Energy xxx (2014) 1–7

Contents lists available at ScienceDirect

Energy

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

Bi-model short-term solar irradiance prediction using support vector regressors

Hsu-Yung Cheng [a,*], Chih-Chang Yu [b,1], Sian-Jing Lin [a]

[a] Department of Computer Science and Information Engineering, National Central University, No. 300, Jhongda Rd., Jhongli City, Taoyuan County, Taiwan
[b] Department of Computer Science and Information Engineering, Vanung University, No. 1 Van-Nung Rd., Chung-Li, Tao-Yuan 32061, Taiwan

ARTICLE INFO

Article history:
Received 22 January 2014
Accepted 22 March 2014
Available online xxx

Keywords:
Solar irradiance prediction
Support vector regression
Clearness index
Ramp-down events
All-sky image

ABSTRACT

This paper proposes an accurate short-term solar irradiance prediction scheme via support vector regression. Utilizing clearness index conversion and appropriate features, the support vector regression models are able to output satisfying prediction results. The prediction results are further improved by the proposed ramp-down event forecasting and solar irradiance refinement procedures. With the help of all-sky image analysis, two separated regression models are constructed based on the cloud obstruction conditions near the solar disk. With bi-model prediction, the behavior of the changing irradiance can be captured more accurately. Moreover, if a ramp-down event is forecasted, the predicted irradiance is corrected based on the cloud cover ratio in the area near the sun. The experiments have shown that the proposed method can effectively improve the prediction accuracy on a highly challenging dataset.

1. Introduction

Renewable energies have grown substantially over the last few decades. Solar energy is one of the green energies that draws much attention of both industries and researchers. Recently, a large number of PV (Photovoltaics) are installed worldwide. The cost of PV arrays is decreasing and their efficiency has been improved substantially. However, the main challenge of solar energy is that the produced electricity is often variable and intermittent [1]. The fluctuation of the supply makes the energy expensive and prevents it from prevalence. It is difficult to integrate the share of intermittent resources into the energy system, especially the electricity supply [2]. To utilize solar energy more effectively, PV grid operators desire mechanisms of forecasting short-term electricity generation. The power generated by PV arrays depends non-linearly on the solar irradiance level and temperature. The short-term change of temperature is usually more predictable because the variation is usually not large within a few minutes. But the solar irradiance can change dramatically and rapidly within a very short period of time. Therefore, the ability to perform accurate short-term forecast on surface solar irradiance is necessary.

Due to the importance of solar irradiance prediction, there are many relative works [3–5]. In the existing works, the prediction accuracy varies a lot given different datasets collected under different sky conditions and at different locations. Normally speaking, the prediction accuracy for cloudy days would be lower compared with clear days, especially for dataset contains more cloud motion. Heinemann et al. [3] reported several different approaches of solar irradiance forecasting in different time scales ranging from 30 min to 6 h. In this work, the reported prediction RMSE ranges from 26% to 54% for various forecast sources. Lorenz et al. [4] analyzed different approaches to refine the ECMWF global model irradiance forecasting. The main improvements include spatial averaging, temporal interpolation, improved clear sky forecast, and post processing with ground data. The GHI forecasting RMSE ranges from 12% for clear skies to 85% for cloudy conditions in this work.

In the above mentioned works, numerical weather database or satellite prediction information have been utilized widely. However, such prediction can only provide rough estimation on hourly or daily bases, which means that the temporal resolution is coarse. For more refined spatial resolution and intra-hour irradiance prediction, ground-plane devices that capture sky images can be exploited. These ground-plane devices include All Sky Imager developed by Japanese Communications Research Laboratory [6], Whole Sky Imager developed by Scripps Institute of Oceanography at the University of California [7,8], Whole Sky Camera developed by Spain's University of Girona [9], and Total Sky Imager by Yankee Environmental Systems [10], [11]. The knowledge of cloud cover,

* Corresponding author. Tel.: +886 3 4227151x35306; fax: +886 3 4222681.
E-mail addresses: chengsy@csie.ncu.edu.tw, breeze.cheng@gmail.com (H.-Y. Cheng).
[1] Tel.: +886 3 4515811.

http://dx.doi.org/10.1016/j.energy.2014.03.096
0360-5442/

cloud motion, and cloud type extracted from all-sky images is very helpful for more accurate prediction. Such information from all-sky images can be obtained using cloud detection, tracking, and classification techniques [12–17]. Researches have shown that within 15-minute time horizon, forecasting with the help of all-sky image information achieved obvious improvement over satellite images [13,14]. For prediction methods incorporating sky imagery, Marquez and Coimbra integrated present and past irradiance, meteorological variables, and cloud cover statistics from images to train artificial neural networks for irradiance forecasting [14]. They reported RMSE ranging from 15% to 22% for different datasets.

In this work, we incorporate history measured irradiance data with features extracted from all-sky images to perform prediction. The bi-model prediction framework is constructed using support vector regressors. Separated models are trained according to cloud obstruction conditions near the solar disk. After getting the estimated forecast from the trained regressor, the prediction is further refined with a correction process. The rest of the paper is organized as follows. Section 2 explains data acquisition. Section 3 elaborates the regression models and the features utilized for the prediction purpose. Section 4 details the procedures to refine the predicted solar irradiance. Section 5 displays and discusses the experimental results. Finally, Section 6 concludes the article.

2. Data

The device used to capture the all sky images is the all sky camera manufactured by the SBIG (Santa Barbara Instrument Group). The detailed specifications of the all sky camera are listed in Table 1. The experimental dataset is collected at a costal site in Taiwan. One all-sky image is captured per minute. The longitude and latitude of the all sky camera are 24°46′34.00″N and E121° 2′40.10″E. The dataset is very challenging since the variation of weather and cloud cover is high with the marine climate at the costal site. The changes of solar irradiance under such cloud conditions are dramatic and fast.

The device used to measure the ground truth solar irradiance is Delta OHM LP RYRA 03. It is a point sensor that is located next to the all sky camera. The sampling interval is 10 s. The measured solar irradiance within each minute is averaged to generate the ground truth irradiance data corresponding to each all-sky image. The specifications of Delta OHM LP RYRA 03 are listed in Table 2.

3. Prediction models and features

The model and the features used for prediction are elaborated in this section.

3.1. Support vector regression

Support vector machine is a popular supervised learning method in machine learning area [18]. The architecture and basic properties of the support vector algorithm can also be applied for regression purpose [19]. Given a set of data $D = \{(\mathbf{x}_1,y_1), (\mathbf{x}_2,y_2),\cdots,(\mathbf{x}_N,y_N)\}$, where $\mathbf{x}_1$ to $\mathbf{x}_N$ are D dimensional feature vectors and y_1 to y_N are scalars, the support vector regression aims at finding a linear function $f = \mathbf{w}\cdot\mathbf{x} + b$ that minimizes the following equation [19]:

$$\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\left(\xi_i + \xi_i^*\right) \tag{1}$$

subject to the constraints

$$y_i - \mathbf{w}\cdot\mathbf{x} - b \le \varepsilon + \xi_i \tag{2}$$

$$\mathbf{w}\cdot\mathbf{x} + b - y_i \le \varepsilon + \xi_i^* \tag{3}$$

$$\xi_i, \xi_i^* > 0 \tag{4}$$

In Eqs. (1)–(3), ξ_i and ξ_i^* are non-negative slack variables that allow soft margin and relax the optimization problem. Introducing slack variables gives control to the sensitivity of SVR to overcome the problems caused by the possible outliers. In many cases, linear regression functions are not sufficient to capture the behavior of the data. If we do not want to be restricted to linear regression functions, it is possible to obtain non-linear SVR solution by applying kernels. The kernels are non-linear mappings that can be used to map the data into a higher dimensional feature space where linear regression is performed. In this work, RBF (radial basis function) kernels are used in SVR to obtain nonlinear regression models.

3.2. Input and output of the regressor

As indicated in Refs. [20], utilizing Perez conversion model in the prediction process can yield better prediction results. According to the Perez conversion model [21], the global horizontal irradiance at surface can be calculated from clearness index, long-term averaged extraterrestrial solar irradiance, mean earth–sun distance, earth–sun distance at the time of interest, and the solar zenith angle at the time of interest. All the terms except clearness index can be determined in advance given the time of interest and the position of the camera. The clearness index is the only term that needs to be predicted. Therefore, we can express the prediction function as

$$\widehat{CI}_{i+\Delta P} = f_{\mathrm{SVR}}(\mathbf{x}_i) \tag{5}$$

In Eq. (5), the function f_{SVR} is the trained support vector regressor. As the input of the regression function f_{SVR}, $\mathbf{x}_i$ is the extracted feature vector at time instance i. And the output $\widehat{CI}_{i+\Delta P}$ is the predicted clearness index for time instance $i + \Delta P$.

The features extracted from all-sky images can be utilized as the input feature vector $\mathbf{x}_i$ of the regression function. In this work, we adopt the five image features that exhibit higher absolute value of the correlation with the solar irradiance to be predicted as suggested in Ref. [20]. These features include the mean intensity level, the mean gradient magnitude, the averaged accumulated intensity of the vertical line of sun, and the mean and variance of cloud cover in a fixed time duration. The details of detecting the vertical line of the sun and determining the cloud cover are described in Ref. [20]. In addition to

Table 1
Specifications of the all sky camera used for experiments.

Specifications of the all sky camera	
Charge-coupled device (CCD)	Kodak KAI-0340 @ 7.4 microns square
Lens	Fujinon FE185C046HA-1
Field of view (FOV)	185°
Resolution	640 × 480 Bitmap Image
Focal length	1.4 mm
FOCAL RATIO RANGE	f/1.4 to f/16
Weight	0.4 kg
Communication Interface	RS-232
Power consumption	12VDC, 600 mA max (2.4 W)

Table 2
Specifications of Delta OHM LP RYRA 03.

Specifications of Delta OHM LP RYRA 03	
Measuring range	0 ∼ 2000 W/m^2
Viewing field	2π sr
Spectral field	335 nm ÷ 2200 nm (95%)
Impedance	13.75 μV (W/m^2)
Weight	0.45 kg