Applied Energy 88 (2011) 368–375

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Modeling and prediction of Turkey's electricity consumption using Support Vector Regression

Kadir Kavaklioglu *

Pamukkale University, Computer Science Department, Denizli, Turkey

ARTICLE INFO

Article history:
Received 21 February 2010
Received in revised form 9 July 2010
Accepted 24 July 2010
Available online 30 August 2010

Keywords:
Electricity consumption
Support Vector Regression
Turkey
Energy modeling
Time series
Prediction

ABSTRACT

Support Vector Regression (SVR) methodology is used to model and predict Turkey's electricity consumption. Among various SVR formalisms, ε-SVR method was used since the training pattern set was relatively small. Electricity consumption is modeled as a function of socio-economic indicators such as population, Gross National Product, imports and exports. In order to facilitate future predictions of electricity consumption, a separate SVR model was created for each of the input variables using their current and past values; and these models were combined to yield consumption prediction values. A grid search for the model parameters was performed to find the best ε-SVR model for each variable based on Root Mean Square Error. Electricity consumption of Turkey is predicted until 2026 using data from 1975 to 2006. The results show that electricity consumption can be modeled using Support Vector Regression and the models can be used to predict future electricity consumption.

1. Introduction

Electricity is one of the main forms of energy that modern life is built upon. It affects a society's quality of living, efficiency and quality of its work, manufacturing and competitiveness in an ever-growing global world. Governments and their related branches in developed and developing countries put major emphasis in modeling and predicting electricity consumption. Forecasting errors for electricity consumption would result in either shortages or excess capacity that are highly undesirable from a financial standpoint. Therefore, modeling electricity consumption with a high degree of accuracy becomes vital in order to avoid aforementioned costly mistakes.

As a developing country and an emerging market, Turkey expects an increasing but somewhat volatile behavior in electricity consumption in the future as its economy moves rapidly to respond to expansions and crises of national and global scale. Because of its limited energy resources, Turkey depends on imported energy sources both for electricity production and for general energy use. For instance, imported oil and gas contribute about 62% of Total Primary Energy Supply (TPES) of Turkey and it imported 77% of its energy needs in 2004 according to WEC–TNC report published in 2006 [1]. Due to extreme dependency on foreign fossil fuels, building models for Turkey's electricity consumption and accurately predicting future electricity consumption are essential. Studies on energy prediction in Turkey are done by Ministry of Energy and Natural Resources (MENR) and State Planning Organization (SPO). The MENR uses the Model for Analysis of Energy Demand (MAED) that historically has not produced reliable intermediate to long term results for Turkey. This has motivated the present research to study a different modeling technique for Turkey's electricity consumption forecasting.

A review of the literature shows that there are numerous studies on the relationship among energy consumption and economy, as it inspired the selection of input variables for the present paper, such as Ebohan in 1996 [2], Kavrakoglu in 1983 [3], Say and Yucel in 2006 [4] Uri in 1980 [5] and Yu and Been in 1984 [6]. Total and sectoral energy modeling and prediction studies have been carried out by many researchers. Gilland [7] developed an energy demand projection of the world for the years 2000 and 2020. Ediger and Tatlidil [8] and Ediger and Camdali [9] proposed an approach that uses the analysis of cyclic patterns in historical curves to predict the primary energy demand in Turkey. Yumurtaci and Asmaz [10] proposed an approach to calculate future energy demand of Turkey, for the period of 1980 and 2050, based on the population

Abbreviations: ANN, Artificial Neural Network; GNP, Gross National Product; MAED, Model for Analysis of Energy Demand; MENR, Ministry of Energy and Natural Resources; RMS, Root Mean Square; RMSE, Root Mean Square Error; SPO, State Planning Organization; SVC, Support Vector Classification; SVM, Support Vector Machine; SVR, Support Vector Regression; TEIAS, Turkish Electricity Transmission Company; TPES, Total Primary Energy Supply.

* Address: Pamukkale University, Computer Science Department, 20070 Kinikli, Denizli, Turkey. Tel.: +90 258 296 3329; fax: +90 258 296 3262.

E-mail addresses: kkavaklioglu@pau.edu.tr, kkavaklioglu@hotmail.com

0306-2619/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.apenergy.2010.07.021

and energy consumption increase rates per capita. Sozen et al. [11] used Artificial Neural Networks (ANN) to predict Turkey's net energy consumption. Toksari [12] developed an ant colony energy demand estimation model for Turkey. Akay and Atak [13] proposed an approach using grey prediction with rolling mechanism (GPRM) to predict the Turkey's total and industrial electricity consumption. Sozen and Arcaklioglu [14] developed the energy sources estimation equations in order to estimate the future projections and make correct investments in Turkey using Artificial Neural Network (ANN) approach. Murat and Ceylan [15] obtained that modeling the energy consumption may be carried out with Artificial Neural Networks (ANN) with a lack of future prediction since ANN models they built are good at solving current data, but are not good for prediction since they do not use any mathematical models. Hamzacebi [16] used ANN's with time series structure to predict Turkish electricity consumption.

The goal of accurate modeling of consumption requires attention to a few extremely important points. The first point is to identify all the necessary variables and parameters that contribute to electricity consumption in a given country. In this paper, the ε-SVR models take time, population, Gross National Product (GNP), import and export values as inputs to model and predict total electricity consumption of Turkey. It is obvious that consumption is related to population as the population increases, more electricity will be consumed. Imports and exports for Turkey are related to manufacturing processes and therefore strongly affect electricity consumption. Finally, GNP is a measure of all economic activities and increasing GNP means improved living standards and thus increased energy use.

The second point is to choose a modeling methodology that can handle the difficulties of the consumption modeling task. One difficulty in this area, similar to many other modeling studies, is the fact that the relationship between the input variables and the output variable is nonlinear and the nature of the nonlinearity is not known very well. Therefore it is not easy to postulate the form and/or order of a mathematical function whose parameters could be estimated through regression type computations. Black box models such as Artificial Neural Networks (ANN) could be applied in these situations since they can handle nonlinearities to an arbitrary degree of accuracy [17]. However, although being the most important aspect, accuracy of the model is never the only concern in any modeling analysis. Models should also be able to handle imperfections in the data such as noise, errors, missing data points, disturbances and short term effects. Another area of concern in modeling is so called local minima. Most modeling techniques compute model parameters by minimizing a cost function which may not have a single local minimum that is also the global minimum. Then getting trapped in one of the local minima avoids reaching the real optimal parameters. Therefore a good modeling method should either be able to guarantee reaching global minimum or should have a single global minimum for its cost function. In summary, the models should exhibit good global optimization and robustness characteristics as well as accuracy in prediction.

Furthermore, the methodology used should allow future prediction by letting the model topology cover one-step or multi-step ahead predictions. This is critical since the eventual goal here is to predict the future values of electricity production that leads to dynamic model structures. It is well-known that time series formalism using nonlinear models is very suitable for these types of problems. There are numerous published papers on successful applications of time series to complicated problems [18–21]. In addition, the methodology should allow rangeability as often the data available for modeling does not cover the ranges of data for the times the predictions need to be done. For instance, population range of Turkey is 40.4–74.6 million for 1974–2006 period; however it is expected to be much higher for year 2026. Then the models will see [40.4,74.6] range during training while they will be presented with much higher values during future prediction.

In the light of the preceding argument, Support Vector Regression methodology was proposed to model Turkey's electricity consumption. The literature search did not yield any study where SVR technique is applied to consumption modeling of Turkey, and therefore, the present paper is original and should contribute significantly to the area of electricity demand forecasting. SVR methodology allows multiple inputs so that it can be used to model consumption as a function of the aforementioned socio-economic indicators. SVR based regression models can also handle any nonlinearity since they use kernel functions that could be chosen as a number of different parametric nonlinear functions. SVR models can also handle data imperfections as it allows an adjustable parameter (ε) that the mismatch between the data and the model output is ignored if it is below that parameter. That means the method will not force the model to go through every point exactly which greatly enhances its overall performance with little tradeoff to accuracy. SVR modeling also has excellent global optimization characteristics since the computation of model parameters becomes a convex optimization problem with a single minimum, as there are no other local minima. Last but not the least, SVR models could handle different data ranges properly through use of data preprocessing and normalization but more importantly through parameters of the kernel functions. The main goal of this research is to develop accurate, robust, globally optimal SVR models for electricity consumption of Turkey. Another equally important goal is to build models that could predict consumption for strategic planning. In order to incorporate prediction capabilities to the eventual system, a separate SVR model is built for each of the input variables using their past values as inputs. This results in four one-step-ahead predictor models for population, GNP, imports and exports. Then, a final one-step-ahead predictor model was built for the electricity consumption as output and the past values of the other variables as the input.

2. Support Vector Regression

The foundations of Support Vector Machines (SVM) have been laid by Vapnik and Chervonenkis [22–24] and the methodology is gaining popularity ever since. SVM's that deal with classification problems are called Support Vector Classification (SVC) and SVM's that deal with modeling and prediction are called Support Vector Regression. There are numerous resources in the form of books, reports and papers that give thorough overview of SVM's. An excellent tutorial on SVC has been published by Burges [25] and an excellent tutorial on SVR has been published by Smola and Schölkopf [26]. This paper exclusively uses modeling and prediction and therefore an overview of the SVR technique is given here.- Let us assume one is trying to model a single output (y) as a function of n input variables ($\boldsymbol{x}$) and is given a training data set of length N: $\mathbf{T} = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots (\boldsymbol{x}_N, y_N)\}$ where $x_k \in \Re^n$ and $y_k \in \Re$, $k = 1, 2, \ldots, N$

In essence, x_k's are n dimensional vectors carrying the values of each input at time step k and y_k's are scalars carrying the values of the output variable at time step k. Now the problem becomes finding a model that explains this training set the best. In the original SVR formulation a linear model is proposed:

$$\hat{y}(x) = \langle w, x \rangle + b \tag{1}$$

where $\hat{y}$ is the estimated output of the model, $\boldsymbol{w}$ is a weight vector, b is a bias term and $\langle \cdot, \cdot \rangle$ denotes vector inner product. The vector $\boldsymbol{w}$ is actually an element of the feature space of the problem. However,