Energy Conversion and Management 137 (2017) 142–154

Contents lists available at ScienceDirect

Energy Conversion and Management

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

A multi-stage intelligent approach based on an ensemble of two-way interaction model for forecasting the global horizontal radiation of India

He Jiang [a,b], Yao Dong [a,b,*], Ling Xiao [c]

[a] School of Statistics, Jiangxi University of Finance and Economics, Nanchang 330013, China
[b] Applied Statistics Research Center, Jiangxi University of Finance and Economics, Nanchang 330013, China
[c] School of Mathematics and Statistics, Lanzhou University, Lanzhou 730000, China

ARTICLE INFO

Article history:
Received 26 September 2016
Received in revised form 4 January 2017
Accepted 14 January 2017

Keywords:
Ensemble learning
Divide and Conquer
Glowworm swarm optimization
LASSO
Global horizontal radiation forecasting

ABSTRACT

Forecasting of effective solar irradiation has developed a huge interest in recent decades, mainly due to its various applications in grid connect photovoltaic installations. This paper develops and investigates an ensemble learning based multistage intelligent approach to forecast 5 days global horizontal radiation at four given locations of India. The two-way interaction model is considered with purpose of detecting the associated correlation between the features. The main structure of the novel method is the ensemble learning, which is based on Divide and Conquer principle, is applied to enhance the forecasting accuracy and model stability. An efficient feature selection method LASSO is performed in the input space with the regularization parameter selected by Cross-Validation. A weight vector which best represents the importance of each individual model in ensemble system is provided by glowworm swarm optimization. The combination of feature selection and parameter selection are helpful in creating the diversity of the ensemble learning. In order to illustrate the validity of the proposed method, the datasets at four different locations of the India are split into training and test datasets. The results of the real data experiments demonstrate the efficiency and efficacy of the proposed method comparing with other competitors.

1. Introduction

Owing to the rapid development of economics, India has become one of the fastest growing countries consuming energy and is expected to be the second-largest contributors by 2030 taking up about 18% of the increase in global energy consumption [1]. India depends heavily on fossil fuels to meet the demand of energy, and coal is primary resource for energy consumption (54.5%) following by crude oil (29.45%). However, burning fossil fuels makes the air quality bad and the increase of respiratory diseases make things worse. Therefore, energy access and energy security force India to be alert to the environmental issues and climate changes. To this end, it is urgent to explore renewable green energy. India owns the fifth largest wind power market in the world [2], and plans to increase approximately 100,000 MW of solar power capacity by 2020 [3]. India is a tropical country which locates along the equatorial belt with latitude lying between 10° and 30°. Its solar energy insolation is quite abundant (about 5000 T kW h per year), which is far richer than its current total primary energy consumption [4]. Accessibility of cheap solar power can avoid the installation of expensive grid lines and provides electricity to those underdeveloped areas. Meanwhile, these solar power plants are self-productive and do not rely on supply of any raw material. It is infinite scope to substitute solar power for fossil fuel energy [5].

The forecasting of global horizontal radiation plays a dominant role in maintaining the security, stability and balance between the supply and demand when connecting the access of solar power to the whole power system [6]. If none of the information about the global horizontal radiation is provided, photovoltaic brownouts will increase when integrating into power grid, which leads to a great reduction to the ability of power consumed by solar power and also the rate of return on investment of solar power plants. Therefore, it is a vital issue to forecast global horizontal radiation and its volatility [7]. However, affected by the weather and climate factors, global horizontal radiation is neither stable nor intensive. It is critically challenging to arrange the plan of solar power generation and scheduling power. The most accurate way of getting global horizontal radiation is to apply ground-based measurements. Nevertheless, the data collected is not very credible in that the measurement equipment is not adequate at many stations [8]. Consequently, global radiation forecast which requires a model including the corresponding meteorological factors has to be taken into account [9].

* Corresponding author at: School of Statistics, Jiangxi University of Finance and Economics, Nanchang 330013, China.
E-mail address: dongyao20051987@126.com (Y. Dong).

http://dx.doi.org/10.1016/j.enconman.2017.01.040
0196-8904/

Nomenclature

DC	Divide and Conquer
F	forecasting part of two-way interaction model
GRESH	group regularized estimation under structural hierarchy
GSO	glowworm swarm optimization
LASSO	least absolute shrinkage and selection operator
L	the number of features in individual subset model in ensemble learning
M	the number of subset models in ensemble learning
N	the number of forecasting data
n	the number of observations
N_t	the parameter to control the number of neighbors
p	the number of main features
R_s	the luciferin sensor radial range
S	the step size
T	training part of two-way interaction model
TI	two-way interaction model
β	Luciferin decay constant
δ	the constant parameter
τ	Luciferin enhancement constant
$\boldsymbol{\omega}$	the weight vector in ensemble learning
$\widehat{\boldsymbol{\Omega}}_{\text{ensemble}}$	the DCGSO-LASSO estimate
$\widehat{\boldsymbol{\Omega}}_{\text{LASSO}}$	The LASSO estimate

There are number of ways to forecast global solar radiation, such as time series models (AR, ARMA, ARIMA and SARIMA), empirical models, artificial intelligent techniques (fuzzy logic, neural networks, genetic algorithm, etc.), support vector machine (SVM). Wu and Chan proposed a novel hybrid model of ARMA and Time Delay Neural Network (TDNN) to predict hourly solar radiation series. The model took the advantages of both approaches and gave excellent results [10]. Mellit et al. developed an adaptive model to predict hourly global, direct and diffuse solar irradiance using relative humidity, air temperature, global, direct and diffuse horizontal irradiance as input variables, it was found that the proposed adaptive model had more accurate results than a Feed-Forward Neural network (FFNN) [11]. Benmouiza and Cheknane found that the proposed combined model based on *k*-means clustering algorithm and artificial neural networks (ANN) owned superiority to the hourly forecast of global horizontal radiation [12]. As the related meteorological factors as input predictors, Jiang et al. verified that the proposed RBF neural network with hard-ridge penalty based on cuckoo search algorithm presented better forecasting performance than RBF [13]. Aybar-Ruiz et al. presented a hybrid neural-genetic approach for global solar radiation prediction. To be specific, a grouping genetic algorithm (GGA) provided feature selection, while an Extreme Learning Machine algorithm (ELM) performed prediction. The developed method had displayed excellent performance in Toledo, Spain [14]. Considering temperatures, extraterrestrial solar radiation and maximum sunshine duration, Belaid and Mellit applied support vector machine (SVM) to forecast daily and monthly global solar radiation on horizontal surface in Ghardaïa [15]. To build an interpretable model, feature selection methods are combined with forecasting models. LASSO is a well-known feature selection method that is used to extract the small number of important variables in the dataset. It employs L_1 penalty to make the nuisance variable exactly zero which provides a parsimonious model. It is also computational efficient in big data computation because of its convex penalty function form. Ziel build a time-varying threshold ARMA-GARCH model for forecasting hourly German electricity load and employed fast iterative LASSO to estimate parameters [16]. Gao et al. used LASSO algorithm to address the disparity between ERA-Interim forecast precipitation data and point-scale meteorological observations, the results proved that LASSO is better than other statistical methods [17]. However, few research papers about its application are well studied in the field of forecasting the global solar radiation. For instance, Yang et al. applied LASSO to automatically shrink and select the most appropriate lagged time series as predictors for forecasting short-term irradiance, it is found that LASSO method outperforms ordinary least squares regression and univariate time series model [18]. Kızıltan used regression methods including Linear, Ridge, LASSO, Gaussian process and so on to estimate solar radiation at 53 location over Turkey with month, altitude, longitude and land surface temperature as input parameters [19].

The main contribution of this paper is proposing the multi-stage intelligent approach with the combination of ensemble learning and LASSO method for feature selection. There are four advantages of the developed method:

- It is adaptive, data-driven and applicable to two-way interaction model with high dimensionality and correlation which are difficult to handle.
- LASSO method is used to select the important features efficiently and optimal regularization parameter is selected by 10-fold cross validation (CV) which fully utilizes the data information through resampling which is especially useful in the data with small sample size.
- The ensemble learning based on Divide and Conquer principle is the main structure of the proposed method. The diversity created by feature selection and parameter selection of the ensemble system is great helpful in increasing the forecasting accuracy. The weight vector applied in the Conquer stage is selected by the Intelligent approach GSO which is shown successfully to achieve the global optimal solution.
- A simple-to-implement and efficient algorithm is designed based on thresholding function. The theoretical analysis reveals the estimation error of the proposed method.

The rest of this paper is organized as follows: Section 2 gives the preliminary on two-way interaction model, LASSO and Θ-estimator , the ensemble learning and glowworm swarm optimization. Section 3 shows the proposed multi-stage intelligent solar radiation forecasting method and its theoretical foundation in detail. Section 4 presents the empirical study and corresponding results analysis. Finally, Section 5 summarizes this study and provides brief concluding remarks.

2. Materials and methods

In this section, to begin with, two-way Interaction model with some notations used throughout the paper are introduced in detail. Then, the ensemble learning, which is the main structure of our proposed method, is also given. The parameter selection method glowworm swarm optimization (GSO) is introduced in the final part.

2.1. Two-way interaction model

Over the last decades, researchers realize that linear additive model including main features only is not sufficient for forecasting. High order terms such as two-way interaction features boost fore-