Energy Conversion and Management 95 (2015) 42–58

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Intelligent optimization models based on hard-ridge penalty and RBF for forecasting global solar radiation

He Jiang [a,b], Yao Dong [b,*], Jianzhou Wang [c], Yuqin Li [d]

[a] Department of Statistics, Florida State University, Tallahassee, FL 32306-4330, USA
[b] School of Mathematics and Statistics, Lanzhou University, Lanzhou 730000, China
[c] School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China
[d] Lanzhou University of Technology, College of Computer and Communication, Lanzhou 730050, China

ARTICLE INFO

Article history:
Received 10 October 2014
Accepted 6 February 2015
Available online 24 February 2015

Keywords:
Global solar radiation forecasting
RBF neural network
Hard-ridge penalty
Cuckoo search algorithm
Differential evolution

ABSTRACT

Due to the scarcity of equipment and the high costs of maintenance, far fewer observations of solar radiation are made than observations of temperature, precipitation and other weather factors. Therefore, it is increasingly important to study several relevant meteorological factors to accurately forecast solar radiation. For this research, monthly average global solar radiation and 12 meteorological parameters from 1998 to 2010 at four sites in the United States were collected. Pearson correlation coefficients and Apriori association rules were successfully used to analyze correlations between the data, which provided a basis for these relative parameters as input variables. Two effective and innovative methods were developed to forecast monthly average global solar radiation by converting a RBF neural network into a multiple linear regression problem, adding a hard-ridge penalty to reduce the number of nodes in the hidden layer, and applying intelligent optimization algorithms, such as the cuckoo search algorithm (CS) and differential evolution (DE), to determine the optimal center and scale parameters. The experimental results show that the proposed models produce much more accurate forecasts than other models.

1. Introduction

As a clean, renewable and sustainable energy resource, solar energy has become a very attractive alternative for modern industrialized society. It has established itself in both small-scale and large-scale power generation. Electricity can be produced from solar radiation using either non-concentrated photovoltaic (PV) modules or concentrated solar thermal (CST) systems [1]. For efficient utilization and conversion of solar power, it is essential to know solar radiation data continuously and accurately. Therefore, accurate solar radiation forecasts have become increasingly significant [2,3].

Many researchers have tried to develop physical and statistical methods to forecast solar radiation. Numeric weather prediction (NWP), as a meteorology model, can be applied to forecast solar radiation using solar zenith angle and clear sky index [4,5]. In fact, solar radiation data can be considered to be time series, and several time series approaches can also be used. Classical statistical models for time series, for example autoregressive integrate moving average (ARIMA) model, have been widely employed to model time series data. Although ARIMA model is very flexible for different types of data with appropriate order, the model requires the time series to be stationary. Thus, it is necessary to convert nonstationary time series into stationary ones before building models [6]. To avoid preprocessing, nonlinear mathematical models, such as artificial neural networks (ANNs) and support vector machines (SVMs), can reveal nonlinear relationships between past and future data. The former can be employed to forecast global solar irradiation [3]. For the latter, a machine-learning algorithm can be used to estimate monthly mean daily solar radiation [7]. In addition, some combined methods can also produce satisfactory forecasting performances. Wavelet-networks, which combine wavelet theory and neural networks, have been proposed to forecast daily total solar radiation at a meteorological station in Algeria [8]. Satellite image analysis and a hybrid exponential smoothing state space with ANN have been employed to forecast hourly solar irradiance time series in Singapore [9]. Based on Bayesian rules, a hybridization of three methods, auto regressive and moving average (ARMA), persistence model and multi-layer perceptron (MLP), enabled better forecasting accuracy than the single method [10]. Although the above models have demonstrated admirable

* Corresponding author. Tel.: +86 931 8912483; fax: +86 931 8912481.
E-mail address: dongyao20051987@aliyun.com (Y. Dong).

http://dx.doi.org/10.1016/j.enconman.2015.02.020
0196-8904/© 2015 Elsevier Ltd. All rights reserved.

Nomenclature

Abbreviation

AIC Akaike information criterion
ARMA auto regressive moving average
ARIMA autoregressive integrate moving average
ANN artificial neural networks
AVGDT average daylight temperatures
AVG-T average 24-h temperatures
AVWS average wind speed
CLDD average cooling degree days
CS cuckoo search
CS-hard-ridge-RBF the optimized model by CS based on hard-ridge and RBF
CST concentrated solar thermal
DE differential evolution
DE-hard-ridge-RBF the optimized model by DE based on hard-ridge and RBF
GMT Greenwich Mean Time
Hard-ridge-RBF the model based on hard-ridge and RBF
H_2O average precipitable water
HTDD average heating degree days
MAPE mean absolute percentage error
Min_sup minimum support
Min_conf Minimum confidence
MAX-T average maximum temperatures
MIN-T average minimum temperatures
MLP multi-layer perceptron
NSRDB National Solar Radiation Database
NWP Numeric Weather Prediction
OLS ordinary least square
OPQ average opaque sky cover
PV photovoltaic
RBF radial basis function
RH average relative humidity
RMSE root mean square error
SNR signal to noise ratio
SSE the sum of the squared error
SVMs support vector machines
TAU broadband aerosol optical depth
TOT average total opaque sky cover

English letter

$\mathbf{c}_j$ the center of hard-ridge-RBF
CR the crossover constant in DE
d the number of predictors
$df(\lambda)$ the degrees of freedom
D the number of parameters in DE, it is equal to $q \times d$
F the mutation factors in DE
G the number of iterations in DE
I set of items including $\{i_1, i_2, \ldots, i_m\}$
m the dimension of input samples
n the number of training samples
N the number of forecasting (test) samples
$Nest$ the number of nests in CS
$Ngen$ the number of maximum iterations in CS
NP population size in DE
$P(\boldsymbol{\theta}; \lambda)$ penalty function
p_a a probability with the egg laid by a cuckoo is discovered by the host bird
$P_{HardRidge}(\boldsymbol{\theta}; \lambda, \eta)$ Penalty function of hard-ridge
q the number of nodes in the hidden layer before dimension reduction
q_1 the number of nodes in the hidden layer before dimension reduction
TD set of transactions
$u_{i,G+1}$ trial vector in DE
$v_{i,G+1}$ mutant vector in DE
$x_{i,G}$ target vector in DE
$\mathbf{y}$ desired output vector

Greek letter

α the step size relating to the size of problem
$\boldsymbol{\varepsilon}$ errors vector between desired output and actual output
η the ridge parameter
$\Theta(\cdot; \lambda)$ a unbound, odd, monotone shrinkage rule
$\boldsymbol{\theta}$ the output weights vector of hard-ridge-RBF
$\hat{\boldsymbol{\theta}}_{HR}$ estimation by hard-ridge penalty
$\hat{\boldsymbol{\theta}}_{OLS}$ estimation by ordinary least square (OLS)
λ the tuning parameter
λ' step length generated by a Lévy distribution.
λ_j the eigenvalues of $\boldsymbol{\Psi}^T\boldsymbol{\Psi}$
σ_j the scale parameter of hard-ridge-RBF
$\boldsymbol{\Psi}$ the kernel function matrix of hard-ridge-RBF

forecasting capacity, none of them has been determined to be superior, because solar radiation can vary significantly in different countries or locations.

Furthermore, owing to the complex equipment and high costs of maintenance, it is critically difficult to observe solar radiation. Therefore, other meteorological parameters are urgently needed for calculating or forecasting solar radiation. In this paper, two novel optimized models for forecasting global solar radiation have been developed by incorporating intelligent optimization algorithms (cuckoo search (CS) [11] and differential evolution (DE) [12]) and a hard-ridge penalty [13] into a RBF neural network [14]. In an application of the proposed models, monthly average global solar radiation and 12 meteorological data were collected from four sites in the United States during the period 1998 to 2010. Based on analyses using statistical methods and a data-mining algorithm, these 12 meteorological parameters that are highly correlated with global solar radiation were applied as input variables to forecast global solar radiation. The empirical results clearly demonstrate that the proposed models outperformed traditional models for both prediction and interpretation. These two effective models have not been used in other fields; therefore we believe they can be innovative and helpful for forecasting solar radiation.

The new contributions of this study are as follows:

- The Pearson correlation coefficient was used to calculate the linear correlation between global solar radiation and each factor.
- Based on the results of Pearson correlation coefficient, Apriori association rules were built to analyze the influence of each factor and the joint effects of 12 factors to global solar radiation.
- This study presents two effective and innovative models that provide a compact solution for multivariate input variables problems, convert a RBF neural network into linear regression model and make use of the properties of hard-ridge penalty to reduce the number of nodes in the hidden layer to perform dimension reduction. In this way, a parsimonious model that is more interpretable was obtained.