Output Power Forecasting of Solar Photovoltaic System Using Meta-cognitive Neuro Fuzzy Inference System

Praynlin Edinson
Department of ECE, V V College of Engineering,
Tisaiyanvillai, Tamil Nadu, India
Praynlin25@gmail.com

ABSTRACT
Forecasting of solar power is in general significant for planning the operations of power plants which convert renewable energies into electricity. Prediction of solar power shows some uncertainties depending on atmospheric parameter such as solar irradiance, atmospheric temperature and relative humidity. This article illustrates Artificial Neural Network (ANN) strategy that is, Meta-cognitive Neuro Fuzzy Inference System (McNFIS) which is used to forecast photovoltaic power from the historical data set collected from different photovoltaic panel located in different location. ANN have some interesting properties that made machine learning very appealing when confronting difficult pattern discovery tasks. ANN has the ability of learning, it can learn from the historic data set and it can predict the future samples. McNFIS self-regulatory learning mechanism that controls the learning process of the cognitive component, by deciding what-to-learn, when-to-learn and how-to-learn from sequential training data, forms the meta-cognitive component. McNFIS realizes the above decision by employing sample deletion, sample reserve and sample learning strategy, respectively. Thus, in this article training and testing are performed by ANN technique- McNFIS. The accuracy of prediction can be described by using various error measurement criteria like Normalised Mean Absolute Error, Weighted Mean Absolute Error, normalised Root Mean Square Error. Finally the performance of neural network is to be noted.

Keywords- Forecasting, McNFIS, Artificial Neural Network.

I. INTRODUCTION

Wind and solar are the renewable energy sources which are received most importance globally in recent years as they are producing clean energy. They are globally accepted as perfect solutions for alternative clean energy sources [6]. Photovoltaic (PV) technology has been rapidly developed in recent years due to their maintenance-free, enduring and environmental-friendly nature. However, power which is obtained as the output of PV system is a non-stationary random process because of unevenness of solar radiation and environmental factors. In the public power grid system, any grid-connected PV system is considered as an uncontrollable and non-scheduling unit and their power output fluctuation will affect the stability of the power system [5].

While implementing the PV system, a significant limitation of PV system is the unpredictability of power from the sun. They will surely affect the standard of the electrical system that they are connected. Therefore, the forecasting of power output of photovoltaic system can help to increase the standard of the power system. To forecast the power output of PV system, particularly for short term forecast applications that can be grouped into two categories, In first method the prediction is based on solar radiation intensity. Another method is to predict the power output of PV system directly. The prediction model based on solar radiation intensity has been considered to be a successful method in practical application and it also takes a lot of meteorological and geographical data to solve some difficult differential equations. Accordingly, different forecasting methods are used. Forecasting methods also depend on the tools and information available to forecaster, such as data from weather stations and satellites, PV system data and outputs from numerical weather prediction (NWP) models.

Forecasting methods [8] [10] was generally classified into two types physical or statistical. The physical approach uses solar and PV models that is available in real time to generate PV
forecasts, whereas the statistical approach relies primarily on past data to “train” models, with little or no reliance on solar and PV models. Nowadays, the most applied techniques are the statistical methods to model the stochastic nature of solar irradiance at the ground level, and thus the power output of PV installations. Regression methods are often made use of to illustrate complex nonlinear atmospheric operations for few a hours ahead of forecast and specific soft computing techniques [3] are used for a few hours of output power forecast.

II. FORECASTING METHODS

In order to estimate the solar power generated several methods has been. The different methods to estimate the solar PV generated is generally categorized into three groups viz, physical methods, statistical methods and artificial intelligence methods [4] [22]. Physical systems use parameterizations based on a detailed physical description of the atmosphere, to reach the best prediction precision. Usually, solar radiation given by the weather service on a coarse grid is transformed to the onsite conditions at the location of the PV system.

Solar power prediction is dominated by Physical methods which can able to predict forecasts in long-term [13]. Weather data with sophisticated meteorological details is used in physical methods that can give more precise prediction for solar radiation forecasting and solar power predictions. A statistical method is different from physical method that aims at finding the relationship of the on-line measured power data. For a statistical model, the historical data of the PV system may be used. Statistical models are cheaper to develop and easy to model compared to the other models [16] [23]. The statistical method has advantages in short-term solar power prediction. The disadvantage with this method is that the prediction error increases as the prediction time increases. Time series-based methods are the leading statistical method [9]. Time series-based methods include the auto regressive (AR), AR with exogenous input (ARX), Auto Regressive Moving Average (ARMA), Auto Regressive Integrated Moving Average (ARIMA), Wavelet Network [15] [19]. Nowadays the most common way to forecast the future values of a time series is the use of machine learning methods, PSO method [1][7], genetic algorithm[7], combined together as GSO. Recently, with the development of artificial intelligence, various artificial intelligence methods for solar power prediction have been developed. The new developed methods include artificial neural network and neuro-fuzzy logic methods. The artificial intelligence methods are self-designing ones that can be automatically adjusted in a changing system.

Soft computing techniques [3] [18] based on ANN are used for a few hours of power output forecast. Thus in statistical approach of forecasting process, artificial neural network plays an important role for photovoltaic energy forecasting. Adaptive neuro-fuzzy inference system (ANFIS) [14] [20] [21] [24] is an integration of neural networks to develop fuzzy inference systems from input-output data sets. ANFIS is a suitable approach for weather and PV power output predictions, which require uncertainty modelling and existing behaviour adaptation for more accurate predictions. To increase the accuracy of solar power forecasting, this paper uses Neural Network method Meta-Cognitive Neuro Fuzzy Inference System [17].

III. ARTIFICIAL NEURAL NETWORK

Artificial neural network is based on intelligent computing with the computer network system simulating biological neural networks. In the neural network algorithm [25] [26], the large amount of data that can be used to provide information, also increases the difficulties of dealing with these data. If we take all of its data as the input of the network, it will hurt the design of the network, and will occupy a lot of storage space and computing time, too many feature inputs and repeated sample training will make the training process time-consuming, hurt the convergence of the network and finally affect the recognition precision of the network [2]. So, it is necessary to pre-process the original data, analyze and extract useful variable features from the large amounts of data, exclude the influences of the related or redundant factors, reduce the feature dimension as far as possible under the premise of not affecting the
solution of the problems. The dimensionality reduction of data as input of neural network, technique will combine data dimensionality reduction with the neural network.

A typical feed forward neural network [11] has an input, a hidden and an output layer. Each component includes a neuron, weights and a transfer function. An input \( x_j \) is transmitted through a connection which multiplies its strength by a weight \( w_{ij} \) to give a product \( x_jw_{ij} \). The product is an argument to a transfer function \( f \) which yields an output \( y_i \) represented by:

\[
y_i = f(x_jw_{ij}) \tag{1}
\]

where \( i \) is a neuron index in the hidden layer, and \( j \) is an index of an input to the neural network. Training is the process of modifying the connection weights in some orderly fashion using a suitable learning method.

Using artificial neural network as an estimation tool has proved its efficiency in predicting different parameters via other parameters where relationship is not specified. So, applying artificial neural networks can be valuable in the prediction of solar radiation [12] [28].

4.1. META COGNITIVE NEURO FUZZY INFERENCE SYSTEM

Meta cognitive component is a self regulatory learning mechanism [17] that controls the learning process of cognitive component. The information flow from the cognitive component to meta-cognitive component is considered monitoring, while the information flow in the reverse direction is considered control. The cognitive component of McNFIS [27] is a three layered feed forward radial basis function network with Gaussian activation function in the hidden layer. The meta-cognitive component contains copy of the cognitive component. When a new training sample arrives, the meta-cognitive components of McNFIS predicts the class label and estimate the knowledge present in the new training sample with respect to the cognitive component.

![Fig. 1. Structure of Cognitive Component.](image-url)
**Cognitive component**: Neuro fuzzy inference system consists of four layers. It consists of an input layer with $m$ nodes, a Gaussian layer with $K$ nodes which forms the rule antecedent and the aggregation layer of an NFIS, a normalization layer with $K$ nodes and an output layer with $n$ nodes. The weight vector connecting normalization layer and output layer forms the rule consequent parameters of an NFIS. The structure is specified in fig 1.

**Input layer**: The input layer is a linear layer with $m$ nodes. It transmits the input features directly to the Gaussian layer.

**Gaussian layer**: The nodes in the Gaussian layer employ the Gaussian activation function to compute the membership value of each input node. The nodes in this layer form the antecedent and does rule aggregation operations of a fuzzy inference system.

**Normalization layer**: This layer provides average firing potential for each rule.

**Output layer**: The output layer is a linear layer with $n$ nodes. The predicted output is obtained.

Based on this information, the meta-cognitive component selects a suitable learning strategy, for the current sample. Thereby, addressing the three fundamental issues in learning process: (a) what-to-learn, (b) when-to-learn and (c) how-to-learn. The Mechanism is given in fig 2.

**Sample delete strategy**: If the new training sample contains information similar to the knowledge present in the cognitive component, then delete the new training sample from the training data set without using it in the learning process.

**Neuron growth strategy**: Use the new training sample to add a new hidden neuron in the cognitive component.
Parameter update strategy: The new training sample is used to update the parameters of the cognitive component.

Sample reserve strategy: The new training sample contains some information but not significant, they can be used at later stage of the learning process for fine tuning the parameters of the cognitive component. These samples may be discarded without learning or used for fine tuning the cognitive component parameters in a later stage.

These are operation steps used during the photovoltaic forecasting operation using this process.

IV. EVALUATION CRITERIA

In order to correctly define the accuracy of the prediction and the related error, it is necessary to define the indexes that can be used to evaluate the performances of the forecasting model. Some of these definitions come from statistics. The error definitions are really different among each other. Also, technical papers present a lot of these indexes, therefore here we report some of the most commonly used error definitions. The starting reference point is the hourly error ‘eh’ defined as the difference between the average power produced (measured) in the h-th hour $P_{m,h}$ and the given prediction $P_{p,h}$ provided by the forecasting model.

Hourly error ($e_h$) - is defined as the difference between the average power produced (measured) $P_{m,h}$ and the given prediction $P_{p,h}$ provided by the forecasting model.

$$e_h = P_{m,h} - P_{p,h}$$

(2)

When the value of hourly error is low, then the accuracy of prediction will be high.

Absolute hourly error ($e_{abs}$) - which is the absolute value of hourly error which is used to calculate the performance.

$$e_{h,abs} = |e_h|$$

(3)

When hourly error is low, then automatically the value of absolute hourly error will be low.

Hourly error percentage ($e_{p,h}$) is based on hourly expected power output $P_{p,h}$

$$e_{p,h} = \frac{|e_h|}{P_{p,h}} . 100$$

(4)

Hourly error percentage ($e_{m,h}$) is based on hourly measured power output $P_{m,h}$

$$e_{m,h} = \frac{|e_h|}{P_{m,h}} . 100$$

(5)

The hourly error percentage based on measured power is generally higher than the expected power output.

Normalised Mean Absolute Error (NMAE%) is based on net capacity of the photovoltaic panel from which the data set is collected, $C$.

$$NMAE\% = \frac{1}{N} \sum_{h=1}^{N} \frac{|P_{m,h} - P_{p,h}|}{C} . 100$$

(6)

where $N$ represents the number of samples.
Weighted Mean Absolute Error (WMAE\%) based on total energy production

\[
WMAE\% = \frac{\sum_{h=1}^{N} |p_{m,h} - p_{p,h}|}{\sum_{h=1}^{N} p_{m,h}} \times 100
\]  
(7)

This error will be greater when NMAE is used during unstable days.

Normalised Root Mean Square Error (nRMSE\%) is based on the maximum observed power output \(p_{m,h}\)

\[
nRMSE\% = \frac{\sqrt{\sum_{h=1}^{N} (p_{m,h} - p_{p,h})^2}}{\max(p_{m,h})} \times 100
\]  
(8)

This error measures the average magnitude of Absolute hourly error and they give higher weight to larger error.

V. DATA SET DESCRIPTION

Solar photovoltaic historical data set is collected from the photovoltaic (PV) panel. The data is collected from the atmosphere for the production of solar power from the solar radiation [5]. The PV panel will collect some parameters such as humidity average, wind speed average, solar irradiance, temperature, cloud coverage, output power, etc. The data is collected for each and every minute time interval. After the first process data collection, data pre processing procedure is carried out to train the neural network. This procedure is done to normalise the data set. Before any other steps, historically measured PV data must be always validated since the untrustworthy data increases the odds of higher error in the forecasting process. Four data sets are collected from different photovoltaic systems for the process.

5.1 RUTLAND DATA SET

The data set which was collected for this project is from 50KW DC ground mount installation located in Rutland. The data set was collected for each fifteen minutes time interval. This data set contains about 1248 data. The data set contains six parameters

- Ambient temperature average
- Panel temperature average
- Transformer temperature average
- Solar irradiance
- Wind speed
- Power output

The data set was collected from 19th January 2010 to 31st January 2010. This was obtained by daily observation of solar irradiance, temperature, humidity and generated power by photovoltaic installation which is located in particular locations.

5.2 GECAD PHOTOVOLTAIC SYSTEM DATA SET

This data is obtained from GECAD photovoltaic system. The capacity of one photovoltaic panel is 200W. This data set was collected for each and every five minutes time interval. This data set was collected for one month, that is, from 1st May 2015 to 31st May 2015. The parameters of the data set can be as follows
Actual solar radiation
Sensor usage
Ambient temperature
Module temperature
Total amount of energy

There is about 8451 data contained in the respective data set collected from GECAD photovoltaic system. Thus, the data set contains the solar radiation average which is the amount of daily irradiance from the sun.

5.3 PV SPOT DATA SET

This data set was collected from the photovoltaic plant located in Netherland. The net capacity of the PV plant is 15.24 KW. This data set is collected for one year, that is, from 1st January 2011 to 31st December 2011. The parameters of the data set can be as follows.
- Low estimation output power
- Best estimation output power
- High estimation output power
- Tilted irradiance average
- Horizontal irradiance average
- Temperature

The low, high and best estimation output values are considered as only one value, that is, by taking the average of three values and considered as total power. This data set contains about 365 data.

5.4 BROADMEADOW DATA SET

This dataset was collected from the photovoltaic plant which was located in Broadmeadow, in Australia. The net capacity of the photovoltaic plant is 3.15KW. This dataset contains about two month data, that is, from 30th November 2015 to 31st December 2015. The dataset contains the following parameters,
- Irradiance average
- System temperature
- Station temperature
- PV yield

The dataset contains about 744 data which is collected for one minute time interval.

VI. NEURAL NETWORK IN PHOTOVOLTAIC FORECASTING

In Implementation of Neural Network for Photovoltaic forecasting process, the initial step is data set collection. Four data sets were collected from various photovoltaic plants located in different locations. In the next step, these data sets are divided in the ratio 70:30, where 70% of the data sets are used for neural network training and the remaining 30% of data sets are used for testing the neural network is specified in Table 1
<table>
<thead>
<tr>
<th>Different data set</th>
<th>Training</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rutland data</td>
<td>874</td>
<td>374</td>
</tr>
<tr>
<td>GECAD data</td>
<td>6330</td>
<td>2121</td>
</tr>
<tr>
<td>PV spot data</td>
<td>255</td>
<td>110</td>
</tr>
<tr>
<td>Broadmeadow data</td>
<td>445</td>
<td>229</td>
</tr>
</tbody>
</table>

So, the first data set Rutland data is divided as 874 for training and 374 for testing. GECAD data set is divided as 6330 data for training and 2121 data for testing purpose, next PV spot data set is divided as 255 and 110 for training and testing respectively. The fourth data set, Broadmeadow is divided as 445 and 229.

These data sets will contain the data which are collected in different climate conditions. Thus, the solar power produced mainly depends on the radiance produced by the sun. Training and Testing of neural network is performed using McNFIS methods.

VII. RESULTS

The training and the testing of neural network for photovoltaic output forecasting are done with four different historical PV data set using McNFIS methods. In performance evaluation process, some error formulas have been introduced during training and testing separately, and thus the error values have been tabulated and plotted separately in the forthcoming sections.

7.1. META COGNITIVE NEURO FUZZY INFERENCE SYSTEM RESULTS

The Meta cognitive Neuro Fuzzy Inference system for genfis1 has been implemented for photovoltaic forecasting and the corresponding error values obtained for seven datasets are tabulated in table 2.

The error values are calculated by substituting in the corresponding formulae to obtain the accurate error values.

<table>
<thead>
<tr>
<th>Different data set</th>
<th>NMAE (Normalised Mean Absolute Error)</th>
<th>WMAE (Weighted Mean Absolute Error)</th>
<th>nRMSE (normalised Root Mean Square Error)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
<td>Testing</td>
<td>Training</td>
</tr>
<tr>
<td>Rutland data</td>
<td>10.74</td>
<td>13.28</td>
<td>131.68</td>
</tr>
<tr>
<td>GECAD data</td>
<td>508.05</td>
<td>487.07</td>
<td>48.74</td>
</tr>
<tr>
<td>PV spot data</td>
<td>0.12</td>
<td>0.0912</td>
<td>38.16</td>
</tr>
<tr>
<td>Broadmeadow data</td>
<td>0.026</td>
<td>0.024</td>
<td>96.36</td>
</tr>
</tbody>
</table>

Thus, by considering five inputs and one output in Rutland data set, four inputs and one output in GECAD data set, three inputs and one output in remaining PV spot, Broadmeadow data sets, the error values such as NMAE, WMAE, nRMSE were calculated. Normalised Mean Absolute Error (NMAE) is based on net capacity of plant from which data set is collected. When NMAE value is lower, then accuracy of prediction will be greater.

Then, hourly error(eh), absolute hourly error(ahe), hourly error based on expected power(heep), and hourly error based on measured power(hemp) can obtain a multiple output which
can be plotted against measured output(Y) and expected output(T). The result obtained for McFIS is plotted in the graphs.

Initially the graph was plotted for Rutland dataset, where five inputs and one output is considered and network training and testing is done with 1248 datasets which is specified in fig 3.

The next graph McNFIS is plotted for GECAD dataset. These graph also plotted against measured output(Y) and expected output(P) during training the network and measured output(Y1) and expected output(P1) obtained during testing the network which is specified in fig 4.

The next graph McNFIS error is plotted for PV spot dataset. These plots are plotted against the measured power and expected power obtained during training and testing of the network which is specified in fig 5.
The next graph McNFIS error for Broadmeadow dataset is plotted for PV spot dataset. These plots are plotted against the measured power and expected power obtained during training and testing of the network which is specified in fig 6.

Thus the training and testing operation of the neural network is performed by using McNFIS method by using four different dataset collected from different photovoltaic plant and the corresponding errors are calculated.

**VIII. CONCLUSION AND FUTURE WORK**

This article reports four predictive models which are applied for prediction of output power in PV systems. A self-regulatory learning mechanism forms the meta-cognitive component of McNFIS. The self-regulatory learning mechanism controls the learning ability of the cognitive component. This method is performed with four different PV data sets. The evaluation of forecasting is performed by calculating some error values such as NMAE, WMAE, nRMSE, etc. The results are plotted and error values are tabulated. The result demonstrates that accuracy of prediction by McNFIS method. Thus, effectiveness of the proposed
forecasting method provides improved accuracy in solar power forecasting. Prediction of solar power output of photovoltaic system gives considerable information for energy managers and all individuals who deal with trading electricity and scheduling of solar projects.

In future, simulated photovoltaic data set can be tried to generate with corresponding procedure. This data set can be implemented in other Hybrid Neuro Fuzzy Systems such as Fuzzy adaptive learning control networks, Generalised Approximate Reasoning based Intelligence Control, fuzzy net, etc. to get better predictive results compared with the proposed methods.

REFERENCE


