Solar photovoltaic power output forecasting using machine learning technique

Dinh Van Tai

Postgraduate Student, National Research Tomsk Polytechnic University, Tomsk, Russia

dinh88@mail.ru

Abstract. Photovoltaic (PV) systems are used around the world to generate solar power. Solar power sources are irregular in nature due to the output power of PV systems being intermittent and depending greatly on environmental factors. These factors include, but are not limited to, irradiance, humidity, PV surface temperature, speed of the wind. Due to uncertainties in the photovoltaic generation, it is critical to precisely envisage the solar power generation. Solar power forecasting is necessary for supply and demand planning in an electric grid. This prediction is highly complex and challenging as solar power generation is weather-dependent and uncontrollable. This paper describes the effects of various environmental parameters on the PV system output. Prediction models based on Artificial Neural Networks (ANN) and regression models are evaluated for selective factors. The selection is done by using the correlation-based feature selection (CSF) and ReliefF techniques. The ANN model outperforms all other techniques that were discussed.

1. Introduction

The global economic policy, climate conditions and security issues of energy are highly affected by the current situation of global warming and the energy crisis over the past few decades due to the excessive consumption of fossil fuels. This situation has been the motivation for the development and use of clean and sustainable energy sources, which can serve as alternatives to the present energy production [1].

The use of solar power energy is rapidly growing, as it is a renewable form of energy. This energy is environmentally friendly because it does not produce pollution. A study done by the European Photovoltaic Industry Association (EPIA) showed that, in 2014, the total in-progress solar power capacity was 177 GW, which demonstrates its rising popularity [2].

Because of the disorganized and unreliable nature of the weather, the power output of PV energy systems is largely uncertain. Due to the discontinuous and uncontrollable nature of solar power, the precise forecast of solar power generation is very important for a grid operator and the solar electric power supply companies. To get the best economic benefit from a PV system, it is essential to design an algorithm for the prediction of the output power of a PV system. The other factors affecting output power being temperature, humidity, wind speed, and dust accumulation [3]. Stability can be achieved by using power prediction techniques and by ensuring approximate production in the future. It helps the power supply companies to make a controller in order to switch between the available energy resources present in a hybrid power plant.

The methods used for predicting solar power are normally categorized as physical or statistical. However, in practice, the lines between these approaches are unclear. Numerical weather prediction (NWP) models or sky images are used in physical approaches as a part of irradiance prediction. Statistical approaches forecast solar irradiance from training and statistically-derived values. In deterministic approaches, output prediction is done by using PV device models obtained through different software such as PVSyst, and System Advisor Model (SAM), among others. Sometimes these prediction methods are not able to assess the variations data. Due to the prevalence of such cases, probabilistic models are commonly used [4]. The model explained in this paper has been confirmed for diverse module types by making use of hourly solar source and meteorological records. These modules are statistical in nature, and they are machine learning ones like Artificial Neural Networks (ANN), Support Vector Machines (SVM), Multiple Linear Regressions (MLR) and, Adaptive Neuro-Fuzzy Inference Systems (ANFIS).

2. The model

The model proposed in this study for power forecasting is established by regression and prediction methods. The methods included in this model are the M5P regression tree, the Gaussian process, and linear regression [4]. In the linear regression model, the input and output parameters of the system have a linear relation. GPR is based on Gaussian process and a computation of the point resemblance, which envisages a value for an unknown point using a training dataset.

The model consists of the following steps (see Figure 1**)**:

1. Data pre-processing: the input variables are processed to extract useful features. The initial dataset is converted into a useful format. It is made free of anomalies. A common practice, for instance, is the removal of outliers. After this, the most relevant features are selected.

2. Training and testing phase: after all the appropriate features are mined, the dataset for testing and training needs to be created.

3. Prediction phase: In this phase of working on the model, we use output responses with known values to train numerous machine learning procedures by making use of the neural network model in Keras.

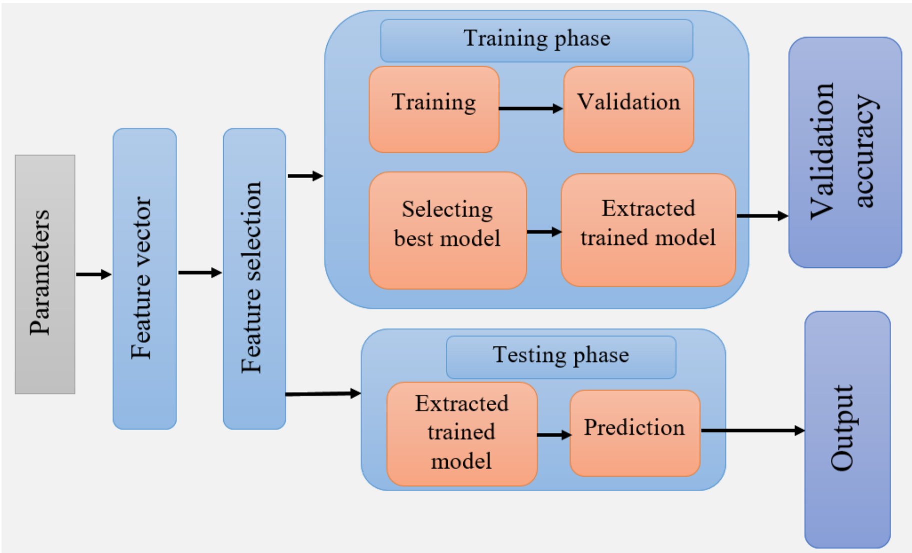

Figure 1. Diagram of proposed model.

The machine learning technique is similar to that of the human brain in the sense that it is based on a layered structure, which includes input layers, output layers, and hidden layers. The ANN model works as non-natural neurons that can carry multiple inputs. The neuron activation level is achieved by applying training input and output pairs [5]. Each function has its pros and cons, and each behaves differently based on the provided datasets. In ML techniques for envisaging the output of a PV, different groups of hidden layers and training functions are tested to identify the best possible integration that can accurately predict the output. The most important aspect of predicting PV-output lies in finding how much the results are biased. The biased forecast is of one of two types: the over-forecast (higher than the actual value), or it is under-forecast (less than the actual value). The techniques used to calculate bias are: the tracking signal technique [6], and Normalized Forecast Metric (NFM) technique.

3. Experiments

The tested PV system is located at the 18th Campus, Institute of Non-destructive Testing, Tomsk Polytechnic University, Tomsk, Russia. The station includes 3 kW solar battery and 2 kW wind-driven electric plant [7, 8]. A view of the system is presented in Figure 2.

Figure 2. A view of the PV system.

The station consists of 10 elements of solar generator ARPS-250, 2 elements of the wind generator 1000 W. This station can operate at temperature, from -50 to 50o C.
Specification:

- generator capacity, no more than 5 kW;
- output voltage ~220 V, 50 Hz;
- surface area of the PV modules, no more than 230 m^2;
- height of wind generator tower 6 m;

Summary of the data of all the environmental parameters are used, in order to predict the output power, is given in Table 1.

Table 1. Input parameters of the model.

Environmental parameters	Min	Max	Unit
PV surface Temperature	-43	40	Degree Celsius
Irradiance	10	1800	W/m^2
Wind speed	0	8	m/s
Power output	0	3500	W

The PV power output forecast is obtained after using environmental data from measurements. The PV power forecast is then compared to the actual value to estimate the accuracy of the models. The examples of this are shown in Figure 3 and Figure 4.

To calculate absolute percent error [9, 10], we are using formula 1. An example of absolute percent error is presented in Figure 5.

IOP Conf. Series: Journal of Physics: Conf. Series **1327** (2019) 012051 doi:10.1088/1742-6596/1327/1/012051

$$APE = \frac{|V_P - V_A|}{V_A} \times 100\%, \tag{1}$$

where, V_P and V_A are the predicted value, and the actual value, respectively.

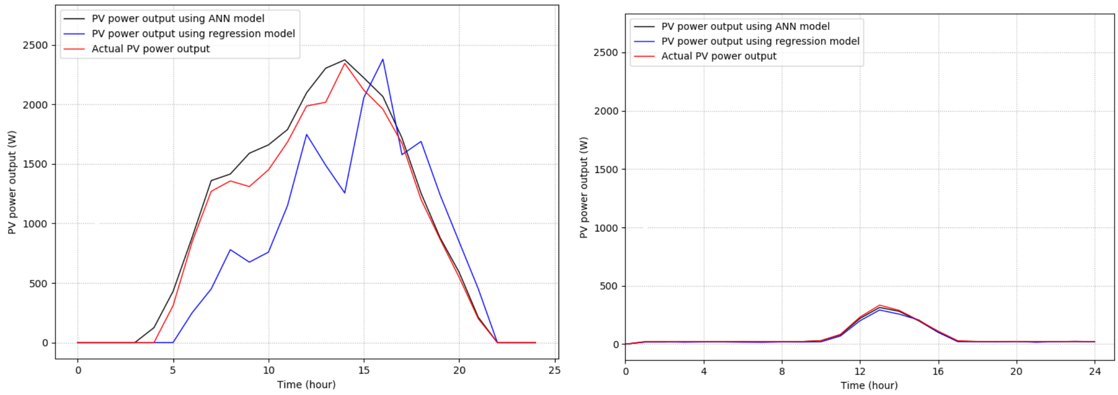

Figure 3. A comparison data between the models and actual data for a day of summer (in June 2019) and a day of winter (in January 2019).

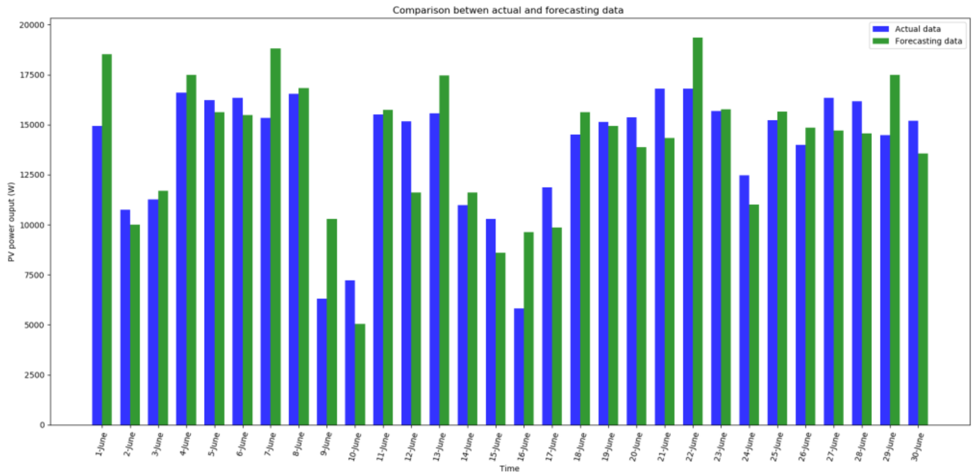

Figure 4. A comparison between forecasting and actual data for June 2019.

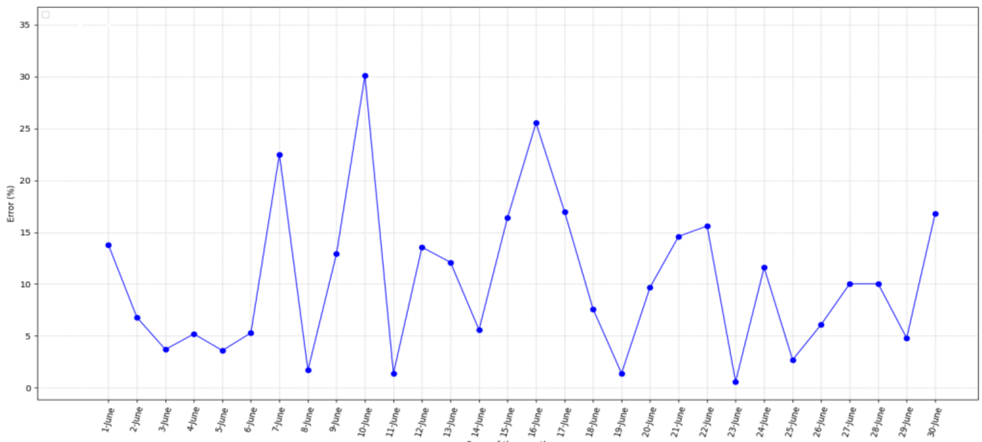

Figure 5. Prediction error for June 2019.

Figure 3 shows the actual daily power generation graph versus the prediction graphs of the forecasting models. It seems that the prediction fits the actual well with using ANN model, especially in winter, when the system did not produce much energy.

Figure 5 shows the PPE values of the prediction model for 30 days in June 2019. Note that the prediction errors on 10^{th} June and 16^{th} June were exceptionally large for the prediction model. This finding may be attributed to the fact that the weather forecast on that day was highly inaccurate. As a result, there are substantial differences in prediction on these 2 days.

4. Conclusion

In summary, the model is designed for forecasting PV power output using regression methods and ANN-based networks using the data gathered by the PV system. For selecting subsets of applicable, high quality and non-redundant features, two methods, namely CFS and ReliefF, are used for the selection of parameters. It is concluded from the data provided above that, compared to the three best regression models (simple linear regression model, M5P decision tree model, and GPR) and the ANN model. It is clear that the ANN model can forecast the output power of the PV system with superior accuracy if we use feature selection methods.

References

[1] Kim J, Kim D, Yoo W, Lee J and Kim Y 2017 Daily prediction of solar power generation based on weather forecast information in Korea *IET Renewable Power Generation* **11** 1268-73

[2] Urbanowicz R J, Meeker M, La Cava, W Olson, R S, & Moore 2018 Relief-based feature selection: Introduction and review *Journal of Biomedical Informatics* **07** 014

[3] Khandakar A 2019 Machine Learning-Based Photovoltaic (PV) Power Prediction Using Different Environmental Parameters of Qatar *Energies* **12(14)** 2782

[4] Puntanen S (2010) Linear Regression Analysis: Theory and Computing by Xin Yan, Xiao Gang Su *International Statistical Review* **78(1)** 144-144

[5] Andrew A 2004 Information Theory, Inference, and Learning Algorithms Information Theory, Inference, and Learning Algorithms *Cambridge: Cambridge University Press*. *ISBN: 0-521-64298-1* **33(7)** 1217-8

[6] Hua C, Fang Y and Chen W 2016 Hybrid maximum power point tracking method with variable step size for photovoltaic systems *IET Renewable Power Generation* **10(2)** 127-32

[7] Dinh V T and Yuhao Yan 2018 Short-Term Forecasting of Electric Energy Generation for a Photovoltaic System *MATEC Web of Conferences* **155** 01033

[8] Yurchenko A V, Zotov L G, Mekhtiev A D, Yugai V V and Tatkeeva G G 2015 Power supply of autonomous systems using solar modules *IOP Conference Series: Materials Science and Engineering* **81(1)** 012112

[9] Bacher P, Madsen H and Nielsen H A 2009 Online short-term solar power forecasting Solar Energy **83(10)** 1772–83

[10] Yurchenko A, Syriamkin V, Okhorzina A and Kurkan N 2015 *IOP Conference Series: Materials Science and PV effectiveness under natural conditions* **81(1)** 012097