Electric Power Systems Research 119 (2015) 138–148

Contents lists available at ScienceDirect

Electric Power Systems Research

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

A hybrid short-term load forecasting with a new data preprocessing framework

M. Ghayekhloo[a], M.B. Menhaj[a,b], M. Ghofrani[c,*]

[a] Department of Electrical and Computer Engineering, Science and Research Branch, Islamic Azad University, Qazvin, Iran
[b] Department of Electrical Engineering, Amirkabir University of Technology, Tehran, Iran
[c] School of Science, Technology, Engineering and Mathematics (STEM), University of Washington, Bothell, USA

ARTICLE INFO

Article history:
Received 15 July 2014
Received in revised form 26 August 2014
Accepted 1 September 2014
Available online 28 September 2014

Keywords:
Bayesian neural network
Correlation analysis
Data preprocessing
Forecasting
Input selection
Standardization

ABSTRACT

This paper proposes a hybrid load forecasting framework with a new data preprocessing algorithm to enhance the accuracy of prediction. Bayesian neural network (BNN) is used to predict the load. A discrete wavelet transform (DWT) decomposes the load components into proper levels of resolution determined by an entropy-based criterion. Time series and regression analysis are used to select the best set of inputs among the input candidates. A correlation analysis together with a neural network provides an estimation of the predictions for the forecasting outputs. A standardization procedure is proposed to take into account the correlation estimations of the outputs with their associated input series. The preprocessing algorithm uses the input selection, wavelet decomposition and the proposed standardization to provide the most appropriate inputs for BNNs. Genetic Algorithm (GA) is then used to optimize the weighting coefficients of different forecast components and minimize the forecast error. The performance and accuracy of the proposed short-term load forecasting (STLF) method is evaluated using New England load data. Our results show a significant improvement in the forecast accuracy when compared to the existing state-of-the-art forecasting techniques.

1. Introduction

Short-term load forecasting is essential for reliable and economic operation of power systems. It is particularly more important for deregulated power systems where the forecast inaccuracies have significant implications for market operators, transmission owners and market participants. An accurate load forecast results in establishing appropriate operational practices and bidding strategies as well as scheduling adequate energy transactions [1].

Load forecasting algorithms are classified into three major categories: classical statistical techniques such as autoregressive (AR), autoregressive moving average (ARMA), semi-parametric and similar-day models [2–5]; computational intelligent methods such as neural networks (NNs) [6,7] and fuzzy systems [8]; and hybrid algorithms. The traditional statistical methods generally use linear models with limited or even no capability to characterize the non-linearity of the load patterns. In addition, the stationary process considered for most of these studies cannot capture the non-stationary features of the load time-series. Modern intelligent methods have been proposed to avoid the inefficiencies of classical techniques and provide a more accurate STLF. However, limitations of the individual intelligent methods called for a hybrid of different techniques to enhance the performance and accuracy of STLF. The data preprocessing for the hybrid methods has been investigated in several studies [9–18]. Ref. [9] studied different preprocessing techniques including regression analysis, Akaike's information criterion (AIC) and correlation analysis. Although the results approved the correlation analysis for the feature extraction, the analysis was limited to daily and weekly periodicities without considering the seasonality of the load patterns as well as their monthly and annual periodicities. Ref. [10] developed a neuro-fuzzy system for STLF in a deregulated and price-sensitive environment. A functional time-series forecasting methodology was introduced in [11] that is mainly based on a similar shape prediction. The proposed method provides a reference load curve using the qualitative and quantitative characteristics of the day considered for the load prediction. The prediction is performed by means of a weighted average of past daily load demands with load shapes similar to that of the reference load. However, qualitative variables such as temperature and wind speed cannot be precisely characterized because of their inherent intermittencies and uncertainties. This questions the accuracy of

* Corresponding author at: UWBB Room 227, 18807 Beardslee Blvd, Bothell, WA 98011, USA. Tel.: +1 425 352 3224; fax: +1 425 352 3775.
E-mail address: mghofrani@uwb.edu (M. Ghofrani).

http://dx.doi.org/10.1016/j.epsr.2014.09.002
0378-7796/© 2014 Elsevier B.V. All rights reserved.

the reference shape selection which is the basis of the proposed method.

Wavelet transform has been incorporated into many hybrid algorithms to better characterize complex load features and improve forecasting accuracy [12–18]. Two strategies were proposed in [12] to combine the discrete WT with neural networks and provide a hybrid forecasting algorithm. A hybrid wavelet-NN method was employed in [13] that uses particle swarm optimization algorithm to adjust the parameters of the network. Ref. [14] proposed a STLF method using wavelet neural networks (WNN) with data pre-filtering. A combination of similar day selection, wavelet decomposition and neural networks was presented in [15] to provide a generic hybrid framework for STLF. A multi-wavelet transform was used in combination with a three-layer feed-forward neural network to extract the training data and predict the load [16]. Refs. [17,18] used wavelet transforms to filter high-frequency components of the load and reduce the computational time for training NNs. All these hybrid WNN methods use several networks to separately predict different components of the load. The individual forecasts are then added up or simply combined to provide the predicted load. The simplistic addition or combination of the individual forecasts is not appropriate as different load components have different weights and contributions in the load prediction. None of these references includes the outputs in their standardization procedures. Therefore, the correlation of the predictions with their associated input series is neglected which reduces the accuracy of the prediction. In addition, relevant information of the load time series is not adequately captured by the data pre-processing methods of the above-mentioned references which call for a comprehensive framework for the input selection and data standardization. This is particularly important for special days such as weekends and holidays where a limited amount of training data imposed more restriction on the accuracy of the forecast leading to significant errors.

This paper proposes a hybrid method for STLF. A data preprocessing algorithm is developed to appropriately select the BNN inputs. The algorithm uses time series and regression analysis, wavelet decomposition and a new standardization. The time series and regression analysis are used to select the best set of inputs for the load forecasting of the special days. The proposed standardization procedure provides an estimation of the predictions for the forecasting outputs whose values are not known during the standardization. The estimations are included in the standardization to take into account the correlation of the outputs with their associated input series and enhance the accuracy of the prediction. Genetic algorithm (GA) is used to optimize the weighting coefficients of different forecast components and minimize the forecast error. This significantly improves the forecast accuracy compared to the existing forecasting methods where the coefficient weighting is either arbitrary or based on practical and experimental estimates, with no attempt at optimizing their parameter values.

Section 2 explains the DWT and BNN. It also describes the developed data preprocessing algorithm including the input selection method, and the proposed standardization procedure. Forecast results and their comparisons with those of the state-of-the-art forecasting methods are given in Section 3. Conclusions are presented in Section 4.

2. Methodology

2.1. Discrete wavelet transform

Wavelet analysis is used to deal with non-stationary features and provide a time-scale representation of electric load time-series. The original time domain signal is decomposed into scales with different levels of resolution to extract the irregular load information and better characterize the load behaviour. Low- and high-pass filters transform the load data into low- and high-frequency components. The low-frequency component is an approximation of the original signal representing its general trend while the high-frequency component provides a detailed representation [13]. Generally, the discrete wavelet transform of a discrete time signal $f(k)$ is defined by [19]:

$$D^c_{m,n} = \sum_k f(k)\psi_{m,n}(k) \quad (1)$$

where $D^c_{m,n}$ is the detail coefficient at scale m and location n; and $\psi_{m,n}(k)$ is the detailed and translated form of the mother wavelet $\psi(k) = \psi_{0,0}(k)$.

Scaling functions for a DWT are given by:

$$\varphi_{m,n}(k) = (2^{-(m/2)})\varphi(2^{-m}k - n) \quad (2)$$

The scaling function with m, $n = 0$ gives the father wavelet as $\varphi(k) = \varphi_{0,0}(k)$.

The approximate coefficient at scale m and location n is calculated by convoluting the scaling functions with the signal:

$$A^c_{m,n} = \sum_k f(k)\varphi_{m,n}(k) \quad (3)$$

The approximation (A_m) and detail (D_j) of the signal at scales m and j are obtained by:

$$A_m = \sum_{n=-\infty}^{n=+\infty} A^c_{m,n} \cdot \varphi_{m,n} \quad (4a)$$

$$D_j = \sum_{n=-\infty}^{n=+\infty} D^c_{j,n}\psi_{j,n} \quad (4b)$$

The original signal is the sum of the approximation and details up to scale M:

$$f = A_M + \sum_{j=1}^{M} D_j \quad (5)$$

The original signal for two consecutive scales of $M = m - 1$ and $M = m$ is calculated by (5) and given by:

$$f = A_{m-1} + \sum_{j=1}^{m-1} D_j \quad (6a)$$

$$f = A_m + \sum_{j=1}^{m} D_j = A_m + D_m + \sum_{j=1}^{m-1} D_j \quad (6b)$$

Comparing (6a) and (6b), we have:

$$A_{m-1} = A_m + D_m \quad (7)$$

This provides a multi-resolution representation to calculate the approximate and detail coefficients at an arbitrary scale using those coefficients at the previous scale as:

$$A^c_{m+1,n} = \sum_i h_i A^c_{m,2n+i} = \sum_i h_{i-2n} A^c_{m,i} \quad (8a)$$

$$D^c_{m+1,n} = \sum_i g_i A^c_{m,2n+i} = \sum_i g_{i-2n} A^c_{m,i} \quad (8b)$$

where h_i and g_i are the coefficients of the low- and high-pass filters used to decompose the signal into the low- and high-frequency components.