Energy 133 (2017) 270–283

Contents lists available at ScienceDirect

Energy

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

Global horizontal radiation forecast using forward regression on a quadratic kernel support vector machine: Case study of the Tibet Autonomous Region in China

He Jiang [a, b], Yao Dong [a, b, *]

[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

ARTICLE INFO

Article history:
Received 16 January 2017
Received in revised form
14 May 2017
Accepted 19 May 2017
Available online 20 May 2017

Keywords:
Global horizontal radiation forecast
Variable selection
Forward regression
Quadratic kernel support vector machine
Support vector machine information criterion

ABSTRACT

Effective and accurate forecasting of solar radiation plays a critically important role in the design of grid-connected photovoltaic installations. However, this is an extremely challenging task because of inconsistencies in variable selection and the prohibitively expensive computational cost as the number of variables increases. Although the support vector machine (SVM) can be applied to forecast solar radiation, it includes a large number of redundant variables. With the intent of establishing an interpretable model, a penalized SVM has been proposed. However, these penalized approaches shrink the estimate, which results in inaccurate results. In order to overcome these drawbacks and improve the accuracy of forecasting, this study develops a novel approach referred to as "forward regression on the quadratic kernel support vector machine" (QKSVM-FR) for building a quadratic regression model using forward regression to select the important variables for forecasting the global horizontal radiation in the Tibet Autonomous Region. A fast and simple-to-implement computational algorithm is derived to perform the variable selection and forecasting tasks simultaneously. Furthermore, the SVM information criterion is utilized to select the kernel parameter to guarantee model consistency. The results of experiments directly confirm the outstanding forecasting performance of the proposed QKSVM-FR method compared to other existing methods.

1. Introduction

With the rapid development of the worldwide economy, the increasing production and consumption of fossil fuel resources such as coal and petroleum have caused growing concern around the world. However, most countries are dependent on thermoelectric power generation, which not only demands increasing consumption of fossil fuels but also emits greenhouse gases, causing air pollution. These problems have motivated scientific researchers to develop sustainable energy applications for power generation. Solar energy, which is considered to be a clean source of sustainable energy, has attracted much attention from researchers in recent years [1], as it is the most abundant form of renewable energy and can be harnessed for power generation using large solar array farms. This has several benefits, as solar energy does not consume raw materials, produces no air pollution, and is reliable and easily maintained [2]. A grid-connected photovoltaic installation is the most efficient way to harness solar energy. Nevertheless, there are many factors limiting the use of solar radiation, including climate and weather changes; it is an inherently time-varying, chaotic, and intermittent source of energy. These drawbacks limit the ability of solar energy generation facilities to manage power uncertainty and voltage stability, and, even more detrimentally, can lead to serious power quality issues. Thus, it remains a challenge to manage a stable energy supply when integrating solar energy facilities into the power grid [3]. At present, the fluctuation of hourly solar radiation has become an important focus for both the transmission of a constant amount of power to the power grid and the determination of the power storage capacity of backup equipment in a solar-based generation system [4]. Accurate forecasting of solar radiation makes it possible to create an effective schedule for power operators and

* Corresponding author. 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.energy.2017.05.124
0360-5442/

Nomenclature

G	Solution path of main effects and interaction terms
$\mathbf{H}$	Data matrix for main effects
$\mathbf{H}_{proj}$	The projection matrix
$\overline{\mathbf{H}}$	Data matrix for interaction terms
$\breve{\mathbf{H}}$	Data matrix for overall variables
$\breve{\mathbf{H}}_{opt}$	The optimal selected model
$\mathbf{K}$	The kernel data matrix
l	The loss function
q	The number of selected components
RSS	Residual Sum of Square
S	Solution path of main effects
T_M	Main effects of true model
T_{MI}	Interaction terms of true model
$\widehat{\beta}_{opt}$	The optimal estimate
γ	The kernel parameter

ensures favorable trading performance and sustainability in the electricity market. However, it is difficult to forecast the available solar radiation because it depends on complex weather or seasonal conditions, as well as the specific location of the solar facility [5].

Scientific researchers have devoted their attention to benchmarking different methods for solar radiation forecasting. There are generally two major categories of forecasting models [6]: (1) time-series models based on present and past samples of solar radiation, and (2) intelligent models based on meteorological variables including temperature, relative humidity, clearness, dust, and wind speed. The former models include the auto-regressive moving average (ARMA) [7], Markov chain-based models [8], and Bayesian model [9]. However, common characteristics of solar radiation such as its intermittence, and high sampling frequency result in inaccurate forecasts using time-series models. Thus, at present, most research is focused on developing intelligent models, including artificial neural network (ANN), support vector machine (SVM), and hybrid forecasting models. Fadare used meteorological and geographical data collected in Nigeria from 1983 to 1993 as inputs to an ANN, using the solar radiation intensity as the output to assess solar radiation potential [10]. Considering multiple meteorological variables as inputs, Kashyap et al. built eight ANNs using three major schemes, including the number of neurons, delay, and activation function, to forecast the solar radiation in India [11]. In ANN applications, the most challenging task is selecting the appropriate number of neurons and delays in each layer, as these approaches do not always generate a unique global solution, and the computational cost of this process is high. The SVM approach was proposed by Vapnik in 1995 for classification and regression in statistical learning theory [12]. It is related to regularization networks and has some advantages over ANNs, namely that the SVM applies the least square method through a set of linear equations based on structural risk minimization and therefore does not require the iterative tuning of model parameters, requires few kernels, and shows good forecasting performance [13]. Chen et al. investigated seven combinations of monthly air temperatures as input features for an SVM model to estimate the monthly solar radiation at the Chongqing meteorological stations of China, and tried different kernel functions. They found that SVM models yielded better results than seven empirical temperature-based models [14]. Zeng and Qiao employed a least-square (LS) SVM with sky cover, relative humidity, and wind speed as predictors to forecast short-term solar power availability. The results demonstrated that the SVM was superior to an auto-regressive (AR) model as well as a radial basis function neural network (RBFNN) [15].

Despite the flexibility of the SVM in handling non-linear inputs, this approach encounters difficulties with non-stationary and training input datasets that contain a wide range of frequencies. Hybrid models not only take full advantage of the SVM but also combine the superiorities of other forecasting models. Deo et al. adopted a wavelet-coupled SVM model to forecast the global incident radiation using the hours of sunshine, temperature, wind speed, precipitation, and evaporation as input variables [16]. Shamshirband et al. developed a coupled SVM-wavelet transform model to estimate the daily horizontal diffuse radiation in Kerman City, Iran [17]. The results revealed that these hybrid models delivered a higher forecasting precision than any single method. In this vein, forward regression offers a subset variable selection approach that utilizes different combinations of variables, adding variables one by one to the model in a given manner and selecting the optimal model using some statistical information criterion [18,19]. To the best of our knowledge, few research papers have detailed the forecasting of solar radiation using forward regression. Lu et al. employed stepwise variable selection to select the relevant meteorological variables in a linear regression model for forecasting solar radiation [20]. Wang et al. chose meteorological variables using multiple stepwise regression to forecast solar radiation [21]. However, these methods focus solely on a linear regression model. The forecasting performance of forward regression when used in conjunction with a nonlinear model is worth investigating. This paper advocates the use of forward regression in the quadratic kernel support vector machine (QKSVM-FR) method to forecast the global horizontal radiation in the Tibet Autonomous Region of China. The main contributions of this paper are as follows:

- A quadratic kernel SVM model using two-way interaction terms is proposed to depict the nonlinear relationships between the variables.
- Forward regression is employed to efficiently select the important variables, and the optimal kernel parameter is selected using the support vector machine information criterion (SVMIC) to ensure forecasting accuracy.
- A simple-to-implement and efficient algorithm is designed, and the optimal forecasting model is selected using the extended Bayesian information criterion (EBIC) after establishing the solution path in forward regression.
- A theoretical analysis of the QKSVM-FR method regarding selection consistency is provided.

This paper is organized as follows: Section 2 describes the penalized kernel support vector machine approaches; the forward regression on a quadratic kernel support vector machine is proposed in Section 3; the theoretical analysis is demonstrated in Section 4; Section 5 investigates the forecasting accuracy based on real data analysis; the conclusions are presented in Section 6.

2. Penalized kernel support vector machine approach

Given the kernel parameter γ and the radical basis function $\mathbf{K}(\mathbf{x}, \mathbf{x}') = \exp\left(-\gamma||\mathbf{x}-\mathbf{x}'||_2^2\right)$, with $\mathbf{x}$ and $\mathbf{x}'$ belonging to a certain domain $\aleph$, the original kernel regularized SVM problem can be written as follows: