Energy Conversion and Management 121 (2016) 380–390

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Univariate and multivariate methods for very short-term solar photovoltaic power forecasting

Mashud Rana [a], Irena Koprinska [b,*], Vassilios G. Agelidis [a]

[a] Australian Energy Research Institute, University of New South Wales, Sydney, NSW 2052, Australia
[b] School of Information Technologies, University of Sydney, Sydney, NSW 2006, Australia

ARTICLE INFO

Article history:
Received 11 February 2016
Received in revised form 2 May 2016
Accepted 9 May 2016
Available online 25 May 2016

Keywords:
Solar power forecasting
Time series prediction
Variable selection
Neural networks
Support vector regression

ABSTRACT

We consider the task of forecasting the electricity power generated by a solar PhotoVoltaic (PV) system for forecasting horizons from 5 to 60 min ahead, from previous PV power and meteorological data. We present a new method based on advanced machine learning algorithms for variable selection and prediction. The correlation based variable selection identifies a small set of informative variables that are used as inputs for an ensemble of neural networks and support vector regression algorithms to generate the predictions. We develop two types of models: univariate, that use only previous PV power data, and multivariate, that also use previous weather data, and evaluate their performance on Australian PV data for two years. The results show that the univariate models performed similarly to the multivariate models, achieving mean relative error of 4.15–9.34%. Hence, the PV power output for very short-term forecasting horizons of 5–60 min can be predicted accurately by using only previous PV power data, without weather information. The most accurate model was univariate ensemble of neural networks, predicting the PV power output separately for each step of the forecasting horizon.

1. Introduction

Solar energy generated from PV systems is one of the most promising and fastest growing renewable energies. Since 2000, the capacity of the installed solar PV systems around the world has grown 100 times, reaching 178 GW in 2014, and it is expected to reach 540 GW in the next five years [1]. This growth is driven by government legislations to increase the electricity supply from renewable sources, by incentives and grants that support the installation of PV systems and also by the decreasing cost of PV systems – their price has declined by 75% in the last 10 years [1]. It is expected that the number of both large capacity PV plants and small scale PV rooftop systems will increase, contributing considerably to the future electricity supply.

However, unlike the power generated from traditional sources, the power generated from PV systems is highly variable as it depends on the solar irradiance and other meteorological conditions. To successfully integrate solar power in the electricity grid, there is a need for accurate forecasting of the solar power output for different forecasting horizons. In this paper we consider forecasting the PV power output from 5 min to 1 h ahead, which is classified as very short-term forecasting, and is especially important for trading solar power in the electricity markets.

We propose a new method for PV power forecasting, based on advanced machine learning algorithms for variable selection and forecasting. Variable selection has received little attention in the previous work on solar power forecasting, although it is essential for building accurate forecasting models. The PV power output depends on different weather variables such as solar irradiance, temperature, humidity, cloud cover and wind speed. It is important to identify the relevant variables and time lags for each data source. Appropriate variable selection leads to improved accuracy, faster training and smaller complexity of the prediction model. We apply the Correlation-based Feature Selection (CFS) algorithm [2] which hasn't been previously used for solar power forecasting.

In conjunction with CFS, as prediction algorithms we apply: (1) an ensemble of Neural Networks (NNs) that we have developed, and (2) Support Vector Regression (SVR) algorithm [3,4]. Although single NNs have been previously applied for solar power forecasting in [5–7], ensembles of NNs are less common. Ensembles of NNs combine the predictions of several NNs in some way and are typically more accurate than a single NNs. Our proposed ensemble of NNs aims to reduce the NN sensitivity to the network architecture and initialization of weights. We chose to use SVR as an alternative prediction algorithm as it is a state-of-the-art algorithm, shown to be successful in many domains.

* Corresponding author.
E-mail addresses: md.rana@unsw.edu.au (M. Rana), irena.koprinska@sydney.edu.au (I. Koprinska), vassilios.agelidis@unsw.edu.au (V.G. Agelidis).

http://dx.doi.org/10.1016/j.enconman.2016.05.025
0196-8904/

We develop two types of prediction models: (1) univariate, that use only previous PV power data, and (2) multivariate, that use previous PV power and weather data from several sources. Our main goal is to compare the performance of these models for very short-term PV power forecasting and especially to investigate if the meteorological data helps to improve the accuracy.

The contribution of this paper is the following:

(1) We present a new method for PV power forecasting that combines feature selection using the CFS algorithm with machine learning prediction algorithms. To select variables, we follow a two-step process that firstly identifies a set of candidate variables and then selects the best subset of variables using the CFS algorithm. To generate predictions we use two algorithms: an ensemble of NNs that we have developed and the state-of-the-art SVR algorithm. We train a separate prediction model for each step of the forecasting horizon.
(2) We develop both univariate and multivariate models. The univariate models use only the previous PV power data while the multivariate models use in addition previous meteorological data – solar irradiance, temperature, humidity and wind speed. A notable advantage of our method is that it uses only variables that are easily obtainable (previous PV power and simple weather data). In comparison to other methods, it does not use sky images that require special equipment to be recorded and processed, or future weather predictions that are not always available for the site of the PV plant.
(3) We conduct a comprehensive evaluation using 5-min PV power data for two years, from a grid-connected PV plant in Australia. We evaluate the performance of our method for forecasting horizons from 5 to 60 min ahead. We compare the performance of the univariate and multivariate models, and also develop and evaluate an iterative multiple step ahead prediction model.

The rest of this paper is organized as follows. Section 2 provides an overview of the previous work on solar power forecasting. Section 3 describes the data used in this study. Section 4 provides a problem statement and Section 5 describes our method for solar power forecasting. Section 6 presents the experimental setup and Section 7 presents and discusses the results. Section 8 summarizes the main results and concludes the paper.

2. Related work

Solar power forecasting has attracted considerable interest in recent years due to the new legislations encouraging the use of solar energy and the large-scale deployment of solar plants. There are two main groups of forecasting approaches: (1) indirect, that firstly forecast the solar irradiance or use forecasts of solar irradiance produced by meteorological centers, and then convert them to PV power output forecasts by considering the characteristics of the PV plant such as area and efficiency, and (2) direct, that directly forecast the PV power output. Our proposed method falls into the second group.

2.1. Group 1: Indirect forecasting

Lorenz et al. [8] predicted the hourly PV power output for up to 2 days ahead based on the weather forecasts for solar irradiance. They firstly applied spatial averaging and polynomial filtering to derive site specific irradiance forecasts which were then converted into PV power output based on the characteristics of the PV system. A detailed evaluation was conducted using German data for 10 months obtaining RMSE of 4%. A similar approach for the same forecasting horizon was followed by Pelland et al. [9]. The solar irradiance forecasts were optimized for the specific site using spatial averaging and Kalman filters which was shown to improve predictive accuracy. The PV power forecasts were produced by using the adjusted solar irradiance forecasts, weather forecasts for temperature and previous PV output data, and the evaluation was conducted using Canadian data for two years.

Mellit and Pavan [5] firstly predicted the hourly solar irradiance for the next day using data from the previous day (mean solar irradiance and temperature) and information about the day of the month, and a backpropagation NN as a prediction algorithm. The PV power output was then calculated by multiplying the predicted solar irradiance with coefficients representing the characteristics of the PV panel.

Yang et al. [10] applied a lasso linear regression model to predict the solar irradiance for very short-term forecasting horizons: from 10 s to 5 min. An extensive evaluation was conducted using American solar irradiance data. The results showed that the lasso linear regression outperform the standard linear regression and Autoregressive Integrated Moving Average (ARIMA) models.

Alonso-Montesinos and Batles [11] predicted the solar irradiance for forecasting horizons from 15 min to 3 h ahead based on satellite images. They predicted three solar irradiation components (beam, diffuse and global irradiance) by identifying the clouds and determining their motion, and then using the ESRA [12] and Heliosat-2 [13] methods for estimating the solar irradiance. The authors considered three sky conditions: clear, partially cloudy and overcast. They obtained normalized RMSE (nRMSE) of 19–21% for all sky conditions together (over all three solar components and all time intervals). In their subsequent work [14] the same group proposed a new method using sky camera images instead of satellite images and a maximum cross-correlation method to determine the cloud motion vectors, obtaining nRMSE = 11–25% for all sky conditions.

Urraca et al. [15] predicted the solar irradiance 1 h ahead based on recorded meteorological data and computed solar variables. They developed two types of models: fixed, that are trained once using all training data, and moving, that build a separate prediction model for each testing instance using a subset of the training data (e.g. only the most similar instances). They applied SVR, random forest, linear regression and nearest neighbor as prediction algorithms, and genetic algorithms for feature selection for the fixed models. The evaluation was done using Spanish data and the results showed that the moving models slightly outperformed the fixed models; the best results were obtained with the moving SVR, MAE = 49–64 W/m^2.

2.2. Group 2: Direct forecasting

Pedro and Coimbra [6] studied four different methods for direct prediction of the solar power produced by a PV plant from previous solar power values: ARIMA, k nearest neighbor, NN trained with the backpropagation algorithm and NN trained with a genetic algorithm. They conducted an evaluation using data for two years from a 1 MW plant in California, predicting the PV power 1 and 2 h ahead. The results showed that the two NN-based methods outperformed the other methods.

Chen et al. [16] predicted the PV power output for the 24 h of the next day using the power output from the previous day and the weather forecast for the next day. They classified the days into sunny, cloudy and rainy, and built a separate prediction model, a Radial-Basis Function Neural Network (RBFNN), for each category. As RBFNN inputs they used the average PV power from the previous day and the weather forecasts for the next day (average daily