Applied Energy 226 (2018) 756–771

Contents lists available at ScienceDirect

Applied Energy

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

A hybrid system for short-term wind speed forecasting

Qingqing He[a], Jianzhou Wang[a,*], Haiyan Lu[b]

[a] School of Statistics, Dongbei University of Finance and Economics, Dalian, China
[b] School of Software, Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia

HIGHLIGHTS

- Develop a novel similarity-based short-term wind speed forecasting system.
- Propose a valid strategy of sample selection.
- Improve the similarity of building sample in the case of continuity.
- Discuss the relationship between the similarity with the modelling requirement.

ARTICLE INFO

Keywords:
Kernel-based fuzzy c-means clustering
Ensemble empirical mode decomposition
Wavelet neural networks
Short-term wind speed forecasting

ABSTRACT

Wind speed forecasting is important for high-efficiency utilization of wind energy. Correspondingly, numerous researchers have always focused on the development of reliable forecasting models of wind speed, which is often noisy, unstable and irregular. Current approaches could adapt to various wind speed data. However, many of these usually ignore the importance of the selection of the modeling sample, which often results in poor forecasting performance. In this study, a hybrid forecasting system is proposed that contains three modules: data preprocessing, data clustering, and forecasting modules. In this system, the decomposing technique is applied to reduce the influence of noise within the raw data series to obtain a more stable sequence that is conducive to extract traits from the original data. To extract the characteristic of similarity within wind speed data, a kernel-based fuzzy c-means clustering algorithm is used in data clustering module. In the forecasting module, a sample with a highly similar fluctuation pattern is selected as training dataset, and which could reduce the training requirement of model to improve the forecasting accuracy. The experimental results indicate that the developed system outperforms the discussed traditional forecasting models with respect to forecasting accuracy.

1. Introduction

With the depletion of conventional resources in recent years and with the lack of enforced uses of emission treatment methodologies, environmental pollution has become a global concern. Therefore, accelerating the exploration and exploitation of alternative renewable energy is a pivotal step in settling the pollution problem [1]. Correspondingly, it has acquired considerable importance for the sake of future sustainable energy development. As one of the most active renewable resources, wind power, owing to its abundance and cleanliness, has attracted widespread attention around the world during the last few decades. Although the widespread use of wind power elicits many benefits, nevertheless, there are also some challenges for wind farms. For example, the development of a scheme for the balanced supply of power based on the demand of electricity is a difficult problem [2]. Furthermore, wind speed plays an important role in wind-power generation [3]. Therefore, efficient applications of wind energy based on accurate wind speed forecasting is a vital component of the operations of wind farms that concurrently provide the solution to the energy supply and demand balance problem [4]. However, wind speed is noisy, irregular, and easily affected by weather and geographical factors that drive the fluctuations of wind speed and increase the difficulties of precise forecasting [5]. Moreover, the accuracy of wind speed forecasting is not only determined by the forecasting cycle and the wind speed traits of the forecasted geographical locations, but is also dependent on the forecasting methods [4]. Many forecasting methods have been proposed in prior studies that have focused on the improvement of the methodological accuracy based on the different forecasting periods and diverse traits of forecasting sites. In general, the literature shows that the mainstream forecasting methods can be roughly divided into two categories, namely, physical models and statistical models [6].

* Corresponding author at: School of Statistics, Dongbei University of Finance and Economics of China, Dalian 116000, China.
E-mail address: wangjz@dufe.edu.cn (J. Wang).

https://doi.org/10.1016/j.apenergy.2018.06.053
Received 12 November 2017; Received in revised form 4 June 2018; Accepted 8 June 2018

The physical forecasting models use the detailed physical properties and historical wind speed data to construct forecasting models. This kind of methods assumes that the historical relationship between wind speed data and related physical information will remain valid in the future [7]. They mainly include mesoscale numerical models and computational fluid dynamic methods, and the input variables of these models are commonly physical or meteorology information. As such, they require considerable amounts of observed data at limited simulation scales, and consume excessive computing resources, which are expensive and difficult to obtain. Conversely, these methods are more suitable for long-term forecasting rather than short-term forecasting [8].

Statistical models have been extensively employed in recent studies for forecasting short-term wind speed. Methods, such as conventional statistical models, and statistical learning methods, have been extensively used in developing wind speed forecasts. Conventional statistical forecasting models, which only use prior wind speed data series, build statistical models based on existing statistical equations [9]. Statistical learning models usually build a high dimensional non-linear function to fit the wind speed, and the factors they influenced by minimizing the training error [10].

The most commonly used conventional statistical models include the autoregressive (AR), autoregressive moving average (ARMA), and the autoregressive integrated moving average (ARIMA) models. AR is a special form of ARMA, and the generalization of the ARMA model is considered to be the ARIMA for the sake of straightforward implementation of ARMA. Schlink and Tezlaff applied the AR for wind speed forecasting tasks at an airport [11,12], and the results showed that the width of intervals produced by AR were narrower than the intervals generated by the persistence model [12]. Lydia et al. forecasted wind speed based on linear and nonlinear ARMA models with and without external variables [13]. In [14], the fractional-ARIMA model was used to forecast wind speed for upcoming two-day periods with a higher accuracy compared to the persistence model. However, the limitations of these time series approaches concluded poor extrapolation effects and narrowed the forecasting scale.

With the rapid development of computer technology, the ability to perform complex calculation has significantly improved. A rapid increase of statistical learning models has been documented during the last few years, and these have led to the development of mature theoretical systems. The well-known ANN is one of extensively used statistical learning methods for wind speed forecasting, which can produce nonlinear maps and perform parallel processing. These approaches mainly include back propagation (BP) [15], radial basis function (RBF) [16], Elman neural network (ENN) [17] and wavelet neural network (WNN) [18], and others. They commonly consist of an input layer, one or more hidden layers, and an output layer. Each layer has some artificial neurons which are connected with the neurons of the previous layer with a connection function. The results show that the different structures of the network lead to different wind speed forecasting performances [19,20]. Thus, the instability and the dependence on data thus become an obvious defect for these models. Statistical machine learning methods emphasize the ability of generalization, the importance of regression perception machines, and the design the learning algorithms based on the generalization indices. They commonly contain support vector machines (SVM) [21], least-squares support vector machines (LSSVM) [22], Gaussian processes [23], and others. These simulation methods can be easily understood and combined with other approaches. In addition, the machine learning approaches elicit a more accurate performance when compared to ANNs, but they usually suffer the perplexity of over-fitting. The fuzzy logic is a research area based on the principles of approximate reasoning and computational intelligence [24]. Using fuzzy relations, the appropriate index is obtained for the assessment of system reliability [25]. The author in Ref. [26] used an adaptive Nero-fuzzy inference system (ANFIS) which was implemented as a hybrid fuzzy logic system with ANN. This system was also employed for Sugeno fuzzy modeling, and had two inputs with four fuzzy rules [27]. Nevertheless, this type of method has a relatively low accuracy and lacks systematization.

However, owing to the inner irregularity, instability and noise of the raw wind speed series, the individual models listed above generally depend on a large amount of historical data to build models that would elicit precise forecasting results. As such, this large data volume could lead to a complex calculation dilemma of the model that causes poor forecasting performance. Conversely, they are often limited in achieving a better fitness to the original data by resorting to tedious and time-consuming trials. As a result, it is very difficult to fit a model to them based on the single use of the conventional physical or statistical methods, in the effort to forecast short-term wind speed from these noisy and irregular datasets.

Therefore, in this study, a powerful hybrid system is developed that comprises three modules and primarily adopts a sample with a highly similar fluctuation pattern to establish forecasting model. The ensemble empirical mode decomposition (EEMD) technique is used in data preprocessing module, and a kernel-based fuzzy c-means clustering (KFCM) algorithm is employed as a data clustering module. Selecting of a sample with distinctive traits of similarity to establish wavelet neural network (WNN) model to conduct the final forecasts defines the forecasting module. Specifically, EEMD technology is efficient to analyze nonlinear and non-stationary data. Furthermore, Shouxiang Wang et al. [28] showed that the hybrid method EEMD-GA-BP is more accurate than the traditional GA-BP model in wind speed forecasting. The KFCM algorithm, which is an extension of a conventional fuzzy c-means clustering method, mainly alters perfectly the input data into a higher dimensional feature space [29], and thus overcomes the drawback of imprecise clustering caused by lower dimensions.

Generally, the innovations of this study can be summarized as follows:

(1) This study proposes a hybrid system that successfully takes advantages of data preprocessing technique and data clustering algorithm to improve the forecasting ability of WNN model on short-term wind speed. The novelty of the EEMD mainly relates to the inference of the decomposition of the wind speed series, and is followed by the removal of high frequency signals to obtain a smoother series. As such, it can be conducive to extract data features, thereby allowing the improvement of the forecasting accuracy of the models.
(2) The common approach adopted is the direct use of the denoising results to train the forecasting models. However, there is an innovation in this proposed methodology that employs the KFCM algorithm to extract data characteristics with similarities before the training.
(3) The data clustering module obtains samples with highly similar fluctuation patterns, which could reduce the training requirements for the forecasting model to improve the accuracy of short-term wind speed forecasting. In addition, this feature has been discussed in Section 5.2.
(4) It is the first time where the relationship between the degree of similarity within the training sample and the forecasting performance of the neural network model is explored, as stated in Section 5.3.
(5) Selecting data points with similar frequencies of fluctuation to produce the inputs of the model discontinuously makes the forecasting results unreasonable. This is caused by the intermittent within input vector. In this case, the settlement in this proposed hybrid system is the choice of the constructed input vectors as the inputs of the model. As a result, the behavior of the clustering approach considers both the similarity within the sample and its inner continuity.

This study is organized as follows. Section 2 outlines the principles