Applied Energy 170 (2016) 22–29

Contents lists available at ScienceDirect

Applied Energy

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

An ensemble approach for short-term load forecasting by extreme learning machine

Song Li *, Lalit Goel, Peng Wang

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

HIGHLIGHTS

- We propose an ensemble STLF method based on the ELM.
- A wavelet-based ensemble scheme is introduced to STLF.
- A parallel model of 24 ELMs is established for 24-h load prediction.
- Both 1-h and 24-h ahead load forecasting are evaluated.
- The proposed method outperforms other techniques on the public datasets.

ARTICLE INFO

Article history:
Received 3 October 2015
Received in revised form 19 February 2016
Accepted 20 February 2016

Keywords:
Ensemble method
Extreme learning machine
Partial least squares regression
Short-term load forecasting
Wavelet transform

ABSTRACT

This paper proposes a novel ensemble method for short-term load forecasting based on wavelet transform, extreme learning machine (ELM) and partial least squares regression. In order to improve forecasting performance, a wavelet-based ensemble strategy is introduced into the forecasting model. The individual forecasters are derived from different combinations of mother wavelet and number of decomposition levels. For each sub-component from the wavelet decomposition, a parallel model consisting of 24 ELMs is invoked to predict the hourly load of the next day. The individual forecasts are then combined to form the ensemble forecast using the partial least squares regression method. Numerical results show that the proposed method can significantly improve forecasting performance.

1. Introduction

Short-term load forecasting (STLF) is a basic requirement in the operation of power systems. The primary application of STLF is to provide load predictions for generation scheduling, such as unit commitment and economic dispatch [1]. For example, if the load demand is known in advance, we can operate the generators at the lowest possible cost. It is shown in [2] that a small increase in forecasting accuracy would save millions of dollars in operation costs. A second application of STLF is for power system security assessment [3]. The forecast results are essential in the detection of future conditions under which the system may be vulnerable. In addition, energy purchasing and bidding also require STLF [4,5].

Various methods have been proposed for STLF in power systems. In the early stages, statistical methods such as regression models and time series methods [6,7], were extensively employed. Nowadays, artificial intelligence techniques such as artificial neural networks (ANN), expert systems and support vector machines [8–12] have been applied to solve the STLF problem. In addition, many hybrid forecasting models have also been developed, which can make use of the advantages of each technique involved. In [13], the forecasting problem was presented in state space form, where the model parameters were fine-tuned by a particle swarm optimizer (PSO). In [14], a hybrid forecasting model combining differential evolution (DE) and support vector regression (SVR) was proposed for load forecasting, where the DE algorithm was used to choose the appropriate parameters for SVR.

As an emerging class of ANN, extreme learning machine (ELM) has a fast learning speed for its iterative-free learning mechanism [15]. In ELM, the input weights and hidden biases are randomly initialized, while the output weights are directly calculated by a least squares method. ELM has been successfully used in load and price forecasting [16,17]. But there is still room for improvement. Like other ANN-based models, the random initialization of weight

* Corresponding author at: Blk S2-B7c-05, 50 Nanyang Avenue, Nanyang Technological University, 639798 Singapore, Singapore. Tel.: +65 98628450.
E-mail address: sli5@e.ntu.edu.sg (S. Li).

http://dx.doi.org/10.1016/j.apenergy.2016.02.114
0306-2619/

parameters may introduce inferior values, leading to unreliable result. Moreover, ELM usually contains hundreds of hidden neurons and the output weights are directly estimated from the training data. Therefore, ELM might suffer from the overtraining problem [18,19].

In this paper, a neural network ensemble model is proposed to alleviate the above problems. An ensemble model consists of a number of ELMs, which are trained for the same load forecasting problem. It has been shown that ensemble models have strengths with respect to single models in terms of increased accuracy and robustness. The reason is that the diverse errors from individual ELMs can cancel out in the aggregating process [20]. Several ensemble methods have been developed for STLF [21–23]. For example, in [23], the regularized negative correlation learning method was used to enhance the prediction capability of the network ensemble.

Wavelet transform has been widely used to capture the inner load features and improve the forecast accuracy [24,25]. However, there is no fixed criterion to select the wavelet parameters: mother wavelet and number of decomposition levels. In this paper, a novel ensemble strategy is introduced, which uses wavelet transform to create the collection of ELM-based predictors. The individual predictors are featured with different combinations of mother wavelet and number of decomposition levels. The rationale is that different wavelet parameters would generate different inputs for the ELM-based predictors, which can promote the diversity of ensemble. In such a way, wavelet parameter selection is avoided without sacrificing the forecast accuracy.

Simple averaging is usually used to combine the individual forecasts [22,26]. But in practice, some predictors in the ensemble are more accurate than others. The ensemble output should be a weighted average of the individual outputs. In this paper, the weight factors are obtained by the partial least squares regression (PLSR) method. Moreover, PLSR can tackle the high degree of correlation between the individual forecasts and therefore generate accurate ensemble output [27].

This paper proposes a novel ensemble method for STLF, in which wavelet transform, extreme learning machine and partial least squares regression are integrated. The proposed wavelet-based ensemble strategy uses different wavelet specifications to generate different input features for the ELM-based predictors. For each sub-component from the wavelet transform, a parallel model of 24 ELMs is allocated to predict the hourly data of the next day. To establish an accurate ensemble forecast, PLSR is used to combine the outputs of individual predictors.

The main contributions of this paper are as follows:

(1) A new hybrid forecast method is proposed to deal with the STLF problem in electric power systems, which can alleviate the overtraining and uncertainty problems and improve the forecasting accuracy.
(2) A novel ensemble strategy based on the wavelet transform is employed in the proposed method. It can avoid the trivial process of wavelet parameter selection. Moreover, the complementary information contained in each set of wavelet parameters can be utilized to improve the forecasting accuracy.
(3) Weighted averaging is used to combine the individual forecast outputs, which can consider the relative accuracy between them. PLSR is adopted to calculate the optimal weight factors.

Although this paper focuses on STLF, the proposed method can be easily extended to electricity price forecasting, which is another important topic in power systems. Note that the input variables should be altered accordingly.

The rest of this paper is organized as follows. Section 2 presents the relevant theories and describes the implementation of the proposed method. Section 3 provides the numerical results and comparisons with other well-established methods. Section 4 outlines the conclusions.

2. Proposed methodology

2.1. Extreme learning machine

ELM accomplishes the learning process in two steps. First, the input weights and hidden biases are initialized with random numbers. Second, the output weights are calculated through an inverse operation on the hidden layer output matrix. The idea of ELM lies in that the neural network learning is turned into a least squares problem, which can be easily solved with much less time [15].

Given a training set of N samples ($\boldsymbol{x}_i$, $\boldsymbol{d}_i$), the single-hidden layer feedforward network (SLFN) in Fig. 1 can be modeled by

$$\sum_{j=1}^{n}\beta_j g(\boldsymbol{w}_j \cdot \boldsymbol{x}_i + b_j) = \boldsymbol{o}_i, \quad i = 1, \ldots, N. \tag{1}$$

where $\boldsymbol{x}_i$ is the input pattern, $\boldsymbol{d}_i$ is the desired output, n is the number of hidden nodes, $g(\cdot)$ is the activation function, $\boldsymbol{o}_i$ is the actual output, $\boldsymbol{w}_j$ is the input weight, b_j is the hidden bias and β_j is the output weight.

If the training error is zero, we can say that there exist β_j, $\boldsymbol{w}_j$ and b_j such that

$$\sum_{j=1}^{n}\beta_j g(\boldsymbol{w}_j \cdot \boldsymbol{x}_i + b_j) = \boldsymbol{d}_i, \quad i = 1, \ldots, N. \tag{2}$$

The compact matrix form of (2) can be given by $\mathbf{H}\boldsymbol{\beta} = \mathbf{D}$, where $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_n]^{\mathrm{T}}$, $\mathbf{D} = [\,\boldsymbol{d}_1, \ldots, \boldsymbol{d}_N]^{\mathrm{T}}$ and $\mathbf{H}$ is called the hidden layer output matrix.

Since the number of hidden nodes is usually less than the number of training samples, $\mathbf{H}$ is not square and ELM cannot approach the zero training error. To train the SLFN, ELM assigns uniformly distributed random numbers for the input weights $\boldsymbol{w}_j$ and hidden biases b_j. Then the network becomes an over-determined linear system and the output weights $\boldsymbol{\beta}$ can be determined by a least squares method. ELM provides a special solution by $\boldsymbol{\beta}^* = \mathbf{H}^{\dagger}\mathbf{D}$, where $\mathbf{H}^{\dagger}$ is the Moore–Penrose (MP) generalized inverse of $\mathbf{H}$.

ELM presents many important properties, which make it an appealing learning algorithm for SLFN. First, ELM exhibits a fast learning speed as it determines the network parameters without iterative adjustments. Second, ELM can achieve good performance because the minimum training error $||\mathbf{H}\boldsymbol{\beta}\text{-}\mathbf{D}||$ is approached with a least squares method. Third, ELM can avoid many problems faced by traditional learning methods, such as local minima, stopping criterion and learning rate.

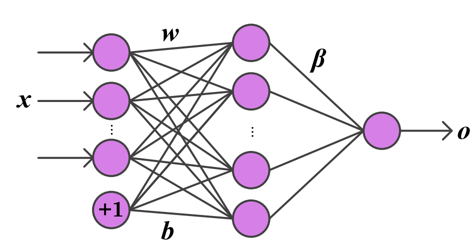

Fig. 1. Single-hidden layer feedforward network.