Solar Energy 155 (2017) 224–232

Contents lists available at ScienceDirect

Solar Energy

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

Mining optimum models of generating solar power based on big data analysis

Yin-Fu Huang *, Shun-Hao Chang

Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, 123 University Road, Section 3, Touliu, Yunlin 640, Taiwan, ROC

ARTICLE INFO

Article history:
Received 31 December 2016
Received in revised form 12 June 2017
Accepted 13 June 2017

Keywords:
Solar power
Big data
Data mining
Machine learning

ABSTRACT

Recently in exploiting green energy, solar power generation is a must-be trend and approach, especially for the countries with nature resource shortage. However, how to build solar power plants with the best power generation efficiency in limited spaces is always a crucial issue. In this paper, the approach of finding the optimum models of generating solar power is proposed to build solar power plants for different environments in Taiwan. First, we collect all the data from existing solar power farms, including (1) design methods of power generation, (2) actual power generation, and (3) surrounding environments. Then, after a series of preprocessing steps and system analysis on them, the optimal models of generating solar power could be mined out. Finally, in the experiments, we evaluate the system from five aspects regarding to input and output parameters. As a result, we observe that using the majority voting strategy improves the system accuracy and helps engineers build solar power plants with the maximum power generation.

1. Introduction

Due to the energy scarcity of the Earth and the pressure of greenhouse gas emission reduction, developing green energy has already been a project that the countries around the world pursued. After the Fukushima Daiichi nuclear disaster, each country doubts about the safety of nuclear power generation, compared with solar power generation. With the development of science and technology, replacing old power generation methods with solar power generation is indeed a correct direction. In Taiwan such a subtropical and island-type country, solar energy is rich so that it is more possible to replace nuclear power generation with solar power generation.

In recent years, there have been quite a number of countries to develop their own solar power generation. Each country has its own national policy to assess the cost and demand of solar power, and explores the current status of solar manufacturers and planning (AlYahya and Irfan, 2016; Behrens et al., 2016; Bett and Thornton, 2016; Campoccia et al., 2014; Comello and Reichelstein, 2016; Dio et al., 2015; Griffiths and Mills, 2016; Ramírez et al., 2017; Rohankar et al., 2016; Strielkowski et al., 2017; Trad and Ali, 2015). The solar power plant settings and different areas of solar radiation affecting the efficiency of solar power generation have been discussed (Mghouchi et al., 2016; Olatomiwa et al., 2015; Peruchena et al., 2016; Sharma et al., 2016). The efficient prediction of solar power generation was also investigated under different weather conditions within 30 days (Alam et al., 2015; Antonanzas et al., 2016). As the current solar technology is increasingly sophisticated, the Taiwan government announced that the solar photovoltaic device capacity targeting from the original 6200 MW will increase to 20,000 MW in 2025, nearly triple increase. Renewable energy plant capacity targeting from 12,513 MW grows to 27,423 MW, also increasing more than twice. In Taiwan, since sunshine duration is long and sunshine deflection angle is small, it is very suitable to develop solar energy. However, Taiwan land area is small, so constructing a land-based solar power plant in a large population is constrained. Thus, building solar power plants in the loose roof space becomes one of the choices in Taiwan. These spaces may refer to the roofs of buildings (e.g., house roof, school roof, walls, or farmhouse roof). How to build solar power plants and create better efficiency in the hardware, cost, and system design were ever addressed issues (Liang et al., 2016; Mermoud and Lejeune, 2010; Patra et al., 2016; Tripathi et al., 2015; Wang et al., 2016). The effects of different altitude angles and azimuth angles on solar power generation were also investigated (Jamil et al., 2016; Mermoud, 2012; Roux, 2016). But, all these studies were for land-based solar power plants. For rooftop solar power plants with limited azimuth angles, it is necessary to design suitable altitude angles. In the study

* Corresponding author.
E-mail address: huangyf@yuntech.edu.tw (Y.-F. Huang).

http://dx.doi.org/10.1016/j.solener.2017.06.035

(Horváth et al., 2016), the solar energy system of city buildings was found to have a unified configuration design, but the optimal power generation design was never considered. Other optimization designs include (1) increasing the electrical output by adopting a better material and a different technique to fabricate the concentrator (Abu-Bakar et al., 2015), (2) selecting the optimum location of thousands of heliostats and the most profitable tower height and receiver size in solar power tower (SPT) systems (Collado and Guallar, 2016), and (3) maximizing the PV electrical production and minimizing relevant system costs using a multi-objective genetic algorithm (Freitas et al., 2015). Besides, some research focused on the parameter identification of PV models using bio-inspired algorithms (Ma et al., 2016; Oliva et al., 2014), and on comparisons among grid-connected power conditioning units (PCUs) with different sizes, technologies and PV system architectures (Spertino and Graditi, 2014).

In this paper, the approach of finding the optimal models of generating solar power is proposed, based on the datasets collected from existing solar power farms. In the experiments, the datasets are provided by Sinogreenergy Sinogreenergy.com (2016), which cover the rooftop solar power designs in Southern Taiwan. The dataset includes solar power plant design, power generation data, and environments. The optimum power generation design for buildings can be found according to azimuth angles, different roof types, and the hardware used. Three methods are proposed to find the optimum power generation design, and then the system will recommend users the optimal models of generating solar power, based on the majority voting strategy.

In summary, we highlight the practical relevance of this paper as follows:

(1) In this study, the approach of finding the optimum models of generating solar power is proposed to build solar power plants for different environments in Taiwan, based on **real big data with diversity such as designs, produced power, and environments**.
(2) Three methods based on **different user requests** are proposed to reach the goals, and the system has been **implemented**. Besides, the conducted experiments are **comprehensive** and verify the **practicability**.
(3) The methodology presented here can be applicable on anywhere or any solar PV systems if their datasets are collected.
(4) Our approach is **complementary** to tools for simulating photovoltaic system performance in building solar power plants. A simulation can be built to observe the system performance **using the I/O parameter values** obtained from the optimal models. Without these I/O parameter values, basic data are probably inputted blindly and then the simulation could not reveal its full values.

The remainder of this paper is organized as follows. First, we introduce the architecture of finding the optimum power generation design in Section 2. Then, in Section 3, the datasets collected from existing solar power farms are described and a series of preprocessing steps are done on them. Next, Section 4 presents three methods for recommending the optimal models of generating solar power, based on user requirement specification. In Section 5, the user interface of the system is introduced and the experiments are conducted to evaluate the system from five aspects regarding to input and output parameters. Finally, we make conclusions in Section 6.

2. System overviews

Building solar power plants with the best power generation efficiency is always a crucial issue. In this paper, the approach of finding the optimal models of generating solar power is proposed, based on the datasets collected from existing solar power farms and the architecture is shown in Fig. 1.

In the preprocessing, PR (i.e., Performance Rating) thresholds can be derived from the original datasets first, and then the data below the PR thresholds are filtered out. Next, the surviving data including all kinds of information are integrated together for facilitating data analysis. Then, we analyze features, find their correlation with power generation, and determine input and output parameters relevant to power generation. In the paper, five input parameters (or features) about solar power environments (i.e., azimuth angle, roof height, inverter size, MPPT (i.e., Maximum Power Point Tracking), and roof material) and four output parameters (i.e., altitude angle, series connection, parallel connection, and module voltage) are determined. Azimuth angle is the angular deviation in degrees from strictly southern orientation with positive values for western deviations. Roof height is the height (in meters) of the roof for modules from the ground. Inverter size is the maximum AC power (in kW) of an inverter. MPPT is the number of independent circuits for maximum power point tracking in an inverter. Roof material is the material used in the roof. Since the input parameters play important roles in power generation, we further find the input parameter weighting using the SAHS (i.e., Self-Adaptive Harmony Search) algorithm. Through the preprocessing, the modified datasets become the training datasets used to build the optimal models of generating solar power.

In the system operation, data are extracted from the training datasets according to predetermined PR thresholds and areas, and then go through normalization. Next, three kinds of classifiers are built for finding the optimal models of generating solar power; i.e., SVM, K-means + SVM, and association rules. In the user interface, users can (1) set PR thresholds, (2) specify areas, and (3) select different combinations of input and output parameters including setting input parameter values to refer to their requirements. In the system analysis, the optimal models of solar power can be found using three proposed classifiers, based on selected input and output parameters. Based on the majority voting strategy, the system will recommend users the optimal models of solar power with efficient altitude angle, series connection, parallel connection, and module voltage.

3. Preprocessing

In this section, we first introduce the datasets collected from existing solar power farms. Then, in order to find the optimal models of generating solar power using data mining or machine learning approaches, the preprocessing should be done on the datasets.

3.1. Datasets

The datasets analyzed in this paper contain the daily solar power records and environments collected from 120 solar power plants. The datasets recorded solar power information from 2014 to 2015. During the period, every 5 min, a solar power plant returned its data to the server while the plant has power generation. For different seasons, sunlight time would be different. In the spring of Taiwan, the sunlight time is 8–9 h per day, so almost 12,000 data could be recorded from plants one-day. Since 2014, 2 million data collected from all the plants in Taiwan have been recorded. How to guarantee our analysis correct, accurate, efficient and fast is dominated by the preprocessing.

3.2. Data filtering and integration

First, the PR thresholds are calculated and determined in this step. The PR thresholds are used to filter out poor genera-