Electrical Load Forecasting Using Support Vector Machines

Belgin Emre Türkay Dilara Demren

Istanbul Technical University
Istanbul, Turkey
demren@itu.edu.tr, turkay@elk.itu.edu.tr

Abstract

In this study, an application with electrical load forecasting - an important topic in the electrical industry - has been carried out by a machine learning method which has recently become popular: Support Vector Machines (SVM). Load forecasting with SVM can model the nonlinear relations with the factors that affect the load in addition to the accurate modelling of the load curve at the weekends and on important calendar days. The data gathered from the Istanbul European Side are used as a sample for the application. In addition to the past load data, daily average temperature, calendar days, holidays and electricity price are considered as an attribute in forecasting. The programme LibSVM is used for modelling the system. It is noted that SVM gave satisfactory results.

1. Introduction

In the power industry, the first step towards making the right decisions is an accurate load forecasting. The electricity demand should be known in order to make profitable investments, build efficient systems, increase the capability of existing systems, schedule energy distribution, etc.

During forecasting, an underestimation in energy demand may result in limited supply of electricity at the consumer end, which leads to energy quality reduction in system reliability. On the other hand, an overestimation may cause unnecessary investments or establishments which run under-capacity and therefore result in uneconomic operating conditions.

Studies about accurate load forecasting gained importance in the late 1960s. However since then, the main focus in load forecasting has been on short term load forecasting field since it is an important tool in the daily operation of utility systems. More recently, with the deregulation of energy markets, more and more attention has also been paid to load forecasts with longer prediction ranges, such as medium-term load forecasts. Medium-term load forecasts are used in scheduling maintenance and fuel supply, as well as in small infrastructure changes. Also, as indicated in [1], medium-term load forecasts enable companies to estimate the load demand for a longer time interval which, for example, helps them in the negotiation of contracts with other companies. Especially in deregulated markets, any deviation from the actual load demand may lead to significant economic costs. Therefore, it is important to make accurate forecasts in order to avoid these issues.

As some of the published survey papers [2-7] indicate, the electrical load forecasting is a quite popular research area. There are two main approaches in this field: the traditional statistical approaches that model the relation between the load and the factors that affect the load (such as time series and regression analysis, etc.) and artificial and computational intelligence approaches. Statistical methods assume the load data follow a pattern and try to forecast the value of the future load by employing different time series analysis techniques. Intelligent systems are derived from the mathematical expressions of human behaviours/ experiences. Especially since the early 1990s, Neural Networks has been considered as one of the most commonly used techniques in the electrical load forecasting field, as it assumes there is a nonlinear function that relates past values and some external variables to future values that may affect the output [2]. The approximation capability of Neural Networks has made it convenient for common usage. As stated in [8], a software based on the Artificial Neural Networks method was used by several electric utilities in the US and Canada for hourly short-term load forecasting.

In recent years, one computational intelligence method involving Support Vector Machines has become commonly used in the electrical load forecasting field. For instance, Chen et al. [9] used a Support Vector Regression technique to solve an electrical load forecasting problem, which was a competition organized by EUNITE network (European Network on Intelligent Technologies for Smart Adaptive Systems). Their approach in fact won the competition. In addition to that, Mohandes [10] used the Support Vector Regression model for short term electrical load forecasting and compared the results with the autoregressive (AR) model. As a result, the performance of the SVMs was much lower than the AR method. The researches [11-12] showed that, with its superior generalization capability, Support Vector Machines are successful in forecasting applications.

In this study, Support Vector Machines are used for the daily peak load forecasting of a monthly period. The main reason for using SVM in solving a medium-term load forecasting problem is that it can easily model the load curve; the relation between the load and the dynamics which change the load demand, such as calendar days, special days (holidays, festive, etc.), temperature, economic and demographic factors.

In this study, the real electrical load values, temperature and electricity price, as an economic factor, between 2006 and 2009, are used to predict April 2010 daily peak load values using SVM.

2. Support Vector Machines

SVM is a powerful technique used in solving main learning problems. It is derived from the Statistical Learning Theory of the 1990s. Basically, it was first used in pattern recognition and classification problems, but it is modified to be used in regression problems.

Generally speaking, for learning problems in the case of regression, training data is given to the learner in order to find out the relation (correlation, mapping or function) between inputs and outputs of a function $f(\mathbf{x})$. Training set;

$$D = \{[\mathbf{x}_i, y_i] \in \Re^n \times \Re, i = 1, \dots, l\} \quad (1)$$

consists of input vectors $\mathbf{x}$, $\mathbf{x} \in \Re^n$, and outputs y, y∈R, which constructs $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_l, y_l)$ pairs.

The main idea of the support vector machines in regression is that the inputs $\mathbf{x}_i$ needs to be mapped into a higher dimensional feature space with the function $\boldsymbol{\phi}$ [13].

$$\boldsymbol{\phi}(\mathbf{x}): \Re^n \rightarrow \Re^f$$
$$\mathbf{x} \in \Re^n \rightarrow \boldsymbol{\phi}(\mathbf{x}) = [\phi_1(\mathbf{x})\ \phi_2(\mathbf{x})\ \dots \phi_n(\mathbf{x})]^T \in \Re^f \quad (2)$$

After the inputs are mapped into higher dimensional space, the problem converts from a nonlinear regression case into a linear regression case, which enables the quadratic programming formulation of the problem. The SVM considers approximating functions, in this case regression hyperplane, of the form [14];

$$f(\mathbf{x},\mathbf{w}) = \mathrm{w}^{\mathbf{T}} \boldsymbol{\phi}(\mathbf{x}) + b \quad (3)$$

$\boldsymbol{\phi}(\mathbf{x})$ mapping function is an advanced defined function. In formulation, parameters $\mathbf{w}$ and b can be found by minimizing expression of the error function R[14];

$$R = \frac{1}{2}\|\mathbf{w}\|^2 + C\left(\sum_{i=1}^{l} |y_i - f(\mathbf{x}_i,\mathbf{w})|_\varepsilon\right) \quad (4)$$

where,

$$|y - f(\mathbf{x},\mathbf{w})|_\varepsilon = \begin{cases} 0 & if\ |y - f(\mathbf{x},\mathbf{w})| \le \varepsilon \\ |y - f(\mathbf{x},\mathbf{w})| - \varepsilon & if\ |y - f(\mathbf{x},\mathbf{w})| > \varepsilon \end{cases} \quad (4)$$

Here, $\|\mathbf{w}\|^2$ is the norm of the weight vector that defines the capacity of the model for an optimum generalization. The term ε, represents the Vapnik's ε-insensitive loss function which defines the ε –zone as shown in Fig. 1. If the forecasted value is within the tube, the approximation error will equal zero. For all other values that lie outside of the tube, the loss is equal to the distance between the data point and the radius of the ε-insensitive tube.

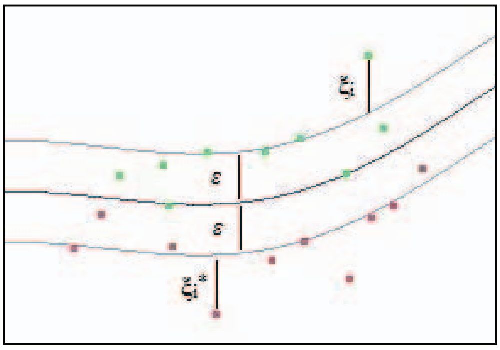

Fig. 1. A Simple Example of Support Vector Regression

The penalty parameter C determines the trade-off between an approximation error and the weights vector norm. An increase in C penalizes larger errors (large ξ and ξ^*) and in this way leads to a decrease in approximation error. However, this can be achieved only by increasing the weights vector norm. At the same time an increase in $\|\mathbf{w}\|$ does not guarantee good generalization performance of a model [14].

As can be seen from the Fig. 1, the training data points that lie outside of the ε-insensitive tube can be written as [14];

$|y - f(\mathbf{x},\mathbf{w})| - \varepsilon = \xi$ for the points above the tube
$|y - f(\mathbf{x},\mathbf{w})| - \varepsilon = \xi^*$ for the points below the tube

where ξ and ξ* are slack variables which are positive values. Substituting the equations above into the R [14];

$$R_{\mathrm{w},\xi,\xi^*} = \left[\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i{}^*)\right] \quad (6)$$

subject to the constraints;

$$\begin{aligned} y_i - \mathbf{w}^T\mathbf{x}_i - b &\le \varepsilon + \xi_i & i = 1, \dots, l \\ \mathbf{w}^T\mathbf{x}_i + b - y_i &\le \varepsilon + \xi_i{}^* & i = 1, \dots, l \\ \xi &\ge 0 \\ \xi^* &\ge 0 \end{aligned} \quad (7)$$

The parameters of the optimal hyperplane $f(\mathbf{x},\mathbf{w})$ can be found with the help of Lagrangian multipliers;

$$\mathbf{w} = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\boldsymbol{\phi}(\mathbf{x}_i) \quad (8)$$

Thus,

$$\begin{aligned} f(\mathbf{x},\mathbf{w}) &= \mathbf{w}^{\mathbf{T}}\boldsymbol{\phi}(\mathrm{x}) + b \\ &= \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\boldsymbol{\phi}(\mathbf{x}_i)\boldsymbol{\phi}(\mathbf{x}) + b \\ &= \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)K(\mathbf{x}_i, \mathbf{x}_j) + b \end{aligned} \quad (9)$$

Here, assigning $K(\mathbf{x}_i, \mathbf{x}_j) = \boldsymbol{\phi}(\mathrm{x}_i)\boldsymbol{\phi}(\mathrm{x}_j)$ is known as the Kernel Trick. These predefined kernel function make the calculations easier as they avoid the high dimensionality of the feature space.

By maximizing the dual Lagrangian, optimal hyperplane can be obtained. To achieve this Karush-Kuhn-Tucker (KKT) conditions for regression are applied [14];

$$\text{Maximize}\ \ L_d(\boldsymbol{\alpha}, \boldsymbol{\alpha}^*) = -\varepsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)y_i - \frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(\mathbf{x}_i, \mathbf{x}_j) \quad (10)$$

Subject to the constraints

$$\begin{aligned} C \ge \alpha_i, \alpha_i^* \ge 0\,, \quad i = 1, \dots, l \\ \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \end{aligned} \quad (11)$$

The parameter b can be found by using the Karush-Kuhn Tucker complementarity conditions [15].

The performance of the regression depends on the parameters the cost of the error C, the width of the tube ε and the chosen kernel function.

Support vector regression can come up with a more generalized solution to the problem than the traditional regression approach by mapping the data into a higher dimensional feature space and setting a width (ε) for the tube.

In this electrical load forecasting study, the programme LibSVM is used for SVM modelling.

3. Data Sets

The relation between load demand and time, meteorological conditions, economic-demographic and random factors are vital for an accurate prediction of an electrical load demand. These factors should be taken into account when determining the data that will be used in load forecasting.

In this study, the data set, which is used in the forecasting process, consists of daily past peak electrical load values, daily average temperature values, calendar days to determine whether it is a weekday/weekend/ holiday and electricity price to project the economic course.

3. 1. Electrical Data

Past daily peak electrical load values gathered from the Istanbul European Side between the years 2006 and 2009 are used in this study. Daily load demand curve is given in Fig.2.

The Istanbul European Side contains both industrial and residential areas. As a result, there is a small difference in respect to electrical load demand between weekdays and weekends (or holidays, special days, etc.). Nevertheless, load demand on Saturdays and Sundays are slightly less than it on weekdays. Fig. 3 clearly shows this distinction.

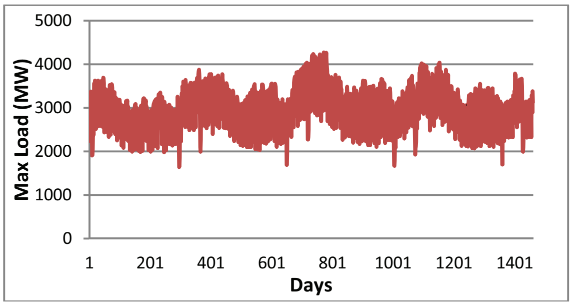

Fig. 2. Daily Peak Load

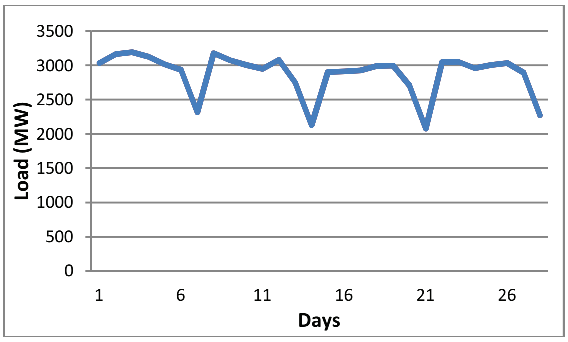

Fig 3. Sample Load Curve from April 2006

3.2. Temperature Values

Temperature is one of the most important factors influencing load demand. Thus, it is included to the data set of this study.

In Fig. 4, the daily average temperature curve between 2006 and 2009 is shown.

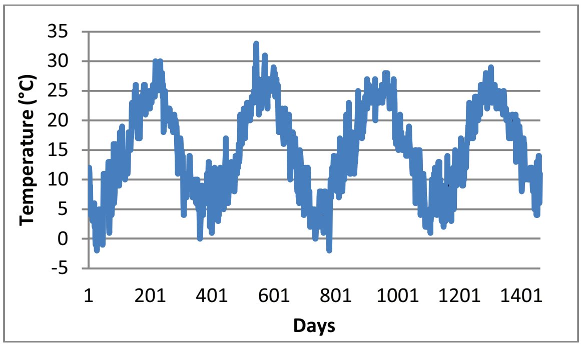

Fig. 4. Daily Average Temperature between the years 2006-2009

It can be observed that the seasonal effect of the temperature is influential on the electrical load demand by comparing Fig. 2 and Fig. 4.

3.3. Calendar Days

Previous studies show that weekends, holidays, special days, such as religious holidays etc. have generally lower load demand than normal days. Thus, these days are considered as a feature in the data set as they may affect the forecasting process.

3.4. Economical Factors

Especially in mid and long term forecasting applications, it is important to consider economic and demographic factors as well as past load values. In this study, electricity price is regarded as an economical factor. Table 1 shows the course of the electricity price in TL/ kWh over April, May and June 2006 through 2010.

Table 1. Electricity Price of Istanbul

Years*	Electricity (TL/kWh)
2006	0.16
2007	0.16
2008	0.19
2009	0.25
2010	0.27

* Only April-May-June

4. Implementation

The SVM training set is constructed from seventeen inputs using past load values, temperature values, calendar days that indicate weekdays/weekends/holidays and electricity price as follows;

- Seven inputs to project the values of electrical loads of the past seven days

- Seven inputs that indicate the day
- One input that indicates whether it is a holiday or not
- One input for the average daily temperature
- One input for the electricity price

The reason for choosing the interval "past seven days" is that the electrical load input accurately projects the periodicity of the load demand. After the related day is predicted, the resulting load is added to the next day's past seven loads and the data set is prepared accordingly. For instance, when predicting April 1, the load values of March 25 through 31 are taken as past load values and a value for April 1 is obtained. Similarly, the value of April 1 is used in the past load values of April 2 and the prediction is repeated until the April 30's load values is found.

The data set includes the load values, temperature values, calendar days and electricity price of April-May and June of the years 2006 through 2009. Thus, the SVM data set is made of the patterns which have the most similar characteristics to the month to be predicted.

The established model has a single output, which is the predicted load. Here, scaling of the input variables is one of the key points of building the SVM model. Scaling of the inputs is achieved by the in-built scaling tool of the programme LibSVM. Therefore, all the inputs are scaled in a range of [0 1].

The predicted electrical load values of April 2010 are compared to the real values of this month. In order to measure the difference as an error metric, Mean Absolute Percentage Error (MAPE) is used.

$$MAPE = \frac{100}{\mathrm{T}} \sum_{\mathrm{t}=1}^{\mathrm{T}} \left| \frac{y_{\mathrm{i}} - \hat{y}_{\mathrm{i}}}{y_{\mathrm{i}}} \right| \qquad (12)$$

In the formula above y_{i} represents the real value whereas $\hat{y}_{\mathrm{i}}$ represents the predicted value [3]. The value T corresponds to the total number of values, in our case this is the total number of predicted days.

5. Results

In SVM modelling, the values of the cost of the error C, the width of the tube ε and mapping function ϕ must be defined. The value of ε -insensitive tube width is defined as 0.5. Radial Basis Function – RBF, one of the most common mapping functions, is chosen as the kernel function. [13-14]

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j) = e^{-\gamma \|x_i - x_j\|^2} \qquad (13)$$

When building the SVM model, 75% of the data is used as training data while the remaining 25% is used for testing data. γ kernel parameter in function (13) and the cost of the error C in regression function is determined using the cross-validation tool in LibSVM. In cross-validation, the training data is divided into random subsets of requested sizes (2.5, 5, 10, etc.) Each subset is used as a test set while the rest of the data is used as a training set. The C ve γ values of the best results of the 5-fold cross validation tests are used to train the training data.

The comparison of the estimated and real values for April, 2010 can be seen in Fig. 5.

The error percentage between the values estimated by SVM training and real values is;

MAPE= % 3.67

Despite the Istanbul European Side's dynamic nature, the SVM achieved a good performance by simply using past load values, average daily temperature and electricity price as an economic factor, especially in respect to projecting the periodicity of the load demand curve.

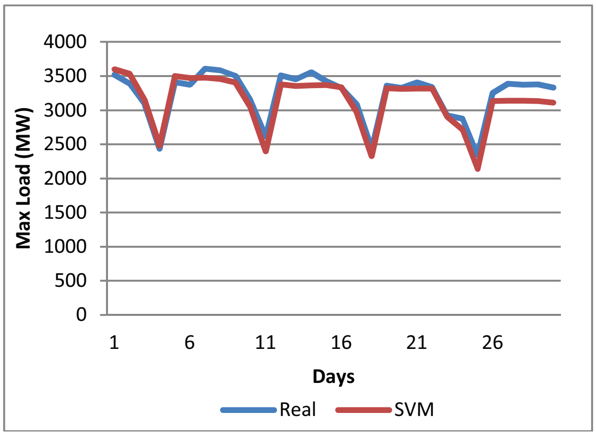

Fig. 5. The comparison of the estimated and real values for April, 2010

6. Conclusion

In this study, the maximum daily load for the Istanbul European Side of April 2010 is estimated with Support Vector Machines using the real maximum daily load, temperature and holiday period values from 2006 to 2009. For this purpose, the programme LibSVM is used. Radial Basis Function, which is one of the most commonly used kernel functions in nonlinear estimation problems, is used in the mapping of input values into higher dimensional feature space. The cost of the error C in regression problem and the γ parameter in the radial basis function is determined using a 5-fold cross validation. The obtained values are used for training the training data and estimating the peak daily load for April 2010. The reliability of established models using SVM is tested using the real values.

The predicted values are clearly projected by the periodicity of the load demand course, which is decreasing during the weekends and increasing during the weekdays. It can be said that when compared to the previous studies on the same field, even with the special days or holidays included in the data set, SVM is showing much more promise as it offers an optimal solution independent from the model, with lesser independent variables and ease of calculations. The fact that variables are determined using trial and error methods and the time consuming nature of model building process are two disadvantages of SVM.

To decrease the error percentage results estimated using SVM training, values of the cost of the error C, ε -insensitive tube width, the type of kernel and the parameters inside the kernel functions must be chosen very diligently. In addition to all these factors, as the size and the consistency of the training set reflects directly on the estimated numbers, determination of these values is of great importance.

7. References

[1] E. Gonzalez-Romera, M.A. Jaramillo-Moran, D. Carmona-Fernandez, "Monthly electric energy demand

forecasting based on trend extraction". *IEEE Transactions on Power Systems*, vol. 21, no. 4, pp. 1946–1953, November 2006.
[2] P.R.J. Campbell, K. Adamson, "Methodologies For Load Forecasting", *3rd International IEEE Conference on Intelligent Systems*, London, England, 2006, pp. 800-806.
[3] H. Hahn, S. Meyer-Nieberg, S. Pickl, "Electric Load Forecasting Methods: Tools For Decision Making.", *European Journal of Operational Research,* vol. 199, no. 3, pp. 902-907, December 2009.
[4] N. Sapankevych, R. Sankar, "Time Series Prediction Using Support Vector Machines: A Survey.", *IEEE Computational Intelligence Magazine*, vol.4, no. 2, pp. 24-38, May 2009.
[5] H.S. Hippert, C.E. Pedreira, R.C. Souza, "Neural Networks For Short-Term Load Forecasting: A Review And Evaluation.", *IEEE Transactions on Power Systems*, vol. 16, no. 1, pp.44-55, February 2001.
[6] E. Kyriakides, M. Polycarpou, "Short Term Electric Load Forecasting: A Tutorial.", in: *Trends in Neural Computation, Studies in Computational Intelligence*, vol. 35. , chapter 16, K. Chen, L. Wang (Eds.), Springer 2007, pp. 391–418.
[7] S. Tzafestas, E. Tzafestas, "Computational Intelligence Techniques For Short-Term Electric Load Forecasting.", *Journal of Intelligent and Robotic Systems*, vol. 31, pp. 7–68, January 2001.
[8] A. Khotanzad, R. Afkhami-Rohani, D. Maratukulam, "ANNSTLF-Artificial Neural Network Short-Term Load Forecaster generation three", *IEEE Transactions on Power Systems*, vol. 13, no. 4, pp. 1413 - 1422, November 1998.
[9] B.-J. Chen, M.-W. Chang, C.-J. Lin, "Load Forecasting Using Support Vector Machines: A Study On Eunite Competition 2001", *IEEE Transactions on Power Systems*, vol. 19, no. 4, pp. 1821-1830, November 2004.
[10] M. Mohandes, "Support vector machines for short-term load forecasting," Int. J. Energy Res., vol. 26, no. 4, pp. 335–345, March 2002.
[11] X.-C. Guo, Y.-C. Liang, C.-G. Wu, H.-Y. Wang, "Electric Load Forecasting Using SVMs", *International Conference on Machine Learning and Cybernetics*, Dalian, China, 2006, pp. 4213 – 4215.
[12] F. Rivieccio, "SVM for an electricity load forecast problem", In: Electricity Load Forcast Using Intelligent Technologies P. Sincak, J. Strackeljan, M. Kolcun, D. Novotný, and P. Szathmáry (Eds), EUNITE: The European Network on Intelligent Technologies for Smart Adaptive Systems, 2002, pp.177-186.
[13] B. Schölkopf, C.J.C. Burges, A.J. Smola, "Advances in Kernel Methods: Support Vector Learning", The MIT Press, Cambridge, London, England, 1999.
[14] V. Kecman, "Learning and Soft Computing: Support Vector Machines, Neural Networks and Fuzzy Logic Models", MIT Press, Cambridge, Massachusetts, England, 2001.
[15] A.J. Smola, B. Schölkopf, "A Tutorial On Support Vector Regression", *NeuroCOLT Technical Report*, NC2-TR-98-030, 1998.
[16] C.-C. Chang, C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm.