Renewable Energy 78 (2015) 599–608

Contents lists available at ScienceDirect

Renewable Energy

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

Short-term wind speed forecasting based on spectral clustering and optimised echo state networks

Da Liu*, Jilong Wang, Hui Wang

School of Economics and Management, North China Electric Power University, Beijing 102206, China

A R T I C L E I N F O

Article history:
Received 16 August 2014
Accepted 10 January 2015
Available online

Keywords:
Short-term wind speed forecasting
Spectral clustering
Echo state network
Wavelet transformation
Genetic algorithm

A B S T R A C T

Predicting the wind speed at multiple time points over a time span between two and 4 h typically requires a multi-input/multi-output model. This study investigates a wind speed forecasting method based on spectral clustering (SC) and echo state networks (ESNs). A wavelet transformation was used to decompose the wind speed into multiple series to eliminate irregular fluctuation. The decomposed series were modelled separately. For every decomposed wind speed series, principal component analysis was used to reduce the number of variables and thus the redundant information among the input variables. SC was used to select similar samples from the historical data to form training and validation sets. An ESN was used to simultaneously predict multiple outputs, and a genetic algorithm was employed to optimise the ESN parameters and ensure the forecast accuracy and the generalisation of the model. The forecasts of the decomposed series were summed to get the wind speed. Tests based on actual data show that the proposed model can simultaneously forecast wind speeds at multiple time points with high efficiency, and the accuracy of the proposed model is significantly higher than that of the traditional models.

1. Introduction

Wind power is a clean, renewable energy source with great potential, and countries around the world have increasingly focused on wind power in recent years. Accurately forecasting the wind power or the wind speed of a wind farm can improve the integration of wind power into the grid and ensure the stability of the electric power system when wind power is included. Accurately forecasting the wind speed over periods of 2–4 h in advance is of great importance for integrating wind power into the grid and maintaining the stability of the power system. Countries such as China and India have proposed standards for short-term wind speed forecasting. For example, the National Energy Administration of China requires that wind farms report wind power forecasts up to 4 h in advance in 15-min intervals, and the forecast error cannot exceed 15%.

Wind speed is easily affected by weather and topographic factors, such as temperature, humidity, air pressure gradient, and evaporation, which means that wind speed can exhibit large variation. In addition, wind speed data contain random fluctuations due to measurement errors and random factors, making accurate wind speed forecasts difficult [1]. The wavelet transform (WT) is often used to reduce the noise contained in wind speed data before modelling is attempted, and this method produces fairly good results [2–4]. The wind speed has a relatively high degree of autocorrelation, and the majority of previous studies relied on historical wind speed data to build models. When using highly correlated historical wind speed data as the input to a model, there is a large amount of redundant information that can make modelling difficult. Many researchers have used dimension reduction techniques such as principal component analysis (PCA) to eliminate redundant information and improve the accuracy of the forecast [5,6].

The wind speed exhibits extremely complex variations, and it is therefore difficult to use a single statistical model such as classic time series analysis (e.g., auto-regressive and moving average model) and soft computing (e.g., artificial neural networks), to capture all wind speed variation patterns [1]. Traditional wind speed forecasting methods typically use all wind speed data as a single sample space for modelling and do not distinguish the data based on their characteristics, making the rationality of these methods questionable. If samples that are similar in terms of fluctuation patterns are chosen to build a model for wind speed forecasting, then the training requirements for the model can be greatly reduced and the forecasting accuracy of the model improved.

* Corresponding author.
E-mail addresses: liuda315@163.com (D. Liu), 65492201@qq.com (J. Wang), wanghuibd@163.com (H. Wang).

http://dx.doi.org/10.1016/j.renene.2015.01.022
0960-1481/

Several studies have proposed the use of the cluster analysis method, which is based on similarity theory, to forecast power system loads and the wind speed [7,8]. Traditional clustering methods such as the K-means method may converge to a local minimum, which affects their applicability to forecasting short-term wind speeds. Spectral clustering (SC) is a new type of clustering method based on spectrum division theory. Compared to traditional clustering algorithms such as K-means, SC has advantages such as low computational complexity, applicability to sample spaces of any shape, the ability to perform cluster analysis on a non-convex distribution, and convergence to a global optimal solution. SC is now widely used in the fields of pattern recognition and forecasting [9,10]. SC has been one of the most popular research topics in the field of machine learning. However, this method has not been used for sample categorisation in the field of short-term wind speed forecasting.

Methods such as the time series and intelligent forecasting have been widely used in developing wind speed forecast models [4,11,12,and13]. The time series forecasting method uses historical time series data to construct forecasting models. However, the existing low-order models are not highly accurate, and the high-order models have difficulty in estimating parameters and cannot maintain accuracy in multi-step forecasts. Because short-term wind speed forecasting requires simultaneous wind speed forecasts at multiple time points, a traditional time series method cannot be applied effectively. Artificial neural networks (ANNs) are widely used as an intelligent forecasting method. This method can produce nonlinear maps and perform parallel processing. Neural networks are therefore suitable for multi-input/multi-output learning and can simultaneously forecast wind speeds at multiple time points [1]. However, ANNs require a complex training process, they may converge to a local minimum, they have difficulty in determining the optimum network structure, and they experience fading memory (FM). Therefore, it is difficult to create a more accurate wind speed model using ANNs. Echo state networks (ESNs) are a novel type of ANN proposed by Jaeger in 2001 and are regarded as the closest representation of the learning process of the human brain [14]. ESNs can effectively solve all of the aforementioned problems with ANNs. The ESN model requires a simple training process and has short-term memory. Thus, this model has found widespread use and success in the field of time series forecasting for applications such as stock prices and power system loads [2,15–18]. An ESN can process multiple inputs and multiple outputs with such high efficiency that it is capable of meeting the requirements for real-time, multi-step, short-term forecasting of wind speeds. This study attempts to combine the WT and PCA for wind speed data processing and then use SC to select a proper sample set with which to construct an ESN model that can predict wind speeds over short periods. In addition, a genetic algorithm (GA) is used to optimise the ESN parameters.

The structure of this paper is as follows. Section 2 explains the basic modelling theory and introduces the principles and advantages of WT, PCA, GA, SC, and the ESN. Section 3 explains the development of the SC-ESN-GA model, and Section 4 describes the numerical examples used to verify the effectiveness of the methods proposed in this study. Section 5 discusses the results and presents conclusions.

2. Modelling theory

2.1. WT, PCA, and GA

The WT is introduced to eliminate noise within the wind speed data, which typically contain a high level of noise. Furthermore, because of the large number of inputs to the wind speed model, PCA is introduced to reduce the dimension of the input data. Through scaling and shifting, a WT performs multi-scale, multi-resolution analysis on signals. This method analyses a signal using a fixed-width time window, although the scale is adjustable. Therefore, a WT can effectively analyse transient and non-stationary signals. Selecting a wavelet function and determining the decomposition layers are vitally important when using a WT to process signals.

The basic concept behind PCA is to transform a given set of correlated variables into a set of uncorrelated variables through a linear transformation. The new variables are chosen based on their variances, which are arranged in descending order, to thus use the fewest number of variables possible to represent the information contained in the original variables.

GAs are heuristic, parallel, global search methods based on the principles of natural selection and genetics. Typically, GAs are used to solve multi-parameter optimisation problems and consist of steps such as parameter initialisation, member selection, and chromosome crossover and mutation. It is critical that appropriate fitness functions be chosen when applying a GA to solve a parameter optimisation problem [12].

2.2. SC

SC transforms a clustering problem into an optimal spectral partitioning problem. Based on an eigenvector matrix, the SC method performs spectral partitioning according to the weights among the vertices. The basic concept behind SC is to cluster individual data points from data sets $\mathbf{D}=\{d_1,d_2,...d_n\}$ at the vertices of an undirected, weighted graph $G(\mathbf{D},\mathbf{W})$. The distances between data points d_i and d_j are then converted into a weighted graph edge w_{ij}. In this fashion, the data clustering problem is transformed into a spectral partitioning problem.

The partitioning strategy greatly affects the results given by SC clustering. Classic partitioning strategies include the normalised-cuts algorithm [19] and the Ng-Jordan-Weiss (NJW) algorithm [20]. Implementations of these algorithms vary slightly, although all are essentially continuous relaxations of the spectral partitioning problem. In short, SC can be summarised by the following three steps:

Step 1: Construct a similarity matrix $\mathbf{Z}$ that represents the sample data.
Step 2: Calculate the first k eigenvalues and eigenvectors and construct the eigenvector space.
Step 3: Cluster the eigenvectors within the eigenvector space by applying a clustering algorithm such as k-means.

2.3. ESNs

A standard ESN is composed of three layers: an input layer, a hidden layer, and an output layer. The network structure is shown in Fig. 1 [21]. The ESN forms an “input-state-output” system. As indicated in Fig. 1, the structure of an ESN is composed of a large number of randomly generated, sparsely connected neurons that form a large dynamic reservoir (DR). This DR is a core component of an ESN. The DR is excited by the input signal and generates a continuous state variable signal (i.e., a network “echo”). Through a linear combination of the echo signal and the target output signal, the ESN output weights can be determined to forecast a chaotic time series.

At time$(n+1)$, $\mathbf{U}(n+1)$ is the input vector, the hidden layer state is $\mathbf{X}(n+1)$, and the network output $\mathbf{Y}(n+1)$ can be obtained using Equations (1) and (2):