Result of All the Models

August 1, 2019

1 Information about Data-set and Models used:

1. In all the models' one-month data has been used.

2. Once some model is performing well by taking a month as a feature, we can build the model for the entire year.

3. AR, MA, ARMA, ARIMA, models are performing well for one unit time ahead forecasting but not as good as classical machine learning model such as SVM.

4. We are getting optimal RMSE Score i.e. .11 when using Time of the Day, Temperature, Energy Produced one Time Stamp before as features, using SVM with Linear Kernel.

No	Features	RMSE	Time	Comment
1.	Moving Average with Previous Time Stamp Values	0.767	One time unit ahead	No comment
2.	Weighted Average of Previous 2 Time Stamp	0.560	One time unit ahead	No comment
3.	Moving Average with Previous Time Stamp Values for Entire Year.	0.900	One time unit ahead	No comment
4.	Weighted Average of Previous 2 Time Stamp for Entire Year	0.789	One time unit ahead	No comment
5.	SVM Using Linear Kernel	1.084	One time unit ahead	No comment
6.	SVM Using RBF Kernel	1.056	One time unit ahead	No comment
7.	Using Neural Network	1.053	One time unit ahead	No comment
8.	Using LSTM model	1.054	One time unit ahead	No comment
9.	Auto Regressive Model	0.845	One time unit ahead	No comment
10.	Moving Average Model	1.394	One time unit ahead	No comment
11.	Using Temperature, Previous Energy Generation as Feature	.413	One time unit ahead	SVM with Linear Kernel
12.	Above Features	1.379	One time unit ahead	SVM with RBF Kernel
13.	Above Features	.308	One time unit ahead	Neural Network
14.	Time of the Day, Energy, Temperature	0.116	One time unit ahead	SVM with Linear Kernel
15.	Above Features	.192	One time unit ahead	SVM with RBF Kernel
16.	Time of the Day, Energy, Temperature, Humidity	0.117	One time unit ahead	SVM with Linear Kernel
17.	Above Features	0.329	One time unit ahead	SVM with RBF Kernel
18.	Using Time of the Day as a Feature and SVM with Linear Kernel	0.490	Any given time of the day	No Comment
19.	Using Time of the Day as a Feature and SVM with Linear Kernel	0.490	Any given time of the day	No Comment
20.	Temperature, Time, Humidity	0.388	Any given time of the day	SVM with Linear Kernel
21.	Temperature, Time, Humidity	0.52	Any given time of the day	SVM with Poly Kernel

Dataset

We use the Solar Energy dataset, that has been collected from 10 different stations inside the IIT Gandhinagar campus. Four stations are having a capacity of 25 KWh and 6 stations are having a capacity of 15 KWh. The dataset has been collected for 1.5 years from January 2018 to July 2019. Each station is generating the data after every 20 minutes i.e one cell in the dataset is representing the total energy produced by a particular station in the last 20 minutes. In one day, a total of 72 data points has been collected for each station i.e. 3 data points in every hour. As there is no sunlight in the night time so all the data points from evening to next morning are having a value of zero. So while evaluating our model we have simply dropped the data points in the time interval of 6 PM to 6 AM. IIT Gandhinagar campus is in one of the remote areas outside the main city. So due to the network connectivity issues (Data was collected with the help of GPRS), sometimes the data collected was faulty. Weather data for the study has been collected from the Dark Sky API, it is containing various meteorological conditions of the site such as temperature, humidity, wind speed etc. Meteorological data was provided after every 1 hour, so in one day we were getting 24 reading.

Baseline

We took time series model and classical machine learning model (SVM, Neural Network) as the baseline model for the solar forecasting without using any external features for the solar forecasting. All these baseline models forecasting was 20 minute ahead. Following was the setup for train and test data for these baseline models.

$\mathrm{X}_{train} \rightarrow Energy_t, y_{train} \rightarrow Energy_{t+20}$

$\mathrm{X}_{test} \rightarrow Energy_t, y_{test} \rightarrow Energy_{t+20}$

Following were the results for all the baseline models.

Model	**RMSE**
Moving Average	0.767
Weighted Average	0.560
Autoregressive Model	0.845
ARMA Model	.578
ARIMA	x
SVM	1.056
Neural Network	1.053

As the forecasting time for all the baseline models was just 20 minutes, so our baseline models should give a fairly good result.

Evaluation Matric

We chose the MAPE (Mean Absolute Percentage Error) as our evaluation metric. Although MAPE sometimes can be misleading, most of the prior work has been done using MAPE so we have chosen MAPE as evaluation metric.

$$\mathrm{MAPE} = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{e_t}{y_t} \right|$$

Experimental Setup

With the help of dark sky API key, we collected the weather data and for the same period we collected the solar data from various stations. As in a country like India, weather conditions fluctuate a lot, so we take month as a feature in the dataset along with weather conditions. Now we trained our model using these meteorological conditions as X value and corresponding solar energy for a particular station as Y value. We performed the testing on the station having similar capacity to that of the trained station capacity.

Result

Result of the various models has been shown in the table 1. For the 20 minute minute ahead forecasting we are getting optimal result i.e. RMSE = 0.116 and For any given time, we are getting optimal RMSE i.e. 0.52. Both the RMSE score are less as compare to our all the baseline models.

Conclusion