Energy 115 (2016) 734–745

Contents lists available at ScienceDirect

Energy

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

Support vector regression with fruit fly optimization algorithm for seasonal electricity consumption forecasting

Guohua Cao, Lijuan Wu*

School of Economics and Business Administration, Chongqing University, Chongqing 400030, China

A R T I C L E I N F O

Article history:
Received 18 July 2015
Received in revised form
8 September 2016
Accepted 10 September 2016

Keywords:
Support vector regression
Fruit fly optimization algorithm
Electricity consumption forecasting
Seasonal mechanism

A B S T R A C T

Accurate monthly electricity consumption forecasting can provide the reliable guidance for better energy planning and administration. However, it has been found that the monthly electricity consumption demonstrates a complex nonlinear characteristic and an obvious seasonal tendency. Support vector regression has been widely applied to handle nonlinear time series prediction, but it suffers from the key parameters selection and the influence of seasonal tendency. This paper proposes a novel approach, which hybridizes support vector regression model with fruit fly optimization algorithm and the seasonal index adjustment to forecast monthly electricity consumption. Besides, in order to comprehensively evaluate the forecasting performance of the hybrid model, a small sample of monthly electricity consumption of China and a large sample of monthly electricity retail sales of the United States were employed to demonstrate the forecasting performance. The results show that the proposed hybrid approach is a viable option for the electricity consumption forecasting applications.

1. Introduction

Electricity consumption forecasting has become an essential step for modern management of electric power systems. Particularly, accurate forecast of the electricity consumption has an importance influence on abundant supply without interruption. In the past few decades, a lot of efforts have been placed on the forecasting issues of electricity consumption by scholars and practitioners. The traditional models mainly are statistical models, such as time-series method [1,2], exponential smoothing [3], Box-Jenkins' ARMA model [4], Kalman filtering [5], Bayesian estimation model [6], etc. However, these methods not only are based on the previous electricity consumption data, but also consider other factors, such as temperature, festival and season.

In recent years, many scholars have also turned to some nonlinear methods such as the artificial neural network (ANN) method [7,8] and support vector regression (SVR) models for electricity consumption forecasting. The most popular ANN model is the back-propagation neural networks (BPNN) [9] due to its simple architecture and powerful problem-solving ability. However, the neural networks suffer from several disadvantages, such as the need for a large number of controlling parameters, difficulty in getting a stable solution and the risk of over-fitting. Because of the introduction of Vapnik's ε-insensitive loss function, support vector regression (SVR) model has become a powerful approach to forecast and solve non-linear systems [10–12]. Meanwhile, SVR has also been applied to forecast electricity consumption with the high precision performance [13–16]. However, the forecast performance of the SVR model is mainly influenced by the approximation degree of its three parameter values, which has been demonstrated by V. Cherkassky and Y. Ma.(2004) [17]. Because inappropriate parameter values of a SVR model would lead to over-fitting or under-fitting, a key step to improve the forecasting performance is how to choose three proper parameter values of the SVR model. However, so far there are still not general guidelines to be used to choose three parameter values of a SVR model [18].

In the past few decades, all kinds of intelligent algorithms have already got rapid development. Some of intelligent algorithms are based on the animals foraging behavior including genetic algorithm (GA) [19,20], ant colony optimization algorithm [21–23], and particle swarm optimization algorithm (PSO) [24–26], which are not only used to solve the optimization problem, but also applied to search the three parameter values of a SVR model in order to enhance the forecasting performance [27–31].

Recently a new evolutionary optimization algorithm based on the food finding process of the fruit fly, named as fruit fly

* Corresponding author.
E-mail address: wulj1989@163.com (L. Wu).

http://dx.doi.org/10.1016/j.energy.2016.09.065

optimization algorithm (FOA), was proposed by Pan [32]. The advantages of this new optimization algorithm are easy to understand due to the shorter program code compared with other optimization algorithms and reaching the global optimal solution fast. As a novel optimization intelligent algorithm, FOA has gained much attention and successfully been applied to various optimization issues, such as autonomous surface vessels control [33], control optimization [34], data mining [35,36], multidimensional knapsack problem [37], power load forecasting [38,39], and traffic flow control [40]. Li et al. (2013) [41] proposed a hybrid model based on the generalized regression neural network (GRNN) and FOA algorithm for the annual power load forecasting, and proved that the prediction performance of the GRNN model optimized by FOA is much better than the GRNN optimized by PSO. Lin et al. (2013) [38] demonstrated that FOA-optimized general regression neural network (FOAGRNN) has a better service satisfaction detection capacity than PSOGRNN. Li et al. (2012) [39] proposed a (least squares support vector machine) LSSVM-based annual electricity load forecast model, in which FOA is applied to automatically determine only those two appropriate parameters (C,γ) for the LSSVM model.

Since seasonality is a significant feature of electricity consumption, it affects the prediction accuracy of electricity consumption and should not be ignored in the modeling process. Therefore, how to deal with seasonal variations of electricity consumption data has always been an important issue in power load analysis [42]. In general, the common seasonal adjustment method is to eliminate the seasonal factor, filtering the raw data by differencing before prediction, such as Seasonal ARIMA (SARIMA), X-11-ARIMA and X-12-ARIMA [43,44]. However, first, the differencing processing is not always an appropriate way to deal with both nonlinearity and seasonality [45]. Second, these models assumed the linear relationship is limited to handle complex nonlinear problems [7]. Third, the techniques require a large amount of previous data, which is not always possible [46]. Due to the above rea$C$,γ,ε sons, in this article the seasonal index adjustment method [47] is presented to handle seasonal variations, which not only adjusts seasonality but also avoids the limitations of above methods since seasonal index adjustment focuses on the predicted values.

In this paper, on one hand, we apply the novel FOA algorithm to automatically choose three approximate parameters () of the SVR model for the monthly electricity consumption forecasting. On the other hand, due to the cyclic climate issue and regular economic activities, the electricity consumption shows a seasonal tendency. However, the SVR model has not been widely explored to deal with seasonal forecasting issues of time series. Therefore, this paper would also attempt to deal with the seasonal tread by using the seasonal adjustment technology. The proposed SFOASVR model is dedicated to improve the forecast performance by capturing the seasonal and non-linear electricity consumption tendencies.

The rest of this paper is organized as follows: Section 2 introduces the proposed SFOASVR forecasting model, including the presentation of SVR, FOA algorithm, and the seasonal adjustment method, besides, an example of finding the global optimization solution using FOA is given. Section 3 presents two numerical forecasting examples, which are a small sample of monthly electricity consumption of China and a large sample of monthly electricity retail sales of the United States. Conclusions are discussed in Section 4.

2. Methodology of SFOASVR model

2.1. Support vector regression(SVR)model

The basic idea of the SVR model: the original data x is nonlinearly mapped to a higher dimensional feature space [48,49]. The training data are denoted as $\{(x_1,y_1),\ldots,(x_n,y_n)\}$, where $x_i \in R^n$, x_i is input vector, y_i is the actual output value. The SVR function can been described as Eq. (1),

$$f(x) = \omega^T \varphi(x) + b \tag{1}$$

Where $\varphi(x)$ is the feature function of inputs; $f(x)$ denotes the output value of forecasting; ω and b are adjustable coefficients. Nonlinear regression of the low-dimensional input space is converted to linear regression of the high-dimensional feature space. By using a penalty function to estimate the values of coefficients ω *and* b, penalty function $R(C)$ becomes

$$R(C) = \frac{1}{2}\|\omega\|^2 + C\cdot\frac{1}{n}\sum_{i=1}^{n}\left|y_i - f(x)\right|_\varepsilon \tag{2}$$

where

$$\left|y - f(x)\right|_\varepsilon = \begin{cases} 0, & |y-f(x)| \le \varepsilon \\ |y-f(x)| - \varepsilon, & otherwise \end{cases}$$

Two slack variables ξ_i and ξ_i^* are introduced to cope with infeasible constraints of the optimization problem Eq. (2), hence the optimization problem becomes

$$min(\omega,\xi,\xi^*) = \frac{1}{2}||\omega||^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{3}$$

$$subject\ to \begin{cases} y_i - \omega^T\varphi(x_i) - b & \le & \varepsilon + \xi_i^* \\ -y_i + \omega^T\varphi(x_i) + b & \le & \varepsilon + \xi_i \\ \xi_i, \xi_i^* & \ge & 0 \end{cases}$$

Using the Lagrange equation, we can get the dual optimization problem:

$$max(\alpha_i,\alpha_i^*) = \sum_{i=1}^{n} y_i(\alpha_i - \alpha_i^*) - \varepsilon\sum_{i=1}^{n}(\alpha_i + \alpha_i^*) - \frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i - \alpha_i^*) \times \left(\alpha_j - \alpha_j^*\right)k(x_i,x_j) \tag{4}$$

$$subject\ to \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0\ and\ \ \alpha_i,\alpha_i^* \in [0,C]$$

To solve the above dual optimization problem, then support vector regression function can be obtained as follows,

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)k(x_i,x) + b \tag{5}$$

where α_i^*, and α_i are the Lagrange multipliers; $k(x_i,x_j)$ is the kernel function. Due to fewer parameters to be set and the nonlinear mapping ability of training data to the infinite dimensional space, the Gaussian radial basis function (RBF) is the most choice [50]. Therefore, in this article we selects the Gaussian RBF function as the kernel function $k(x_i,x)$, namely

$$k(x_i,x) = exp\left(-\gamma\|x - x_i\|^2\right), \gamma > 0 \tag{6}$$

where γ is the bandwidth of the Gaussian RBF function.

Consequently, there are three parameters that need to be chosen in the SVR model, which are the penalty parameter C, the