Electric Power Systems Research 122 (2015) 96–103

Contents lists available at ScienceDirect

Electric Power Systems Research

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

Short-term load forecasting by wavelet transform and evolutionary extreme learning machine

Song Li*, Peng Wang, Lalit Goel

School of Electrical and Electronic Engineering, Nanyang Technological University, 639798 Singapore, Singapore

ARTICLE INFO

Article history:
Received 29 August 2014
Received in revised form
28 November 2014
Accepted 4 January 2015

Keywords:
Artificial bee colony
Extreme learning machine
Short-term load forecasting
Wavelet transform

ABSTRACT

This paper proposes a novel short-term load forecasting (STLF) method based on wavelet transform, extreme learning machine (ELM) and modified artificial bee colony (MABC) algorithm. The wavelet transform is used to decompose the load series for capturing the complicated features at different frequencies. Each component of the load series is then separately forecasted by a hybrid model of ELM and MABC (ELM-MABC). The global search technique MABC is developed to find the best parameters of input weights and hidden biases for ELM. Compared to the conventional neuro-evolution method, ELM-MABC can improve the learning accuracy with fewer iteration steps. The proposed method is tested on two datasets: ISO New England data and North American electric utility data. Numerical testing shows that the proposed method can obtain superior results as compared to other standard and state-of-the-art methods.

1. Introduction

Short-term load forecasting (STLF) is always essential for electric utilities to estimate the load power from one hour ahead up to a week ahead. Load forecasting can be used for power generation scheduling, load switching and security assessment. Accurate forecast results help to improve the power system efficiency, reduce the operating cost and cut down the occurrences of power interruption events. Load forecasting becomes more important because of the development of the deregulated electricity markets and the promotion of the smart grid technologies.

Many statistical methods have been used for STLF, including exponential smoothing [1], Kalman filters [2] and time series methods [3]. These methods are highly attractive because some physical interpretation can be attached to their components. However, they cannot properly represent the nonlinear behavior of the load series. Hence, artificial intelligence techniques have been tried out such as neural networks (NNs), fuzzy logic and support vector machines [4–7]. In particular, NNs have drawn the most attention because of their capability to fit the nonlinear relationship between load and its dependent factors.

Recently, extreme learning machine (ELM) has been proposed to train single-hidden layer feedforward neural networks (SLFNs), which can overcome the drawbacks (e.g. time-consuming and local minima) faced by the gradient-based methods [8]. In ELM, the input weights and hidden biases are initialized with a set of random numbers. The output weights of hidden layer are directly determined through a simple inverse operation on the hidden layer output matrix. ELM has been verified to obtain good performance in many applications, including electricity price and load forecasting [9,10].

This paper presents a hybrid STLF model based on the ELM. Two improvements are carried out to tackle the two key issues in load forecasting: the nonstationary behavior of load series and the robustness of forecast model [6]. First, the wavelet transform is an efficacious treatment to handle the nonstationary load behavior, because it can provide an in-depth time-frequency representation of the load series [11,12]. We use wavelets to decompose the load series into a set of different frequency components and each component is then separately forecasted. In such a way, we do not handle all the frequency components by a single forecaster but treat them differently. Second, it is found that ELM may yield unstable performance because of the random assignments of input weights and hidden biases [13]. To alleviate this problem, the modified artificial bee colony (MABC) algorithm is developed to look for the optimal set of input weights and hidden biases. MABC is a swarm-based optimization algorithm, which simulates the intelligent forging behavior of honey bee swarm [14]. MABC can be easily employed and does not require any gradient information. Furthermore, MABC can probe the unknown regions in the solution space and look for the global best solution. This hybrid learning method can be named as ELM-MABC, which makes use of the merits of ELM and MABC.

* Corresponding author at: Blk S2-B7c-05, 50 Nanyang Avenue, Nanyang Technological University, 639798 Singapore, Singapore. Tel.: +65 83286906.
E-mail address: sli5@e.ntu.edu.sg (S. Li).

http://dx.doi.org/10.1016/j.epsr.2015.01.002
0378-7796/

The proposed method is tested on two datasets: ISO New England data and North American electric utility data. Section 2 describes wavelet transform, extreme learning machine, artificial bee colony algorithm and the proposed STLF method. Simulations are presented in Section 3. Section 4 provides discussion and Section 5 outlines conclusion.

2. Methodology

2.1. Wavelet transform

The multiple frequency components in load series are always the challenging parts in forecasting [15]. A single forecaster cannot handle them appropriately and we can treat them differently with the help of wavelet transform. Wavelet transform can be used to decompose a load profile into a series of constitutive components [16]. These constitutive components usually have better behaviors (e.g. more stable variance and fewer outliers) and therefore can be forecasted more accurately [15].

Wavelet transform makes use of two basic functions: scaling function $\varphi(t)$ and mother wavelet $\psi(t)$. A series of functions are derived from the scaling function $\varphi(t)$ and the mother wavelet $\psi(t)$ by

$$\varphi_{j,k}(t) = 2^{j/2}\varphi(2^j t - k) \tag{1}$$

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k) \tag{2}$$

where j and k are integer variables for scaling and translating [17]. The wavelet functions $\psi_{j,k}(t)$ and scaling functions $\varphi_{j,k}(t)$ can be used for signal representation. Then a signal $S(t)$ can be expressed by

$$S(t) = \sum_k c_{j_0}(k)2^{j_0/2}\varphi(2^{j_0}t - k) + \sum_k \sum_{j=j_0}^{\infty} d_j(k)2^{j/2}\psi(2^j t - k) \tag{3}$$

where j_0 is the predefined scale, $c_{j_0}(k)$ and $d_j(k)$ are the approximation and detail coefficients, respectively. It is seen that wavelet decomposition is done to compute the above two sets of coefficients. The first term on the right of (3) gives a low resolution representation of $S(t)$ at the predefined scale j_0. For the second term, a higher resolution or a detail component is added one after another from the predefined scale j_0 [18].

A demonstration of two-level decomposition for load series is given by

$$S(t) = A_1(t) + D_1(t) = A_2(t) + D_2(t) + D_1(t). \tag{4}$$

The load signal S is broken up into a set of constitutive components. The approximation A_2 reflects the general trend and offers a smooth form of the load signal. The terms D_2 and D_1 depict the high frequency components in the load signal. Specifically, the amplitude of D_1 is very small, which carries information about the noise in the load signal.

Three issues must be considered before using the wavelet transform: type of mother wavelet, number of decomposition levels and border effect. In this paper, the trial and error method is used to choose the mother wavelet and number of decomposition levels. Three popular wavelet families: Daubechies (db), Coiflets (coif) and Symlets (sym) [16] are investigated for decomposing the load signal. The combinations of 12 mother wavelets (db2–db5, coif2–coif5 and sym2–sym5) and 3 decomposition levels (1–3) have been tested. It is found that the combination of coif4 and 2-level decomposition can produce the best forecasting performance. In addition, the border distortion will arise if the transform is performed on finite-length signals, which would degrade the performance. The signal extension method in [19] is adopted in this paper, which appends the previous measured values at the beginning of the load signal and forecasted values at the end of it.

2.2. Modified artificial bee colony (MABC) algorithm

The ABC algorithm, introduced by Karaboga, simulates the intelligent foraging behavior of honey bees [14]. The swarm in ABC is divided into three groups: employed bees, onlookers and scouts. The position of a food source represents a solution to the target problem while the nectar amount stands for the fitness value of that solution. An employed bee may update its position in case of finding a new food source. If the fitness of the new source is higher than that of the old one, the employed bee chooses the new position over the old one. Otherwise, the old position is retained. After all the employed bees finish search missions, they share the information (i.e. positions and nectar amounts) of the food sources with the onlookers in the hive. An onlooker bee will choose a food source based on the associated probability value p_i, which is given by

$$p_i = \frac{fit_i}{\sum_{j=1}^{SN} fit_j} \tag{5}$$

where fit_i is the fitness value of ith food source and SN is the number of food sources.

The basic ABC generates a new solution v_{ij} from the old one u_{ij} by:

$$v_{ij} = u_{ij} + \theta_{ij}(u_{ij} - u_{kj}) \tag{6}$$

where i and k are the solution indices and j is the dimension index. The index k has to be different from i and θ_{ij} is a uniformly random number within the range $[-1, 1]$. The old solution u_{ij} will be replaced by the new one v_{ij}, provided that v_{ij} has a better fitness value.

If a food source cannot be improved for many cycles, this source is abandoned. The number of cycles for abandonment is called *limit*, which is a control parameter in ABC. The employed bee related to the abandoned food source becomes a scout. The scout discovers the new food position by

$$u_{ij} = u_{\min,j} + rand(0, 1)(u_{\max,j} - u_{\min,j}) \tag{7}$$

where $u_{\min,j}$ and $u_{\max,j}$ are the lower and upper bounds for the dimension j, respectively. The random number in (7) follows the uniform distribution.

It has been pointed out that the search equation given by (6) is good at exploration but poor at exploitation [20]. To balance these two capabilities and improve the convergence performance, a modified search equation is proposed as follows:

$$v_{ij} = w \cdot u_{best,j} + \theta_{ij}(u_{best,j} - u_{ij}) \tag{8}$$

where u_{best} is the best solution in current population, and w is the inertia weight. The search equation (8) uses the information of the best solution to direct the movement of population. The new solution is driven towards the best solution of the previous cycle. The coefficient w controls the impact from the best solution u_{best}. A large weight encourages the global exploration, while a small one speeds up the convergence to optima. In this paper, the inertia weight w is chosen to be 0.1. Hence, the modified equation is able to improve the exploitation capability and accelerate the convergence speed. The search process of MABC will end if a stop criterion is satisfied. Normally, a maximum cycle number (MCN) is used to terminate the algorithm.