Application of nuSupport Vector Regression in Short-Term Load Forecasting

Adnan Omidi  
Faculty of Electrical and Computer  
Sistan and Baluchestan University, Zahadan, Iran  
Adnanomidi@gmail.com

S. Masoud Barakati  
Faculty of Electrical and Computer  
Sistan and Baluchestan University, Zahadan, Iran  
smbaraka@ece.usb.ac.ir

Saeed Tavakoli  
Faculty of Electrical and Computer  
Sistan and Baluchestan University, Zahadan, Iran  
tavakoli@ece.usb.ac.ir

Abstract—Short-term load forecasting (STLF) of electric power systems plays an essential role in the optimal operation of power systems. Economic performance and reliability of a power system is substantially dependent on the load prediction. STLF is a complex process in electric grid due to having many non-linear factors, such as daily and weekly cyclical changes. Support vector regression has a good ability to estimate non-linear equations. In this paper, a new support vector machine model called nu support vector regression (nu-SVR) is proposed for electrical load forecasting. Results of the proposed method are compared with forecasting results achieved using an artificial neural network (ANN). Results show that the nu-SVR is a proper method for STLF.

Keywords—Short-term load forecasting, support vector regression, multilayer perceptron (MLP) neural networks

I. INTRODUCTION

Load forecasting is a very important problem for the operational planning of electric power systems. In terms of lead time, load forecasting is divided into four classes: Long-term load forecasting, Medium-term load forecasting, short-term load forecasting (STLF), and very short-term load forecasting [1]. STLF targets the prediction of electric loads for a period of one day to seven days. STLF have a significant impact on the efficiency of electrical utilities operation and it is used for unit commitment, fuel allocation, economic dispatch, maintenance scheduling, and load management. Many STLF methods have been suggested during the last few decades [1, 2]. The methods for STLF divided in two categories: statistic techniques and artificial intelligence techniques. In statistic techniques, equations can be achieved showing the relationship between load and its variables, but artificial intelligence techniques have been trying to mimic way of human thinking that how to learn from past experience and anticipated future load [1]. Statistic techniques include time series [3, 4], regression analysis [5, 6], and etc. Artificial intelligence techniques include expert systems [7, 8], fuzzy logic [9, 10], neural networks [11, 12] and etc.

Support vector machine (SVM) was developed by Vapnik [13, 14]. SVM is a novel type of machine learning method, which is a type of artificial intelligent method and has been applied to regression estimation, signal processing, and pattern recognition. Mohandes in [15] was applied support vector machine method to STLF, and compared the results with the auto regression method. The results showed that the auto regression method has superiority to support vector regression (SVR) in results accuracy point of view. Chen and at al. used SVR model to predict the load demand every day for a month [16]; their program won EUNITE competition in 2001 [17]. In this paper, a new method based on nu support vector regression (nu-SVR) for the electric power system STLF is introduced. Case study that used for performance evaluation of the proposed method is energy consumption of the Iranian south-east power grid.

In the following sections, epsilon support vector regression (epsilon-SVR) and nu-SVR methods are explained. Section 2 is dedicated to introduce the model selection. The accuracy of the proposed method is compared with that of the multilayer perceptron neural network method, in section 4. Last section gives conclusion.
A. Support vector machine

Support vector machine is a new machine learning method proposed for classification and regression problems. The main idea of SVR is to use a kernel function to map the initial input data into a high-dimensional space for a given set of the training data \( \{(x_i, y_i)\} (i = 1, 2, \ldots, N) \). \( N \) is sample size, \( x_i \) is input vector, and \( y_i \) is corresponding desired output.

SVR uses a nonlinear function to map input space to high dimensional space, and fits data in high dimensional feature space. The estimating function of using SVR is:

\[
f(x_i) = w\phi(x_i) + b,
\]

where \( \phi \) is a nonlinear transformation from input space to high dimensional feature space, \( b \) and \( w \) denotes a constant bias and weight vector, respectively.

Some methods are developed, such as Epsilon-SVR, to obtain \( f(x) \) in (1) using optimization problem. The \( nu \)-SVR is a kind of support vector regression improves upon epsilon-SVR with introducing a new parameter. In the following, these methods are explained.

B. Epsilon support vector regression

Epsilon-SVR solves the following optimization problem:

\[
\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} (\xi_i^+ + \xi_i^-),
\]

Subject to

\[
\begin{align*}
&f(x_i) - y_i + \varepsilon + \xi_i^- \geq 0, \\
&-f(x_i) + y_i + \varepsilon + \xi_i^+ \geq 0, \\
&\xi_i^-, \xi_i^+ \geq 0, i = 1, 2, \ldots, N,
\end{align*}
\]

where, the terms \( 1/2 \|w\|^2 \) and \( \sum_{i=1}^{N} (\xi_i^+ + \xi_i^-) \) denote model complexity and empirical risk, respectively. The constant \( C \) indicates the tradeoff between the empirical risk and model complexity.

Using Lagrange function and dual theory, one can get dual optimization problem. So, the dual form of the SVR can be expressed as following:

\[
\begin{align*}
&\min \left\{ \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\alpha_i^+ - \alpha_i^-)(\alpha_j^+ - \alpha_j^-)(\phi(x_i)\phi(x_j)) \right. \\
&\quad - \sum_{i=1}^{n} (\alpha_i^+ - \alpha_i^-) f(x_i) \left. \right\},
\end{align*}
\]

Subject to

\[
\begin{align*}
&\sum_{i=1}^{n} (\alpha_i^+ - \alpha_i^-) = 0, 0 \leq \alpha_i^+ \leq C, 0 \leq \alpha_i^- \leq C, i = 1, 2, \ldots, n.
\end{align*}
\]

C. \( nu \) support vector regression

The \( nu \) support vector machine (\( nu \)-SVM) is a kind of support vector machine for classification and regression introduced by Schölkopf et al. \([1]\). This algorithm introduced a new parameter \( \nu(\nu \in (0,1)) \) which can control the number of support vectors and training errors. Essentially \( nu \) support vector regression (\( nu \)-SVR) improves upon epsilon-SVR by allowing the tube width to adapt automatically to the data. Then \( u \)-SVR method like epsilon-SVR method is used for regression problems. The difference \( nu \)-SVR method compared with epsilon-SVR is the optimization problem. In this method, the objective function is as follow:

\[
\min w, b, \xi_i^+, \xi_i^-, \varepsilon \quad \frac{1}{2} \|w\|^2 + C(\nu \varepsilon + \frac{1}{N} \sum_{i=1}^{N} R_{emp}),
\]

Subject to

\[
\begin{align*}
&f(x_i) - y_i + \varepsilon + \xi_i^- \geq 0, \\
&-f(x_i) + y_i + \varepsilon + \xi_i^+ \geq 0, \\
&\xi_i^-, \xi_i^+ \geq 0, i = 1, 2, \ldots, N,
\end{align*}
\]

where \( \frac{1}{2} \|w\|^2 \) denotes model complexity and \( R_{emp} \) is empirical risk that can be obtained from equation (10). The constant \( C \) determines the tradeoff between the empirical risk and model complexity.

\[
R_{emp} = \sum_{i=1}^{N} L_{\varepsilon}(f(x_i), y_i),
\]

where \( \xi_i^+ \) and \( \xi_i^- \) are positive and negative slack variables, respectively, \( \varepsilon \) is the permissive error, and \( L_{\varepsilon} \) is called \( \varepsilon \)-insensitive loss function that was introduced by Vapnik, as follows:

\[
L_{\varepsilon}(f(x_i), y_i) = \begin{cases} 
|f(x_i) - y_i| & \text{if } |f(x_i) - y_i| < \varepsilon \\
0 & \text{otherwise}
\end{cases}
\]

Then equation (10) can be rewritten as follows:

\[
\frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} (\xi_i^+ + \xi_i^-),
\]

If \( x_i \) is not in the tube \( [f(x_i) - y_i] \), there is an error named \( \xi_i^+ \) or \( \xi_i^- \). Figure 1 and Figure 2 illustrates this situation graphically.
Using Lagrange function and dual theory, we can get dual optimization problem. So, the dual form of the SVR can be expressed as following:

\[
L_d = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (\alpha_i^+ - \alpha_i^-)(\alpha_j^+ - \alpha_j^-)k(x_i, x_j) - \sum_{i=1}^{n} (\alpha_i^+ - \alpha_i^-)f(x_i),
\]

Subject to

\[
\sum_{i=1}^{n} (\alpha_i^+ - \alpha_i^-) = 0,
\]

\[
\sum_{i=1}^{n} (\alpha_i^+ - \alpha_i^-) \leq C \nu,
\]

\[
0 \leq \alpha_i^+ , \alpha_i^- \leq \frac{C}{n}, \quad i = 1, 2, \ldots, n,
\]

where \(\alpha_i^+\) and \(\alpha_i^-\) are Lagrange multipliers. \(\alpha_i^+\) and \(\alpha_i^-\) are the solutions of the dual optimization problem. This problem can be solved by quadratic programming techniques.

Value of the kernel equals to inner product of two vectors, \(\phi(x_i)\) and \(\phi(x_j)\). According to the Mercer’s condition \([15]\), the kernel function can be set as following:

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

So, the regression function formulated as:

\[
f(x_i) = \sum_{i=1}^{n} (\alpha_i^+ - \alpha_i^-)k(x_i, x) + b,
\]

The vectors that its \(\alpha_i^+\) and \(\alpha_i^-\) values is not equivalent to zero are the support vectors and they can determine the regression function.

Architecture of SVR is shown in Figure 3.

Computation of bias value \(b\) can be done by exploiting Karush-Kuhn-Tucker (KKT) conditions \([11, 17]\). The bias value is mean of bias values for all of support vectors. So, the bias value computes as flow:

\[
b = \frac{1}{s} \sum_{i=1}^{s} f(x_i) - w^T x_i - \sin g(\alpha_i^+ - \alpha_i^-) \varepsilon,
\]

where \(s\) is support vector number.

II. FORECASTING MODEL

The flowchart of proposed algorithm is illustrated in Figure 4. In the flowing this flowchart is explained.
A. Normalization

In the proposed algorithm all values in the dataset have to be normalized. Normalization implies that all values from the data set should take values in the range \([0, 1]\). This causes load data large amount do not effect on calendar data with small amount. Normalization can be done using the following formulation:

\[
\text{Normalized value} = \frac{x}{x_{\text{max}}},
\]

Where \(x\) is the actual value and \(x_{\text{max}}\) is maximum value of each data set.

B. Parameter selection

The three parameters \(C, \varepsilon,\) and \(\nu\) effect on the accuracy of the model. By trial and error, the parameters are chosen as showing Table 1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\varepsilon)</td>
<td>(\cdot, \cdot)</td>
</tr>
<tr>
<td>(C)</td>
<td>(\cdot, \cdot)</td>
</tr>
<tr>
<td>(\nu)</td>
<td>(\cdot, \cdot)</td>
</tr>
</tbody>
</table>

C. Kernel function

Type of Kernel function has significant effect on the accuracy of the model. In this paper, to transfer data to a higher dimensional feature space, a kernel function called normalized poly function is used. This function is defined as follows:

\[
K(u, v) = \left(\frac{u^T v}{\sqrt{(u^T u)(v^T v)}}\right)^3,
\]

D. Simulation results

To evaluation of the proposed method in short-term load forecasting, a case study based on power consumptions of Iranian south-east power grid is chosen. Power demands of two years are used for training and one year used for test. Mean absolute percentage error (MAPE) is used as evaluation index. MAPE calculates for each hour of day based on the following formula:

\[
\text{MAPE} = \left(\frac{1}{N} \sum_{i=1}^{N} \left|\frac{\hat{y}_i - y_i}{y_i}\right|\right) \times 100,
\]

where \(\hat{y}_i\) denotes actual load value at time \(i\) and \(y_i\) denotes forecasting load value at time \(i\).

Table 1 shows average of MAPE error for months of a year using nu-SVR, and ANN method. The ANN is a multi layers perceptron (MLP) with \(\cdot, \cdot\) neurons in hidden layer. Also mean error is calculated for the four seasons of the year is shown in Table 2. As can be seen, the average error of nu-SVR method has better performance in most of the months and seasons. Forecasting using nu-SVR, and ANN methods for a normal day are illustrated in Figure 5. As one can see, the proposed method has better performance.

<table>
<thead>
<tr>
<th>month</th>
<th>nu-SVR</th>
<th>ANN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2,436</td>
<td>3,058</td>
</tr>
<tr>
<td>2</td>
<td>2,784</td>
<td>3,299</td>
</tr>
<tr>
<td>3</td>
<td>2,881</td>
<td>3,414</td>
</tr>
<tr>
<td>4</td>
<td>2,880</td>
<td>3,399</td>
</tr>
<tr>
<td>5</td>
<td>2,110</td>
<td>2,978</td>
</tr>
<tr>
<td>6</td>
<td>2,450</td>
<td>2,749</td>
</tr>
<tr>
<td>7</td>
<td>2,749</td>
<td>3,434</td>
</tr>
<tr>
<td>8</td>
<td>2,646</td>
<td>3,483</td>
</tr>
<tr>
<td>9</td>
<td>2,441</td>
<td>3,799</td>
</tr>
<tr>
<td>10</td>
<td>2,441</td>
<td>3,799</td>
</tr>
<tr>
<td>11</td>
<td>2,119</td>
<td>2,959</td>
</tr>
<tr>
<td>12</td>
<td>2,034</td>
<td>2,735</td>
</tr>
</tbody>
</table>
Table 7. Mean of MAPE error for different seasons

<table>
<thead>
<tr>
<th>season</th>
<th>nu-SVR</th>
<th>ANN</th>
</tr>
</thead>
<tbody>
<tr>
<td>spring</td>
<td>3.91</td>
<td>3.45</td>
</tr>
<tr>
<td>summer</td>
<td>3.78</td>
<td>3.28</td>
</tr>
<tr>
<td>Fall</td>
<td>3.18</td>
<td>3.09</td>
</tr>
<tr>
<td>winter</td>
<td>3.07</td>
<td>3.04</td>
</tr>
</tbody>
</table>

Figure 8. Short term load forecast by ANN and nu-SVR

III. CONCLUSION

In this paper, a new method is proposed for short-term load forecasting using nu support vector regression. Nu support vector regression trained and tested using a load power data of a practical case study. For short-term load forecasting of 24 hours a day, 4 different models are trained. Using the criteria of mean absolute percentage error for different months and seasons the proposed method is compared with results obtained from ANN method. The results reveal that the new support vector regression has better performance compared with multilayer perceptron neural network.

REFERENCES