Renewable Energy 89 (2016) 268–284

Contents lists available at ScienceDirect

Renewable Energy

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

A novel hybrid model based on artificial neural networks for solar radiation prediction

Yujie Wu [a], Jianzhou Wang [b, *]

[a] School of Mathematics and Statistics, Lanzhou University, Lanzhou 730000, China
[b] School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China

ARTICLE INFO

Article history:
Received 27 July 2015
Received in revised form
26 November 2015
Accepted 28 November 2015
Available online xxx

Keywords:
Time series strategy
Multi-step-ahead predict
OPELM
Energy prediction

ABSTRACT

As a kind of clean, substantial and renewable energy, solar energy can reduce environmental pollution with an extensive application potential. Precise prediction of global solar radiation has great significance for the design of solar energy systems and management of solar power plants.

In this paper, a new hybrid model combining the SOM-OPELM with time series strategies is presented for predicting the global solar radiation on the horizon. In this model, the SOM divides the original data into distinct clusters and the OPELM establishes the prediction model. Subsequently, three population time series strategies, (i.e. Recursive strategy, DirRec strategy and MISMO strategy) are adopted to accomplish the multi-step prediction. A comparison between the proposed SOM-OPELM model and other conventional methods is carried out to demonstrate its efficiency and estimation performance. The simulation results show that the proposed SOM-OPELM model with DirRec strategy or MISMO strategy outperforms the following models: Recursive-BP, DirRec-BP, MISMO-BP, Recursive-SOM-OPELM and ARIMA.

1. Introduction

Due to the depletion of the conventional energy resources, the demand for renewable alternatives has soared, among which, the pollution-free solar energy has gained worldwide attention.

Precise prediction of solar radiation is important for the design of solar power plants and other systems. Besides, it has been applied to our lives, such as the design for agriculture, water resources and passive heating of buildings [1].

In view of the high correlation between weather conditions and the solar radiation received on earth, it is necessary to predict with several climatic parameters, including relative humidity, daily clear-sky global radiation, total cloud cover, maximum ambient temperature, average temperature, average cloudiness and average wind velocity, etc. [2]. The information about geographic position, such as latitude, longitude and altitude, is usually a part of the prediction model as well.

As a prediction technique, the artificial neural network (ANN) has been widely applied in the solar power field [2–7], such as the global solar radiation prediction [8] and the estimation of photovoltaic energy generation [9].

K. Kadirgama et al. [3] have developed the quick propagation algorithms Artificial Neural Network (ANN) models to estimate the solar radiation in Pekan located in Pahang. The results showed that the MAPE by the presented model is less than 7.74% and R-squared values are approximately 98.9%.

Cyril Voyant et al. [4] proposed a hybrid ARMA/ANN model for the estimation of hourly global radiation in five places in five Mediterranean locations. Results demonstrate that, the RMSE with this hybrid model is 14.9% compared with 18.4% with the original ANN model.

Yingni Jiang [5] applied a feed-forward back-propagation algorithm to estimate mean daily global solar radiation by month of eight typical cities in China. The results (MPE = 0.48, MBE = −0.021,

* Corresponding author.
E-mail address: wjz@lzu.edu.cn (J. Wang).

http://dx.doi.org/10.1016/j.renene.2015.11.070
0960-1481/

RMSE = 0.867) achieved high similarity with observations of global solar radiation.

In Cao research [6], back-propagation (BP) network and wavelet analysis technology have been combined to forecast solar radiation. To validate it, the historical record of daily solar irradiance of Shanghai was used. The results indicated that the proposed model is much more accurate than the ANN model alone.

Mohamed Beng Hanem [7] et al. investigated Radial Basis Function network (RBF) for predicting the daily global solar radiation recorded during 1998–2002 in Al-Madinah, Saudi Arabia. In this investigation, the parameters inputs were temperature, sunshine duration, and relative humidity. By selecting optimal input, the best RBF model achieves a high correlation coefficient (0.9880).

In addition to the traditional neuron network, Huang et al. [10,11] have recently proposed a simpler and more efficient learning algorithm designed for single-hidden layer feed forward neural networks (SLFNs). They call extreme learning machine (ELM). Compared with the conventional neural network methods, the miscellaneous hyper-parameters of which need to be adjusted, quite often leading to heavy computation, ELM has not only a relatively simple structure but also a comparable accuracy.

The main problem in the application of the ELM method is determining the optimum number of the hidden neurons to use. On the one hand, too many hidden neurons can significantly increase the computational complexity. On the other hand, too few would influence the accuracy. Many researchers have managed to handle this problem via neuron pruning [12]. Based on this approach, Miche et al. have proposed a modification, named optimally pruned ELM (OPELM). In this method, Multiresponse Sparse Regression (MASR) [13] is usually regarded as an extension of LARS [14] to prune the useless neurons. Subsequently, Leave-One-Out (LOO) validation method is adopted to determine the optimal number of the hidden neuron. More details about OPELM are explained in Section 3.

One notable fact is that most of the methods mentioned above are utilized to predict the solar irradiance in one step hour by hour or day by day and cannot predict in multiple steps by themselves. Given this situation, time series strategies have garnered much international attention recently for their good prediction performance in fields ranging from engineering to biology and to economics.

The predictive step of these strategies can be roughly classified into two categories: one-step ahead prediction and multi-step ahead prediction. The main difference between these strategies lies in how the historical values are utilized to estimate the future value. Generally speaking, due to the lack of information and the accumulation of errors, multi-step-ahead prediction cannot achieve the same accuracy as one-step ahead prediction [15].

Currently, in addition to traditional statistical methods, e.g., autoregressive (AR), moving average (MA) and autoregressive moving (ARMA) models, more and more prediction strategies have emerged [16], for example, the Recursive strategy, Direct strategy, DirRec strategy, MIMO strategy and MISMO strategy [17,18]. Research from Ref. [18] supports conclusions that the DirRec strategy can outperform the Recursive one. However [19], shows that Recursive strategy gives smaller SMAPE values compared with the DirRec strategy and that the latter provides a worsen prediction than others overall (Detailed descriptions of these strategies are given in Section 2).

To our knowledge, as the prediction strategies continue to proliferate, which strategy outperforms the others has not been confirmed yet. The work by Ref. [19] supports the assertion that prediction accuracy is influenced by the degree of dependence between input and output. When the dependence is distinct, the MISMO strategy performs better than the MIMO and Direct strategy. Therefore, the best strategy can be contingent on the selected prediction data.

In this study, a hybrid model based on the SOM and OPELM models, referred to as SOM-OPELM, is proposed to explore the optimum prediction strategy for solar radiation and improve the estimation accuracy. We refer to it as SOM-OPELM. Recursive, DirRec and MISMO strategies as distinct instances of the same prediction methods are used. To begin, the SOM network is used to split the historical data into clusters dynamically. Afterwards, an optimally pruned extreme learning machine (OPELM) is adopted to establish the prediction model. Subsequently, multi-step-ahead prediction can be directly or indirectly implemented by combining with time series strategies.

This paper is organized as follows. Section 2 provides general descriptions of time series strategy. In Section 3, the descriptions of SOM and OPELM model are presented, and the overall structure of SOM-OPELM model will be given as well. Section 4 illustrates the numerical results from two experiments, and discussion the predicted performance of our hybrid models. Finally, the conclusions are presented in Section 5.

2. Time series strategies

This section gives a brief overview of three time series prediction strategies used in our study. For narrative coherence and to give readers a better understanding of the nature of each method, we explain another two types of time strategies: Direct strategy and MIMO strategy and show that the MISMO strategy likely to replace of both. That's why only MISMO strategy implemented in our experiment.

2.1. Recursive strategy

The Recursive strategy, as an intuitive and venerable prediction strategy [20] [21], has appeared on many statistics studies. This method uses historical data $\{y_t, ..., y_{N-r}\}$, where r is the numbers of previous observations that can be determined by cross-validation. N is the number of data samples, to estimate the next one-step output. Then this output value is fed back to predict subsequent values, and this process iterates until reaching the horizon H. Here, H is the prediction horizon, which is a hyper-parameter specified by actual needs. Mathematically, this is expressed by:

$$y_{t+1} = f(y_t, \cdots\cdots, y_{t-r+1}) + w \quad (1)$$

with $t \in \{r, ..., N-1\}$. The algorithm description is shown as follows: