Solar Architect - Intelligent Recommendation Engine for Solar Energy Battery Use

Shreyansh Kakadiya  
Computer Engineering  
Dwarkadas J. Sanghvi College of Engineering

Drumil Bakhai  
Computer Engineering  
Dwarkadas J. Sanghvi College of Engineering

Ronak Jain  
Computer Engineering  
Dwarkadas J. Sanghvi College of Engineering

Dharmesh Mistry  
Computer Engineering  
Dwarkadas J. Sanghvi College of Engineering

ABSTRACT
Few years ago, a 10 Gigabyte of data was considered to be a large amount of data. However, with the enhancement of ease of use of internet and social networking, data collection boosted to more than a terabyte of data in a day. With this increase in the data size, data storage devices and algorithms to mine the data increased paving way to a vast field of Data Mining. A large set of data gave rise to many application domain. Internet of Things and Sensor data can be used in data mining to broaden the scope. This paper describes a project undertaken to do predictive analysis using Data mining tools. A solar panel transmits data every two minutes. Suppose a Kb of data is transmitted in one shot, then 30 Kb of data is accumulated in an hour. Thousands of solar panel in a field can generate many gigabytes of data opening ways for data mining. This paper collects data from a set of solar panel (batteries replace solar panel in the project) and feed it to the system where it can processed and analyzed all-together to produce efficient energy reports. Hadoop is a tool for data mining that has different clustering and classification algorithms. Sensor data is sent to hadoop and data mining techniques are used to contribute to Green Energy Efficiency. This data about the generated voltage from the solar panel, will be used to estimate the approximate working time of the house appliances. Weather feeds will help to predict the weather, upcoming weather is favorable for high voltage production or not, thereby helping to plan efficient use of available voltage. The paper aims to solve the real world problem of Energy saving using the computational methods of data mining of the sensor data in the domain of Internet of Things.

General Terms  
Link Analysis Algorithms, Web Structure Mining

Keywords  
Web Mining, Web Structure Mining, Link Analysis, PageRank, Weighted PageRank, Hypertext Induced Topics Search

1. INTRODUCTION
Solar energy availability is never guaranteed and differs from places to places depending upon the location of the users as well as upon the timing of the day. Solar energy depends upon the very obvious, given time of any day and also prevailing weather conditions. The meteorological departments provides with constant real time updates of the weather and therefore their predictions can be used to for further strategic planning activities that requires solar energy.

The following three interesting questions arises. (1) Is it reliable to predict the energy consumption for tomorrow using the information about standard weather updates that are present today ? (2) Till what extent the prediction can be made based on our own measured isolation ? (3) Can the most optimal approach be determined using the two predictions source ?

The answer for these questions are very much based on the practical importance and not only just solely related to the academic knowledge.

2. METHODOLOGY
2.1 Data Collection
2.1.1 Data Acquistion
The solar powered panels will result into the charging of the batteries connected to the generators and will eventually emit the charged levels of the batteries to the Microcontrollers that are attached to the computers via the MCO cards. Data needs to transmitted serially over the connection using UART.

Preprocessing of the data will be performed in case of the inconsistent data and buggy appo
Data received on the computer will be processed using the help of port communication programs in JAVA. The code combining the concepts of UART and Java will help to create a file where the transmitted data will be written onto. Along with the conversion from the digital signals to computer readable format, the java code will also be able to perform cleaning and formatting of the data. Libraries on the COM port of the Java application are available which will help for the further communication of the programs.

Data sets can be created using the help of the files where the Java UART performs the write operations.

Now the data sets created will be fed into the Hadoop eco systems with the help of the component of the Hadoop known as the HIVE. Hive will allow us to preprocess the data and modify the data in such a manner that it will help to store the raw data into the structural format that can be understood by the database. Whenever the data is converted into structured sets, it will be presented in the format of tables consists of rows and columns.

We can perform all the CRUD operations i.e Create, Read, Update, Delete operation. The schema of the table will be organized in such a manner that it will consist of the parameters like Temperature, Battery levels, and other attributes which we will use for the prediction analysis. In this manner the data from the solar panels is converted and formatted in such a manner that it can be given as an input to the Hadoop ecosystems and used to perform the fuzzy decision trees.

2.1.2 Weather forecast

The battery levels will be fed from the text file created by java uart and then imported into the hadoop ecosystems and the live temperature are often recorded from the ground level source of information. Here in our project we analyze the real time climate updates provided by the various meteorological departments from around the corner of the world. The other parameters will include the location of the user's house i.e latitudes and longitudes and the current climate update related to that nearby location. Whenever the user wants to check about the battery levels of the solar panels, simply logs onto the application and can get real time update of the charged levels.

For the most accurate match with the data forecast and the current conditions, we implemented the usage of the default 48 standard categories. As well in order to provide most reliable information the analysis of the Twitter feed is been done and provided to the database for further precise prediction of the energy consumption. In order to minimize the fallacious data we choose to predict the weather updates just before the sunrise.

3. ENERGY AND WEATHER PREDICTION.

A major challenge that is faced after collection of data is prediction of energy that will be generated during the day. This prediction is to be made before the day begins, so as to inform the user about energy generation prior to occurrence of any problem. Different methods are used to do the same. All these methods can be broadly classified into two categories: direct and indirect. Direct method uses black box algorithm while the indirect one first predicts the weather class after which the average electricity is predicted.

This paper demonstrates use of the indirect algorithm, where we take data from the national weather forecasting services. Here we do not apply
any prediction to this data rather predict the energy. This formulation gives us a range of algorithms that can be applied. Here, we choose to use Fuzzy Decision tree. Fuzzy Decision tree can deal with numerical and symbolic data simultaneously as well as provides explanation to prediction.

3.1 Fuzzy Decision Tree

Classical decision tree can be extended to form Fuzzy Decision Tree (FDT). FDT, introduced in machine learning, helps to handle training sets that has fuzzy numerical data. FDT helps to develop soft boundary data rather than a crisp data. Degree of membership are provided as a result.

A training set \( S = \{a_1, \ldots, a_n\} \) constructs FDT based on ID3 or CART algorithm. The comparison attribute is used to sequentially partition the tree from the root to leaves i.e. dividing the set \( S \) in subsets. This makes the nodes of the tree.

Suppose example \( e \) from set \( S \), given by attributes \( A = \{A_1, \ldots, A_n\} \), such that attributes can take fuzzy, numerical or symbolic value. An attribute can be represented in the form of a set of attribute and value pair. The class and value of an attribute is associated with a membership value.

While the method proceeds, an attribute is selected based on measure of discrimination using theories that order the attributes with increasing correlation to class. The attribute with highest discrimination power is selected to form the node. The FDT can not only classify example in \( S \) but also examples outside \( S \). Classification takes place by following the path from the root to the node. On the path all membership values are given to the example \( e \) outside the set \( S \). This leads to membership of example \( e \) such that it belongs to class \( c \) (\( c \) belongs to the set of classes).

3.2 Baseline Prediction

Measuring the improvement of project demands to define a baseline and distance measure. To extent to which this prediction is accurate can be determined by taking the average of absolute values of differences between predicted and observed energy. Three baselines can be as follows:

- Constant average prediction: Assumptions are made that average energy for a region and for a period of time remains constant. However it is constant for all period. This is an ideal point that cannot be achieved in practical.
- Energy tomorrow equals the one produced today: Standardized method used in weather forecast prediction and time series.
- Pure weather forecast based prediction: Weather forecast is used as a predicted class. It uses the natural way that humans use: “If predictions say today is a sunny day, records says that on a sunny day energy \( E \) was produced then we can conclude that today \( E \) amount of energy will be generated.”

3.3 Bayesian Predicting Model

Bayesian Prediction Model is based on Bayes’ theorem. Bayes’ theorem basically calculates the posterior probability using the likelihood table. We first assume a frequency table. A likelihood table is then created based on the frequency table. This likelihood table provides all value of the probabilities. This means given the individual probabilities the baye’s theorem calculates probability of any event happening given an already occurred case.

Let \( P(c) \) be the probability of ‘c’ occurring. Let \( P(x) \) be the probability that ‘x’ occurs. \( P(x|c) \), therefore is the probability that ‘x’ occurs knowing ‘c’ has already occurred. Now, occurrence of ‘c’ can be predicted knowing that ‘x’ has turned up is \( P(c|x) = \frac{P(x|c) \cdot P(c)}{P(x)} \)

Example,

This technique can help in prediction of suitable weather for solar energy generation. We first construct a frequency table as follows.

Figure 2.
probability that the weather is Suitable knowing that is Sunny is,

\[ P(\text{Suitable}|\text{Sunny}) = \frac{0.44 \times 0.60}{0.33} = 0.8. \]

Thus, probability that a sunny weather is suitable is 0.8 for the above example. This represents the Bayesian prediction model.

4. CONCLUSION AND FUTURE SCOPE

Weather forecast are wrong upto forty percent and are based upon predictor's categories- therefore it is necessary to use efficient machine learning algorithm to uncover the secret underlying rules as well as it is necessary to manually add coherence for accurate systems. The Takagi Sugeno models are reliable but without a proper study, the controller's performance reduces because of the intricate relationship that exists between weather forecast and energy predictions. The tradeoff between accuracy and speed has to establish depending upon the implementation scenarios.

One of the most important future work should include the focus on developing the application which accurately predicts the cost of the electricity consumption made by a particular customer. The smart bill generation use the existing system for knowing the energy consumption and based on regional energy prices, it should calculate cost for the energy bills and suggest user some valuable feedback to reduce the same.

Other future work can focus on more reliable sources of gathering information such as sun-tracking panels and close integrations with efficient used batteries. The sun tracking will not only enable the sensors to gather the most of the light but it will also allow efficiency in which the data is collected. However the induction of the sun tracking panels will lead into significant increase in the cost of the entire systems and generally will increase overall complexity.

The current topic does not implement any methodology that gathers the data about the sunlight received by a particular area. For the father iterations of the project we will include any reliable source of the Thermal Imagining that provides the most accurate details of the amount of light received by a particular region.

By considering the humongous data that will be generated by the solar parks it is essential to improve the computation power of the computers. For doing so we need robust architecture that is less prone to failures and thereby an entire cluster of Hadoop eco-systems can be implemented. Hadoop Clusters can process the data in batch processing based systems and allow processing upto the size of three terabytes on a single day if a 20 node clusters is maintained. Solar energy being a large source of renewable energy gives us a solution to save resources for the future generations. Thus, this project paves the way. Implementation of such techniques will help to save resources for future generations as well as help us to make sustainable development.

5. REFERENCES


