Weather Forecasting Using Predictive Analysis for Determination of Solar Output of PV Cells

Supriya Bachal¹, Shubhada Mone², Aparna Ghodke³ & Ujwala Ahirrao⁴

¹, ³ & ⁴Student of Dept. of Computer Engineering, Marathawada Mitramandals College Of Engineering, Pune, Maharashtra, India.
²Professor, Dept. of Computer Engineering, Marathawada Mitramandals College Of Engineering, Pune, Maharashtra, India.

Abstract: Weather is an ever changing phenomenon. It has various components such as temperature, rainfall, cloud cover, humidity etc. Weather affects every sphere of human life, whether it is business or Entertainment. Weather Forecasting is one of the most technologically challenging tasks. Weather forecasting using Data mining is an innovative way of predicting the weather with minimal cost and maximum accuracy.

1. Introduction

The need for weather forecasting has been amply justified over the years. Weather forecasting has existed since times unknown. The method for weather forecasting and the technology used has since massive improvement over the years. In ancient times the weather cock was used to determine the direction of the wind, the color of the clouds determined the likelihood of precipitation and so on. As new and profound technology replaced these methods the forecasts became more accurate and precise. Weather forecasting has been one of the most scientifically and technologically challenging problems around the world in the last century.

Various methods such as climatology, numerical weather forecasting and analog methods are preliminarily used for weather forecasting. The common thread between all these methods is that the prediction is made on basis of the past occurrences of the characteristic of weather. It is thus noted that weather records for multiple hundreds of years are available for our perusal. Since such humungous amounts of historical data is available to us using data mining to predict the weather seems to be a smart option. Predictive analysis is a sub type of data mining which is centered on pattern mining. Data mining to predict weather is a smart, effective and accurate way to get optimum results for predicting the weather.

Today, solar power has become part of our daily lives. Appliances like solar notebooks, solar air-conditioners, solar cars, etc. demonstrate the use of the sustainable power of the sun. As the adverse effects of burning of fossil fuels and the depletion rate of non-renewable energy sources increase, the future of solar energy looks bright. The problem with renewable sources of energy is that they are not easily predictable in advance and vary based on both weather as well as site specific conditions.

1.1 Approach

In this paper we have used an altered version of the KNN algorithm. It is based on the fact that weather shows a repetitive pattern. To elaborate this consider the example of the monthly weather, it often said that summer sets in may or the second half of June will experience a certain amount of precipitation or terms such as “October heat” are derived from the fact that in the past India has experience similar weather at those times. It is thus safe to conclude that the weather conditions of a certain place at a certain time show the characteristic for repetitiveness. We exploit this quality to form our algorithm as well as for fine tuning the output of our system.
1.2 Strategies for Better Accuracy

It is often seen that weather related data is noisy and contains many outliers. We observed that when the value for parameter to be predicted for the day in ant of the previous years is a garbage value or unavailable the algorithm fails to predict the value for the parameter. To combat this problem we used a two way approach. To combat noisy data the dataset was first preprocessed and the missing values were replaced with a global constant. When we encounter a situation where the data to be considered is a missing value, the preliminary check performed identifies the global constant and the data from the previous day is used instead.

The second approach we adopted was to set a maximum and min criteria for the parameter to be predicted. We analyzed data for the past 5 years and we set values for the max and min occurrences for every month, if the system encounters a value which does not lie between the range of the maximum and minimum value it is scrapped as garbage value. In its place the previous day's data is used.

Predictions are always tentative and many factors such as global warming affect the climate. If a prediction is made by the system which falters we adopt the feedback method. The predictions that we make are cross checked with the prediction from an authentic source and if the predictions are not found admissibly accurate the actual accurate predictions are placed in the system instead of the faltering predictions.

2. ALGORITHM

In this paper we have used an altered version of the KNN algorithm. It is based on the fact that weather shows a repetitive pattern. To elaborate this consider the example of the monthly weather, it often said that summer sets in may or the second half of June will experience a certain amount of precipitation or terms such as “October heat” are derived from the fact that in the past India has experience similar weather at those times. It is thus safe to conclude that the weather conditions of a certain place at a certain time show the characteristic for repetitiveness. We exploit this quality to form our algorithm as well as for fine tuning the output of our system.

2.2 Algorithm

1. Define k = 4 (group of 5 nearest similar days)
2. Identify day of month x to be predicted
3. for each parameter p to be predicted
   4. for each previous year’s month m
5. Identify value x1 for x day of month m for parameter p
6. for other days d of month m
7. Identify value x2 for parameter p
8. Identify distance for d and x using vectors x1 & x2 and add to set distances as
   Dig=SQRT(∑xij-xjk)^2
9. End for
10. Sort distances set in increasing order
11. Select first k days from distances as k nearest neighbors to x day from other days and add to set nun
12. End for
13. Store
14. Identify frequency of other days in set nun
15. Identify day y with highest frequency in set nun
16. Set predicted output for day x = parameter value for day y of current month for the previous year.
17. Display predicted output for day x.

2.1 Elaboration of Algorithm

The first step of the algorithm defines k=4 which means we will find 5 nearer neighbors of the chosen value. In the second step we take input from the user about the date for which we will be predicting the weather and also the parameters for which the user needs the prediction. Parameters can be temperature, wind speed, cloud cover etc. Then we consider the same day in the same month of the previous year wrt the year for which the prediction is to be made. The value for the parameter to be predicted is taken into consideration. Then we apply the KNN algorithm by comparing the distance (Euclid distance) between the values of the parameter on all the days in the same month for the past 20 years. We store the date with the corresponding values. These values are arranged in the increasing order. The first 5 entries for the values are taken into consideration. We check the corresponding dates for these values. The date occurring most frequently is identified. The value for the parameter on that day for the same month is displayed if the most frequently occurring day has already passed else if the date is in the future the value for the parameter for the day is set to the value for the same parameter for the most frequently occurring day of the previous

3. Determination of Solar Output

The weather forecasts obtained are used for determination of the solar output of the solar cells often called PV cells. One of the most important factors that affect the productivity of the solar cell is the solar irradiance hence in order to predict the output of the solar cells we have to first predict the solar irradiation. Then the Energy that can be produced by the solar cell and then we calculate the De rating factors are taken into account. Prediction
of the Energy a solar cell will generate is a threefold process.

### 3.1 Calculations for Solar Radiation Using Zhang Huang Solar Model.

\[
I = \frac{[I_0 \cdot \sin(h) \cdot (c_0 + c_1 \cdot CC + c_2 \cdot CC^2 + c_3 (T_n - T_n - 3) + c_4 \cdot \phi + c_5 \cdot V_w) + \alpha]}{k}
\]

Where

- \( I \) = estimated solar radiation, W/m²
- \( I_0 \) = global solar constant, 1355 W/m²
- \( h \) = solar altitude angle, i.e., the angle between the horizontal and the line to the sun

The following are the altitude angles for Agartala in India.

<table>
<thead>
<tr>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
</tr>
</thead>
<tbody>
<tr>
<td>50°</td>
<td>48°</td>
<td>66°</td>
<td>74°</td>
<td>82°</td>
<td>90°</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>82°</td>
<td>74°</td>
<td>66°</td>
<td>58°</td>
<td>50°</td>
<td>42°</td>
</tr>
</tbody>
</table>

- \( CC \) = cloud cover
- \( \phi \) = relative humidity, %
- \( T_n, T_n-3 \) = dry-bulb temperature at hours \( n \) (current) and \( n-3 \), respectively
- \( V_w \) = Wind speed, m/s
- \( c_0, c_1, c_2, c_3, c_4, c_5, d, k \) = regression coefficients

The constants were determined by analysis from measured data and are as follows:

- \( c_0 = 0.5598 \)
- \( c_1 = 0.4982 \)
- \( c_2 = 0.7672 \)
- \( c_3 = 0.02842 \)
- \( c_4 = 0.00317 \)
- \( c_5 = 0.014 \)
- \( d = -17.853 \)
- \( k = 0.843 \)

### 3.2 Calculating the Solar Energy Generated By Solar Panels.

Maximum (kWp) power out of a solar cell

\[
K (\text{WP}) = \frac{n \cdot WP}{1000}
\]

\( n \) = no of solar cells

\( WP \) = the max power of each cell

Then,

\[
E_p = I \cdot K (\text{WP}) \cdot \eta
\]

\( I \) = solar radiation (from Zhang Huang model)

\( K (\text{WP}) \) = from equation 1

\( \eta \) = efficiency of solar cell

### 3.3 Derating Factors

- Temperature

  The solar cell is affected by the temperature. The surface temperature of the solar cell should be ideally 25 degrees if the temperature increases the output of the solar cell is affected negatively. The effect of temperature on the solar cell depends upon the material which the solar cell is made up of.

  Monocrystalline cells: temperature coefficient of 
  \(-0.45\% / °C\) (for every degree above 25 °C, the power should be derated by 0.45%)

  Polycrystalline modules: temperature coefficient of
  \(-0.5\% / °C\)

  Amorphous modules: temperature coefficient of
  \(-0.2\% / °C\) (these modules have a much lower coefficient due to a different cell temperature characteristic).

- Dust

  Dust is present in the environment at all times. The efficiency of solar cells depends on the environment in which they are present if the pollution and dust in the environment is very high then they under perform. Generally derating for dirt is considered—5%.

### 4. Characteristics Of The System

The User Classes and characteristics

<table>
<thead>
<tr>
<th>User</th>
<th>The user first register in the system. The primary goal of user is to enter the “number of days” (max 1 week ahead) for which he/she wants the weather prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Admin</td>
<td>The admin will have to register/login in order to view/edit. The admin will have the right to feed all the data about the weather for training purposes as well as previously fed data.</td>
</tr>
<tr>
<td>Prediction</td>
<td>According to the trained datasets, the results of the prediction will be provided to the users.</td>
</tr>
</tbody>
</table>

### 4.1 Assumptions and Dependencies

i. Users

Users submit number of day’s data.

Users feel comfortable for providing data

ii. Admin

Content is available and provided by other admin.

Admin submit previous data correctly

Admin feel comfortable posting and annotating weather data

iii. Developers

The requirements are met, and clearly defined. The stakeholders are available for clarification and questions.

Technology standards for the system do not drastically change.

iv. Software Engineering Department

Users make effective use of the system.

Admin make effective use of the system.

The system enhances the learning experience.
v. Software development based on open source technology that is readily available

vi. Weather’s prediction Systems

Weather data
Current standards
JAVA
MySQL

4.2 System Features

1. Predicts the weather attributes such as temperature precipitation, humidity etc
2. Based on these results predicts the solar output of the pv cell.
3. It is a system that uses adaptive learning.
4. The system uses the predicted data to train the model further.
5. The admin is given access to the database to modify any noisy data or outliers.
   To make sure that accuracy is maintained.
7. The system makes use for registration and login for users and admin and validates password and username.

4.3 System Architecture

5. Conclusion

The paper sheds light on a data mining technique which can be used for weather prediction and forecasting, this weather forecast is used by the system to predict the output of the solar cells at a specific location. In totality the system helps to estimate the solar energy produced by a solar cells in a certain area by using its own weather forecast. It helps to promote solar energy and gives an impetus to development of a hybrid model using solar energy and conventional electricity sources.

6. References

11. Mellit, A.; Massi Pavan, A. 24-h forecast of solar irradiance using artificial neural network:
17. Yona, A.; Senjyu, T.; Saber, A.Y.; Funabashi, T.; Sekine, H.; Kim, C.-H. Application of Neural Network to One-day-ahead 24 hours Generating Power Forecasting for Photovoltaic System. In
29. Proceedings of the International Conference on Intelligent Systems Applications to Power Systems,
36. Yao Wang and Banavar Sridhar “Connective Weather Forecasts Accuracy Analysis At Center and Sector Level”, NASA Ames Research Center, Moffett Field, California.
40. 1064–1069.