Energy and Buildings 126 (2016) 94–103

Contents lists available at ScienceDirect

Energy and Buildings

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

Time series forecasting for building energy consumption using weighted Support Vector Regression with differential evolution optimization technique

Fan Zhang [a], Chirag Deb [b,*], Siew Eang Lee [b], Junjing Yang [b], Kwok Wei Shah [b]

[a] Institute of Systems Science, National University of Singapore, 119615, Singapore
[b] Department of Building, School of Design and Environment, National University of Singapore, 117566, Singapore

A R T I C L E I N F O

Article history:
Received 29 March 2016
Received in revised form 10 May 2016
Accepted 11 May 2016
Available online 12 May 2016

Keywords:
Building energy consumption
Energy forecasting
Support vector regression
Differential evolution algorithm
Institutional building

A B S T R A C T

Electricity load forecasting is crucial for effective operation and management of buildings. Support Vector Regression (SVR) have been successfully used in solving nonlinear regression and time series problems related to building energy consumption forecasting. As the performance of SVR heavily depends on the selection of its parameters, differential evolution (DE) algorithm is employed in this study to solve this problem. The forecasting model is developed using weighted SVR models with nu-SVR and epsilon-SVR. The DE algorithm is again used to determine the weights corresponding to each model. A case of time series energy consumption data from an institutional building in Singapore is used to elucidate the performance of the proposed model. The proposed model can be used to forecast both, half-hourly and daily electricity consumption time series data for the same building. The results show that for half-hourly data, the model exhibits higher weight for nu-SVR, whereas for daily data, a higher weight for epsilon-SVR is observed. The mean absolute percentage error (MAPE) for daily energy consumption data is 5.843 and that for half-hourly energy consumption is 3.767 respectively. A detailed comparison with other evolutionary algorithms show that the proposed model yields higher accuracy for building energy consumption forecasting.

1. Introduction

Energy efficiency in buildings is identified as one of the five measures to secure long-term decarbonisation of the energy sector by the International Energy Agency[1] [1]. Along with environmental benefits, building energy efficiency also presents vast economic opportunities. Buildings with efficient energy systems and management strategies have much lower operating costs. For effective management and operation on a daily basis, accurate forecasting models are of vital importance. These models can set boundary conditions for the building facility managers and owners within which the buildings energy consumption should ideally fall on a daily, weekly, monthly, and annual basis respectively. Such models can also be combined with other computer simulation models like EnergyPlus and can be used to derive operational factors like occupancy etc. [2]. Due to the ability of the forecasting model to learn from previous energy consumption usage patterns, a gradual increase in the forecasted energy consumption values are indicative of the maintenance aspects of the building systems. In addition, accurate load forecasting plays a vital role in cost effective purchase of electricity. Several time series forecasting methods have been employed for this purpose in recent years [3–5]. Artificial Neural Network (ANN) and Support Vector Machines (SVM) have been the most popular due to their capability in modelling non-linear time series. Karatasou et al. [6] discussed the application of ANN in predicting building energy consumption in combination with statistical analysis. Gonzalez and Zamarreno [7] used feedback ANN to predict short term electricity load. Deb et al. [8,9] developed a feed-forward ANN to forecast the cooling load electricity consumption for three institutional buildings in Singapore. Azadeh et al. [10] used a feedforward neural network with back propagation algorithm to forecast long-term electricity consumption in energy intensive manufacturing industries in Iran. Similarly, Hong [11] developed Support Vector Regression (SVR) model with immune algorithm (IA) to forecast a regional annual electric load in Taiwan. Dong et al. [12] applied SVM to predict building energy consumption in

* Corresponding author.
E-mail addresses: chirag.deb@u.nus.edu, chirag.arian@gmail.com (C. Deb).

[1] © OECD/IEA 2015 World Energy Outlook Special Report, IEA Publishing. License: [http://www.iea.org/t&c/termsandconditions/]

http://dx.doi.org/10.1016/j.enbuild.2016.05.028

the tropical region where weather data, including monthly mean outdoor dry-bulb temperature, relative humidity and global solar radiation are taken as the three input features. Xuemei et al. [13] proposed Least Square Support Vector Machine (LS-SVM) for cooling load forecasting for an office building in Guangzhou, China by taking hourly climate data and building cooling load for five months as inputs to develop the model.

Wang et al. [14] noted that the SVM models may be superior to that of ANN and ARIMA models. They applied SVM and DE to develop a forecasting model for annual electricity consumption for Beijing city. However, the SVM model development depends on the type and parameters of the Kernel function and also the penalty factor. In practice, it is found that such selection of model parameters can be quite difficult. In order to overcome this, many hybrid models have been developed that combine SVM models with evolutionary learning algorithms that aid in determining model parameters. Ogliari et al. [15] applied proposed a hybrid model, Neural Network with Genetical Swarm Optimization for PV power forecasting. Pai and Hong [16] coupled SVM with simulated annealing (SA) algorithms to forecast Taiwan's annual electric load. In relation to optimization techniques, differential evolution (DE) has gained momentum in recent years as a novel evolutionary computation technique. Vesterstrom and Thomsen [17] have noted that DE outperforms other optimization techniques like Genetic Algorithm (GA) and Particle Swarm Optimization (PSO). De has also been applied to many real work applications relating to pattern recognition [18,19] and classification [20].

The forecasting models are normally classified as short-term, medium-term and long-term forecasting models based on the time interval between subsequent data points. Short-term ranges for prediction for 24 h to one week [21]. For medium-term forecasting, the forecasting time ranges from weeks to months [22]. Long-term forecasting usually involves yearly predictions of regional or national electricity consumption [23]. However, a model for short term electricity consumption cannot be applied to medium term and long term forecasting. This study proposes a novel hybrid model using weighted SVM with DE algorithm that can forecast both short and medium term electricity consumption. A numeric example is employed using half-hourly and daily energy consumption data from an institutional building in Singapore. This single model can predict the half-hourly as well as daily energy consumption through the 'weighted' multi SVM approach.

1.1. SVM model

Support vector machines (SVM) method is put forward by Vapnik [24]. SVMs essentially consist of two components namely—kernel and optimizer algorithm. Kernel divides non-linear data into high-dimensional space and makes the data linearly separable. The learning takes place in the feature space, and the data points only appear inside dot products with other points. The second component of SVM namely, the optimizer algorithm is applied to solve the optimization problem. Since SVM is based on the structural risk minimization (SRM) inductive principle, it seeks to minimize an upper bound of the generalization error consisting of the sum of the training error and a confidence level. This makes SVM superior to commonly used empirical risk minimization (ERM) principle, which only minimizes the training error. Based on such induction principle, SVM usually achieves higher generalization performance than other machine learning techniques. Vapnik proposed the concept of Support Vector Regression (SVR) to solve function fitting problems, which is called the epsilon-SVR (eps-SVR). Suppose that a data set

$G = \left\{(x_i, d_i)\right\}$, $i = 1\ldots N$, $x_i \in R^n$ is an n dimension input vector, $d_i \in R^1$ is the corresponding target output, 'N' expresses the total number of pattern records. The linear regression estimating function can be expressed as is Eq. (1).

$$y = f(x) = w\psi(x) + b \tag{1}$$

In which, $\psi(x)$ is a non-linear mapping from the input space to a high dimensional feature space. 'w' is a weight vector and 'b' is a threshold value, which can be estimated by minimizing the regularized risk function as in Eqs. (2) and (3).

$$R(C) = \left(C/N\right) \Sigma_{i=1}^{N} L_\varepsilon(d_i, y_i) + \|W\|^2/2 \tag{2}$$

In which,

$$L_\varepsilon(d, y) = \begin{cases} 0 |d - y| \le \varepsilon \\ |d - y| - \varepsilon \ otherwise \end{cases} \tag{3}$$

$\|W\|^2/2$ measures the smoothness of the function, and the function $L_\varepsilon(d, y)$ is called ε –insensitive loss function. 'C' specifies the trade-off between the empirical risk and the model smoothness. 'ε ' expresses the Vapnik's linear loss function zone to measure empirical error. When two slack variables 'ζ' and 'ζ^* ' are introduced to represent the distance from actual values to the corresponding boundary values of ε –tube, Eq. (2) can be transformed to Eq. (4) as:

$$R\left(w, \zeta, \zeta^*\right) = \frac{1}{2}\|w\|^2 + C\Sigma_{i=1}^{l} L_\varepsilon\left(\zeta_i + \zeta_i^*\right) \tag{4}$$

$$\text{subject to} \begin{cases} d_i - w \cdot \psi(x_i) - b_i \le \varepsilon + \zeta_i, i = 1, 2, \ldots, N \\ w \cdot \psi(x_i) + b_i - d_i \le \varepsilon + \zeta_i^*, i = 1, 2, \ldots, N \\ \zeta_i, \zeta_i^* \ge 0, i = 1, 2, \ldots, N \end{cases}$$

Scholkopf et al. [25] proposed an improved SVR model, known as nu-SVR which uses a parameter 'ν' to control the number of support vectors and training errors. The range of 'ν' is from 0 to 1. After replacing 'ε' by 'ν', Eq. (4) is transformed as shown in Eq. (5).

$$R\left(w, \zeta_i, \zeta_i^*, \varepsilon\right) = \frac{1}{2}\|w\|^2 + C\left(\nu\varepsilon + \frac{1}{N}\Sigma_{i=1}^{l}\left(\zeta_i + \zeta_i^*\right)\right)$$

$$\text{subject to} \begin{cases} d_i - w \cdot \psi(x_i) - b_i \le \varepsilon + \zeta_i, i = 1, 2, \ldots, N \\ w \cdot \psi(x_i) + b_i - d_i \le \varepsilon + \zeta_i^*, i = 1, 2, \ldots, N \\ \zeta_i, \zeta_i^* \ge 0, i = 1, 2, \ldots, N \end{cases} \tag{5}$$

This constrained optimization problem can be solved by the primal Lagrangian form and the Karush-Kuhn-Tucker conditions. These conditions can be obtained as shown in Eq. (6).

$$\vartheta\left(\beta_i, \beta_i^*\right) = \Sigma_{i=1}^{N} d_i\left(\beta_i - \beta_i^*\right) - \frac{1}{2}\Sigma_{i=1}^{N}\Sigma_{j=1}^{N}\left(\beta_i - \beta_i^*\right)\left(\beta_j - \beta_j^*\right) K\left(x_i, x_j\right)$$

$$\text{Subject to} \Sigma_{i=1}^{N}\left(\beta_i - \beta_i^*\right) = 0, \tag{6}$$

$$\sum_{i=1}^{N}(\beta_i - \beta_i^{\ *}) \le C.\nu,$$

$$0 \le \beta_i \le \frac{C}{l}, 0 \le \beta_i^* \le \frac{C}{l}, i = 1, 2, \ldots, N$$

Here, $\beta_i\beta_i^* = 0$ and an optimal desired weight vector of the regression hyperplane is as shown in Eq. (7).

$$w^* = \Sigma_{i=1}^{N}\left(\beta_i - \beta_i^*\right)\psi(x) \tag{7}$$

The final regression function is as shown in Eq. (8).

$$f\left(x, \beta, \beta^*\right) = \Sigma_{i=1}^{N}\left(\beta_i - \beta_i^*\right) K(x, x_i) + b \tag{8}$$

where $K\left(x_i, x_j\right)$ is the inner product of two vectors in the feature space $\psi(x_i)$ and $\psi\left(x_j\right)$. $K\left(x_i, x_j\right) = \psi(x_i)\psi\left(x_j\right)$ is called the kernel