Solar Radiation Prediction using Adaboost Algorithm

D.Kamalasri[1] J.Arun Prasath[2] R.Thandaiah Prabu[3]
[1]P.G. Student [2,3]Assistant Professor
[1,2,3]Prathyusha Institute of Technology and Management, Thiruvallur

Abstract— Predictions of incoming solar energy are acquiring more importance, because of strong increment of solar power generation. Predictions is very useful in solar energy applications because it permits to generate solar data for locations where measurements are not available. In existing systems, solar radiation is predicted using fuzzy logic and neural networks separately. So that Mean absolute percentage error is greater than 10%. Fuzzy logic and neural networks are combined together using Takagi Sugeno Kang (TSK) method. TSK method is very efficient than mamdani method. Previous year solar radiation data is collected from National Environmental Agency and using this values neural network was trained. The graph between measured and predicted data values was plotted .Error is calculated using the difference between desired and output value. Prediction using combination of fuzzy and neural network model having Mean Absolute Percentage Error 6% was achieved. But in order to reduce Absolute Percentage Error value, we need to check the validity of the input data, so, that Adaboost algorithm is introduced in our proposed method. Adaboost algorithm is one of the best method of classification of weak learners. The algorithm classifies the training and testing data and also produces the corresponding errors. After finding the errors, it will be neglected from input data to make the predicted with more accuracy and less error. So that Mean Absolute Percentage Error -2.33% was achieved.

Key words: Solar Radiation Prediction, Adaboost

I. INTRODUCTION

Radiation from the sun sustains life on earth and determines climate. The energy flow within the sun results in a surface temperature of around 5800 K, so the spectrum of the radiation from the sun is similar to that of a 5800 K blackbody with fine structure due to absorption in the cool peripheral solar gas.

A. Solar Constant and Sun Value

The irradiance of the sun on the outer atmosphere when the sun and earth are spaced at 1 AU - the mean earth sun distance of 149,597,890 km is called the solar constant. Currently accepted values are about 1360 W m^{-2} .The World Metrological Organization promotes a value of 1367 W m^{-2}. The solar constant is the total integrated irradiance over the entire spectrum .The irradiance falling on the earth's atmosphere changes over a year by about 6.6% due to the variation in the earth/sun distance. Solar activity variations cause irradiance changes of up to 1%. For Solar Simulators, it is convenient to describe the irradiance of the simulator in suns. One sun is equivalent to irradiance of one solar constant.

B. Extraterrestrial Spectra

The spectrum of the solar radiation outside the earth's atmosphere. The range 200 - 2500 nm, includes 96.3% of the total irradiance with most of the remaining 3.7% at longer wavelengths. Many applications involve only a selected region of the entire spectrum. In such a case, a 3 sun unit has three times the actual solar irradiance in the spectral range of interest and a reasonable spectral match in this range.

C. Terrestrial Spectra

The spectrum of the solar radiation at the earth's surface has several components. Direct radiation comes straight from the sun, diffuse radiation is scattered from the sky and from the surroundings. Additional radiation reflected from the surroundings depends on the local albedo. The total ground radiation is called the global radiation. The direction of the target surface must be defined for global irradiance. For direct radiation the target surface faces the incoming beam.

II. BOOSTING

A. Boosting

Boosting is a machine learning ensemble meta-algorithm for reducing bias primarily and also variance in supervised learning, and a family of machine learning algorithms which convert weak learners to strong ones. Boosting is based on the question posed by Kearns and Valiant. A weak learner is defined to be a classifier which is only slightly correlated with the true classification. In contrast, a strong learner is a classifier that is arbitrarily well-correlated with the true classification.

When first introduced, the hypothesis boosting problem simply referred to the process of turning a weak learner into a strong learner. Informally, the hypothesis boosting problem asks whether an efficient learning algorithm that outputs a hypothesis whose performance is only slightly better than random guessing a weak learner implies the existence of an efficient algorithm that outputs a hypothesis of arbitrary accuracy a strong learner. Algorithms that achieve hypothesis boosting quickly became simply known as boosting. Freund and Schapires arcing Adaptive Resampling and Combining, as a general technique, is more or less synonymous with boosting.

B. Boosting Algorithm

While boosting is not algorithmically constrained, most boosting algorithms consist of iteratively learning weak classifiers with respect to a distribution and adding them to a final strong classifier. When they are added, they are typically weighted in some way that is usually related to the weak learners' accuracy. After a weak learner is added, the data is reweighted examples that are misclassified gain weight and examples that are classified correctly lose weight. Thus, future weak learners focus more on the examples that previous weak learners misclassified.

There are many boosting algorithms. The original ones, proposed by Robert Schapire and Yoav Freund , were not adaptive and could not take full advantage of the weak learners. However, Schapire and Freund then developed AdaBoost, an adaptive boosting algorithm that won the prestigious Prize. Only algorithms that are provable boosting algorithms in the probably approximately correct learning formulation are called boosting algorithms. Other algorithms that are similar in spirit to boosting algorithms are sometimes called leveraging algorithms, although they are also sometimes incorrectly called boosting algorithms.

The main variation between many boosting algorithms is their method of weighting training data points and hypotheses. AdaBoost is very popular and perhaps the most significant historically as it was the first algorithm that could adapt to the weak learners.

Boosting algorithms are used in Computer Vision, where individual classifiers detecting contrast changes can be combined to identify Facial Feature.

III. ADABOOST ALGORITHM

AdaBoost, short for Adaptive Boosting, is a machine learning meta-algorithm formulated by Yoav Freund and Robert Schapire who won the prestigious Godel Prize in 2003 for their work. It can be used in conjunction with many other types of learning algorithms to improve their performance. The output of the other learning algorithms is combined into a weighted sum that represents the final output of the boosted classifier. AdaBoost is adaptive in the sense that subsequent weak learners are tweaked in favor of those instances misclassified by previous classifiers. AdaBoost is sensitive to noisy data and outliers. In some problems, however, it can be less susceptible to the over fitting problem than other learning algorithms. The individual learners can be weak, but as long as the performance of each one is slightly better than random guessing, their error rate is smaller than 0.5 for binary classification, the final model can be proven to converge to a strong learner.

While every learning algorithm will tend to suit some problem types better than others, and will typically have many different parameters and configurations to be adjusted before achieving optimal performance on a dataset, AdaBoost is often referred to as the best out-of-the-box classifier. When used with decision tree learning, information gathered at each stage of the AdaBoost algorithm about the relative 'hardness' of each training sample is fed into the tree growing algorithm such that later trees tend to focus on harder to classify examples.

A. *Algorithm*

With:

1) Samples $x_1, \ldots\ldots x_n$ (1.1)
2) Desired outputs $y_1, \ldots\ldots y_n$, $y \in \{-1,1\}$ (1.2)
3) Initial weights $w_{1,1} \ldots. w_{n,1}$ set to $\frac{1}{n}$ (1.3)
4) Errorfunction $E(f(x),y,i) = e^{-yif(x_i)}$ (1.4)
5) Weak learners $h:x \rightarrow [-1,1]$ (1.5)

For t in 1….T:

1) Choose $f_t(x)$
2) Find weak learner $h_t(x)$ that minimizes ϵ_t, the weighted sum error for misclassified points
$$\textstyle\sum_i w_{i,t} E(h_t(x),y,i) \quad (1.6)$$
3) Choose $\alpha t = \frac{1}{2}\ln(\frac{1-\epsilon t}{\epsilon t})$ (1.7)
4) Add to ensemble:
$$F_t(x) = F_{t-1}(x) + \alpha_t h_t(x) \quad (1.8)$$
5) Update weights:
$$w_{i,t+1} = w_{i,t}\, e^{-y_i \alpha_t h_t(x_i)} \quad \text{for all } i \quad (1.9)$$
6) Renormalize $w_{i,t+1}$ such that
$$\textstyle\sum_i w_{i,t+1} = (1.10)$$
$$\frac{\sum h_{t+1}(x_i)=y_i w_{i,t+1}}{\sum h_{t+1}(x_i)\neq y_i w_{i,t+1}} = \frac{\sum h_t(x_i)=y_i w_{i,t}}{\sum h_t(x_i)\neq y_i w_{i,t}}$$
7) It can be shown that at every step, which can simplify the calculation of the new weights.choosing α_t : α_t is chosen as it can be analytically shown to be the minimizer of the exponential error function for discrete adaBoost.
8) Minimize
$$\textstyle\sum_i w_i\, e^{-y_i h_i \alpha_t} \quad (1.11)$$

Using the convexity of the exponential function, and assuming that we have
$$= \left(\frac{1+\epsilon_t}{2}\right) e^{\alpha_t} + \left(\frac{1-\epsilon_t}{2}\right) e^{-\alpha_t} \quad (1.12)$$

We then differentiate that expression with respect to α_t and set it to zero to find the minimum of the upper bound:
$$\left(\frac{1+\epsilon_t}{2}\right) e^{\alpha_t} - \left(\frac{1-\epsilon_t}{2}\right) e^{-\alpha_t} = 0 \quad (1.13)$$
$$\alpha_t = \frac{1}{2}\ln\left(\frac{1-\epsilon_t}{1+\epsilon_t}\right) \quad (1.14)$$

Note that this only applies when $h_i \in \{-1,1\}$, though it can be a good starting guess in other cases, such as when the weak learner is biased ($h(x) \in \{a,b\}, a \neq -b$), has multiple leaves ($h(x) \in \{a,b,\ldots.n\}$) or is some other function $h(x) \in R$. In such cases the choice of weak learner and coefficient can be condensed to a single step in which $f_t = \alpha_t h_t(x)$ is chosen from all possible α, h as the minimizer of $\sum_i w_{i,t}\, e^{-y i f_t(x_i)}$ by some numerical searching routine.

B. *Direct Normal Irradiance*

It is the amount of solar radiation received per unit area by a surface that is always held perpendicular (or normal) to the rays that come in a straight line from the direction of the sun at its current position in the sky. Typically, you can maximize the amount of irradiance annually received by a surface by keeping it normal to incoming radiation. This quantity is of particular interest to concentrating solar thermal installations and installations that track the position of the sun.

C. *Direct Normal Irradiance Measurements*

1) *Pyranometer*

Pyranometers act as solar energy *transducers*, in that they collect irradiance signals and transform them into electrical information signals. That information is passed on to a data logger and computer, and then we either present the data in short bursts (1 second) or integrate and average the data over longer periods of 1 minute to 1 hour.

2) *Pyrheliometer*

If we wished to measure only the direct component of down welling irradiation, we would use a pyrheliometer. The device is a combination of a long tube with a thermopile at the base of the tube and a two-axis tracking system to always point the aperture of the device directly normal to the surface of the Sun. A measure of irradiance from a pyrheliometer is therefore called Direct Normal Irradiance **(**DNI**)** data. An Eppley Normal Incidence Pyrheliometer is displayed below on the left, while an Eppley Solar Tracker is displayed on the right.

D. Mean Absolute Percentage Error

Mean Absolute Error is a value regularly used to measure the accuracy of predictions against actual conditions. This statistic is determined by calculating the difference between all of the predicted values and all of the actual observed values and then finding the average of all the resulting values.

The mean absolute percentage error (MAPE), also known as mean absolute percentage deviation (MAPD), is a measure of accuracy of a method for constructing fitted time series values in statistics, specifically in trend estimation. It usually expresses accuracy as a percentage, and is defined by the formula:

$$M = \frac{1}{n}\sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right. \quad (1.15)$$

Where A_t is the actual value and F_t is the forecast value.

The difference between A_t and F_t is divided by the Actual value A_t again. The absolute value in this calculation is summed for every fitted or forecasted point in time and divided again by the number of fitted points n. Multiplying by 100 makes it a percentage error.

IV. SOLAR RADIATION PREDICTION USING ADABOOST

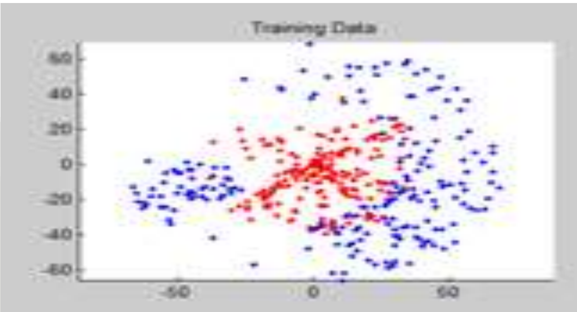

Fig. 1.1: Training Data

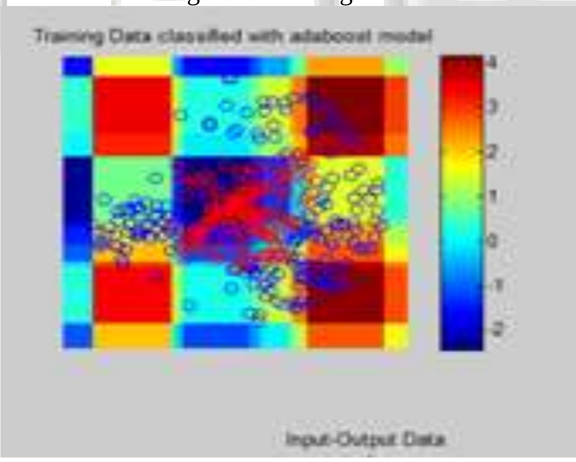

Fig. 1.2: Data Classification

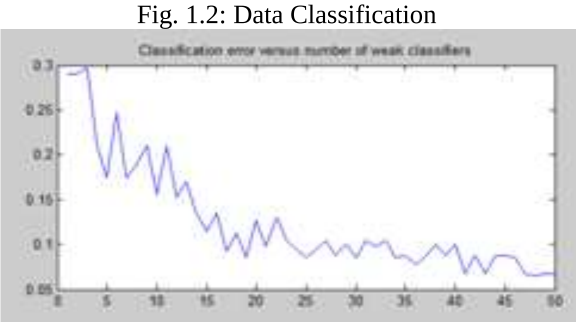

Fig. 1.3: Classification Error versus Number of Weak Classifiers

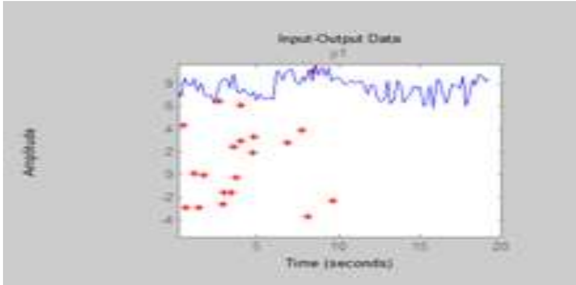

Fig. 1.4: Input – Output Data

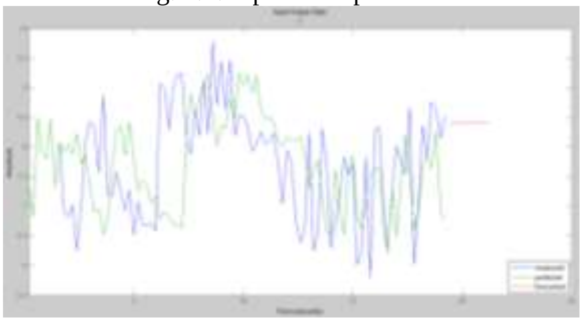
Fig. 1.5: Predicted Output

V. CONCLUSION

Solar radiation was forecasted using adaboost algorithm in this project. AdaBoost, short for Adaptive Boosting, is a machine learning algorithm. In fuzzy logic and neural networks based solar radiation prediction, Mean Absolute Percentage Error was greater than 3% achieved. In order to reduce MAPE value, adaboost algorithm is used to check the validity of the input. After check the validation, it will find the weak learners. Then it neglect the error from input, it will give the accurate predicted value. The proposed method can be used for both hourly and day-to-day solar radiation forecast. Once the sky and temperature information for the future period can be obtained from NEA, the solar radiation can be forecasted by the proposed method and with reasonably good accuracy. The quantitative results of this project show that the different sky conditions will not affect the accuracy of the forecasted results very much. The radiation forecast results are more accurate compared with those of other forecast methods. So that, using this project MAPE value was reduced to 2.33%.

REFERENCES

[1] A. Ahmed and M. El-Nouby Adam, "Estimate of Global Solar Radiation by Using Artificial Neural Network", Journal of Clean Energy Technologies, Vol. 1, No. 2, 2013
[2] Chen S, Gooi H, Wang M. "Sizing of energy storage for microgrids". IEEE Trans Smart Grid, 3(1):142e51., 2012
[3] Chen SX, Gooi HB. "Jump and shift method for multi-objective optimization". IEEE Trans Ind Electron, 58(10): 4538e48, 2011