.21	37.20	35.02
Optimal parameter 1 (Sigma)	436.81	223.32	171.48	106.49
Optimal parameter 2 (Gamma)	9042.72	2916.76	2179.52	817.32

GA-SVR model in the ‘Worldwide EUNITE Network Competition’ dataset, achieving a lower MAPE and MW. Complete EUNITE network competition reports can be found at the EUNITE website (<http://neuron.tuke.sk/competition/index.php>).

The comparison results in various generations for GA-SVR and HGA-SVR are shown in Table 6. The best model is marked in bold style fonts. In all models, the best model is the poly kernel function with 7.84 RMSE, 0.81 MAPE, and 23.67 maximum forecasting error. The optimal values which were obtained by HGA-SVR are quite astounding. In our previous experience, the RBF seemed to be the best choice for the type of SVR kernel function for non-linear forecasting. However, our research results reveal that besides the RBF

Table 6

Comparison results of GA-SVR and HGA-SVR in various generations

Optimal kernel	Generations			
	50 generations		100 generations	
	GA-SVR (RBF only)	HGA-SVR ^a (Optimize all)	GA-SVR (RBF only)	HGA-SVR ^d (Optimize all)
	RBF	Poly	RBF	RBF
Optimal RMSE	9.68	7.84	9.70	9.44
Optimal MAPE	0.86	0.81	0.85	0.85
Optimal max. error	38.47	23.67	38.21	34.28
Optimal parameter 1	436.81	4.55	223.32	87.43
Optimal parameter 2	9042.72	192.85	2916.76	457.44

Notes: GA-SVR only optimize the parameter values with RBF kernel; and HGA-SVR^{a,d} optimize all parameters (i.e. type of kernel function and all kernel function parameter values).

kernel function, the HGA-SVR found that the Poly kernel function also performed well in the electricity load forecasting problem, but only if it has optimal values. Another interesting point is the fact that the local optimal values can be found in only a few generations (in this case 50 generations). We tried to increase the number from 50 generations to 100 generations, but the forecasting error did not decrease significantly.

Based on the results obtained by HGA-SVR in Table 6, we found that the optimal kernel function type of SVR is Poly and the optimal parameters are 4.55 and 192.85 in the electricity loading dataset. In the next experiment, we tried to limit the range of the first parameter in SVR from 0 to 5 in order to obtain more precise optimal values. The results of HGA-SVR are shown in Table 7. Two extra models are implemented (HGA-SVR^b and HGA-SVR^d) in this experiment. The HGA-SVR^b and HGA-SVR^d are optimized with a lower range of parameters of SVR. The new limited HGA-SVR models are run in 50 generations and 100 generations in order to compare them with the results of the HGA-SVR models (HGA-SVR^a and HGA-SVR^d) in Table 6.

The improvement in reducing the forecasting error via HGA-SVR is shown in Table 8. Compared with our previous work, GA-SVR, the proposed HGA-SVR can lower the forecasting error further. The optimal RMSE, MAPE and maximum error by HGA-SVR is 7.73 (a decrease of 1.73), 0.76 (a decrease of 0.09), and 20.88 (a decrease of 14.14), respectively. The HGA-SVR also found all the optimal values—type of kernel function (i.e. Poly) and optimal values for parameters 1 and 2 to be 4.42 and 184.98, respectively.

Although most research results point out that the RBF kernel outperforms any other kinds of kernel function in a non-linear case, the fact is that our proposed HGA-SVR found that the Poly kernel function is not only good for the non-linear case but that it also performs well, even better than the RBF kernel function in this electronic loading forecasting problem.

Table 7

Results of HGA-SVR in various generations

	Generations			
	50 generations		100 generations	
	HGA-SVR ^a	HGA-SVR ^b	HGA-SVR ^c	HGA-SVR ^d
Range of parameter 1	0–10000	0–5	0–10000	0–5
Range of parameter 2	0–10000	0–200	0–10000	0–200
<i>Optimal values</i>				
Optimal kernel	Poly	Poly	RBF	Poly
Optimal RMSE	7.84	7.73	9.44	7.77
Optimal MAPE	0.81	0.76	0.85	0.75
Optimal max. error	23.67	20.88	34.28	26.34
Optimal parameter 1*	4.55	4.42	87.43	4.0
Optimal parameter 2*	192.85	184.98	457.44	186.34

Table 8

Improvement of forecasting error of HGA-SVR

	Generations			
	50 generations		100 generations	
	EUNITE winner	GA-SVR	HGA-SVR	Forecasting error
	(Model A)	(Model B)	(Model C)	
<i>Optimal values</i>				
Optimal kernel	RBF	RBF	Poly	(B)–(C)
Optimal RMSE	–	9.46	7.73	↓ 1.73
Optimal MAPE	2.0	0.85	0.76	↓ 0.09
Optimal max. error	50–60	35.02	20.88	↓ 14.14
Optimal parameter 1	–	106.49	4.42	
Optimal parameter 2	–	817.32	184.98	

Notes: The winning SVM model in EUNITE was proposed by Chen et al. (2004). Parameter 1 for the RBF kernel is sigma, and for the poly kernel it is d ; and Parameter 2 for the RBF kernel is gamma, and for the poly kernel it is p .

4.6. Discussions

The performance of our proposed HGA-SVR approach has been tested and compared with that of the traditional SVR model, other neural network approaches, and GA-SVR. During the competition other researchers tried other artificial neural network approaches, besides SVR. Various ideas were employed for the different proposed solutions to improve the accuracy, when they approached the selection of input variables and splitting data.

Among all the models on EUNITE network published, our approach provides a better generalization capability and a lower prediction error than the neural network approaches, traditional SVM models, and GA-SVR without variable selection and data segmentation. Our HGA-SVR model shows that the STLf can be improved by setting proper values for all parameters (parameter values and type of kernel function) in the SVR model. In addition to the RBF

kernel function, this study also found that the Poly kernel function may be an appropriate choice of SVR kernel function in forecasting daily electricity loading. The research results reveals that the Poly kernel function may outperform the RBF kernel function in a non-linear electricity loading forecasting problem. According to previous studies (Clements & Galvao, 2004), a non-linear model usually shows superior results in more accurate short-horizon forecasts. We believe that our proposed non-linearity model can be applied to other complex forecasting problems in the future.

In addition, the structural risk minimization principle (SRM), shown to be superior to the traditional empirical risk minimization principle (ERM) employed by the traditional neural networks, was embodied in SVM. SRM is able to minimize an upper bound of the generalization error as opposed to ERM that minimizes the error on training data (Tian & Noore, 2004). Thus, the solution of SVM may be a global optimum while other neural network models tend to fall into a local optimal solution, and overfitting is unlikely to occur with SVM (Hearst, Dumais, Osman, Platt, & Scholkopf, 1998; Cristianini et al., 1999; Kim, 2003). Therefore, most traditional neural network models yield an acceptable predictive error for training data, but when out-of-sample data are presented to these models, the error becomes unpredictably large, which yields limited generalization capability (Tian & Noore, 2004).

5. Conclusions

This study proposed a novel hybrid genetic algorithm for dynamically optimizing all the essential parameters of SVR. Our experimental results demonstrated the successful application of our proposed new model, HGA-SVR, for the complex forecasting problem. It demonstrated that it increased the electricity load forecasting accuracy more than any other model employed in the EUNITE network competition. Specifically, the new HGA-SVR model can successfully identify all the optimal values of the SVR parameters with the lowest prediction error values, MAPE, in electricity load forecasting.

Acknowledgement

This work was supported by National Science Council of the Republic of China under Grant No. NSC 95-2416-H-147-005.

References

- Adewuya, A.A. (1996). New methods in genetic search with real-valued chromosomes. Master's thesis, Cambridge: Massachusetts Institute of Technology.
- Alba, E., & Dorronsoro, B. (2005). The exploration/exploitation tradeoff in dynamic cellular genetic algorithms. *IEEE Transactions on Evolutionary Computation*, 9(2), 126–142.
- Alba, E., & Dorronsoro, B. (2005). The exploration/exploitation tradeoff in dynamic cellular genetic algorithms. *IEEE Transactions on Evolutionary Computation*, 9(2), 126–142.
- Aurnhammer, M., & Tonnie, K. D. (2005). A genetic algorithm for automated horizon correlation across faults in seismic images. *IEEE Transactions on Evolutionary Computation*, 9(2), 201–210.
- Campbell, C. (2002). Kernel methods: A survey of current techniques. *Neurocomputing*, 48(1–4), 63–84.
- Cao, L. (2003). Support vector machines experts for time series forecasting. *Neurocomputing*, 51(1–4), 321–339.
- Cao, Y. J., & Wu, Q. H. (1999). Optimization of control parameters in genetic algorithms: A stochastic approach. *International Journal of Systems Science*, 30(5), 551–559.
- Chen, B. J., Chang, M. W., & Lin, C. J. (2004). Load forecasting using support vector machines: A study on EUNITE competition 2001. *IEEE Transactions on Power Systems*, 19(4), 1821–1830.
- Christiani, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines*. Cambridge, England: Cambridge University Press.
- Clements, M. P., & Galvao, A. B. (2004). A comparison of tests of nonlinear cointegration with application to the predictability of US interest rates using the term structure. *International Journal of Forecasting*, 20(2), 219–236.
- Cristianini, N., Campell, C., & Taylor, J. S. (1999). Dynamically adapting kernels in support vector machines. *Advances in Neural Information Processing Systems*, 11(2), 204–210.
- Darwen, P. J., & Xin, Y. (1997). Speciation as automatic categorical modularization. *IEEE Transactions on Evolutionary Computation*, 1(2), 101–108.
- Dastidar, T. R., Chakrabarti, P. P., & Ray, P. (2005). A synthesis system for analog circuits based on evolutionary search and topological reuse. *IEEE Transactions on Evolutionary Computation*, 9(2), 211–224.
- Duan, K., Keerthi, S. S., & Poo, A. N. (2003). Evaluation of simple performance measures for tuning SVM hyperparameters. *Neurocomputing*, 51(1–4), 41–59.
- Fogel, D. B. (1994). An introduction to simulated evolutionary optimization. *IEEE Transactions on Neural Networks*, 5(1), 3–14.
- Goldberg, D. E. (1989). *Genetic algorithms in search, optimization and machine learning*. Reading, MA: Addison-Wesley.
- Harvey, A. C., & Koopman, S. J. (1993). Forecasting hourly electricity demand using time-varying splines. *Journal of American Statistical Association*, 88(424), 1228–1236.
- Haupt, R. L., & Haupt, S. E. (1998). *Practical genetic algorithms*. Wiley Interscience Publication.
- Hearst, M. A., Dumais, S. T., Osman, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Expert Intelligent Systems and Their Applications*, 13(4), 18–28.
- Hippert, H. S., & Pedreira, C. E. (2001). Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on Power Systems*, 16(1), 44–55.
- Hokey, M., Hyun, J. K., & Chang, S. K. (2006). A genetic algorithm approach to developing the multi-echelon reverse logistics network for product returns. *OMEGA: The International Journal of Management Science*, 34(1), 56–69.
- Holland, J. H. (1975). *Adaptation in natural and artificial system*. Ann Arbor, MI: University of Michigan Press.
- Hsu, C.C., Wu, C.H., Chen, S.J., & Peng, K.L. (2006). Dynamically optimizing parameters in support vector regression: An application of electricity load forecasting. In Proceedings of the Hawaii International Conference on System Science (HICSS39), January 4–7.
- Huang, Y. P., & Huang, C. H. (1997). Real-valued genetic algorithms for fuzzy grey prediction system. *Fuzzy Sets and Systems*, 87(3), 265–276.
- Kim, K. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, 55(1), 307–319.
- Kulkarni, A., Jayaraman, V. K., & Kulkarni, B. D. (2003). Control of chaotic dynamical systems using support vector machines. *Physics Letters A*, 317(5), 429–435.
- Li, F., & Aggarwal, R. K. (2000). Fast and accurate power dispatch using a relaxed genetic algorithm and a local gradient technique. *Expert Systems with Applications*, 19, 159–165.
- Mattera, D., & Haykin, S. (1999). Support vector machines for dynamic reconstruction of a chaotic system. In B. Schölkopf, C. J. C. Burges, & A. J. Smola (Eds.), *Advances in kernel methods – Support vector learning* (pp. 211–242). Cambridge, MA: MIT Press.
- McCall, J., & Petrovski, A. (1999). A decision support system for cancer chemotherapy using genetic algorithms. In Proceedings of the international conference on computational intelligence for modeling, control and automation, pp. 65–70.
- McCall, J. (2005). Genetic Algorithms for Modelling and Optimization. *Journal of Computational & Applied Mathematics*, 184(1), 205–222.
- Min, J. H., & Lee, Y. C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications*, 28(4), 603–614.
- Mukherjee, S., Osuna, E., & Girosi, F. (1997). Nonlinear prediction of chaotic time series using a support vector machine. In Proceedings of the NNSP'97.
- Müller, K.-R., Smola, A., Rätsch, G., Schölkopf, B., Kohlmorgen, J., & Vapnik, V. (1997). Predicting time series with support vector machines. In Proceedings of the ICANN'97, 999 pp.
- Pelckmans, K., Suykens, J.A.K., Van Gestel, T., De Brabanter, J., Lukas, L., Hamers, B., et al. (2002). LS-SVMlab Toolbox User's Guide version 1.4, November, 2002. Software available at <http://www.esat.kuleuven.ac.be/sista/lssvmlab/>.
- Ramanathan, R., Engle, R., Granger, C. W. J., Vahid-Araghi, F., & Brace, C. (1997). Short-run forecast of electricity loads and peaks. *International Journal of Forecasting*, 13(2), 161–174.
- Shin, S. Y., Lee, I. H., Kim, D., & Zhang, B. T. (2005). Multiobjective evolutionary optimization of DNA sequences for reliable DNA computing. *IEEE Transactions on Evolutionary Computation*, 9(2), 143–158.
- Suykens, J. A. K., Van Gestel, T., De Brabanter, J., De Moor, B., & Vandewalle, J. (2002). *Least squares support vector machines*. World Scientific.
- Tay, Francis E. H., & Cao, L. (2001). Application of support vector machines in financial time series forecasting. *OMEGA The International Journal of Management Science*, 29(4), 309–317.
- Taylor, J. W., & Buizza, R. (2003). Using weather ensemble predictions in electricity demand forecasting. *International Journal of Forecasting*, 19(1), 57–70.
- Tian, L., & Noore, A. (2004). A novel approach for short-term load forecasting using support vector machines. *International Journal of Neural Systems*, 14(5), 329–335.
- Vapnik, V. (1998). *Statistical learning theory*. New York: Wiley.
- Venkatraman, S., & Yen, G. G. (2005). A generic framework for constrained optimization using genetic algorithms. *IEEE Transactions on Evolutionary Computation*, 9(4), 424–435.
- Waters, D. C., & Sheble, G. B. (1993). Genetic algorithm solution of economic dispatch with valve point loading. *IEEE Transactions on Power Systems*, 8(3), 1325–1332.

- Yaochu, J., & Branke, J. (2005). Evolutionary optimization in uncertain environments—a survey. *IEEE Transactions on Evolutionary Computation*, 9(3), 303–317.
- Zhang, Q., Sun, J., & Tsang, E. (2005). An evolutionary algorithm with guided mutation for the maximum clique problem. *IEEE Transactions on Evolutionary Computation*, 9(2), 192–200.