Energy Conversion and Management 50 (2009) 105–117

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Chaotic particle swarm optimization algorithm in a support vector regression electric load forecasting model

Wei-Chiang Hong *

Department of Information Management, Oriental Institute of Technology, 58 Sec. 2, Si-Chuan Road, Panchiao, Taipei 220, Taiwan

ARTICLE INFO

Article history:
Received 18 June 2007
Received in revised form 17 January 2008
Accepted 26 August 2008
Available online 9 October 2008

Keywords:
Support vector regression (SVR)
Chaotic particle swarm optimization (CPSO) algorithm
Electric load forecasting
Forecasting support system

ABSTRACT

Accurate forecasting of electric load has always been the most important issues in the electricity industry, particularly for developing countries. Due to the various influences, electric load forecasting reveals highly nonlinear characteristics. Recently, support vector regression (SVR), with nonlinear mapping capabilities of forecasting, has been successfully employed to solve nonlinear regression and time series problems. However, it is still lack of systematic approaches to determine appropriate parameter combination for a SVR model. This investigation elucidates the feasibility of applying chaotic particle swarm optimization (CPSO) algorithm to choose the suitable parameter combination for a SVR model. The empirical results reveal that the proposed model outperforms the other two models applying other algorithms, genetic algorithm (GA) and simulated annealing algorithm (SA). Finally, it also provides the theoretical exploration of the electric load forecasting support system (ELFSS).

1. Introduction

1.1. Electric load management and traditional forecasting approaches

Accurate electric load forecasting can provide those export oriented economies (like Taiwan) advantages through saving and efficiently distributing limited energy resources. For inaccurate electric load forecasting, it may increase operating costs [1,2]. For example, over estimation of future electric load results in unnecessary spinning reserve, wastes limited energy resources, even leads to distribution inefficiency, and, furthermore, is not accepted by international energy networks owing to excess supply. In contrast, under estimation of load causes failure in providing sufficient reserve and implies high costs in the peaking unit, which discourage any economic and industrial developments. Thus, the accuracy of future electric demand forecasting have received growing attention, particularly in the areas of electricity load planning, energy expenditure/cost economy and secure operation fields, in regional and/or national systems.

However, the electric load forecasting is not easily to conduct, primarily due to the various influences, such as climate factors, social activities, and seasonal factors. During previous decades, numerous investigations had been proposed to improve the accuracy of electricity load forecasting. The famous approach is, weather insensitive, employing historical load data to forecast future electric load, such as Box-Jenkins' ARIMA models [3], exponential smoothing models [4], multiplicative autoregressive (AR) model [5], Bayesian estimation model [6], and the state space and Kalman filtering technology [7]. The second approach is regression model, which is based on the cause-effect relationships between electric load and relevant independent variables (weather, holiday, temperature, wind conditions, humidity, and so on), such as linear regression [8]. For those models mentioned above, electric load was decomposed into weather insensitive and weather sensitive components, respectively; they were all based on linear assumption. Thus those models could not play the excellent role in forecasting because electric load is known to be nonlinear.

To improve the performance of nonlinear electric load forecasting, artificial intelligence techniques are employed. Knowledge based expert system (KBES) approach [9] extracts rules from received relevant information (e.g., daily temperature, day type, load from the previous day, and so on), then, derives training rules and transforms the information into mathematical equations. In addition, artificial neural networks (ANNs) approach [7,10,11] is superior to traditional forecasting approach. However, the training procedure of a KBES model is time consuming; and, for ANNs models, it is possible to get trapped in local minima and subjectively in selecting the model architecture [12].

1.2. Support vector regression with evolutionary algorithms in electric load forecasting

Proposed by Vapnik [13], support vector machines (SVMs) are one of the significant developments in overcoming shortcomings of ANNs mentioned above. Rather than by implementing the

* Tel.: +886 2 7738 0145x327#55; fax: +886 2 7738 6310.
E-mail address: samuelhong@ieee.org

0196-8904/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.enconman.2008.08.031

empirical risk minimization (ERM) principle to minimize the training error, SVMs apply the structural risk minimization (SRM) principle to minimize an upper bound on the generalization error. SVMs could theoretically guarantee to achieve the global optimum, instead of trapping local optimum like ANNs models. Thus, the solution of a nonlinear problem in the original lower dimensional input space could find its linear solution in the higher dimensional feature space. For more detailed mechanisms introduction of SVMs, it is referred to [14], among others.

SVMs have found wide application in the field of pattern recognition, bio-informatics, and other artificial intelligence relevant applications. Particularly, along with the introduction of Vapnik's ε-insensitive loss function, SVMs also have been extended to solve nonlinear regression estimation problems, which are so-called support vector regression (SVR). SVR have been successfully employed to solve forecasting problems in many fields, such as financial time series (stocks index and exchange rate) forecasting [15–17], engineering and software field (production values and reliability) forecasting [18], atmospheric science forecasting [19,20], and so on. Meanwhile, SVR model had also been successfully applied to forecast electric load [21,22]. The empirical results indicated that the selection of the three parameters (C, ε, and σ) in a SVR model influences the forecasting accuracy significantly. Although, numerous publications in the literature had given some recommendations on appropriate setting of SVR parameters [23], however, those approaches do not simultaneously consider the interaction effects among the three parameters. There is no general consensus and many contradictory opinions, thus, evolutionary algorithms are employed to determine appropriate parameter values.

1.3. Chaotic particle swarm optimization algorithm in parameters determination

Although, both SVR with genetic algorithm and SVR with simulated annealing are superior to other competitive forecasting models (ARIMA and ANNs), however, genetic algorithms (GA) and simulated annealing algorithm (SA) are lack of knowledge memory or storage functions, while previous knowledge of the problem is destroyed once the population (GA) or the temperature changes (SA). Thus, these drawbacks of GA and SA would lead to time consuming and inefficiency in the searching the suitable parameters of a SVR model. Recently, inspired by the social behavior of organisms such as fish schooling and bird flocking, Kennedy and Eberhart [24] first introduced particle swarm optimization (PSO). In which, it is also initialized with a population of random solutions. Each individual, namely particle, is assigned with a randomized velocity flown through hyperspace to look for the optimal position to land. Compared with GA and SA, PSO has memory to store the knowledge of good solutions by all particles, in addition, particles in the swarm share information with each other. Therefore, due to the simple concept, easy implementation and quick convergence, nowadays PSO has gained much attention and wide applications in solving continuous nonlinear optimization problems [25]. However, the performance of PSO greatly depends on its parameters, and similar to GA and SA, it often suffers from being trapped in local optimum [26,27].

With the easy implementation and special ability to avoid being trapped in local optimum [28], chaos and chaos-based searching algorithms have aroused intense interests [27,29]. This investigation presented in this paper is motivated by a desire to improve the inefficient disadvantages of searching algorithms mentioned above in determining the three free parameters in the SVR model. Therefore, the chaotic particle swarm optimization (CPSO) method proposed by [27] is employed in a SVR model, namely SVRCPSO, to provide good forecasting performance in capturing nonlinear electric load changes tendency. Two other forecasting approaches, SVR with GA (namely the SVMG model) and SVR with SA (namely the SVMSA model) consequently, were used to compare the forecasting accuracy of electric load.

The remainder of this paper is structured as follows. The fundamental principle of SVR and its formulation are presented in Section 2. In Section 3, the standard PSO and the chaotic particle swarm optimization algorithms are overviewed, which are used to select the parameters of the SVR model. Numerical simulations to demonstrate the forecasting performance of the proposed models and corresponding comparison results with the other two search algorithms are provided in Section 4. Sections 5 and 6 provide some discussions and conclusions.

2. Support vector regression

The brief ideas of SVMs for the case of regression are introduced. A nonlinear mapping $\varphi(\cdot):\Re^{n}\to\Re^{n_h}$ is defined to map the input data (training data set) $\{(\mathbf{x}_i,y_i)\}_{i=1}^{N}$ into a so-called high dimensional feature space (which may have infinite dimensions), $\Re^{n_h}$ (Fig. 1a and b). Then, in the high dimensional feature space, there theoretically exists a linear function, f, to formulate the nonlinear relationship between input data and output data. Such a linear function, namely SVR function, is as Eq. (1),

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

where $f(\mathbf{x})$ denotes the forecasting values; the coefficients $\mathbf{w}$ ($\mathbf{w}\in\Re^{n_h}$) and b ($b\in\Re$) are adjustable. As mentioned above, SVM method one aims at minimizing the empirical risk,

$$R_{\mathrm{emp}}(f)=\frac{1}{N}\sum_{i=1}^{N}\Theta_{\varepsilon}(y_i,\mathbf{w}^{\mathrm{T}}\varphi(\mathbf{x}_i)+b) \tag{2}$$

where $\Theta_\varepsilon(\mathbf{y},f(\mathbf{x}))$ is the ε-insensitive loss function (as thick line in Fig. 1c) and defined as Eq. (3),

$$\Theta_{\varepsilon}(\mathbf{y},f(\mathbf{x}))=\begin{cases}|f(\mathbf{x})-\mathbf{y}|-\varepsilon, & \text{if } |f(\mathbf{x})-\mathbf{y}|\geqslant\varepsilon\\ 0, & \text{otherwise}\end{cases} \tag{3}$$

In addition, $\Theta_\varepsilon(\mathbf{y},f(\mathbf{x}))$ is employed to find out an optimum hyper plane on the high dimensional feature space (Fig. 1b) to maximize the distance separating the training data into two subsets. Thus, the SVR focuses on finding the optimum hyper plane and minimizing the training error between the training data and the ε-insensitive loss function.

Then, the SVR minimizes the overall errors,

$$\mathrm{Min}_{\mathbf{w},b,\xi^*,\xi}R_{\varepsilon}(\mathbf{w},\xi^*,\xi)=\frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}+C\sum_{i=1}^{N}(\xi_i^*+\xi_i) \tag{4}$$

with the constraints

$$\begin{array}{ll}\mathbf{y}_i-\mathbf{w}^{\mathrm{T}}\varphi(\mathbf{x}_i)-b\leqslant\varepsilon+\xi_i^*, & i=1,2,\ldots,N\\ -\mathbf{y}_i+\mathbf{w}^{\mathrm{T}}\varphi(\mathbf{x}_i)+b\leqslant\varepsilon+\xi_i, & i=1,2,\ldots,N\\ \xi_i^*\geqslant 0, & i=1,2,\ldots,N\\ \xi_i\geqslant 0, & i=1,2,\ldots,N\end{array}$$

The first term of Eq. (4), employed the concept of maximizing the distance of two separated training data, is used to regularize weight sizes, to penalize large weights, and to maintain regression function flatness. The second term penalizes training errors of $f(\mathbf{x})$ and $\mathbf{y}$ by using the ε-insensitive loss function. C is a parameter to trade off these two terms. Training errors above ε are denoted as ξ_i^*, whereas training errors below $-\varepsilon$ are denoted as ξ_i (Fig. 1b).

After the quadratic optimization problem with inequality constraints is solved, the parameter vector $\mathbf{w}$ in Eq. (1) is obtained,

$$\mathbf{w}=\sum_{i=1}^{N}(\beta_i^*-\beta_i)\varphi(\mathbf{x}_i) \tag{5}$$

where β_i^*, β_i are obtained by solving a quadratic program and are the Lagrangian multipliers. Finally, the SVR regression function is obtained as Eq. (6) in the dual space,