ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

IJESRT
INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

Survey on Machine Learning Approaches for Solar Irradiation Prediction

U.Divya*, Dr.Chitra Pasupathi
* Computer Science and Engineering, Prathyusha Institute of Technologyand Management, India

Abstracts

Renewable energy technologies are clean sources of energy that have a much lower environmental impact than conventional energy generation methods. Researches focusing on different energy generation techniques are gaining much importance worldwide, to manage exponential increase in the energy requirements. Solar energy is used in various applications like solar charged sensor nodes, solar charged vehicles, agriculture, electricity production etc. This solar energy can be harnessed using a range of technologies such as solar heating, solar photovoltaic cells, solar thermal electricity, solar architecture and artificial photosynthesis. The need for solar energy requires the estimation of solar energy production at various atmospherical conditions. This estimation involves the prediction of solar irradiation. Machine learning techniques based on Support Vector Machine (SVM), Neural Networks, Multilayer Perception(MLP), etc as well as Gaussian Process Regression method are normally applied for learning and predicting solar parameters. These models make use of parameters like air temperature, wind direction, relative humidity, and total rainfall as input to predict the temperature for a particular day. This paper highlights on the features of these different approaches for prediction and various metrics that are normally used for measuring the accuracy of the prediction process.

Keywords: solarradiation prediction, support vector machine, Multi-Layer perception, Gaussian process regression.

Introduction

The solar irradiation data provides information on how much energy is present at a specific location on the earth during a specific time period[1]. These data are required for designing solar energy systems. Solar energy is available throughout the year and also it is secure and clean. So the cleaner renewable solar energy can be replaced in various applications using fossil fuels when there is a shortage or raise in cost of fossil fuels like coal, oil and natural gas[1]. To make use of available solar energy efficiently, the energy can be harnessed using a range of technologies such as solar heating, solar photovoltaic cells, solar thermal electricity, solar architecture and artificial photosynthesis. There will be continuous need of energy for the applications using solar energy. This calls for the estimation of available solar energy and the estimation involves the prediction of solar energy.

Various models have been developed to predict daily global solar irradiation using the parameters collected by the meteorological centre[2]. Machine learning techniques based on Neural Networks, Support Vector Machines(SVM) are used to predict the global solar irradiation. The parameters used in the prediction model are average air temperature, average wind speed, current wind direction, average relative humidity, total rainfall, maximum peak wind gust, current evaporation, average absolute barometer, average solar radiation[3]. But the Machine Learning techniques based on Neural Networks, Support Vector Machine(SVM) are found to provide less accurate prediction values[4,7]. Alternatively, Gaussian Process Regression method[2] provides more robust and accurate daily global solar irradiation prediction values.

Daily solar irradiation prediction

The solar irradiation can be predicted using a wide range of meteorological data. The collected solar irradiation data may contain too much of information out of which only a few data are required for daily global solar irradiation prediction[3]. Also the use of noisy data will lead to failure in the prediction process. The noisy, irrelevant and redundant data from a wide-range of datasets can be reduced and filtered using various techniques [8],[9] like kalman filter. Now, the preprocessed data can be used to predict the daily global solar irradiation for more accurate results.

An advanced model for estimation of surface solar irradiance from satellite (AMESIS) has been developed to estimate the incident solar radiation at the surface from the spinning enhanced visible and infrared imager (SEVIRI) satellite measurements[10]. Nowadays, Photovoltaic systems are widely used in electricity production. Numerical weather prediction

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

model(NWM) has been used to predict the solar energy available for photovoltaic systems[11].

In Hidden Markov Model (HMM) with Pearson R model was utilized for the extraction of shape based clusters from the input meteorological parameters and it was then processed by the Generalized Fuzzy Model (GFM) to accurately estimate the solar radiation[18]. The Coral Reef Optimization algorithm has been designed in such a way that the Extreme Learning Machine solves the prediction problem, whereas the Coral Reef Optimization evolved the weights of the neural network, in order to improve the solutions obtained[19].

Machine learning techniques

Machine Learning focuses on prediction, based on known properties learned from the training data. Machine learning is employed in a range of computing tasks where designing and programming explicit, rule-based algorithms is infeasible. Machine learning tasks can be of several forms,

- In supervised learning, the computer is presented with example inputs and their desired outputs.
- In unsupervised learning, no labels are given to the learning algorithm, leaving it on its own to groups of similar inputs(clustering), density estimates or projections of high-dimensional data that can be visualized effectively.
- In reinforcement learning, a computer program interacts with a dynamic environment in which it must perform a certain goal.

Support vector machine

Support Vector Machine is a supervised machine learning technique, which is used for binary classification of the input into different sets after training the model with previous data. It can work in multiple dimensions represented by different variables. The Support Vector Machine has been used particularly in the classification of two different categories of patterns[21]. Support Vector Machine works by building a model which is used for classification[12]. It is based on a kernel method and is used for classification / regression, reducing dimensionality or clustering. In a simple linear regression technique such as minimizing least square error, the decision to select a regression function for the input data is based on fitting a line that minimizes the total square error for the entire data-set. Its goal is to minimize the average distance between the two clusters and the hyperplane which is equivalent to identifying a hyperplane having the maximum distance from the nearest data point from both the clusters[13].

Support Vector Machine algorithm has been used in the prediction of solar energy, wind energy by using different variables gathered at meteorological station. The input to the model served as combinations of maximum and minimum values of that gathered variable and other meteorological variables like air temperature, atmospheric transmissivity in a novel 2D form[13]. The prediction of power output from photovoltaic plant was done based on Support Vector Machine with wavelet analysis and similar data by decomposing the solar radiation signals[22].

Given some training data D, a set of *n* points of the form

$$D=\left\{(x_i,y_i)\middle| x_i \epsilon R^p, y_i \epsilon \{-1,1\}\right\}_{i=1}^{n} \quad (1)$$

where the y_i is either 1 or −1, indicating the class to which the point x_i belongs. Each x_i is a *p*-dimensional real vector. Maximum-margin hyperplane has been identified that divides the points having $y_i = 1$ from those having $y_i = -1$. Any hyperplane can be written as the set of points X satisfying

$$W . X - b = 0 \quad (2)$$

where . denotes the dot product and W the (not necessarily normalized) normal vector to the hyperplane. The parameter $\frac{b}{\|w\|}$ determines the offset of the hyperplane from the origin along the normal vector W.

Artificial neural networks

This technique is based on the simulation of a biological neural network, which consists of artificial neurons. By training the artificial neurons over a period of time (with historical data) one can predict future output. It is based on studying how the neurons connect (represented by connection weights) and pass information to each other for different inputs. Neurons collect the input from neighboring neurons and based on the magnitude of this input decide whether to fire or not[14].

Neural networks can be classified based on their structures,

Feed forward network

Signals flow always from the input layer to the output layer through unidirectional connections, the neurons being connected from one layer to the next, but not it in same layer.

Recurrent network

In recurrent network, signals can flow in both forward and backward directions. The outputs of some neurons are feed back to the same neurons or to neurons in preceding layers.

The network consists of three layers: an input layer, hidden layer and output layer. The input layer consist of all the input factors: information from the input layer is then processed in the course of one hidden layer; the output vector is then computed in the final or output layer[23].

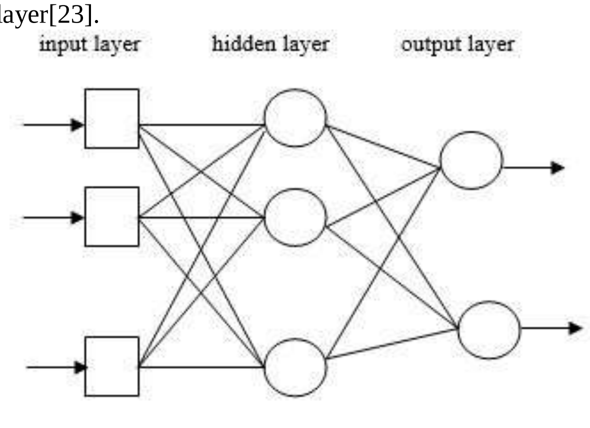

Fig.1:Multi-layer Perceptron

daily global solar predictions in many research works[15]-[17].

Gaussian process regression

Gaussian process regression is a method of interpolation for which the interpolated values are modeled by a Gaussian process governed by prior covariance. In this method, new data points can be constructed within the range of a discrete set of known data points. Under suitable assumptions on the priors, Gaussian process regression gives the best linear unbiased prediction of the values.

The main idea is to predict the value of the function at a given point by computing a weighted average of the known values of the function in the neighborhood of the point.

Gaussian regression process can handle the limitation in the availability of training data[4]. It involves parametric and non-parametric approach. The parametric approach compares two parametric values. It finds covariance between any two samples and uses the covariance estimation to predict the daily global solar irradiation[5],[7].The Temporal GPR approach is said to be more robust and more accurate than the other Machine Learning techniques as it finds covariance between two samples based on time-series[6].

Earlier, the Gaussian Process Regression was used to predict the data by finding covariance between two variables. Later, the prediction is done by finding covariance between two variables based on time series[2].

Prediction analysis

The predicted solar irradiation values are compared with actual or observed solar irradiation values for accuracy based on root mean square error or root mean square deviation (RMSD), mean absolute error(MAE) and mean bias error(MBE)[2],[20].

***RMSD*:**

The root-mean-square deviation (RMSD) or root-mean-square error (RMSE) is a frequently used measure of the differences between predicted values and actually observed values.

$$RMSD = \sqrt{\frac{\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}{n}} \qquad (3)$$

The RMSD of predicted values $\hat{y}_t$ for times t of a regression's dependent variable y is computed for n different predictions as the square root of the mean of the squares of the deviations.

***MAE*:**

The mean absolute error (MAE) is a quantity used to measure how close predictions are to the eventual outcomes. The mean absolute error is given by

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| = \frac{1}{n}\sum_{i=1}^{n}|e_i| \qquad (4)$$

The mean absolute error is an average of the absolute errors $e_i=|f_i-y_i|$, where f_i is the prediction and y_i the true value.

***MBE*:**

The mean bias error of an estimator is the mean difference between the estimator's expected value and the true value of the parameter being estimated.

$$MBE = \frac{(\sum_{i=1}^{N}(y_i - x_i))}{N} \qquad (5)$$

where y_i is the estimator's expected value and x_i is the estimated parameter value and N is the number of predictions.

Conclusion

In this paper, the machine learning techniques used for solar irradiation prediction was discussed. From this study, it was clear that Gaussian process regression method based on time series provides more accurate, bias and robust values when compared with other machine learning techniques like Support Vector Machine(SVM) and Artificial Neural Networks(ANN). And the accuracy can be estimated based on root mean square deviation(RMSD), mean absolute error(MAE) and mean bias error(MBE).The Gaussian process regression provides advantage in predicting solar irradiation with the limited collection of parameters. The parameters used for prediction are Air temperature, Relative humidity, Atmospheric pressure, Total rainfall, wind speed, wind direction, etc. The predicted values can be used in various applications using solar energy. Based on this study, Gaussian process regression based

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

on time series is recommended for accurate prediction of daily global solar irradiation.

References

1. T.Khatib, A.Mohamed, and K.Sopian , M.Mahmoud, "Assessment of Artificial Neural Networks for Hourly Solar Radiation Prediction" *International journal of Photoenergy.*,2012.
2. Sancho Salcedo-Sanz, Casanova-Mateo, Jordi Munoz-Mari" Prediction of Daily Global Solar Irradiation Using Temporal Gaussian Processes"IEEE *Geosci.Remote Sens.*,vol.11,No.11,2014.
3. Md Rahat Hossain, Amanullah Maung Than Oo, A B M Shawkat Al, "The Combined Effect of Applying Feature Selection and Parameter Optimization on Machine Learning Techniques for Solar Power Prediction" *American Journal of Energy Research*, Vol. 1, No. 1, 7-16,2013.
4. L.Pasolli, F.Melgani, and E.Blanzieri, "Gaussian process regression for estimating chlorophyll concentration in subsurface waters from remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 3, pp. 464- 468, Jul. 2010.
5. J.Verrelst, L.Alonso, G.Camps-Valls, J.Delegido, and J.Moreno, "Retrieval of vegetation biophysical parameters using Gaussian process techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, pt. 2, no. 5, pp. 1832–1843, May 2012.
6. J. Verrelst, L. Alonso, J. P. Rivera, J.Moreno, and G. Camps-Valls, "Gaussian process retrieval of chlorophyll content from imaging spectroscopy data," *IEEE J. Sel. Topics Appl.Earth Observ. Remote Sens.*, vol. 6, no. 2,pp. 867–874, Apr. 2013.
7. M. Lázaro-Gredilla, M. K. Titsias, J. Verrelst, and G. Camps-Valls, "Retrieval of biophysical parameters with heteroscedastic Gaussian processes,"*IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 4, pp. 838–842,Apr. 2014.
8. Omar W. Ahmed , Rami Qahwaji , Tufan Colak ,Paul A. Higgins , Peter T.Gallagher, D. Shaun Bloomfield " Solar Flare Prediction Using Advanced Feature Extraction, Machine Learning, and Feature Selection"*Solar Phys*,2011.
9. Jingying Fu , Dong Jiang , Yaohuan Huang , Dafang Zhuang and YongWang,"A Kalman Filter-Based Method for Reconstructing GMS-5 Global Solar Radiation by Introduction of *In Situ* Data"*energies*,2013.
10. E.Geraldi, F.Romano, and E.Ricciardelli, "An advanced model for the estimation of the surface solar irradiance under all atmospheric conditions using MSG/SEVIRI data," *IEEE Trans. Geosci. Remote Sens.*, vol. 50,no. 8, pp. 2934–2953, Aug. 2012.
11. E.Lorenz, J.Hurka, D.Heinemann, and H.G.Beyer, "Irradiance forecasting for the power prediction of grid-connected photovoltaic systems," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 2, no. 1, pp. 2–10, Mar. 2009.
12. J. L.Chen, H. B. Liu,W.Wu, and D. T. Xie, "Estimation of monthly solar radiation from measured temperatures using support vector machines—A case study," *Renew. Energy*, vol. 36, no. 1, pp. 413–420, Jan. 2011.
13. J.Zeng and W.Qiao, "Short-term solar power prediction using a support vector machine," *Renew. Energy*, vol. 52, pp. 118–127, Apr. 2013.
14. A. S.Dorvlo, J. A.Jervase, and A.Al-Lawati, "Solar radiation estimation using artificial neural networks," *Appl. Energy*, vol. 71, no. 4, pp. 307–319, Jul. 2002
15. Emad A.Ahmed and M.El-Nouby Adam, " Estimate of Global Solar Radiation by Using Artificial Neural Network in Qena, Upper Egypt",*Journal of Clean Energy Technologies, Vol. 1, No. 2, April 2013*.
16. M.A.AbdulAzeez,"Artificial Neural Network Estimation of Global Solar Radiation Using Meteorological Parameters in Gusau, Nigeria" , Archives of Applied Science Research, 2011.
17. Chukwu, S.C and Nwachukwu, A.N " Analysis of some meteorological parameters using artificial neural network method for Makurdi, Nigeria", African Journal of Environmental Science and Technology Vol. 6(3), pp. 182-188, March 2012.
18. S.Bhardwaj, V.Sharma, S.Srivastava, O.S.Sastry, B.Bandyopadhyay, S.S. Chandel, and J.R.Gupta, "Estimation of solar radiation using a combination of hidden Markov model and generalized fuzzy model," *Sol.Energy*, vol. 93, pp. 43–54, Jul. 2013.
19. S. Salcedo-Sanz a , C. Casanova-Mateo b, A. Pastor-Sa´nchez a, M. Sa´nchez-Giro´n," Daily global solar radiation prediction based on a hybrid Coral Reefs Optimization –

Extreme Learning Machine approach",Solar Energy,2014.
20. M. Ant´on , R. Rom´an, A. Valenzuela , F. J. Olmo3,4, and L. Alados-Arboledas, "Direct-sun total ozone data from a spectroradiometer: methodology and comparison with satellite observations" , Atmos. Meas. Tech., 6, 637–647, 2013.
21. Jianwu Zeng, Wei Qiao, "Support Vector Machine-Based Short-Term Wind Power Forecasting", *IEEE PES Power System Conference and Exposition*, held in Phoenix, Arizona, U.S.A., March 20-23, 2011.
22. Xiyun Yang, Feifei Jiang, Huan Liu, “Short-Term Power Prediction of Photovoltaic Plant Based on SVM with Similar Data and Wavelet Analysis”, ISSN 0033-2097, R. 89 NR 5/2013.
23. Rasheed H. AL-Naimi, Ali M. AL-Salihi, Dher I. Bakr, “Neural network based global solar radiation estimation using limited meteorological data for Baghdad, Iraq”, International Journal of Energy and Environment (IJEE), Volume 5, Issue 1, 2014.