Renewable Energy 90 (2016) 542–553

Contents lists available at ScienceDirect

Renewable Energy

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

Machine learning for solar irradiance forecasting of photovoltaic system

Jiaming Li [a,*], John K. Ward [b], Jingnan Tong [b], Lyle Collins [b], Glenn Platt [b]

[a] CSIRO CCI, Australia
[b] CSIRO Energy Technology, Australia

ARTICLE INFO

Article history:
Received 7 September 2015
Received in revised form
24 November 2015
Accepted 31 December 2015
Available online 18 January 2016

Keywords:
Virtual power station
Renewable energy optimization
Genetic algorithm
Hidden Markov model
SVM regression
Neural networks

ABSTRACT

Photovoltaic generation of electricity is an important renewable energy source, and large numbers of relatively small photovoltaic systems are proliferating around the world. Today it is widely acknowledged by power producers, utility companies and independent system operators that it is only through advanced forecasting, communications and control that these distributed resources can collectively provide a firm, dispatchable generation capacity to the electricity market. One of the challenges of realizing such a goal is the precise forecasting of the output of individual photovoltaic systems, which is affected by a lot of factors. This paper introduces our short-term solar irradiance forecasting algorithms based on machine learning methodologies, Hidden Markov Model and SVM regression. A series of experimental evaluations are presented to analyze the relative performance of the techniques in order to show the importance of these methodologies. The Matlab interface, the Weather Forecasting Platform, has been used for these evaluations. The experiments are performed using the dataset generated by Australian Bureau of Meteorology. The experimental results show that our machine learning based forecasting algorithms can precisely predict future 5–30 min solar irradiance under different weather conditions.

1. Introduction

The world has abundant solar energy resources. Photovoltaic (PV) technology has become one of several promising alternatives for use in energy technology [1]. Yet many critics of the widespread use of solar energy cite its intermittency, or the challenges around predicting the future output of a solar generator. The Virtual Power Station (VPS) [2–4] conducted by CSIRO aims to address such concerns by combining a large number of geographically disperse, and technically diverse, small scale renewable energy generators that will allow them to present to the electricity market as a single reliable dispatchable entity. The aggregated energy of the VPS can be sourced from a large number of small energy generation and storage systems, such as roof-mounted solar PV panels, and associated grid-connected battery systems installed in individual domestic houses. These individual systems are then aggregated together, to form a “virtual power station”, with one coordinated response, of benefit to the wider electricity network. However, integration of large amounts of PV into the electricity grid poses technical challenges due to the fluctuating characteristics of available solar energy sources. PV output is not easily predictable in advance and varies based on both weather conditions and site-specific conditions. Such variability of solar energy resources at ground level thus raises concerns regarding how to manage and integrate output from the VPS to the power grid.

Given the issues above, there is increasing interest in more precise modeling and forecasting of solar power. Irradiance is a measurement of solar power and usually measures the power per unit area. Most researches consider the solar irradiance forecasting at a site, which is essentially the same problem as forecasting solar power. The ability to forecast solar irradiation will enable power grid operators to be able to ensure the quality and control of solar electricity supplies in an environment of greater solar panel usage, allow them to better accommodate highly variable electricity generation in their scheduling, dispatching, and regulation of power. In particular, the possibility to forecast solar irradiance can became fundamental in making power dispatch plans, and also a useful reference for improving the control algorithms of battery charge controllers. Ultimately, the development of more accurate

* Corresponding author.
E-mail address: jiaming.li@csiro.au (J. Li).

http://dx.doi.org/10.1016/j.renene.2015.12.069
0960-1481/

methods for modeling and forecasting solar irradiance remains a key requirement of our future energy system.

Different solar irradiance forecast methodologies have been proposed for various time horizons. Some of them forecast up to 24 h or even more [5]. Whilst useful for long-term forecasts, such techniques don't meet the demands of many electricity markets, for example the Australian electricity market uses 5-min dispatch price and a 30-min trading price. Thus, accurate short-term forecast are essential for energy market participation, both due to forward contracting and the need for a predictable, stable and smooth supply. Accurately forecasting direct normal irradiance or global horizontal irradiance in the seconds-to-minutes time-frame ultimately enables finely-tuned dynamic operational schedules that can reduce fuel costs, increase network stability or maximize system lifetimes.

Machine learning methods have been used to solve complicated practical problems in various areas [6–11] and are becoming more and more popular nowadays. Several machine learning based methodologies, such as genetic algorithm (GA) [12] and neural networks (NNs) [13,14], have been proposed and applied for modeling and forecasting of solar irradiance [15–22]. Quaiyum et al. [15] presented a neural networks and genetic algorithm model to predict the solar irradiance data from both endogenous and exogenous variables. Mellit et al. [16] developed a neural network-based genetic algorithm model for generating the sizing curve of stand-alone photovoltaic systems. Mellit et al. [17] developed a multilayer model to forecast the solar irradiance 24 h ahead. The inputs of proposed model use mean daily irradiance and mean daily air temperature and the output is solar irradiance data 24 h ahead. Mellit et al. [18] also developed radial basis function based neural network model for prediction solar radiation data. Kemmoku et al. [19] used a multistage neural network to predict irradiance of the next day. The input data to the network are the average atmospheric pressure, predicted by another neural network and various weather data of the previous day. Irradiance forecast by the multi-stage and the single-stage neural networks are compared with measured irradiance. Sfetsos et al. [20] used neural network to make one-step predictions of hourly values of global irradiance and to compare them with linear time series models that work by predicting the clearness index. They introduced an approach for forecasting hourly solar irradiance using various neural network based techniques and also investigated other meteorological variables such as temperature, wind speed, and pressure. Mihalakakou et al. [21] developed a total solar irradiance time series simulation model based on neural network and applied it in Athens. The neural network was identified as the model with the least error. Hocaoglu et al. [22] incorporate multistage to time-delay neural network models for the prediction of hourly solar radiation. But the problems for these methods are as follows.

- Genetic algorithm (GA) [10] is an optimum search technique based on the concepts of natural selection and survival of the fittest, has been successfully applied to many difficult problems [9–11]. But for our solar irradiance forecasting issue, the first and one of the most difficult questions is the physical model definition, which describes the physical state and dynamic motion of the atmosphere defined by mathematical equations. Current GA based solar irradiance forecasting algorithms [15,16] couldn't give such physical properly.
- Neural Networks (NNs) are an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurones) working in unison to solve specific problems. The same as GA, NN has been successfully applied in a lot areas [22,23]. But for our application it cannot easily be determined which variables are the most important contributors to future solar irradiance, what novel structure of the network is optimal, and a neural network model may contain a number of unimportant input variables that the developer fails to appreciate.

To avoid the problems of GA and NNs, this paper introduces our short-term solar forecasting algorithms using Hidden Markov Model and SVM regression based on the dataset gathered from the Australian Bureau of Meteorology (BOM). In order to evaluate the generalizability of the forecasting algorithms, the evaluations are performed under different weather conditions for forecasting future 5–30 min solar irradiance. The paper is organized as follows. Section 2 briefly presents our Matlab interface, the Weather Forecasting Platform, for evaluating our developed solar irradiance forecasting algorithms. Section 3 introduces the BOM dataset used for comparison. Section 4 introduces our solar forecasting algorithms in details. Section 5 provides a series of experimental results to evaluate solar irradiance forecasting performance. Finally, a conclusion is drawn in Section 6.

2. Solar irradiance forecasting platform

We developed a Matlab interface – the Weather Forecasting Platform (WFP), which is shown in Fig. 1. The WFP is designed so that historical and forecast data - data that is necessary to perform a forecast - can be delivered to the different forecasting algorithms (an example of which is a neural networks algorithm) in a consistent manner. Another primary purpose of the WFP, is to ensure the forecasting algorithms act in a causal manner, so algorithms don't accidentally 'cheat' by looking at future data (from the perspective of the WFP). The arrows in Fig. 1 represent the flow of the data – not simply communication between the modules, as communication between all of the connected modules is bidirectional.

The WFP is designed to work as follows:

1. The user application provides the WFP with a forecast request. The request includes such details as:
   a. Which of the forecasting algorithms is to perform the forecast. Note that potentially all of the forecasting algorithms can be requested to perform the forecast, because of the common forecast output format from the WFP, makes a comparative analysis of the differing forecasting algorithms easy.
   b. The weather parameter to be forecasted – e.g. irradiance, temperature. It is irradiance that is the focus of the remainder of the paper.
   c. The present time. By 'present' it is meant the time for which the WFP can consider to be the current time, such that any request to get data from a future time is inhibited. This is necessary to enforce the real-world requirement that the forecasting algorithms be causal.
   d. Any forecasting algorithm parameters of interest. Parameters such as frequency of the forecast data, forecast horizon, or any other configuration parameters are all valid parameters.
2. The WFP passes on the relevant data from the request to the one or many forecasting algorithms identified in (1a).
3. The algorithms then make independent decisions on what data is necessary for them to perform their requested forecasting task. The forecasting algorithms then respond to the forecast request by sending a data request to the WFP.