Accepted Manuscript

Title: A Computational Intelligence Scheme for the Prediction of the Daily Peak Load

Authors: Jawad Nagi, Keem Siah Yap, Farrukh Nagi, Sieh Kiong Tiong, Syed Khaleel Ahmed

PII: S1568-4946(11)00266-3
DOI: doi:10.1016/j.asoc.2011.07.005
Reference: ASOC 1252


Received date: 9-11-2009
Revised date: 21-4-2011
Accepted date: 17-7-2011


This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
A Computational Intelligence Scheme for the Prediction of the Daily Peak Load

Jawad Nagi a*, Keem Siah Yap b, Farrukh Nagi c, Sieh Kiong Tiong b, Syed Khaleel Ahmed b

a Dalle Molle Institute for Artificial Intelligence (IDSIA), CH-6928 Manno–Lugano, Ticino, Switzerland
b Department of Electronics and Communication Engineering, University Tenaga Nasional, 43000 Kajang, Selangor, Malaysia
c Department of Mechanical Engineering, University Tenaga Nasional, 43000 Kajang, Selangor, Malaysia

Abstract

Forecasting of future electricity demand is very important for decision making in power system operation and planning. In recent years, due to privatization and deregulation of the power industry, accurate electricity forecasting has become an important research area for efficient electricity production. This paper presents a time series approach for mid-term load forecasting (MTLF) in order to predict the daily peak load for the next month. The proposed method employs a computational intelligence scheme based on the self-organizing map (SOM) and support vector machine (SVM). According to the similarity degree of the time series load data, SOM is used as a clustering tool to cluster the training data into two subsets, using the Kohonen rule. As a novel machine learning technique, the support vector regression (SVR) is used to fit the testing data based on the clustered subsets, for predicting the daily peak load. Our proposed SOM-SVR load forecasting model is evaluated in MATLAB on the electricity load dataset provided by the Eastern Slovakian Electricity Corporation, which was used in the 2001 European Network on Intelligent Technologies (EUNITE) load forecasting competition. Power load data obtained from (i) Tenaga Nasional Berhad (TNB) for peninsular Malaysia and (ii) PJM for the eastern interconnection grid of the United States of America is used to benchmark the performance of our proposed model. Experimental results obtained indicate that our proposed SOM-SVR technique gives significantly good prediction accuracy for MTLF compared to previously researched findings using the EUNITE, Malaysian and PJM electricity load datasets.

Keywords: Computational intelligence; Mid-term load forecasting; Daily peak load; Self-organizing map; Support vector machine.

* Corresponding author. Tel.: (+41) (0)767-7548-75
E-mail addresses: jawad@idsia.ch, yapkeem@uniten.edu.my, farrukh@uniten.edu.my, siehkiong@uniten.edu.my, syedkhaleel@uniten.edu.my.
1. Introduction

Load forecasting is a key instrument in power system operation and planning. Many operational decisions in power systems such as: unit commitment, economic dispatch, automatic generation control, security assessment, maintenance scheduling, and energy commercialization depend on the future behavior of loads. In particular, with the rise of deregulation and free competition of the electric power industry all around the world, load forecasting has become more important than ever before [1].

Along with the power system privatized and deregulated, the issue of accurately forecasting electricity load has received more attention in recent years. The error of electricity load forecasting increases operational costs [2–4]. Overestimation of future load results in excess supply, which is not welcome to the international energy network. In contrast, underestimation of load leads to failure in providing enough reserve and implies high costs. Thus, adequate electricity production requires each member of the global co-operation being able to forecast its demands accurately [5].

During the last four decades, a wide variety of techniques have been used for the problem of load forecasting [6–8]. Such a long experience in dealing with the load forecasting problem has revealed time series modeling approaches based on statistical methods and artificial neural networks (ANNs). Statistical models include moving average and exponential smoothing methods such as: multi-linear regression models, stochastic process, data mining approaches, autoregressive moving average (ARMA) models, Box-Jenkins' methods, and Kalman filtering-based methods [9–15]. Since, load time series are usually nonlinear functions of exogenous variables; therefore, to incorporate non-linearity, ANNs have received much attention in solving problems of load forecasting [16–19]. ANN based methods have reported fairly good performances in forecasting. However, two major risks in using ANN models are the possibilities of less or excessive training data approximation, i.e. under-fitting and over-fitting, which increase the out-of-sample forecasting errors. Hence, due to the empirical nature of ANNs their application is cumbersome and time consuming.

Recently, new machine learning techniques such as the support vector machines (SVMs) have been used for load prediction and electricity price forecasting, and have achieved good performances [20],[21]. SVM, namely, support vector regression (SVR) is a powerful machine learning technique used for regression, which is based on recent advances in statistical learning theory [22]. Established on the structural risk minimization (SRM) principle (estimate a function by minimizing an upper bound of the generalization error), SVMs have shown to be very resistant to the under-fitting and over-fitting problems caused by ANNs [20].

In recent times, a Multi-Layer Perceptron SVM (MLP-SVM) was introduced in literature [23]. Compared to standard SVM, MLP-SVM or Hidden Space Support Vector Machine (HSSVM) [24] can adopt to more kinds of kernel functions that are not satisfied by Mercer’s conditions, because the positive definite property of the kernel function is not a necessary condition [24]. From the viewpoint of Mercer’s condition, MLP-SVM’s are less attractive because they are not sufficiently understood for which values of the hidden layer parameters the condition is satisfied [23]. Moreover, the drawbacks of the MLP-SVM in comparison to a standard SVM approach are their higher computational
complexity, the problem of tuning a large number of parameters in the hidden layer and the problem of selecting the optimum number of hidden units [23]. Due to these reasons MLP-SVM’s and their higher computation complexity, the choice of using standard SVM, namely SVR is considerably more suitable for the problem of MTLF.

In this paper, we present our approach on the problem of MTLF in order to predict the daily peak load for the next month. Our study develops a computational intelligence scheme of the Self Organizing Map (SOM) and Support Vector Regression (SVR) using the electricity load data from the 2001 European Network on Intelligent Technologies (EUNITE) competition [25]. In order to evaluate the performance of our forecasting technique, power load data obtained from (i) Tenaga Nasional Berhad (TNB) for peninsular Malaysia and (ii) PJM for the eastern interconnection grid of the United States of America is used to benchmark the performance of the proposed SOM-SVR model.

In our proposed model the SOM is applied to cluster the training data into separate subsets using the Kohonen rule. According to the similarity of time series samples the SOM clustered data is further used to fit the SVR model. Comparisons of recently proposed MTLF results as reported by authors on the EUNITE dataset have been conducted. The theoretical parts are addressed in Sections 2 to 5. Section 6 presents the development and implementation of the SOM-SVR load forecasting model. Section 7 shows experimental results and Section 8 presents concluding remarks.

2. Recent Approaches to Load Forecasting

In the last few decades, there have been widespread references with regards to efforts improving the accuracy of forecasting methods. One of these methods is a weather-insensitive approach which uses historical load data. It is famously known as the Box-Jenkins’ Autoregressive Integrated Moving Average (ARIMA) method discussed in [14] and [26–28]. Christianse [29] and Park et al. [30] proposed exponential smoothing models by Fourier series transformation to forecast electricity load. Douglas et al. [31] considered verifying the impacts of a forecasting model in terms of temperature. To avoid variable selection problems, Azadeh et al. [32] employed a fuzzy system to provide an ideal rule-base to determine the type of ARMA models that can be used. Wang et al. [33] proposed a hybrid Autoregressive and Moving Average with Exogenous variables (ARMAX) model with Particle Swarm Optimization (PSO) to efficiently solve the problem of trapping into local minimum caused by exogenous variables.

To achieve better accuracy of load forecasting, state space and Kalman filtering methods have been developed to reduce the difference between the actual load and the predicted load [34–36]. Moghram and Rahman [37] proposed a model based on historical data to construct the periodic load. Recently, Al-Hamadi and Soliman [38] employed a fuzzy rule-based logic by utilizing a moving window of current values of weather data, and historical load data to estimate the optimal fuzzy parameters for the hourly load of the day. Amjady [39] proposed a hybrid model of the Forecast-Aided State Estimator (FASE) and the Multi-Layer Perceptron Neural Network (MLPNN), to forecast the short-term bus load of power systems.
Regression models construct a causal-effect relationship between electricity load and independent variables. The most popular model is the linear regression, proposed by Asbury [40], considering the weather variable into his model. Papalexopoulos and Hesterberg [10] added holiday and temperature factors into their proposed model. Soliman et al. [41] proposed a multivariate linear regression model for load forecasting, including temperature, wind and humidity factors. Mirasgedis et al. [42] incorporated weather meteorological variables such as: relative humidity, heating and cooling to forecast the electricity demand in Greece. In contrast, Mohamed and Bodger [43] employed economic and geographic variables such as: GDP, electricity price and population to forecast the electricity consumption in New Zealand. Recently, Tsekouras et al. [44] introduced a non-linear multivariable regression approach to forecast the annual load by considering correlation analysis to select appropriate input variables.

In the recent decade, many researchers have tried to apply machine learning techniques to improve the accuracy of load forecasting. Rahman and Bhatnagar [45] constructed a Knowledge-Based Expert System (KBES) approach to electricity load forecasting by simulating the experiences of the system operators [46]. Recently, applications of the Fuzzy Inference System (FIS) and fuzzy set theory have received attention in load forecasting. Ying and Pan [47] introduced an Adaptive Neuro-Fuzzy Inference System (ANFIS) by looking for the mapping relation between the input and output data to. In addition, Pai [48] and Pandian et al. [49] employed fuzzy approaches to obtain superior performance in terms of load forecasting accuracy.

ANNs have been applied to improve the accuracy of load forecasting. Park et al. [50] proposed a Back-Propagation Neural Network (BPNN) for daily load forecasting. Novak [51] applied the Radial Basis Function Neural Network (RBFNN) to forecast electricity load. Darbellay and Slama [52] applied ANNs to predict the regional electricity load in Czechoslovakia. Abdel-Aal [53] proposed an abductive network to conduct a one-hour ahead load forecast for five years. Hsu and Chen [54] employed an ANN model to forecast the regional electricity load in Taiwan. Recently, load forecasting applications of ANNs hybrid with statistical methods and other intelligence techniques have received a lot of attention. These include ANN models combined with: Bayesian inference [55,56], Self-organizing Map (SOM) [57,58], Wavelet transform [59,60], PSO [61] and Dynamic mechanism [62].

Other machine learning techniques such as Support Vector Machines (SVMs), are a promising technique for classification problems. SVMs implement the Structural Risk Minimization (SRM) principle [63], which minimizes the training error and maximizes the confidence interval, resulting in a good generalization performance [64]. With the introduction of Vapniks’ ξ-insensitive loss function [65], SVMs have been extended to solving nonlinear regression estimation problems, and can be considered as successful tools for forecasting problems.

SVMs have been recently applied by researchers to solve load forecasting problems. Cao [66] used the SVM experts for time series forecasting. Cao and Gu [67] proposed a Dynamic SVM (DSVM) model to deal with non-stationary time series problems. Tay and Cao [68] used SVMs for forecasting the financial time series. Hong and Pai [69] applied SVMs to predict engine reliability. For electricity load forecasting, Chen et al. [20] are the pioneers for proposing the SVM model, which was the winning entry of the 2001
EUNITE competition [25]. Pai and Hong [70] employed the concepts of Jordan recurrent neural networks to construct a recurrent SVR model for regional long-term load forecasting (LTLF) in Taiwan. Similarly, Pai and Hong [71] proposed a hybrid model composed of the SVR and the Simulated Annealing (SA) algorithm to forecast Taiwan’s LTLF.

In more recent times, Yuancheng et al. [72] used the Least-Squares SVM (LS-SVM) for 24 hour ahead short-term load forecasting (STLF). Wu and Zhang [73] used LS-SVM with the Chaos theory for load forecasting. SVM based on grey relation analysis was used by Nui et al. [74] for daily load forecasting. In addition, Nui et al. [75] used a combination of SVM and ANN and found that the combination of SVM and ANN produced better results when combined together. SVMs with reduced input dimensions used for load forecasting were proposed by Tao et al. [76]. A comparison of STLF on the ability of SVM verses BPNNs is given in by Zhang [77], concluding the superiority of SVM in load forecasting problems.

3. EUNITE Data Analysis

3.1. EUNITE Competition Data

In 2001, the European Network on Intelligent Technologies (EUNITE) organized a competition on the problem of electricity load forecasting [25]. The information used in the competition contains data from the Eastern Slovakian Electricity Corporation for two years, i.e., from January 1, 1997 until December 31, 1998, which is as follows:

- Half-hourly electricity load
- Daily average temperature
- Annual holidays

![Graph showing half-hourly load data](image)

**Fig. 1.** Half-hourly load data of March 21, 1997. Half-hourly load data represents consumption over an entire day, i.e. 48 consumption values.
The load, temperature and holiday information obtained from the EUNITE competition dataset [25] is listed in Table 1. The objective of the competition is to predict the daily peak electricity load for the 31 days of January 1999 using the given historical data for the preceding two years, i.e. 1997 and 1999. For representation of the EUNITE load data, the half-hourly load of March 21, 1997 is indicated in Figure 1. Since 48 consumption values of half-hourly load do not contribute effectively for prediction of the daily peak load, therefore, the maximum half-hourly load of each day was selected as the daily peak load, as indicated in Table 1.

3.2. Data Analysis

Observations regarding the EUNITE competition data were investigated to determine the relationship between the load data and other information such as, temperature and annual holidays. The following observations are concluded for the given data.

3.2.1. Load Data

Table 1 Data from the 2001 EUNITE competition [25]. Data was provided by the Eastern Slovakian Electricity Corporation.

<table>
<thead>
<tr>
<th>Data</th>
<th>Date</th>
<th>Half-hourly load</th>
<th>Daily peak load</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Year</td>
<td>Month</td>
<td>Day</td>
</tr>
<tr>
<td>Train (Load)</td>
<td>97</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>98</td>
<td>12</td>
<td>31</td>
</tr>
<tr>
<td>Predict (Load)</td>
<td>99</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>99</td>
<td>1</td>
<td>31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Date</th>
<th>Year</th>
<th>Month</th>
<th>Day</th>
<th>Temp (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train (Temperature)</td>
<td>97</td>
<td>1</td>
<td>1</td>
<td>-7.6</td>
</tr>
<tr>
<td></td>
<td>98</td>
<td>12</td>
<td>31</td>
<td>-8.7</td>
</tr>
<tr>
<td>Predict (Temperature)</td>
<td>99</td>
<td>1</td>
<td>1</td>
<td>-10.7</td>
</tr>
<tr>
<td></td>
<td>99</td>
<td>1</td>
<td>31</td>
<td>-6.0</td>
</tr>
</tbody>
</table>

Through simple analysis of the graphs representing the yearly load data, it is observed that the electricity load follows seasonal patterns, i.e. high demand of electricity in the winter (September through March) while low demand in the summer (April through August) [20]. This pattern implies the relationship between the electricity usage and weather conditions in different seasons, as indicated in Figure 2 and Figure 3. Secondly,
another load pattern is observed, where load periodicity exists in the profile of every week, i.e., the load demand on the weekend (Saturday and Sunday) is usually lower than that on weekdays [20] (Monday through Friday), as shown in Figure 4. In addition, electricity demand on Saturday is a little higher than that on Sunday, and the peak load usually occurs in the middle of the week, i.e., on Wednesday.

![Graph of daily peak load from January 1, 1997 until December 31, 1998. Load data represents a period of two years — EUNITE data.](image)

3.2.2. Temperature Data

By analyzing the temperature data, it is observed that the load data has seasonal variation, which indicates a great influence in climatic conditions. A negative correlation between the daily peak load and daily temperature is observed from the two year historical data. The correlation coefficient is found to be -0.8676 [20] as indicated in Figure 5. This observation concludes that electricity consumption in the summer is lower than in the winter (due to the use of heaters in the winter), as lower temperature causes higher electricity demands.

3.2.3. Holiday Effects

Local events, including holidays and festivities, also affect the load demand. These events may lead to a higher load demand for the extra usage of electricity. From the two year historical load data, it is observed that the load demand usually reduces on holidays. Further analysis reveals that the load demand also depends on the type of holiday. On some major public holidays such as Christmas or New Year, the demand for electricity is affected more compared to other holidays [20].
Fig. 3. Daily temperature from January 1, 1997 until December 31, 1998. Temperature data represents a period of two years — EUNITE data.

3.3. Mean Absolute Percentage Error

The accuracy of load forecasting depends upon the error metric, the Mean Absolute Percentage Error (MAPE) of the predicted result. MAPE is defined by the following expression [20]:

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{L_{ia} - L_{ip}}{L_{ia}} \right| \times 100$$

(1)

where \(L_{ia}\) and \(L_{ip}\) are the actual and the predicted values of the daily peak load on the \(i\)th day of January 1999 respectively, and \(N\) is the number of days in January 1999.

3.4. Magnitude of Maximum Error

The Magnitude of Maximum Error (MAX) is an additional error metric used in the EUNITE competition, for comparison of the prediction accuracy. MAX is defined by the following expression [25]:

$$\text{MAX} = \max \left( \left| L_{ia} - L_{ip} \right| \right)$$

(2)

where \(L_{ia}\) and \(L_{ip}\) are the actual and the predicted values of the daily peak load on the \(i\)th day of January 1999 respectively for \(i = 1,2,3...31\) and \(\max\) represents the maximum value.
Fig. 4. Daily peak load for the month of January 1997 and January 1998. Load data for January represents a period of 31 days — EUNITE data.

Fig. 5. Correlation between daily peak load and the daily temperature from January 1, 1997 until December 31, 1998. Correlation coefficient is: -0.8676 [20] — EUNITE data.
4. Self-Organizing Map

The self-organizing map (SOM) introduced by Kohonen in the 1980s is also referred to as a Kohonen Map [78]. The SOM is an unsupervised artificial neural network (ANN), which learns the distribution of a set of patterns without any class information. SOMs non-linearly project high-dimensional data onto a low-dimensional grid [79–82]. SOMs have been used in numerous applications since they were initially proposed. SOM initially focused on applications such as: engineering [81], image processing [82–85], process monitoring and control [86,87], speech recognition [88,89], and flaw detection in machinery [90]. Recently, SOM applications in other fields have emerged including business and management, such as: information retrieval [91,92], medical diagnosis [93,94], time series prediction [95,96], optimization [97], as well as financial forecasting and management [98–100].

The SOM projected data preserves the topological order in the input space; hence, similar data patterns in the input space are assigned to the same map unit or nearby units on the trained map. The core process of the projection is first, for each input pattern, determining its best matching unit (BMU) from the map units. The BMU is the unit that is most similar to the input pattern. In short, the two key steps in training an SOM are: (1) Determining the BMU and, (2) Updating the BMU and its neighbours [101].

Specifically, consider the SOM algorithm [102] as a nonlinear, ordered and smooth mapping of an \( n \)-dimensional input space to an \( m \)-dimensional output space. In the output space, each neuron or unit of the SOM is represented by a codebook vector, \( m_i = [m_{i1}, m_{i2}, ..., m_{in}]^T \in \mathbb{R}^n \), where \( m_{ik} \) is the value of the \( k \)th component. The projection of an input pattern \( x = [x_1, x_2, ..., x_n]^T \in \mathbb{R}^n \) is achieved by assigning \( x \) to the closest codebook vector \( m_c \) with respect to a general distance measure \( d(x, m_i) \) [101], where \( m_c \) is the winning neuron [103].

The method for determining the BMU, \( c \), with respect to an input pattern \( x \) is to identify the unit that is most similar to \( x \). The distance function is employed to measure the similarity. The smaller the distance is, the more similar the neurons are. Formally, the BMU of an input \( x \) is defined as [101]:

\[
    c = \arg \min_i \{d(x, m_i)\}
\]

where \( m_i \) is a unit on the map. The typical method for computing the distance \( d(x, m_i) \) is by using the Euclidean distance function, defined by:

\[
    d(x, m_i) = \|x - m_i\| = \left(\sum_{k=1}^{n} (x_k - m_{ik})^2\right)^{1/2}
\]

During SOM training, the winning neuron and all the neurons in its neighbourhood may adapt their codebook vectors. In the sequential SOM algorithm, the rate of adaptation is steered by a monotonically decreasing and time-dependent function \( \alpha(t) \), which is known as the incremental learning rate [103]. The neighbourhood function \( h_c(t) \) is centered on the best-matching neuron. Both \( \sigma(t) \) and \( h_c(t) \) are monotonically
decreasing functions of time. A typical smooth neighbourhood function is the Gaussian function, is defined by:

\[ h_{ci}(t) = \exp\left[ -\frac{\|r_c - r_i\|^2}{2\sigma(t)^2} \right] \] (5)

where \(0 < \sigma(t) < 1\) is the width of the Gaussian kernel representing the size of the neighbourhood and \(\|r_c - r_i\|^2\) is the distance between the winning neuron and the neuron \(i\), with \(r_c\) and \(r_i\) representing the 2D positions of neurons \(c\) and \(i\) on the SOM grid. Thus, the main steps of the sequential SOM training are as follows [103]:

1. Initialize the codebook vectors \(m_i(0)\) of all neurons.
2. Determine the winning neuron \(m_c(t)\) for the input \(x\), using the Euclidean distance measure \(d(x, m_i)\).
3. Update the codebook vectors:

\[ m_i(t + 1) = m_i(t) + \alpha(t)h_{ci}(t)[x(t) - m_i(t)] \] (6)

4. Repeat the last two steps until a predefined number of steps are reached.

The sequential SOM training is usually performed in two phases. The first phase is characterized by choosing large learning rate and neighbourhood radius parameters, and the second phase is used to fine-tune the codebook vectors by setting smaller starting values for \(\sigma(t)\) and \(h_{ci}(t)\).

5. Support Vector Regression

Support vector machines (SVMs) introduced by Vapnik in the 1960s, are based on the foundation of the statistical learning theory [63]. SVMs are a set of related supervised machine learning methods used for classification, which have recently become an active area of intense research with extensions to regression and density estimation [64]. SVMs employ the structural risk minimization (SRM) principle to overcome intrinsic limitations of ANNs. Support vector regression (SVR) in SVMs can be used for time series prediction, which is useful for problems characterized by non-linearity, high dimension and local minima. SVRs have been successfully employed to solve regression problems such as: time series modelling [66, 67], financial forecasting [68], electricity load forecasting [70–77] and non-linear control systems [104].

The basic concept of SVR is to map the input data, \(x\), non-linearly into a higher dimensional feature space. Hence, given a set of data \(G = \{\{x_i, d_i\}\}\) for \(i = 1, \ldots, N\), where \(x_i\) is the \(i\)th input vector to the SVR, \(d_i\) is the actual \(i\)th output value, and \(N\) represents the total number of data patterns, the SVR function is defined by:

\[ y = f(x) = w\phi(x) + b \] (7)
where \( \phi(x) \) is the feature that is non-linearly mapped from the input space, \( x \). The coefficients \( w \) and \( b \) are the support vector (SV) weight and bias respectively, which are calculated by minimizing the following regularized risk function:

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

(8)

where,

\[
L_{\varepsilon}(d, y) = \begin{cases} 
0, & |d - y| \leq \varepsilon \\
|d - y| - \varepsilon, & \text{otherwise}
\end{cases}
\]  

(9)

Here, parameters \( C \) and \( \varepsilon \) are prescribed parameters. In (8) \( L_{\varepsilon}(d, y) \) is the \( \varepsilon \)-insensitive loss function, as illustrated in Figure 6. The loss equals zero if the forecasted value is within the \( \varepsilon \)-tube [105,106], as indicated in (9). The second term \( \|w\|^2 / 2 \) in (8), measures the flatness of the function.

![Fig. 6. The linear \( \varepsilon \)-insensitive loss function of \( \varepsilon \)-SVR.](image)

Therefore, \( C \) specifies the trade-off between the empirical risk and the model flatness. Both \( C \) and \( \varepsilon \) are user-determined parameters. Two positive slack variables \( \xi \) and \( \xi^* \), represent the distance from the actual values to the corresponding boundary values of the \( \varepsilon \)-tube, are introduced as shown in Figure 7. Then the regularized risk function in (8) is transformed into the following constrained form:

\[
R(w, \xi, \xi^*) = \|w\|^2 / 2 + C \left( \sum_{i=1}^{N} \left( \xi_i + \xi_i^* \right) \right)
\]  

(10)

subject to the constraints,

\[
\begin{align*}
    w\phi(x_i) + b_i - d_i & \leq \varepsilon + \xi_i^*, \quad i = 1, 2, \ldots, N \\
    d_i - w\phi(x_i) - b_i & \leq \varepsilon + \xi_i^*, \quad i = 1, 2, \ldots, N \\
    \xi_i, \xi_i^* & \geq 0, \quad i = 1, 2, \ldots, N
\end{align*}
\]
Fig 7. The intensive band ±ε and slack variables \( \xi_i \) and \( \xi_i^* \) for ε-SVR.

This constrained optimization problem in (10) is solved using the following primal Lagrangian form:

\[
L(w, b, \xi, \xi^*, \alpha, \alpha^*, \beta, \beta^*) = \frac{1}{2} \|w\|^2 + C \left( \sum_{i=1}^{N} (\xi_i + \xi_i^*) \right) - \sum_{i=1}^{N} \beta_i \left[ w\phi(x_i) + b - d_i + \varepsilon + \xi_i \right] - \sum_{i=1}^{N} \beta_i^* \left[ d_i - w\phi(x_i) - b + \varepsilon + \xi_i^* \right] - \sum_{i=1}^{N} (\alpha_i \xi_i + \alpha_i^* \xi_i^*)
\]

(11)

where (11) is minimized with respect to the primal variables \( w, b, \xi \) and \( \xi^* \) and maximized with respect to the non-negative Lagrange multipliers \( \alpha_i, \alpha_i^*, \beta_i \) and \( \beta_i^* \). Therefore, (12–15) are obtained:

\[
\frac{\partial L}{\partial w} = w - \sum_{i=1}^{N} (\beta_i - \beta_i^*) \phi(x_i) = 0,
\]

(12)

\[
\frac{\partial L}{\partial b} = w - \sum_{i=1}^{N} (\beta_i^* - \beta_i) = 0,
\]

(13)

\[
\frac{\partial L}{\partial \xi_i} = C - \beta_i - \alpha_i = 0
\]

(14)

\[
\frac{\partial L}{\partial \xi_i^*} = C - \beta_i^* - \alpha_i^* = 0
\]

(15)

Finally, the Karush-Kuhn-Tucker (KKT) conditions are applied to the regression and (10) yields the dual Lagrangian by substituting (12–15) into (11). Then, the dual Lagrangian (16) is obtained by the following expression:
\begin{equation}
\mathcal{L}(\beta, \beta^*) = \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_{j=1}^{N} (\beta_i - \beta_i^*) (\beta_j - \beta_j^*) K(x_i, x_j)
\end{equation}

subject to the constraints,

$$
\sum_{i=1}^{N} (\beta_i - \beta_i^*) = 0
$$

$$
0 \leq \beta_i \leq C, \quad i = 1, 2, ..., N
$$

$$
0 \leq \beta_i^* \leq C, \quad i = 1, 2, ..., N
$$

The Lagrange multipliers in (16) satisfy the equality \((\beta_i \cdot \beta_i^*) = 0\). Lagrange multipliers \(\beta_i\) and \(\beta_i^*\) are calculated and an optimal desired weight vector of the regression hyperplane is given by the following expression:

$$
w^* = \sum_{i=1}^{N} (\beta_i - \beta_i^*) \phi(x)
$$

Hence the regression function is expressed by:

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

In (18), \(K(x, x_i)\) is called the \textit{kernel function}. The value of the kernel is equal to the inner product of the two vectors \(x\) and \(x_i\) in the feature space \(\phi(x)\) and \(\phi(x_i)\), i.e., \(K(x, x_i) = \phi(x) \cdot \phi(x_i)\). Four of the popular SVM kernel functions used in this experiment are [107] as follows:

1. Linear kernel:
   \(K(x, x_i) = x^T x_i\)

2. Polynomial kernel:
   \(K(x, x_i) = (\gamma x^T x_i + r^d), \quad \gamma > 0\)

3. Radial basis function (RBF—Gaussian) kernel:
   \(K(x, y) = \exp\left(-\gamma \|x-x_i\|^2\right), \quad \gamma > 0\)

4. Sigmoidal kernel:
   \(K(x, y) = \tanh(\gamma x^T x_i + r), \quad \gamma > 0\)

In general, it is difficult to determine the type of kernel function to use for specific data patterns [106,108]. However, any function that satisfies \textit{Mercer’s condition} by Vapnik [63] can be used as a kernel function in SVMs.

The selection of three (3) parameters, \(\varepsilon\), \(C\) and \(\gamma\) of \(\varepsilon\)-SVR are important to the forecasting accuracy. As an example, if \(C\) is chosen as too large (approximated to
infinity), then the objective is to minimize the empirical risk, $L_\varepsilon(d, y)$ without the model flatness in the optimization formulation in (10). The parameter $\varepsilon$ controls the width of the $\varepsilon$-insensitive loss function, which is used to fit the training data. Large $\varepsilon$ values result in a more flat regression estimated function. The parameter $\gamma$ controls the width of Gaussian function, which reflects the distribution range of $x$ values in the training data. There are a lot of existing practical approaches for parameter selection of $\varepsilon$, $C$ and $\gamma$ such as, user-defined based on prior knowledge and experience, cross-validation (CV), asymptotical optimization [109] and grid search [107].

For experiments in this paper, LIBSVM [110], a library for support vector machines, is used. The selection of optimal $\varepsilon$-SVR parameters $\varepsilon$, $C$ and $\gamma$ is achieved using the Grid Search method suggested by Hsu et al. in [107], which is discussed in Section 6.7.

6. Methodology

6.1. Attribute Selection

Our proposed load forecasting model is based on the past daily peak loads (historical consumption data) as one of the candidate input variables. The best input features for our forecasting model are those which have the highest correlation with the output variable (i.e. peak load of the next day) and the highest degree of linear independency. Thus, the most effective candidate inputs with minimum redundancy are selected as the model attributes. For the purpose of attribute selection, a two step correlation analysis employed in [114] is used in this research study, which is described as follows:

a) In the first step, the correlation between each candidate input and output variable is computed, where higher correlation indicates more effective candidates. If the correlation index between a candidate variable and the output feature is greater than a prespecified value then this candidate is retained for the next step; else, it is not considered any further [114].

b) In the second step, for the retained candidates, a cross-correlation analysis is performed. If the correlation index between any two candidate variables is smaller than a prespecified value then both variables are retained; else, only the variable with the largest correlation with respect to the output is retained, while the other is not considered any further [114].

This correlation analysis is applied to the candidate inputs, i.e. the past daily peak loads and daily temperatures, excluding the major calendar indicators: type of day and annual holidays. The type of days and annual holidays are excluded from the two step correlation analysis, since these indicators are composed using binary variables and have low correlation with the output feature, even in the normalized form. However, the indicator can be still useful to separate different kinds of days.
Retained candidates after the two steps of the correlation analysis are selected as the input features of the forecasting method. The proposed two step correlation analysis resulted in the best feature selection and forecast accuracies, due to consideration of linear independency of candidate inputs in addition to their correlation. The four attributes used in the SOM-SVR modelling process are mentioned below.

6.1.1. Daily Peak Load

Since, past load demand affects and implies the future load demand, therefore, including the past daily peak load as a key attribute, will greatly influence improvement in the forecasting performance.

6.1.2. Daily Temperature

As previously indicated through our data analysis in Figure 5, the electricity load and the temperature have a causal relationship (high correlation) between each other. Therefore, the daily temperature is used as an attribute in the forecasting model.

6.1.3. Type of Day

Weekly periodicity of the electricity load is noticed through load data analysis as shown in Figure 4. As the electricity demand on holidays is observed to be lower than on non-holidays, therefore, encoding information of the type of day (calendar indicator) into the forecasting model will benefit performance of the model.

6.1.4. Annual Holidays

As it is commonly understood that annual holidays, local events and festivities contribute significantly towards the electricity load demand, hence, inclusion of annual holidays as a calendar indicator in the forecasting model will have a significant impact on the prediction accuracy.

6.2. Data Normalization

The daily peak power load and the daily temperature data needs to be represented in a normalized scale for SVR training and prediction. Thus, all feature data are linearly scaled in the range of \([0, 1]\) using the following expression:

\[
X_N = \frac{(X - X_{\text{min}})}{(X_{\text{max}} - X_{\text{min}})}
\]  

(23)
where $X$ is the load or temperature feature data, $X_{\text{min}}$ represents the smallest value in the feature data $X$, and $X_{\text{max}}$ represents the largest value in the feature data $X$.

6.3. Time Series Modelling for Weekly Periodicity

The daily peak electricity load for the preceding two years is introduced into the forecasting model with the concept of a time series. If $x_i$ for $i = 1, 2, \ldots, 730$ is the historical peak load data of two years from January 1, 1997 until December 31, 1998 and $y_i$ is the respective target value for training, then the target vector $y_i$ includes several next load values based on **weekly periodicity**. Therefore, $x_i$ can be represented as a training load matrix ‘$R$’, in the form:

$$
R = \begin{bmatrix}
  x_i(1) & x_i(2) & x_i(3) & \cdots & x_i(7) \\
  x_i(2) & x_i(3) & x_i(4) & \cdots & x_i(8) \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  x_i(722) & x_i(723) & x_i(724) & \cdots & x_i(728) \\
  x_i(723) & x_i(724) & x_i(725) & \cdots & x_i(729)
\end{bmatrix}
$$

where $R$ consists of $723 \times 7$ load values. The first column in the $R$ corresponds to the historical load data for the preceding two years, excluding the last week. The second column in the $R$ is shifted one day ahead with respect to the first column. The third column is shifted two days ahead with respect to the first column, and so on with the last column being shifted six days ahead, such that all 7 columns contain 723 load values. Similarly, the target vector, $y_i$, is represented by the vector $S$, in the form:

$$
S = \begin{bmatrix}
  x_i(8) \\
  x_i(9) \\
  \vdots \\
  x_i(729) \\
  x_i(730)
\end{bmatrix}
$$

where $S$ consists of 723 load data values. The target vector $S$ is shifted one day ahead with respect to the last column in the training load matrix, $R$, i.e., the vector $S$ is a continuation of the one day ahead day-shifting sequence from the matrix, $R$. Therefore, $S$ represents a **weekly periodicity** consumption pattern in accordance with $R$, for the last 723 days of the preceding two years.

6.4. Data Representation

After formulation of the training matrix ‘$R$’ and target vector ‘$S$’, all feature attributes are selected in a proper combination to prepare the training dataset. In the training dataset,
a training sample [i.e., \( x_i \) in (10)] encoded for a particular \( i \)th day is in the following format:

\[
(\text{peak load, temperature, type of day, annual holiday})
\]

(26)

where \textit{peak load} represents the training matrix ‘\( R \)’, \textit{temperature} represents the daily temperature, \textit{type of day} represents the day in the week, and \textit{annual holidays} represents the public holidays. To represent the calendar information, seven binary attributes are used to encode the ‘\textit{day type}’ from Monday through Sunday, where the current day is represented by 1 and all other digits are set to 0. Similarly, ‘\textit{annual holidays}’ are represented using one binary digit, where 1 is used to represent a public holiday and 0 represents no public holiday.

6.5. Load Forecasting Model

A computational intelligence scheme based the SOM and SVR is applied to reconstruct the dynamics of electricity load forecasting using a time series approach. The proposed SOM-SVR load forecasting model is shown in Figure 8.
Training matrix clustered into two subsets
Subset 1: $L_1, T_1, D_1, H_1, V_1$
Subset 2: $L_2, T_2, D_2, H_2, V_2$

**Euclidean distance comparison**
Cluster centers of the training data and testing data

Two cluster centres selected randomly based on the training data subsets

$\varepsilon$-SVR1
Training using Subset 1
Prediction

$\varepsilon$-SVR2
Training using Subset 2
Prediction

SVR models

Predicted electricity load for the 31 days of January

**Fig. 8.** Flowchart of the proposed SOM-SVR load forecasting model. Parameter subscript *train* denotes the ‘training data’ and subscript *test* denotes the ‘testing data’.

In Figure 8, the subscripts *train* and *test* represent the training and testing data respectively. The parameter $L$ represents the daily peak load matrix, $T$ represents the daily temperature, $D$ represents the type of day, $H$ represents the annual holidays, and $V$ represents the target vector in correspondence with $L$. 
The load forecasting model proposed in Figure 8 is based on hybrid architecture of the SOM and the ε-SVR. The SOM is used as a clustering tool to cluster the training data into two subsets, using the Kohonen rule. The reason for using two subsets is obtained from the fact that, two years of historical data is used to construct the forecasting model. After clustering is complete, two ε-SVRs (one for each training subset) are employed to fit the clustered data appropriately for SVR training. The Euclidean distance between the cluster centres and testing samples is used as a measure to select the appropriate ε-SVR model for predicting the peak daily load.

The forecasting model presented in this research study is developed using MATLAB R2009b. The computer used for testing is a 2.40 GHz Quad-core processor with 4 GB of RAM. The SOM is implemented using the MATLAB Neural Network Toolbox and the SVR was implemented using LIBSVM [110].

6.6. SOM Clustering

As indicated in Figure 8, in the first stage the SOM is employed as a gating network to identify the switching or piece-wise stationary characteristics of the training data. Firstly, two cluster centres for the two year historical training data are randomly selected based on the training data space. The SOM is applied to cluster the space of the two year historical training data into two different subsets with similar properties, as shown in Figure 8. This process leads to the decomposition of the normalized training data into two subgroups as indicated in Figure 9. Each clustered subset is considered as a stationary time series data. The SOM parameters selected for the clustering process are indicated in Table 2.

Table 2  SOM parameters selected for the clustering of training data. Parameters are selected by iterating different combinations using trial and error.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_i$</td>
<td>Initial learning rate</td>
<td>0.96</td>
</tr>
<tr>
<td>$w_p$</td>
<td>Winning neuron criteria</td>
<td>$10\exp(10)$</td>
</tr>
<tr>
<td>$E$</td>
<td>Training iterations</td>
<td>55</td>
</tr>
</tbody>
</table>

As the incremental learning rate, $\alpha_f$ of a SOM (based on the size of the neighbourhood radius) calculates the winning neuron, the following expression is used to calculate $\alpha_f$ on each iteration:

$$\alpha_f(e) = \alpha_i^{(-0.02e)} \quad (27)$$
Fig. 9. Two year historical data clustered into two subsets in a supervised manner using the Kohonen rule – EUNITE data.

6.7. Model Training

Each training sample consists of 17 features, which includes the following: 8 normalized peak load values (7 normalized peak load data from the training load matrix, \( R \) and 1 normalized peak load data from the target vector, \( S \)), 1 normalized temperature data, 7 binary digits representing the type of day and 1 binary digit representing the annual holiday, in the format, illustrated in eq. (26). Thus, the training matrix is composed of \( 723 \times 17 \) normalized feature values (17 features from the first 723 days of the preceding two years).

SVR training involves training two \( \varepsilon \)-SVR models (\( \varepsilon \)-SVR1 and \( \varepsilon \)-SVR2) with the training matrix. The reason for selecting ‘two \( \varepsilon \)-SVR’ models is due to the fact that the historical data is clustered into two subsets. The two \( \varepsilon \)-SVRs select the clustered training data appropriately between themselves for training, based on the two cluster centres.

Training data for the \( \varepsilon \)-SVRs is indicated in Table 3. Optimal SVR parameters, \(( C, \gamma )\) were selected using the Grid Search method proposed by Hsu et al. in [107]. Grid Search is implemented by generating exponentially growing sequences of parameters \(( C, \gamma )\) and iterating them for all possible combinations to find the optimal SVR hyper-parameter set. The best ‘\( \varepsilon \)’ parameter for the SVR models was determined using the highest 10-fold cross-validation (CV) accuracy obtained. CV is used as a measure to evaluate the fitting provided by each parameter value set during grid search in order to avoid over-fitting.
Table 3 Training data for the ε-SVR. Training data corresponds to the training matrix, represented in the format illustrated in eq. (26).

<table>
<thead>
<tr>
<th>Input</th>
<th>Parameter</th>
<th>Detail description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-7</td>
<td>(L_{\text{train}})</td>
<td>Daily peak load data matrix, (R) in eq. (24) obtained using the historical load data of the preceding two years.</td>
</tr>
<tr>
<td>8</td>
<td>(T_{\text{train}})</td>
<td>Daily temperature vector obtained using the temperature data of the preceding two years.</td>
</tr>
<tr>
<td>9-15</td>
<td>(D_{\text{train}})</td>
<td>Type of day binary matrix representing calendar information for the preceding two years with seven binary attributes.</td>
</tr>
<tr>
<td>16</td>
<td>(H_{\text{train}})</td>
<td>Annual holiday vector for the preceding two years representing the public holidays with one binary attribute.</td>
</tr>
<tr>
<td>17</td>
<td>(V_{\text{train}})</td>
<td>Target vector (S), in eq. (25) obtained using the historical load data of the preceding two years.</td>
</tr>
</tbody>
</table>

6.8. Model Testing

During the prediction phase, model attributes for January 1999 such as temperature data, type of day and annual holidays need to be encoded in the testing data. The type of day information is obtained from a calendar and annual/public holidays for January 1999 are provided in the 2001 EUNITE competition data.

The temperature information for January 1999 is not provided. In order to encode the temperature data in our testing entries, we need to predict or estimate the temperature of January 1999. Temperature forecast is not easy, especially with limited data. We employ a straightforward idea for temperature prediction which is taken from [20]. This involves using the average of the past temperature data for the estimation. The daily temperature of the past four years for the 2001 EUNITE competition is provided by the Eastern Slovakian Electricity Corporation. Hence, the temperature of each of the 31 days in January 1999 is estimated by averaging the past January daily temperature data from 1995 to 1998.

Each testing sample consist of 16 features, which includes the following: 7 normalized peak load values (7 normalized peak load data of the last 7 days from the preceding two years), 1 normalized temperature data, 7 binary digits representing the type of day and 1 binary digit representing the annual holiday, in the format illustrated in eq. (26). Thus, the testing matrix is composed of \(7 \times 16\) normalized feature values. Feature information in the testing matrix includes:

1. Initially for predicting the first day of January 1, 1999, the daily peak load data consists of the last week load of the preceding two years, i.e. from December 25, 1998 until December 31, 1998.
3. The calendar information comprises of all the type of days, for the 31 days of January 1999.
4. The annual holidays include all public holidays for the month of January 1999.
The training load matrix, $R$ includes the training vector, $S$ for SVR training (see Table 3). For SVR testing, the testing vector in the testing matrix was set to 0, i.e. $V_{test} = 0$ (see Table 4). This is because no likely prediction values for the testing data are currently known or can be estimated. Testing data for the $\varepsilon$-SVRs is indicated in Table 4.

Table 4  Testing data for the $\varepsilon$-SVR. Testing data corresponds to the testing matrix represented in the format illustrated in eq. (26).

<table>
<thead>
<tr>
<th>Input</th>
<th>Parameter</th>
<th>Detail description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–7</td>
<td>$L_{test}$</td>
<td>Daily peak “historical” and/or “currently predicted” load data vector for one week data.</td>
</tr>
<tr>
<td>8</td>
<td>$T_{test}$</td>
<td>Daily temperature for the current day of January 1999 to be predicted.</td>
</tr>
<tr>
<td>9–15</td>
<td>$D_{test}$</td>
<td>Type of day vector representing calendar information for the current day of January 1999 to be predicted.</td>
</tr>
<tr>
<td>16</td>
<td>$H_{test}$</td>
<td>Annual holiday vector representing public holidays for the current day of January 1999 to be predicted.</td>
</tr>
<tr>
<td>-</td>
<td>$V_{test}$</td>
<td>Target vector for the testing matrix, set to 0 (no current values for the testing data are known).</td>
</tr>
</tbody>
</table>

6.9. Daily Peak Load Prediction

The steps involved for predicting the electricity load of the 31 days of January 1999 with the testing data (outlined in Table 4) are as follows:

1. Two Euclidean distances ($d_1, d_2$) are computed between the two cluster centres and for each testing sample (see Figure 8).
2. The logical comparison of the Euclidean distances provides the decision of selecting the appropriate $\varepsilon$-SVR trained model to fit the testing sample.
3. If distance $d_1 < d_2$ then $\varepsilon$-SVR1 is selected for prediction of the daily peak load, else if distance $d_2 <= d_1$ then $\varepsilon$-SVR2 is selected for prediction.
4. After selection of the appropriate $\varepsilon$-SVR, the electricity load for the 31 days of January 1999 is predicted on a daily basis.
5. For the first prediction of January 1, 1999, the last week load data from the preceding two years (December 25, 1998 until December 31, 1998) is used with the temperature and the calendar information of January 1, 1999. This is the first SVR testing sample.
6. To predict the peak load of January 2, 1999, the load data from December 26, 1998 until December 31, 1998 and the predicted peak load of January 1, 1999 is used with the temperature and the calendar information of January 2, 1999. This is the second SVR testing sample.
7. The peak load of January 3, 1999 is predicted in a similar way as for the case of the first two days of January. The historical load data from December 27, 1998 until December 31, 1998 and the predicted peak load of the first two days of January 1999 are used with the temperature and the calendar information of January 3, 1999. This is the third SVR testing sample.
8. The ‘one day shifting’ sequence in the load data, as indicated in steps (5), (6) and (7), follows the same procedure for predicting the peak load for remaining days of January 1999.

9. The last prediction, January 31, 1999 is predicted using the last seven predicted days of January 1999. This means that peak load data from January 24, 1999 until January 30, 1999, is used with the temperature and the calendar information of January 31, 1999. This is the last SVR testing sample.

10. The 31 daily peak loads predicted by the two ε-SVRs, contribute to the predicted electricity load for the month of January 1999.

7. Experimental Results

7.1. SVR Kernel Selection

The behaviour of four different SVM kernels, namely: linear, polynomial, radial basis function (RBF) and sigmoidal on the accuracy of prediction (MAPE) is observed using 10-fold CV. Experimentally, by iterating various parameter ε values, the best value obtaining the highest CV accuracy was found to be ε = 0.12.

![Fig. 10. Comparison of the predicted electricity load for January 1999 using different SVM kernels — EUNITE data.](image)

In this experiment, default parameter values for all SVM kernels defined in (19–22) were used and results obtained are given in Table 5. Results obtained indicated that the best prediction accuracy was obtained using the RBF kernel, resulting in a MAPE of 1.32%. From this point onwards, all experiments were performed using the RBF kernel. The comparison plot of the predicted electricity load of January 1999 using different kernels is shown in Figure 10.
Table 5 Prediction accuracy of forecasting model using different SVM kernels.

<table>
<thead>
<tr>
<th>Kernel type</th>
<th>Kernel parameters</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>No parameters</td>
<td>2.19%</td>
</tr>
<tr>
<td>Polynomial</td>
<td>C = 100, γ = 0.001, r = 0, d = 3</td>
<td>1.77%</td>
</tr>
<tr>
<td>RBF</td>
<td>C = 100, γ = 0.001</td>
<td>1.32%</td>
</tr>
<tr>
<td>Sigmoidal</td>
<td>C = 100, γ = 0.001, r = 0</td>
<td>2.33%</td>
</tr>
</tbody>
</table>

7.2. SVR Parameter Optimization

Using the RBF kernel, two parameters \((C, γ)\) need to be determined in SVM. For this study, the Grid Search method proposed by Hsu et al. in [107] is used in LIBSVM [110] for SVR hyper-parameter optimization.

In the Grid Search, exponentially growing sequences of parameters \((C, γ)\) were used to identify SVR parameters obtaining the highest prediction accuracy (lowest MAPE). Sequences of parameters, \(C = [2^1, 2^2, 2^3, \ldots, 2^{20}]\) and \(γ = [2^{-1}, 2^{-2}, 2^{-3}, \ldots, 2^{-20}]\) were used for \(100 \times 100 = 10,000\) combinations. For each pair of \((C, γ)\) the validation performance was measured using 10-fold CV accuracy. Experimentally it was found that: \(C_1 = 1068.84, γ_1 = 0.0017, C_2 = 252.73\) and \(γ_2 = 0.0058\) for \(\varepsilon\)-SVR1 and \(\varepsilon\)-SVR2 respectively obtained the highest 10-fold CV accuracy of 96.47%, for a MAPE of 0.97% and MAX of 13.50 MW.

7.3. Model Evaluation Using EUNITE Data With Other Techniques

A comparative study of our proposed model with other machine learning techniques was performed using the EUNITE data. Our proposed SOM-SVR method was compared with two different prediction techniques: (1) SVR and (2) ML-BPNN. The prediction accuracy of our proposed model compared with the different prediction techniques for predicting the electricity load of January 1999 is shown in Table 6. Figure 11 shows the comparison plot of the predicted electricity load of January 1999 using the SOM-SVR, SVR and ML-BPNN techniques.

Results obtained in Table 6 indicate that the prediction accuracy of the ML-BPNN is not satisfactory. This is due to the problems of local minima and over-fitting associated with ANNs, which tends to decrease the generalization performance for unseen data. Our proposed SOM-SVR model proves to be superior in terms of the MAPE compared to the SVR model. This is due to the presence of the SOM in our model, which fits the clustered training data into the appropriate \(\varepsilon\)-SVRs based on the Euclidean distance.

Table 6 Comparison of the forecasting accuracy using different prediction techniques.

<table>
<thead>
<tr>
<th>Method</th>
<th>MAPE</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML-BPNN</td>
<td>3.31%</td>
<td>Comparison technique</td>
</tr>
<tr>
<td>SVR</td>
<td>1.89%</td>
<td>Comparison technique</td>
</tr>
<tr>
<td>SOM-SVR</td>
<td>0.97%</td>
<td>Proposed technique</td>
</tr>
</tbody>
</table>
Fig. 11. Comparison plot of predicted electricity load for January 1999 using SOM-SVR, SVR, and ML-BPNN techniques — EUNITE data.

To further evaluate performance of our proposed model, results from the 2001 EUNITE competition were compared with that of our proposed model. Table 7 compares the results of our proposed SOM-SVR model with the results of the winner of the competition. The results in Table 7 indicate that our proposed forecasting model outperforms the winning model in the EUNITE competition, by achieving a lower MAPE and a lower MAX. The winning model was proposed by Lin et al. in [20, 111], as shown in Figure 12 [25]. Figure 12 indicates the results of the 2001 EUNITE competition. Complete reports of the competition can be found in [25].

Fig. 12. Results of the 2001 EUNITE competition indicating MAPE and the MAX metrics [25]. Winning model was proposed by Lin et al. in [20].
Table 7  Comparison of our proposed SOM-SVR forecasting model with the 2001 EUNITE competition winner.

<table>
<thead>
<tr>
<th>Method</th>
<th>MAPE</th>
<th>MAX</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVR</td>
<td>1.95%</td>
<td>51.42</td>
<td>[20], [111]</td>
</tr>
<tr>
<td>SOM-SVR</td>
<td>0.97%</td>
<td>13.50</td>
<td>Proposed method</td>
</tr>
</tbody>
</table>

For the purpose of a comparative study, recently proposed load forecasting techniques as reported by authors on the 2001 EUNITE data are indicated with their results in Table 8. From the tabulated, it is indicated that our proposed SOM-SVR load forecasting model outperforms other load forecasting approaches in [73,76] and [112–121] with promising results. The results obtained indicate that, better forecasting accuracy is possible when the training data is clustered and a distance measure is used for the selecting a non-linear regression model to fit the testing samples based on similar properties. A comparison plot of the actual and predicted electricity load for January 1999 is given in Figure 13.

Table 8  Comparison of proposed forecasting model with recently proposed load forecasting techniques using the EUNITE data.

<table>
<thead>
<tr>
<th>No.</th>
<th>Method</th>
<th>MAPE</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Locally Linear Model Tree</td>
<td>1.98%</td>
<td>[121]</td>
</tr>
<tr>
<td>2.</td>
<td>SVM-GA</td>
<td>1.93%</td>
<td>[112]</td>
</tr>
<tr>
<td>3.</td>
<td>Autonomous ANN</td>
<td>1.75%</td>
<td>[118]</td>
</tr>
<tr>
<td>4.</td>
<td>Floating Search + SVM</td>
<td>1.70%</td>
<td>[76]</td>
</tr>
<tr>
<td>5.</td>
<td>MLP-NN + Levenberg–Marquardt</td>
<td>1.60%</td>
<td>[114]</td>
</tr>
<tr>
<td>6.</td>
<td>Feedforward and Feedback ANN</td>
<td>1.58%</td>
<td>[116]</td>
</tr>
<tr>
<td>7.</td>
<td>Auto-Regressive Recurrent ANN</td>
<td>1.57%</td>
<td>[115]</td>
</tr>
<tr>
<td>8.</td>
<td>Local Prediction Framework + SVR</td>
<td>1.52%</td>
<td>[117]</td>
</tr>
<tr>
<td>9.</td>
<td>Feedforward ANN</td>
<td>1.42%</td>
<td>[113]</td>
</tr>
<tr>
<td>10.</td>
<td>SOFNN + Bilevel Optimization</td>
<td>1.40%</td>
<td>[119]</td>
</tr>
<tr>
<td>11.</td>
<td>LS-SVM + Chaos Theory</td>
<td>1.10%</td>
<td>[73]</td>
</tr>
<tr>
<td>12.</td>
<td>DLS-SVM</td>
<td>1.08%</td>
<td>[120]</td>
</tr>
<tr>
<td></td>
<td>SOM-SVR</td>
<td>0.97%</td>
<td>Proposed model</td>
</tr>
</tbody>
</table>
Fig. 13. Comparison of the actual and predicted electricity load for January 1999 using the proposed SOM-SVR model — EUNITE data.

7.4. Model Evaluation Using Malaysian Data

To benchmark the performance of our proposed SOM-SVR load forecasting model, the load data for peninsular Malaysia was obtained from Tenaga Nasional Berhad (TNB), Malaysia’s sole electric utility company. The power data provided by TNB contains the half hourly consumption over a period of two years, i.e., from January 1, 2003 until December 31, 2004. The daily peak load from the two year historical Malaysian data was calculated in similar context to the EUNITE data, as illustrated in Section 3.1. The daily peak load profile of the two year historical data is shown in Figure 14.

Fig. 14. Daily peak load from January 1, 2003 until December 31, 2004. Load data represents a period of two years — Malaysian data.
Similarly as with the EUNITE data in Section 3.2.1, load periodicity for the Malaysian data exists in the profile of every week, i.e., the load demand on weekends is usually lower than that on weekdays and the peak load usually occurs during the middle of the week, as indicated in Figure 15.

![Figure 15. Daily peak load for the month of January 2003 and January 2004. Load data for January represents a period of 31 days — Malaysian data.](image)

Using the Malaysian power load data for the two years of 2003 and 2004, we apply a similar procedure as with the case of the EUNITE data, i.e. to predict the daily peak electricity load for the 31 days of January 2005. Our proposed SOM-SVR model requires additional attributes other than the load data, such as temperature, type of day and annual holiday information. The type of day and Malaysian annual/public holidays information for the two years was retrieved from online calendar systems.

The data for the daily temperature of peninsular Malaysia for the two years of 2003 and 2004 is not available. As peninsular Malaysia is positioned in the equatorial zone, this guarantees it a classic tropical climate with relative humidity levels. The weather in Malaysia is fairly hot, averaging around 30°C (86°F) throughout the year. Due to the unavailability of Malaysian daily temperature data and considering temperature (climate) stability in Malaysia throughout the year, we have considered to omit the temperature information from the model attributes as it will not have a significant impact the forecasting accuracy. Hence, the remaining attributes selected to build the SOM-SVR model for the Malaysian power load data are:

\[
\text{(peak load, type of day, annual holiday)}
\]  

(28)
Fig. 16. Comparison of the actual and predicted electricity load for January 2005 using the proposed SOM-SVR model — Malaysian data.

Our proposed SOM-SVR model using the feature attributes in eq. (28) was evaluated using the Malaysian load data with the RBF kernel and Grid Search method indicated in Section 6.7. The best value of MAPE for the proposed model using the Malaysian power data was found to be 1.04% using $\varepsilon = 0.2$. To obtain this lowest value of MAPE, the SVR parameters are: $C_1 = 541.65$, $\gamma_1 = 0.0134$, $C_2 = 1452.78$ and $\gamma_2 = 0.00034$ for $\varepsilon$-SVR1 and $\varepsilon$-SVR2 respectively. This significantly low value of MAPE supports our interest in justifying that our proposed model has the capability to produce good prediction results for MTLF problems.

7.5. Model Evaluation Using PJM Data

PJM which operates an electricity transmission and distribution system is part of the eastern interconnection grid of the United States of America, which is the world’s largest competitive wholesale electricity market [122]. PJM considers forecasts of load growth and additions of demand response as interconnection requests for new and planned retirements of existing generating plants, and solutions to mitigate congestion on the transmission and distribution system.

The load data for the PJM Mid-Atlantic Region (PJM-E) was obtained online from the PJM website [123]. The online PJM power data contains the hourly consumption load from 1998 until the present day. For the case of experiments in this paper, we acquired the latest PJM-E power data over a period of two years, i.e. from January 1, 2009 until December 31, 2010. The daily peak load from the two year historical PJM-E data was calculated in similar context to the EUNITE data, as indicated in Section 3.1. The daily peak load profile of the two year historical data is shown in Figure 17. As observed from Figure 17, the daily peak load for the year of 2010 indicates a significant growth from the year of 2009.
Fig. 17. Daily peak load from January 1, 2009 until December 31, 2010. Load data represents a period of two years — PJM-E data.

Similarly as with the EUNITE data (see Figure 4) and the Malaysian data (see Figure 15), a form of load periodicity for the PJM-E data exists in the profile of every week, as shown in Figure 18, which is discussed in Section 3.2.1 earlier.

The PJM-E power load data in Figure 17 for the two years of 2009 and 2010 is used with the proposed SOM-SVR model in Figure 8 to predict the daily peak electricity load for the 31 days of January 2011, similar to the case of the Malaysian and EUNITE datasets. Since our SOM-SVR model requires additional attributes such as temperature, type of day and annual holiday information, the type of day and annual/public United States holiday information for the two years was retrieved from online calendar systems.

Fig. 18. Daily peak load for the month of January 2009 and January 2010. Load data for January represents a period of 31 days — PJM-E data.
The daily temperature data of the United States is available for the years of 2009 and 2010. However, since the PJM eastern interconnection grid serves a large area of eastern United States, i.e. from Michigan and Pennsylvania in the north to Tennessee in the south, the temperature and climate between in the northern and central parts of the United States have significant differences. Moreover, the PJM interconnection areas near the coast lines, i.e. from and New Jersey and Delaware in the north to North Carolina in the south, have significant temperature changes as compared to the non-coastal areas. Due to these limitations, we have considered omitting the temperature information from the model attributes. So, the remaining attributes selected to build the SOM-SVR model for the PJM-E power load data are the same as for the Malaysian data, as represented by eq. (28).

![Figure 19](image)

**Fig. 19.** Comparison of the actual and predicted electricity load for January 2011 using the proposed SOM-SVR model — PJM-E data.

Our proposed SOM-SVR model using the feature attributes in eq. (28) was evaluated using the PJM-E data with the RBF kernel and Grid Search method illustrated in Section 6.7. The best value of MAPE for the proposed model using the PJM-E power data was found to be 1.07% using $\varepsilon = 0.15$. To obtain this lowest value of MAPE, the SVR parameters are: $C_1 = 372.86$, $\gamma_1 = 0.00682$, $C_2 = 125.64$ and $\gamma_2 = 0.0653$ for $\varepsilon$-SVR1 and $\varepsilon$-SVR2 respectively. This low value of MAPE supports our interest in justifying that our proposed model gives can produce good prediction results for MTLF problems.

**Table 9** Comparison of our SOM-SVR model’s prediction results using different electricity load datasets.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Training Data</th>
<th>Data to Predict</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern Slovakian Electricity Corp. (EUNITE)</td>
<td>1997, 1998</td>
<td>Jan. 1999</td>
<td>0.97%</td>
</tr>
<tr>
<td>Peninsular Malaysia Power Load</td>
<td>2003, 2004</td>
<td>Jan. 2005</td>
<td>1.04%</td>
</tr>
<tr>
<td>PJM Mid-Atlantic Region (PJM-E)</td>
<td>2009, 2010</td>
<td>Jan. 2011</td>
<td>1.07%</td>
</tr>
<tr>
<td><strong>Average MAPE</strong></td>
<td><strong>-</strong></td>
<td><strong>-</strong></td>
<td><strong>1.02%</strong></td>
</tr>
</tbody>
</table>
From the results obtained in Section 7.3, 7.4 and 7.5, Table 9 summaries the results of our proposed SOM-SVR model, which indicates that our proposed model gives an average MAPE of approximately 1% after testing with three different electricity load datasets. Since our proposed model is a simple implementation for MTLF applications, i.e., by fixing two centres of the SOM and without spending much time on optimization of unknown SOM and SVR parameters, our proposed model has shown to achieve relatively good prediction results for MTLF.

8. Conclusion

Mid-term load forecasting (MTLF) is a newer aspect of power system load forecasting. Different kinds of forecasting such as prediction of the daily peak load, daily energy consumption, and monthly electricity consumption, and annual peak load have been proposed previously by researchers. This paper focuses on daily peak load prediction for the next month.

In this paper a computational intelligence scheme based the SOM and SVR is applied to reconstruct the dynamics of electricity load forecasting using a time series approach. The SOM is used as a clustering tool to cluster the training data into two subsets, using the Kohonen rule. The reason for using two subsets is obtained from the fact that, two years of historical data is used to construct the forecasting model. After clustering is complete, two ε-SVRs (one for each subset) are employed to fit the clustered data appropriately for SVR training. The Euclidean distance between the cluster centres and testing samples is used as a measure to select the appropriate ε-SVR model for predicting the peak daily load. The proposed SOM-SVR technique is applied on the EUNITE competition data [25] to predict the month-ahead electricity load of January 1999, which demonstrates the effectiveness and efficiency of the clustering and prediction technique in contrast to others.

In order to evaluate the performance of our forecasting technique, power load data obtained from (i) Tenaga Nasional Berhad (TNB) for peninsular Malaysia and (ii) PJM for the eastern interconnection grid of the United States of America is used to benchmark the performance of the proposed SOM-SVR model.

The empirical results obtained reveal the capability of the proposed model for load prediction. Experimental results reveal that by introducing the load data as a time series and using calendar attributes (type of day and annual holidays) in the forecasting model, the future load demand can be predicted more accurately. Other prediction techniques using the SVR and the ML-BPNN were used to compare the prediction accuracy with the proposed model. Results obtained indicated that the proposed SOM-SVR model outperforms the other two approaches in terms of the MAPE. Furthermore, comparison with the winning model of the EUNITE competition proposed by Lin et al. in [20,111] and other load forecasting techniques using the EUNITE data in [73,76] and [112–121] prove that our proposed forecasting model has better prediction accuracy compared to the other approaches. Experimental results obtained from testing the SOM-SVR model with the EUNITE, Malaysian and PJM datasets achieves low MAPE values of 0.97%, 1.04% and 1.07% respectively, giving an average MAPE of approximately 1%,
which indicates that our proposed model gives significantly good prediction accuracy for MTLF compared to previously researched findings.

Although forecasting of electricity load has reached a comfortable state of performance with errors of 1-3% in recent years [8], this problem is still considered a difficult task that comprehensive and general solutions are far from easy. For a specific system, the best performances can only be achieved if a deep investigation of the inherent characteristics of the system has been carried out, as in our proposed work.

Acknowledgements

The authors would like to thank EUNITE for bringing the problem of MTLF to our attention, which inspires us to conduct research on load forecasting. We would like to express our gratitude the Eastern Slovakian Electricity Corporation and PJM for providing the data to be useful in this research study. In addition, we would like to thank Tenaga Nasional Berhad (TNB) Malaysia, for providing the power load data of peninsular Malaysia in order to support this research study. Lastly, the authors would like to thank the editor and reviewers for their careful review of the paper and their constructive remarks and suggestions.

References


