Dynamically Optimizing Parameters in Support Vector Regression: An Application of Electricity Load Forecasting

Chin-Chia Hsu\textsuperscript{a}, Chih-Hung Wu\textsuperscript{b}, Shih-Chien Chen\textsuperscript{c}, Kang-Lin Peng\textsuperscript{d}

\textsuperscript{a}Department of Information Management, National Taiwan University, Taiwan
\textsuperscript{b}Department of Business Administration, St. John’s University, Taiwan
\textsuperscript{c}Department of Business Administration, Takming College, Taiwan
\textsuperscript{d}Department of Hotel Management, Ming Hsin University of Science and Technology, Taiwan

\texttt{a}hill@mail.sjsmit.edu.tw, \texttt{b}ericwu@takming.edu.tw, \texttt{c}chen@takming.edu.tw, \texttt{d}pkl1218@must.edu.tw

Abstract

This study develops a novel model, GA-SVR, for parameters optimization in support vector regression and implements this new model in a problem forecasting maximum electrical daily load. The real-valued genetic algorithm (RGA) was adapted to search the optimal parameters of support vector regression (SVR) to increase the accuracy of SVR. The proposed model was tested on a complicated electricity load forecasting competition announced on the EUNITE network. The results illustrated that the new GA-SVR model outperformed previous models. Specifically, the new GA-SVR model can successfully identify the optimal values of parameters of SVR with the lowest prediction error values, MAPE and maximum error, in electricity load forecasting.

Keywords: Support vector regression (SVR); Real-valued genetic algorithm (RGA); Parameter optimization; Electrical load forecasting; Forecasting accuracy

1. Introduction

Support vectors machines (SVMs) have been successfully applied to a number of applications, 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 [29]. Although SVMs have become more widely used to forecast time series data [3, 17, 27] and dynamically reconstruct of chaotic systems [18, 19, 22, 23], it is difficult to build a highly effective model before the parameters of SVMs are carefully determined [9]. Min and Lee [21] stated that the optimal parameter search on SVM plays a crucial role to build a prediction model with high prediction accuracy and stability. To make an efficient SVM model, two extra parameters: sigma squared and gamma have to be carefully predetermined. The kernel-parameters are the few tunable parameters in SVMs, controlling the complexity of the resulting hypothesis [8]. However, seldom articles have been devoted on the study of parameter optimization of SVMs. This study proposes a new method called GA-SVR which can automatically optimize SVR parameters via the real-valued genetic algorithm with little additional computational cost. In addition, variety range of approaches includes time-varying splines [12], multiple regression models [25], judgemental forecasts and artificial neural networks [15] have been used to forecast electricity load. There is no concord as to the perfect approach to electricity demand forecasting [28]. Thus, the proposed method then applied to predict maximum electrical daily load and its performance analyzed.

2. Literature Review

2.1 Support Vector Regression (SVR)

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

\[
\min_{\omega,b,\xi^+,\xi^-} \frac{1}{2}\omega^T\omega + C \sum_{i=1}^{l} (\xi_i + \xi_i^-) \tag{1}
\]

Subject to \( y_i - (\omega^T \phi(x_i) + b) \leq \varepsilon + \xi_i^+, \tag{2} \)
\((\omega^T \phi(x_i) + b) - y_i \leq \varepsilon + \xi_i^-, \tag{3} \)
\( \xi_i^+, \xi_i^- \geq 0, i = 1, ..., l \tag{4} \)
Where \( l \) denotes the number of sample size, \( x_i \) are mapped to a higher dimensional space, \( \xi_i \) represents the upper training error, and \( \xi_i^- \) is the lower training error subject to \( \mathcal{E} \). Three parameters control SVR quality: error cost \( C \), width of tube, and the mapping function (also called kernel function). The basic idea in SVR is to map the data \( x \) into a high-dimensional feature space via nonlinear mapping. Kernel functions perform the nonlinear mapping between input space and a feature space. The approximating feature map for the Mercer kernel performs the nonlinear mapping. In machine learning theories, popular kernel functions are Gaussian (RBF) kernel:

\[
k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2 \sigma^2}\right)
\]

(5)

Polynomial kernel:

\[
k(x_i, x_j) = (1 + x_i \cdot x_j)^d
\]

(6)

Linear kernel:

\[
k(x_i, x_j) = x_i \cdot x_j
\]

(7)

In equation 5, \( \sigma^2 \) denotes the variance of the Gaussian kernel.

2.2 Parameter Optimization

To design an effective model, values of two parameters in SVR have to be chosen carefully in advance [9]. These parameters include the following. Regularization parameter \( C \), which determines the tradeoff cost between minimizing the training error and minimizing model complexity. Parameter sigma \( \sigma \) (or \( d \)) of the kernel function which defines the nonlinear mapping from the input space to some high-dimensional feature space. This investigation only considers the Gaussian kernel that is namely sigma square \( (\sigma^2) \) is the variance of the kernel function. Generally, model selection by SVMs 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, Cristiani, Campell and Taylor [8] were proposed the Kernel-Adatron Algorithm which can automatically perform model selection without testing on a validation. Unfortunately, this algorithm is ineffective if the data has a flat ellipsoid distribution [2]. Therefore, one possible way is to consider the data distribution.

2.3 Real-valued genetic algorithm (RGA)

Recently, genetic algorithms (GA) have been widely and successfully applied to various types of optimization problems [3, 10, 11]. Although many optimization methods have been proposed (e.g. Nelder-Mead simplex method), GA is well suited to the concurrent manipulating of models with varying resolutions and structures since they can search non-linear solution spaces without requiring gradient information or a priori knowledge about model characteristics [20]. In addition, the problem existing in the 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 operation. This is particularly the case when a lot of parameters are needed to be adjusted in the same problem and a higher precision is required for the final result. This is also the major problem while initializing values of parameters of SVM in advance. To overcome the inefficient occupation of the computer memory, the underlying real-valued crossover and mutation algorithms are employed [16]. Different from the binary genetic algorithm (BGA), the real-valued genetic algorithm (RGA) uses real value as a parameter of the chromosome in populations without coding and encoding process before calculating fitness value [13]. Consequently, the RGA is more straightforward, faster and efficient than the BGA.

3. The GA-SVR Approach

The SVM model can be represented as following. The non-linear objective function maximize

\[
W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j \in \mathcal{V}} \alpha_i \alpha_j y_i y_j k(x_i, x_j)
\]

(8)

Subject to \( 0 \leq \alpha_i \leq C, \quad i = 1, \ldots, l \)

(9)

\( \sum_{i=1}^{l} \alpha_i y_i = 0 \)

(10)

The optimal weight \( w \) and bias are determined by solving the quadratic programming problem.

\[
w^* = \sum_{i=1}^{l} \alpha_i^* y_i x_i
\]

(11)

\[
b^* = y_i - w^* \cdot x_i
\]

(12)

The optimal decision function is as follows:

\[
f(x) = \text{sign}\left(\sum_{i=1}^{l} \alpha_i^* k(x_i, x) + b^*\right)
\]

(13)

Cristiani et al. [6] proposed the Kernel-Adatron Algorithm which can automatically select models without testing on a validation data. Unfortunately, this algorithm is ineffective if the data have a flat ellipsoid distribution [2]. This might be often happened in the real world case. Unlike the Kernel-
Adatron Algorithm, this study developed a new method named GA-SVR to optimize the two parameters of SVM simultaneously. The real-valued genetic algorithm (RGA) was adapted to determine the optimal values of SVR parameters to increase SVR accuracy. In the proposed GA-SVR model, the SVR parameters are dynamically optimized via RGA evolution process and the SVR model then performs the prediction task using these optimal values. The process of GA-SVR was illustrated in Fig. 1. The RGA tries to search for the optimal values to enable SVR to fit various datasets. The GA-SVR was developed and coded in the MATLAB v6.5 environment. The major SVR training and validation tool used in this study was developed by [24, 26]. Based on this tool, the proposed model is able to handle huge data sets and easily be combined with the real-valued genetic algorithm in the MATLAB environment.

Unlike traditional BGA, in RGA for optimization problems, all of the corresponding parameters or variables are directly coded to form a chromosome. Hence, the representation of the chromosome is straightforward in RGA. The two parameters, gamma and sigma, of SVR were directly coded to form the chromosome in the present approach. Consequently, the chromosome \( X \) was represented as \( X = \{p_1, p_2\} \), where \( p_1 \) and \( p_2 \) denote the gamma (the regularization parameter C) and sigma \( \sigma \) (the parameter of the kernel function), respectively.

### 3.2 The fitness function

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

### 3.3 Genetic operators

The real-valued genetic algorithm uses selection, crossover, and mutation operators to generate the offspring of the existing population. The proposed GA-SVR model incorporated with 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 in the next generation. The chromosomes that survive in the next generation then are 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 overload frequently occurs while using RGA to optimize values. This study used the method proposed by Adewuya [1] in genetic algorithm with real-valued chromosomes to avoid post-crossover overload problem. The mutation operation follows the crossover to determine whether or not a chromosome should mutate in the next generation. In this study, uniform mutation was designed in the presented model.

Uniform mutation
\( x_{old} = \{x_1, x_2, \ldots, x_n \}, \) \hspace{1cm} (14)

\( x_{k\ new} = LB_k + r \cdot (UB_k - LB_k), \) \hspace{1cm} (15)

\( x_{new} = \{x_1, x_2, \ldots, x_{k\ new}, \ldots, x_n \} \) \hspace{1cm} (16)

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 bound of the parameter, respectively. \( LB_k \) and \( UB_k \) denote the low and upper bound in location \( k \). \( x_{old} \) represents the population before mutation operation; \( x_{new} \) represents the new population after mutation operation.

### 4. Experiment

In this section, the proposed GA-SVR model was tested on a forecasting problem which was announced on the ‘Worldwide Competition within the EUNITE Network'\(^1\). The set problem was to predict maximum daily electricity load for January 1999 based on 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 [28]. In addition, the winning model, SVM, demonstrated superior predictive accuracy to traditional neural network models (e.g. functional network [30], Back-propagation ANN [31], adaptive logic networks [32]). Based on these reasons, our proposed GA-SVR model is used to predict maximum daily values of electricity load values and its prediction performance is compared to other models in the previous EUNITE competition.

#### 4.1 Description of Competition Data and Structure


#### 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 analyses the data to ensure all essential variables were be included in the GA-SVR model. Only if all essential variables are included can the model result 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 applying the proposed model to them. The maximum electrical loads were strongly influenced by temperature factor with a negative correlation existing between the two, as shown in Fig. 2. Specifically, people require more electricity loads to keep warm in cold weather. Despite daily temperature change, the maximum loads data also showed a seasonal pattern in Fig. 3. Repeated, high peak of electricity demand occurred in winter and low peak in summer. Based on previous study, the distribution of temperature has gaussian character. The indexes for gaussian curve are: \( a = 20.85 \), \( b = 196.04 \), \( c = 64.85 \) [31].

---

\(^1\) http://neuront.tuke.sk/competition/index.php
4.2.2 Maximum Load and Holiday Effect

Fig. 4 displays a nonlinear pattern showed in the maximum electricity loads from 1997 to 1998. Meanwhile, the descriptive statistical information of maximum loads was summarized in Table 2. The descriptive statistical information revealed that the lowest peak of electricity demand was 464 and the highest peak of electricity demand was 876 during 1997 to 1998. Moreover, average demand was 670.8 with a high volatility. The data sets also offered holiday information to help predict the maximum electricity loads because early work on this area noted that the holiday might influence maximum load demand. According to public holiday information, electricity load demand generally is lower on holidays and affects by types of holidays.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum:</td>
<td>464</td>
</tr>
<tr>
<td>Maximum:</td>
<td>876</td>
</tr>
<tr>
<td>Mean:</td>
<td>670.8</td>
</tr>
<tr>
<td>Std. :</td>
<td>93.54</td>
</tr>
<tr>
<td>Range:</td>
<td>412</td>
</tr>
<tr>
<td>Skewness:</td>
<td>-.043</td>
</tr>
<tr>
<td>Kurtosis:</td>
<td>-1.235</td>
</tr>
</tbody>
</table>

4.3 Modeling

Kernel and variable selection are an important step for SVR modeling. Since electricity load is a non-linear function of weather variables [28] and other variables (see Fig.5), a nonlinear RBF kernel function (Equation. 4) seemed more properly used here than others to fit the electricity load data. Hence, this study choose the RBF kernel function as the kernel function for data mapping function and obtained the GA-SVR parameters (i.e. sigma and gamma) by RGA 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 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. Additionally, lagged demands such as day-head inputs which might be useful in short term demand forecasting were not included in the input variables in this short term forecasting problem. Extra variable information was not used for modeling. Restated, this work adopted the same variables which 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 SVM models in the ‘Worldwide Competition within the EUNITE Network’, this work evaluated the GA-SVR model based on the same criteria in the previous EUNITE competition.

1. Magnitude of MAPE error

\[
MAPE = 100 \times \frac{1}{n} \sum_{i=1}^{n} \left| \frac{L_{Ri} - L_{Pi}}{L_{Ri}} \right|
\]  

(14)

Where \(L_{Ri}\) denotes the real value of maximum daily electrical load on day “\(i\)” of 1999, \(L_{Pi}\) represents the predicted maximum daily electrical load on the “\(i\)” day of the year 1999, and \(n\) is the number of days in January of 1999, hence \(n=31\).

2. Magnitude of Maximum Error

\[
M = \max \left( \left| L_{Ri} - L_{Pi} \right| \right)
\]  

(15)

\(i\) represents the day in January of 1999, where \(i=1,2,\ldots,31\)

4.5 Design of parameters and fitness function

Some parameters have to be determined in advance before using GA-SVR to forecast electricity loads. Table 3 summarizes all GA-SVR training parameters. The values of individual parameters and the fitness value of the fitness function were based on prior experiences of GA-SVR training and on problem type. Moreover, the fit function design was based on the formula of the first criterion (equation 14), MAPE, and its value is taken as the fitness value in this GA-SVR.

<table>
<thead>
<tr>
<th>Table 3 GA-SVR Training Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
</tr>
<tr>
<td>Population Size:</td>
</tr>
<tr>
<td>Generations :</td>
</tr>
<tr>
<td>Gamma Range:</td>
</tr>
<tr>
<td>Sigma Range:</td>
</tr>
<tr>
<td>Selection Method:</td>
</tr>
<tr>
<td>Mutation Method:</td>
</tr>
<tr>
<td>Snoise:</td>
</tr>
<tr>
<td>Elite :</td>
</tr>
<tr>
<td>Mutation Rate:</td>
</tr>
<tr>
<td>Problem Type:</td>
</tr>
</tbody>
</table>

From Table 3, 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 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 generations were increased to ensure that the global optimum values of the two parameters, gamma and sigma, can be found. Fig. 5 illustrates the whole optimization process of MAPE in the proposed GA-SVR.

![Fig. 5. Optimization Process of MAPE in GA-SVR](image)

The goal of the problem was to predict the real maximum electricity loads in January 1999. Fig. 6 shows the results of the GA-SVR conducted here. Although the real values fluctuated sharply during January 1999, our prediction values (dashed line) are still very close to the real values (solid line).
In the proposed model, the best MAPE was 0.8501 and the maximum error (MW) was 35.02. The optimal values of parameters of SVM for sigma and gamma are 106.49 and 817.32, respectively. Comparing the results obtained by GA-SVR with previous results, the best MAPE generated by the winning SVM model in the previous EUNITE network competition was 2.0. Table 5 lists the results of the proposed GA-SVR versus the results of the previous winner. The new GA-SVR model outperformed the previous winning model in the ‘Worldwide EUNITE Network Competition’, achieving a lower MAPE and MW. Fig. 7 shows the original results of the EUNITE network competition. Complete reports can be found at the EUNITE website (http://neuron.tuke.sk/competition/index.php).

<table>
<thead>
<tr>
<th>Generations</th>
<th>RMSE</th>
<th>MAPE</th>
<th>Max. Error</th>
<th>Sigma*</th>
<th>Gamma*</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>9.68</td>
<td>0.8551</td>
<td>38.47</td>
<td>436.81</td>
<td>9042.72</td>
</tr>
<tr>
<td>100</td>
<td>9.70</td>
<td>0.8540</td>
<td>38.21</td>
<td>223.32</td>
<td>2916.76</td>
</tr>
<tr>
<td>200</td>
<td>9.60</td>
<td>0.8519</td>
<td>37.20</td>
<td>171.48</td>
<td>2179.52</td>
</tr>
<tr>
<td>500</td>
<td>9.46</td>
<td>0.8501</td>
<td>35.02</td>
<td>106.49</td>
<td>817.32</td>
</tr>
</tbody>
</table>

*Note: * denotes optimal value

### Table 5 Results of GA-SVR versus previous champion model

<table>
<thead>
<tr>
<th>GA-SVR</th>
<th>Previous Champion Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAPE :</td>
<td>0.8501</td>
</tr>
<tr>
<td>Max. Error:</td>
<td>35.02</td>
</tr>
</tbody>
</table>

Note: Exact numbers were not specified in Chen, Chang, & Lin (2004)

2 The winning SVM model was proposed by Chen, Chang, & Lin (2002).

5. Conclusions

This study proposed a real-valued genetic algorithm for dynamically optimizing the two parameters of the support vector regression. Experimental results demonstrate the feasibility of successfully applying this new model, GA-SVR, to the complex forecasting problem and in doing so increasing the electricity load forecasting accuracy more than previous winner of the EUNITE network competition. Based on previous studies [7], nonlinear model is usually shown superior results in more accurate short-horizon forecasts. We believe that the novel nonlinearity model can be applied to other complex forecasting problems in the future. In addition, the structural risk minimization principle (SRM) has been shown to be superior to traditional empirical risk minimization principle (ERM) which employed by conventional neural networks, was embodied in SVM. SRM is able to minimize an upper bound of generalization error as opposed to ERM that minimizes the error on training data. Thus, the solution of SVM may be global optimum while other neural network models tend to fall into a local optimal solution, and overfitting is unlikely to occur with SVM [6, 14, 17].
References


