Journal of Wind Engineering & Industrial Aerodynamics 175 (2018) 136–143

Contents lists available at ScienceDirect

Journal of Wind Engineering & Industrial Aerodynamics

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

Data mining-assisted short-term wind speed forecasting by wavelet packet decomposition and Elman neural network

Chuanjin Yu, Yongle Li, Huoyue Xiang *, Mingjin Zhang

Department of Bridge Engineering, Southwest Jiaotong University, 610031 Chengdu, Sichuan, PR China

ARTICLE INFO

Keywords:
Wind speed forecasting
Wavelet packet decomposition (WPD)
Gradient boosted regression trees (GBRT)
Density-based spatial clustering of applications with noise (DBSCAN)
Elman neural network (ENN)

ABSTRACT

On the basis of data-mining technology, a hybrid method of short-term wind speed forecast is proposed by the wavelet packet decomposition, density-based spatial clustering of applications with noise, and the Elman neural network (WPD-DBSCAN-ENN). First, the WPD is applied to decompose a raw wind speed series into several sub-series. The gradient boosted regression trees (GBRT) algorithm is then applied to determine the structure of the ENNs for each sub-wind series. Next, the training dataset is clustered by the DBSCAN to select the representative data for the ENNs. A key parameter in the DBSCAN is chosen through a new method. Finally, the wind speed forecast is conducted by the ENNs. Case studies are adopted to validate the accuracy of the hybrid methods. The results are compared with those obtained using the WPD-ENN hybrid method and a single ENN via four general error criteria. The performance of the WPD-DBSCAN-ENN hybrid method outperformed those of the other methods indicated above.

1. Introduction

Bridges and trains are susceptible to wind action, especially in the presence of strong winds (Zhang et al., 2015). As a result, lighter train materials and higher running speeds produce more significant problems. Wind alarm systems are necessary to control train speeds in high-wind situations (Delaunay et al., 2003, 2006). The wind forecast model is a vital part of the system (Hoppmann et al., 2002). Accurate wind speed predictions can aid bridge managers in generating better traffic management. However, it is difficult to obtain satisfactory forecast results because wind speed is a non-stationary signal.

In recent years, wind speed forecasting has been extensively examined (Tascikaraoglu and Uzunoglu, 2014). Wind speed forecasting can be categorized by four major methods; namely, statistical methods, physical methods, intelligent methods, and hybrid models (Ma et al., 2009). The autoregressive integrated moving average model (ARIMA) is a popular statistical method. However, it cannot handle nonlinear problems. With the help of physical considerations, physical models can accurately predict the future wind speed. However, physical models are computationally expensive and are not suitable for short-term predictions. In terms of intelligent methods, the artificial neural network (ANN) is one of the most widely used intelligent models, although it requires more historical data compared to statistical methods. Hybrid models have increasingly become the most popular models as they utilize the features of different forecasting methods to generate more accurate predictions (Tascikaraoglu and Uzunoglu, 2014).

In terms of hybrid models with statistical methods, Cadenas and Rivera (2010) proposed an ARIMA-ANN model, wherein the original forecasting error generated by ARIMA was assumed to be a nonlinear component and was obtained by the ANN given that the ARIMA is a linear methodology. The results were added to the final forecasting. Liu et al. (2012) proposed two new models: ARIMA-ANN and ARIMA-Kalman. In the hybrid ARIMA-ANN model, the ARIMA model was utilized to develop the structure of the ANN model. In the hybrid ARIMA-Kalman model, the ARIMA model was employed to initialize the Kalman measurement and the state equations of a Kalman model. Liu et al. (2010) modified the Taylor-Kriging model properly to forecast a wind speed time series, which outperformed the ARIMA method. In terms of hybrid models with physical methods, short-term wind speed forecasting and observations from a global numerical weather prediction model were integrated for the fifth-generation mesoscale model (MM5). In addition, its outputs were processed by a neural network to obtain a specific wind speed forecast (Salcedo-Sanz et al., 2009). Kalman filtering was used in the numerical weather prediction (NWP) to forecast the wind speed (Cassola and Burlando, 2012). Two high-resolution regional atmospheric systems were employed to provide accurate local wind

* Corresponding author.
E-mail addresses: 7818534@qq.com (C. Yu), lele@swjtu.edu.cn (Y. Li), hy@swjtu.edu.cn (H. Xiang), Minjin@swjtu.edu.cn (M. Zhang).

https://doi.org/10.1016/j.jweia.2018.01.020
Received 23 November 2016; Received in revised form 13 January 2018; Accepted 13 January 2018

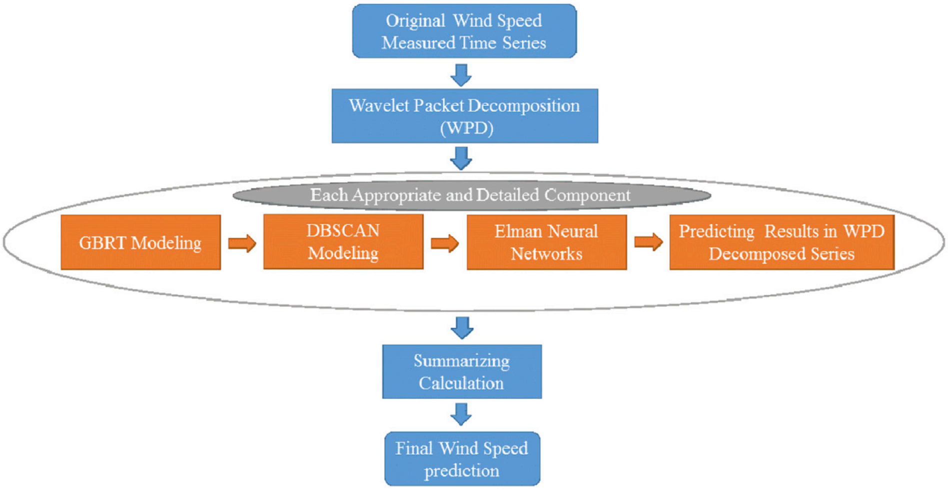

Fig. 1. Framework of the proposed WPD-DBSCAN-ENN hybrid model.

forecasts (Stathopoulos et al., 2013). In terms of the hybrid models with intelligent methods, the wind speed series was first decomposed by the wavelet packet decomposition (WPD), after which the detailed components were decomposed by the fast ensemble empirical mode decomposition (FEEMD) method. All the sub-series were modeled by the Elman neural networks (ENNs) to generate the final forecast (Liu et al., 2015a,b,c).

Data mining is an additional powerful way to reinforce hybrid methods. Data mining generates many applications, especially in terms of clustering (Colak et al., 2012). Lorenzo et al. (2011) portioned an input dataset into *k* clusters using a K-means technique and built a set of corresponding multilayer perceptrons (MLPs) to predict wind power. Liu et al. (2014) used three clustering methods, specifically K-means, a self-organizing map (SOM), and spectral clustering (SC), to improve the accuracy of short-term wind power forecasts. Based on previous observations, the cluster number is essential and must be carefully and reasonably chosen. Koksoy et al. (2015) employed association rule mining (ARM) to classify an NWP training data set into several groups, from which all the model forecasts were combined into a single final forecast based on their determined weights. Liu et al. (2015a,b,c) employed SC to divide sub-series into several groups, which were firstly decomposed by the wavelet transform (WT). Echo state networks (ESNs) were then employed to generate predictions. Liu et al. (2015a,b,c) used the adaptive boosting (AdaBoost) algorithm to divide the original wind speed data into several groups and to generate some appropriate MLPs to make forecasts.

Based on the above reviews, the clustering method helped improve wind forecasting by gathering representative data into several groups and building corresponding prediction models. However, this method generates a significant margin prediction error when new input data is sorted into a fake cluster, thereby deeming the applied model inappropriate (Lhermitte et al., 2011). In addition, the cluster number is vital for forecasting but is very difficult to determine, which presents another obvious drawback. An increase in the cluster number generates an increase in the number of models for prediction and the calculation cost. Although neural network models are often fitted to generate wind speed predictions, it is not easy to determine their structure, which is vital for forecast accuracy.

Therefore, the present study proposes a new hybrid method named the wavelet packet decomposition, density-based spatial clustering of applications with noise, and the Elman neural network (WPD-DBSCAN-ENN). Firstly, a raw wind speed series was decomposed by the WPD. An ENN was then built for each wind sub-series to independently generate predictions. The ENN structure was determined by the gradient boosted regression trees (GBRT) algorithm (Friedman, 2001) to overcome the weakness of the ARIMA methods. In addition, training samples of each ENN were processed by the DBSCAN to select representative ones. Finally, the ENN was employed to forecast the wind speed.

The innovations of the hybrid method are further explained as follows. 1. To our knowledge, the DBSCAN has not been investigated in wind speed forecasting. As compared to other clustering methods, such as the SC, K-means, and SOM, the DBSCAN only has one prediction model for each sub-series decomposed by the WPD. However, the selected ENN data may also be better represented by the DBSCAN. The presented process is effectively executed by new concepts of "notable percentage" and "notable radius", which are defined in the study. The use of the hybrid method is assumed to improve the forecast precision. 2. The ARIMA poorly handles nonlinear components and is sensitive to outliers. The GBRT overcomes this weakness and is used to identify the structure of the ENN by "cumulative importance", which is also a new definition presented in this study.

Certain real wind speed time series obtained from different places were employed to validate the performance of the proposed method. The results were compared with those using the hybrid method WPD-ENN and the single ENN via four general error criteria, namely the mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE), and the maximum absolute error (MaxAE). The

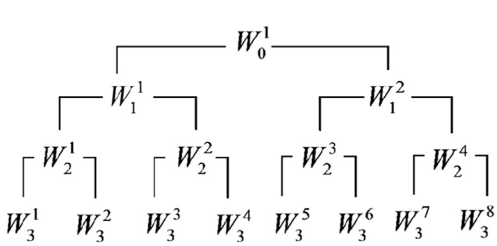

Fig. 2. Node numbers for the three-space decomposition by the WPD.

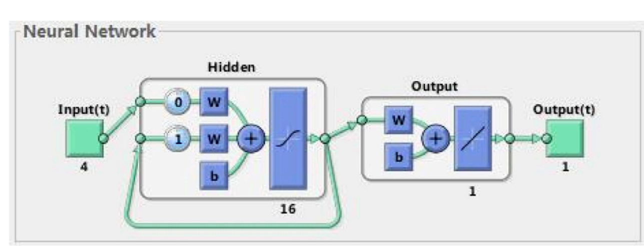

Fig. 3. Typical structure of the applied Elman neural network from the MATLAB Neural Network Toolbox (Beale et al., 2014).