Neurocomputing 151 (2015) 1362–1373

Contents lists available at ScienceDirect

Neurocomputing

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

Hybridization of seasonal chaotic cloud simulated annealing algorithm in a SVR-based load forecasting model

Jing Geng [a], Min-Liang Huang [b], Ming-Wei Li [a], Wei-Chiang Hong [c,*]

[a] College of shipbuilding engineering, Harbin Engineering University, Harbin 150001, Heilongjiang, PR China
[b] Department of Industrial Management, Oriental Institute of Technology, 58, Section 2, Sichuan Rd., Panchiao, New Taipei City 220, Taiwan, ROC
[c] Department of Information Management, Oriental Institute of Technology, 58, Section 2, Sichuan Rd., Panchiao, New Taipei City 220, Taiwan, ROC

ARTICLE INFO

Article history:
Received 15 August 2013
Received in revised form
17 May 2014
Accepted 25 October 2014
Communicated by M.T. Manry
Available online 4 November 2014

Keywords:
Support vector regression (SVR)
Cloud theory
Cat mapping
Chaotic cloud simulated annealing algorithm (CCSA)
Seasonal mechanism
Electric load forecasting

ABSTRACT

Support vector regression with chaotic sequence and simulated annealing algorithm in previous forecasting research paper has shown its superiority to effectively avoid trapping into a local optimum. However, the proposed chaotic simulated annealing (CSA) algorithm in previous published literature as well as the original SA algorithm could not realize the mechanism of temperature decreasing continuously. In addition, lots of chaotic sequences adopt Logistic mapping function which is distributed at both ends in the interval [0,1], thus, it could not excellently strengthen the chaotic distribution characteristics. To continue exploring any possible improvements of the proposed CSA and chaotic sequence, this paper employs the innovative cloud theory to be hybridized with CSA to overcome the discrete temperature annealing process, and applies the Cat mapping function to ensure the chaotic distribution characteristics. Furthermore, seasonal mechanism is also proposed to well arrange with the cyclic tendency of electric load, caused by economic activities or climate cyclic nature. This investigation eventually presents a load forecasting model which hybridizes the seasonal support vector regression model and chaotic cloud simulated annealing algorithm (namely SSVRCCSA) to receive more accurate forecasting performance. Experimental results indicate that the proposed SSVRCCSA model yields more accurate forecasting results than other alternatives.

1. Introduction

Electric systems play the infrastructural backbone role for modern economic development. Their expected performance is ensured to reach equilibrium between the supply and the demand for optimal electric energy allocation. The balancing cost should be seriously considered by electric suppliers due to low margin profit. Bunn and Farmer [1] indicate that an increase in the load forecasting error by 1% will lead to 10 million Pound (£, GBP) higher operation costs. Hobbs et al. [2] also estimate that a decrease in the load forecasting error by 1% will lead to save almost 1.6 million USD per year. Thus, accurate electric load forecasting is essential for any electric systems to prevent over-estimation or underestimation, particularly for those developing countries with limited capability in electricity import and export [3]. Load forecasting is a complex task for collecting necessary historical data, pre-processing useful data, and building feasible models. The load data itself may illustrate nonlinear characteristics which are caused by historical or exogenous factors [4,5], such as weather conditions, electric price, social activities, etc. There are various techniques based on several statistical methods to forecast future electric load, including Box-Jenkins models [6], exponential smoothing models [7], Bayesian estimation model [8], state space and Kalman filtering technologies [9], and regression models [10]. These approaches are difficultly to receive satisfied performance in nonlinear electric load forecasting. Since 1980s, due to super nonlinear mapping capability, artificial neural network (ANN) has begun its successful applications in load forecasting and has received significant improvements [11,12]. Recently, several novel intelligent approaches are hybridized with ANN, such as fuzzy theory, for example, adaptive network based fuzzy inference system (ANFIS) with RBF Neural Network [13], hybrid evolutionary algorithms, for example, fuzzy neural network with chaotic genetic algorithm and simulated annealing algorithm [14], wavelet transformation, for example, wavelet fuzzy neural network (WFNN) [15], and so on. The principal drawback of the ANN model is subjective determination of model structures [16]. An insight review of the developments of load forecasting by ANN is shown in the references [17,18].

Support vector regression (SVR) [19] has been successfully employed in many forecasting applications, such as financial forecasting [20], traffic flow forecasting [21], atmospheric science

* Corresponding author. Tel.: +886 2 7738 8000x5316; fax: +886 2 7738 6310.
E-mail address: samuelsonhong@gmail.com (W.-C. Hong).

http://dx.doi.org/10.1016/j.neucom.2014.10.055
0925-2312/

forecasting [22], and electric load forecasting [23–25]. Based on previous studies, to obtain better forecasting performance, all three hyper-parameters of a SVR model have to be determined carefully. Usually these hyper-parameters are determined by data re-sampling techniques, these techniques are almost intuitive and computational time consuming. The author has proposed several trials in hybridization of chaotic sequence with evolutionary algorithms for hyper-parameters determination to improve forecasting accuracy level. The low forecasting accuracy is almost caused by the theoretical limitations from original evolutionary algorithms (i.e., premature convergence and trapped in local optimum) [21,23–25]. In author's previous studies, the chaotic sequence is used to transform these hyper-parameters to the chaotic space to richen the searching ergodically over the whole space. However, lots of chaotic sequences adopt Logistic mapping function which is distributed at both ends in the interval [0,1], it could not excellently strengthen the chaotic distribution characteristics. By comparing with the analysis on chaotic distribution characteristics after mapping hyper-parameters into chaotic space, the Cat mapping function is with good ergodic uniformity in the interval [0,1] and is not easily to fall into minor cycle [26]. This paper decides to apply the Cat mapping function to map hyper-parameters into the chaotic space. On the other hand, the chaotic sequence is also only employed to transform the three hyper-parameters of a SVR model from the solution space to the chaotic space, any variable in this kind of chaotic space can travel ergodically over the whole space of interest to find out the improved solution eventually. In addition, several disadvantages embedded in these evolutionary algorithms are required to be improved to get more accurate forecasting performance of a SVR model. For example, based on the operation procedure of SA, shrewd and careful treatments in the annealing schedule are required, such as the degree of the temperature decreasing steps during annealing. Particularly, the temperature of each state is discrete and unchangeable, which does not meet the requirement of continuous decreasing in actual physical annealing processes. Furthermore, SA is easily to accept degenerated solution with high temperature, and it is hardly to escape from local minimum trap with low temperature [27].

Cloud theory is a model of the uncertainty transformation between quantitative representation and qualitative concept using language value [28]. It is successfully used in intelligence control [29,30], data mining [31], spatial analysis [32], intelligent algorithm improvement [33], and so on. In case of SA algorithm, its basic operation procedure requires shrewd and careful treatment during annealing schedule, such as the temperature decreasing degree. Especially, many applications disregard the fact that the temperature should be decreased continuously and cannot be fixed for each state in actual physical annealing processes. Moreover, based on the theoretical definition of the SA algorithm, in the high temperature stage, it is easily to accept worsened solution, then, lead to converge to local minimum while decreasing to low temperature [27]. Based on the mechanism of the SA algorithm, along with the temperature decreasing, the annealing process, likes a fuzzy system, will let these hyper-parameters move from large scale to small scale randomly as the temperature decreasing. Cloud theory can successfully realize the transformation between a qualitative concept in words and numerical representation [28]. Therefore, it is suitable to be employed to solve the problem of temperature decreasing discretely. This investigation tries to apply the chaotic sequence (mapping by the Cat mapping function) with cloud theory hybridized into the original SA algorithm (namely CCSA) to determine the values of these hyper-parameters in a SVR model to escape from stagnation.

In the meanwhile, as shown in the existed literatures [34–36] that electric loads also demonstrate a seasonal tendency caused by the difference in demand from month to month and season to season. The applications of SVR models have not been widely explored to deal with seasonal trend time series, however. Therefore, this paper also attempts to apply the seasonal mechanism [36,37] to deal with seasonal tendency time series problem.

The proposed seasonal SVR with CCSA algorithm, namely SSVRCCSA model, will compare the forecasting performances with two other existed forecasting approaches, ARIMA and TF-ε-SVR-SA models proposed by Wang et al. [36]. The structure of the paper is organized as follows. Section 2 proposes how to model the proposed SSVRCCSA model. The basic formulation of SVR, the Cat mapping, the CSA algorithm, and the seasonal mechanism will all be illustrated. Section 3 provides a numerical example and compares the forecasting performances among all alternatives. Finally, the conclusions are shown in Section 4.

2. Methodology of SSVRCCSA model

2.1. Support vector regression (SVR) model

The brief ideas of a SVR model are introduced as followings. 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), $\Re^{n_h}$. Then, in the feature space, there theoretically exists a linear function, f, to formulate the nonlinear relationship between input data and output data. Such a linear function, is called as the SVR function and is shown 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 ($b \in \Re$) can be adjusted. The SVR method aims to minimize the empirical risk as Eq. (2),

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

where $\Theta_\varepsilon(y_i, \mathbf{w}^T\varphi(\mathbf{x}_i) + b)$ is the ε-insensitive loss function and defined as Eq. (3),

$$\Theta_\varepsilon(y_i, \mathbf{w}^T\varphi(\mathbf{x}_i) + b) = \begin{cases} |\mathbf{w}^T\varphi(\mathbf{x}_i) + b - y_i| - \varepsilon, & \text{if} \quad |\mathbf{w}^T\varphi(\mathbf{x}_i) + b - y_i| \ge \varepsilon \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

In addition, $\Theta_\varepsilon(y_i, \mathbf{w}^T\varphi(\mathbf{x}_i) + b)$ is employed to find out an optimum hyper-plane on the feature space to maximize the distance separating the training data into two subsets. Thus, the SVR model focuses on finding the optimum hyper-plane and on minimizing the training error between the training data and the ε-insensitive loss function, i.e., the SVR model minimizes the overall errors, as shown in Eq. (4),

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

with the constraints

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

The first term of Eq. (4) 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. The parameter C is used to trade off these two terms. Training errors above ε are denoted as ξ_i^*, on the contrary, the training errors below $-\varepsilon$ are denoted as ξ_i.