Renewable Energy 115 (2018) 575–584

Contents lists available at ScienceDirect

Renewable Energy

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

Short-term wind power forecasts by a synthetical similar time series data mining method

Gaiping Sun [a, b], Chuanwen Jiang [a, *], Pan Cheng [c], Yangyang Liu [a], Xu Wang [a], Yang Fu [b], Yang He [d]

[a] School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[b] School of Electric Power Engineering, Shanghai University of Electric Power, Shanghai 200090, China
[c] Shanghai Aircraft Design & Research Institute, Shanghai 201210, China
[d] State Grid Henan Electric Power Corporation, Zhengzhou 450052, China

ARTICLE INFO

Article history:
Received 3 November 2016
Received in revised form
27 May 2017
Accepted 26 August 2017
Available online 28 August 2017

Keywords:
Wind power forecasts
Hybrid clustering method
Similarity measure
Wavelet neural network

ABSTRACT

As the aggravating influence of growing wind power, wind power forecasting research becomes more important in economic operation and safety management of power system. A novel short-term wind power forecasting methodology consists of a hybrid clustering method and a wavelet based neural network is introduced. The clustering similar measure function combines the Euclidean Distance and Angle Cosine together, aims to identify the similar wind speed days which are close in space distance and have similar variance trend synthetically. Then similar daily samples as the predicting days are treated as training samples of an improved particle swarm optimization based wavelet neural network. The proposed forecasting strategy is applied to two real wind farms in China. The results demonstrate that the strategy can identify the similar time series and improve the predicting accuracy effectively, compared with some other forecasting models.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Wind power generation has been growing fast in the last few decades since the governments take more concern on the global warming problem, which may be caused by the traditional energy [1]. Wind power could be easily fluctuating by the stochastic wind. Therefore, the random wind power may affect the power quality of the power grid, resulting in the power grid voltage fluctuations, flicker and harmonic problems. It will also make affection on the stability of the power grid and reduce the reliability of the power system operation [2–5]. An accurate prediction of short-term wind power output will be advantageous for the independent system operators (ISO) to adjust the scheduling plan in time, make the correct decision, reduce the standby capacity, cut down the operation cost of the power system and weak the adverse effect of the wind power fluctuation.

Wind power forecasting methods are mainly divided into three categories at present [6,7]. The first one is physical approach based on meteorological, many physical factors should be considered at building a physical forecasting model such as temperature, humidity, terrain, surface roughness, hub height, etc. Statistical model is the second method aims to find the relationship between the input variables and the output wind power by using history data sets. Such as Autoregressive Integrated Moving Average Model (ARIMA) [8], Kalman filter [9], Bayesian [10], Support Vector Machine (SVM) [11] and artificial neural network (ANN) [12–14]. Physical methods always take advantages in long-term and large scale forecast while statistical method is good at very short term forecast. In order to improve the predicting accuracy and time horizon, researchers always use a third method which combines physical method and statistical method together. For example, a statistical model employs the numerical weather prediction (such as predicted values of wind speed) as part of input variables. Many literature approve utilizing hybrid approaches may perform a better forecasting result [15–18].

The selection of training data has a great influence on the establishment of a statistical forecasting model. For example, in a daily wind power forecasting approach, nonlinear relationship between the input variables and output wind power could be easier to construct if the training samples are similar as the predicting day.

* Corresponding author.
E-mail address: jiangcw@sjtu.edu.cn (C. Jiang).

http://dx.doi.org/10.1016/j.renene.2017.08.071
0960-1481/© 2017 Elsevier Ltd. All rights reserved.

Literature [19] clusters the historical data sets by two different horizon: one day and continuous time series (half a month), then the whole year is divided into several categories, the classified data is used in prediction for the same time period of the next year. Literature [20] extracted similar data from large amounts of data to create training samples. The similarity judgment is decided by the ratio of each linear segment. Then a support vector machine (SVM) is used to predict the output wind power. All of the results show that a great improvement is achieved compared with other models without filtering history data.

However, in previous literature, researchers just focus on one side of time series segments similarity analysis. Actually, similarity of samples is multi-side: it could represent time series are close to each other in space distance, or have the same curve shape. In this paper, both of the similar measures are considered in identifying the similar wind speed time series. A novel short-term wind power forecasting strategy is proposed, and the mainly contributions are provided as follows:

1) A hybrid clustering method is carried out to identify the similar historical daily wind time series as the predicting days', since the training data sets affect the accuracy of a statistical forecasting model output to some extent. The clustering similar measure function combines the Euclidean Distance and Angle Cosine together, aim to identify the similar wind speed days which are close in space distance and have similar variance trend synthetically. The weight of each variation feature is calculated by a ReliefF algorithm. Then an improved particle swarm optimization (IPSO) based K-means algorithm is applied to accelerated clustering process and achieve the optimal clustering result.
2) The clustering similar daily samples as the predicting days' are treated as training samples of a novel predicting model. The forecasting model is composed of an improved particle swarm optimized algorithm mentioned before and a wavelet neural network (IPSO-WNN). The forecasting model performs a great predicting capacity with great global search ability of IPSO and powerful learns ability of wavelet.

The paper is organized as follows: in section 2, the proposed short-term wind power forecasting model is introduced in detail. In section 3, the data sets of two wind farms in China are provided and the benchmark model and evaluation index are described. Section 4 illustrates the performance of the proposed wind power strategy by comparing the obtained result with other benchmark models. Section 5 concludes this paper.

2. Proposed forecasting model

The forecasting strategy is composed of a hybrid clustering method and a wavelet neural network. Fig. 1 shows the structure of the proposed forecasting strategy. The hybrid clustering component consists of a hybrid similarity measure, ReliefF based attribute weight, and IPSO-K-means algorithm, is introduced in 2.1. The IPSO-WNN forecasting method is presented in 2.2.

2.1. Novel hybrid time series clustering method

2.1.1. Similarity measure function

Based on the previous introduction, it is advantageous to cluster days with similar wind speed variation of daily wind power prediction. The function of similarity measure estimates the similarity degree of two samples. The expression of similarity measure could be varied in different applications [21]. Take time series for example, similarity measures commonly include space distance

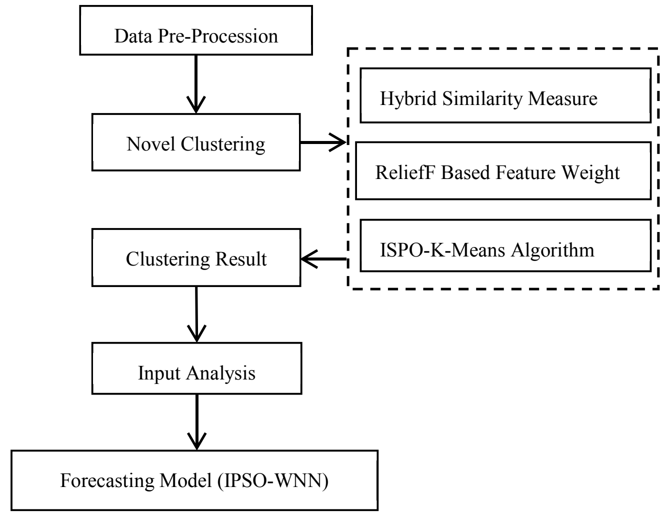

Fig. 1. Structure of the proposed forecast model.

method and similar coefficient method.

Euclidean distance (ED) is one of the most classic similarity distance measure in the data mining problem [22]. Multidimensional data are assigned to each cluster based on the space distance calculated by ED. Consider a set of data samples $X = \{x_{ij} | i = 1, 2, \ldots, n, j = 1, 2, \ldots, m\}$, n is the number of the samples, m is the dimension of the samples. The ED between two samples is defined as:

$$d_{euc}(x_i, x_k) = \left(\sum_{j=1}^{m} \left(x_{ij} - x_{kj} \right)^2 \right)^{\frac{1}{2}} \quad (1)$$

The value represents the closeness of two samples in the multidimensional space. The smaller of the value is, the more similar of the two samples are.

Angle cosine (AC) is applied as a similar coefficient method in the clustering similarity measure [22]. Compared with ED, AC is focused on the curve shape of time series. And the function of AC is shown as:

$$d_{cos}(x_i, x_k) = \frac{\sum_{j=1}^{m} x_{ij} \cdot x_{kj}}{\sqrt{\sum_{j=1}^{m} \left(x_{ij}\right)^2} \cdot \sqrt{\sum_{j=1}^{m} \left(x_{kj}\right)^2}} \quad (2)$$

The value range calculated by the AC is [-1, 1]. The direction of the two samples coincides when the angle cosine value is 1, the direction of the two samples exactly the opposite when the angle cosine value is -1. The result indicates that the bigger of the value is, the more similar of the two samples are. In order to make a same conclusion with ED method, AC method is modified to one minus the cosine of the included angle between sample vectors in this paper.

Since the analysis of ED and AC represent different characteristic of similarity, different clustering results may be produced. Given a dataset $\{x_1, x_2, x_3, x_4\}$, they represent 4-dimensional samples shown in Fig. 2. $x_1 = (2, 3, 2, 3)$, $x_2 = (3, 2, 3, 2)$, $x_3 = (4, 6, 4, 6)$, $x_4 = (6, 4, 6, 4)$. When ED is applied as the similarity measure and the dataset are classified to 2 categories, x_1, x_4 are similar, and x_2, x_3 are similar. When angle cosine is used as the metric function, x_1, x_3 are similar and x_2, x_4 are similar. Obviously, different clustering results are caused since two similar metric reflect different essences. ED mainly reflects the space distance of Multidimensional

Download English Version:

https://daneshyari.com/en/article/6765255

Download Persian Version:

https://daneshyari.com/article/6765255

Daneshyari.com