Power Consumption Analysis and Prediction of a Smart Home Using ARIMA Model

Gopikrishna P.B.
MSc. Computer Science
St. Thomas College (Autonomous), Thrissur
Email:gopikrishnapb8@gmail.com

Dr. Jiju A. Mathew
Assistant Professor
St. Thomas College (Autonomous), Thrissur
Email: jijuamathew@gmail.com

Muhammed Arif P
Software Developer
Ernst & Young Global Shared Service Centre, Trivandrum
Email:hisajarif@gmail.com

Abstract

Smart home is an important application of Internet of Things (IoT) that utilizes the Internet to monitor and control appliances in a home automation system by providing security, energy efficiency, low operating costs, convenience *etc*. The installation of smart products allows the user to monitor and control energy usage in new ways which leads to reduce energy wastage of the system. It is possible to analyze the usage of electricity in a smart home and creating an efficient power consumption analysis system using the data retrieved. Proposed system analyses the real-time electrical energy consumption of a smart home and will forecast the future power consumption using ARIMA model which gives better results for the time-series data analysis. This system aims is to predict the future electricity consumption for some services in smart homes in an efficient way.

Introduction

IoT refers to a self-configuring wireless network which enables the Internet to reach out into the real world of physical objects. It is the need of the hour to communicate remotely with the day-to-day interacting appliances using a portable device like a smartphone with internet connectivity and was made possible through IoT. A smart home is an application of IoT that utilizes the Internet to monitor and control appliances using a home automation system. Smart home systems achieved a great popularity in the last decades as they increase the comfort as well as quality of life through smartphones and microcontrollers. Energy consumption in the residential sector represents an important part of the total electricity demand. In this context, a proper prediction of energy demand in housing sector is very important [1]. Predictive analysis is an approach to forecast data on the basis of historical data. Time-series data is the order of historical data, which resembles the group or observation of the data that have been collected over time according to the continuous period of time. It consists of four components: trend, seasonal effect, cyclical, and irregular effect [2]. The analysis of

Electronic copy available at: https://ssrn.com/abstract=3819512

a time series used forecasting techniques is to identify and built models from the past data. According to the trend of data, we can forecast the future events.

Predictive analytics is a category of data analytics which aimed towards creating predictions about future outcomes based on historical data and analytics techniques such as statistical modelling and machine learning. It will generate future insights with a significant degree of precision. Predictive analytics uses many techniques from the areas of data mining, statistics, modelling, machine learning, and artificial intelligence for analysing current data to make predictions about future and involves in extracting data from existing data sets with the goal of identifying trends and patterns.

Many models are implemented to deal with time series data, ARIMA is one of the models which can be used for predictive analysis. ARIMA stands for AutoRegressive Moving Average; is actually a class of models that explains a given time-series data based on its own past values, *i.e.* its own lags and the lagged will forecast the errors, so that using equations, it is possible can forecast future values. ARIMA model uses the dependency between an observation and residual errors from a moving average model applied to lagged observations and are applied in cases where data shows some non-stationarity in it [3].

In this paper, future power consumption of the smart home is predicted using ARIMA model which is one of the best models to deal with time-series data. This power consumption analysis and prediction helps to reduce the usage of electrical energy in a smart home which results in energy saving as well as in implementation of an efficient energy analysis system.

ARIMA Model

Predictive analysis deals with many steps to get an accurate solution for the given dataset. ARIMA model is a class of model that captures a suite of different standard temporal structures in time series data, which is popular and widely used statistical method for time series forecasting. ARIMA stands for AutoRegressive Integrated Moving Average, and it is an integrated generalization of the simpler AutoRegressive Moving Average model [3].

- **AR**: *Autoregression*. A model that uses the dependent relationship between an observation and some number of lagged observations.
- **I**: *Integrated*. The use of differencing of raw observations (e.g. subtracting an observation from an observation at the previous time step) in order to make the time series stationary.
- **MA**: *Moving Average*. A model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations.

Electronic copy available at: https://ssrn.com/abstract=3819512

Each of these components are explicitly specified in the model as a parameter and a standard notation is used in ARIMA (p, d, q) where the parameters are substituted with integer values which quickly indicate the specific ARIMA model being used.

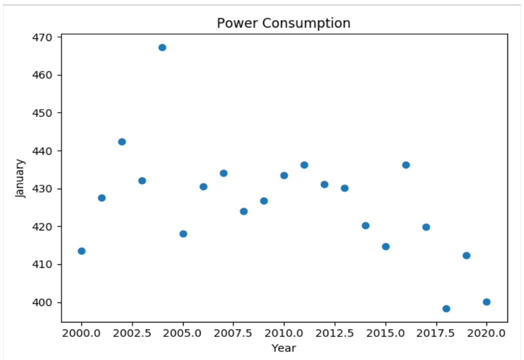
The parameters of the ARIMA model are defined as follows:

- **p**: Lag order; The number of lag observations included in the model.
- **d**: Degree of differencing; The number of times that the raw observations are differenced.
- **q**: moving average; The size of the moving average window.

Adopting an ARIMA model for a time series data assumes that the underlying process that generated the observations is an ARIMA process which seem obvious, but helps to motivate the need to confirm the assumptions of the model in the raw observations and in the residual errors of forecasts from the model.

Scatter plot of dataset

The very first step of predictive analysis is to create a scatter plot of given dataset. In this system, dataset is already taken from IoT devices and previously stored dataset is used for the prediction. Scatter plots are used to plot data points on a horizontal and a vertical axis to show how much one variable is affected by another. Each row in the data table is represented by a marker whose position depends on its values in the columns set on the X and Y axes. Scatter plots are important in predictive analysis because they can show the extent of correlation, if any, between the values of observed quantities or phenomena. Scatter plot of power consumption dataset is shown below.

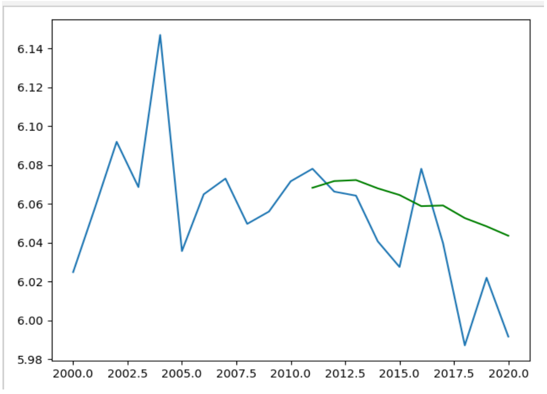


Identifying Trend

Trend is a component of a time series data that represents variations of low frequency, the high and medium frequency fluctuations that having been filtered out. It is a smooth, general, long-term,

Electronic copy available at: https://ssrn.com/abstract=3819512

average tendency which can be either increasing or decreasing. A trend is not always necessary that the increase or decrease is in the same direction throughout the given period of time for a dataset. Here the green line shows the trend of dataset. Seasonality and fluctuations may occur in a time series data and this can be identified by analyzing the trend of the data.



Stationarity

A stationary time series is a series with its statistical properties such as the mean, variance and autocorrelation are all constant over time whereas non-stationary series is a series with its statistical properties change over time. An algebraic equivalent linear function, which is not a constant one and the value of a linear function changes as $\boldsymbol{x}$ grows, but the way it changes remains constant — it has a constant slope and one value that captures that rate of change [4]. A trended series is called non-stationary and with unit root and on the other hand non-trended series is a stationary series characterized by without unit root. In formal way the series is called stationary if it satisfies three conditions;

i) Mean of Yt (E(Yt)) remain same over time or time invariant. i.e. E(Yt) = u, ∀ t

Where the symbol ∀, is use for all and (u) is any scalar

ii) Variance of Yt (V (Yt)) is time invariant. i.e. V(Yt) = σ^2, ∀ t

iii) Cov of Yt and Yt-s (cov (Yt, Yt-s) is time invariant, but can be depend upon the lag length. i.e. Cov (Yt, Yt-s) = γs

If the above conditions are not true for a series then the series is non-stationary. There is no link between previous values in a stationary series. While the regression on trended (non-stationary) variables are meaningless they provide misleading and spurious results. Normally most of the time series data gives us meaningless results until appropriate econometrics and statistical tools were not applied. An important question is how to know the data is stationary or non-stationary. There are many formal and informal tests as well to test the data for unit root.

Electronic copy available at: https://ssrn.com/abstract=3819512

Augmented dickey Fuller Test

Dickey Fuller test examines the null hypothesis of an autoregressive integrated moving average (ARIMA) against stationary and alternatively. Due to the error term unlikely to be white noise and to eliminate the problem of autocorrelation, simple Dickey Fuller test extended their test by including extra lagged in terms of the dependent variables. An Augmented Dickey–Fuller test (ADF) is an extended version of simple Dickey Fuller and is a formal testing method which tests the null hypothesis that a unit root is present in a time series sample to check whether a dataset is stationary or not. Normally we use Augmented Dickey-Fuller test instead of simple Dickey-Fuller test. By including the lagged values of dependent variable to the existing model, and continue this procedure up till where the autocorrelation eliminated. It can be illustrated as:

Yt = β1 + β2 Yt + εt……………………………………………………………………(1)

Yt = β1 + β2 Yt + β3 Yt-1 + εt………………………………….………………. ……… (2)

Yt = β1 + β2 Yt + β3 Yt-1 + β4 Yt-2 + εt………………………….…………… ……… (3)

Now,

ΔYt = γ Yt-1 + β1 ΔYt-1 + εt……………………………………………………. ……… (4)

ΔYt = γ Yt-1 + β1 ΔYt-1 + β2 ΔYt-2………… + βp ΔYt-p + εt………………...……….. (5)

Continue this process up till, where autocorrelation eliminated. This expression could be written as:

$\Delta Yt = \gamma Yt - 1 + \sum_{i=1}^{p} \beta 1 \ \Delta Yt - 1 + \varepsilon t$………………………………………………….. (A)

$\Delta Yt = \alpha + \gamma Yt - 1 + \sum_{i=1}^{p} \beta 1 \ \Delta Yt - 1 + \varepsilon t$…………………………………………... (B)

$\Delta Yt = \alpha + \beta t + \gamma Yt - 1 + \sum_{i=1}^{p} \beta 1 \ \Delta Yt - 1 + \varepsilon t$……………………………………...(C)

Some common assumptions of ordinary least square (OLS) are discussed here:

1. ε must be independent

2. There should be no heteroskedasticity, should homogeneity.

3. There should no structural break, co-efficient should stable.

4. Error term should be normally distributed.

Testing for stationarity in Augmented Dickey-Fuller test follows the same procedure as in simple Dickey-Fuller test. First, stationary is checked at level than at first difference and finally on second difference. At first difference the equation will be as follows:

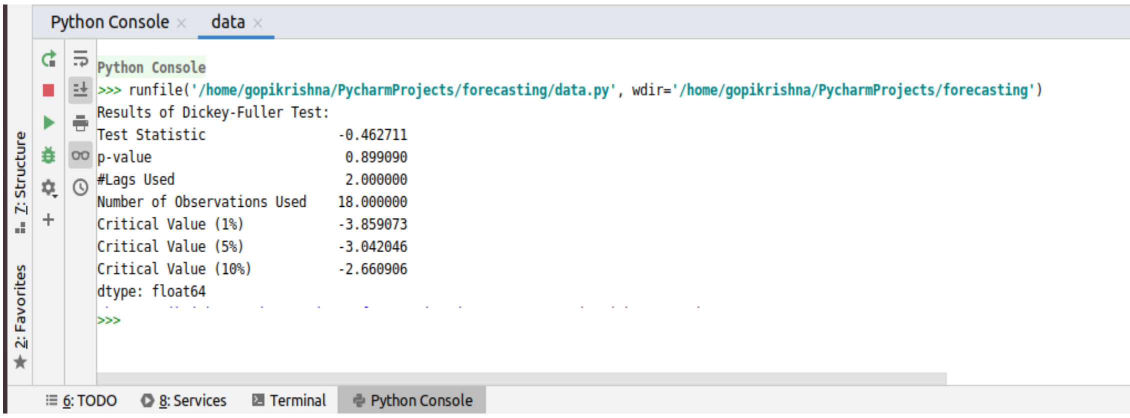
$\Delta 2Yt = \gamma Yt - 1 + \sum_{i=1}^{p} \beta 1 \ \Delta 2Yt - 1 + \varepsilon t$…………………………………………………...(a)

Electronic copy available at: https://ssrn.com/abstract=3819512

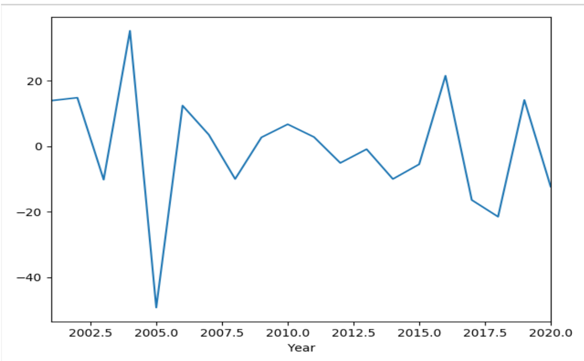
$$\Delta 2Yt = \alpha + \gamma Yt - 1 + \sum_{i=1}^{p} \beta 1 \ \Delta 2Yt - 1 + \varepsilon t \text{...(b)}$$

$$\Delta 2Yt = \alpha + \beta t + \gamma Yt - 1 + \sum_{i=1}^{p} \beta 1 \ \Delta 2Yt - 1 + \varepsilon t \text{..(c)}$$

If series is still not-stationary we can use same equations by replacing Δ^2 with Δ^3 rest process will be same. If the value of statistical value is less than critical value, then the dataset is stationary; otherwise not.



Here stationarity of the dataset is tested using ADF test and if statistics is greater than critical value then the dataset is stationary otherwise not. Mean and covariance is identified to make the dataset stationary.



Prediction using ARIMA model

Fitting ARIMA model for the stationary dataset will give accurate prediction results. The values of p, d and q are estimated by calculating ACF and PACF. ACF is an auto-correlation function which gives us values of auto-correlation of any series with its lagged values [5]. PACF is a partial auto-correlation function. Basically, instead of finding correlations of present with lags like ACF, it finds correlation of the residuals with the next lag value hence ‘partial’ and not ‘complete’ as we remove already found variations before we find the next correlation. The values of p and q are estimated by ACF and PACF while d is the number of times that differentiation takes place. Predicted values may

Electronic copy available at: https://ssrn.com/abstract=3819512

change according to the values of coefficients. The forecasting equation for an AR (1) model for a series Y with no orders of differencing is:

$$\hat{Y}_t = \mu + \phi_1 Y_{t-1}$$

If the AR (1) coefficient ϕ_1 in this equation is equal to 1, it is equivalent to predicting that the first difference of Y is constant--i.e., it is equivalent to the equation of the random walk model with growth:

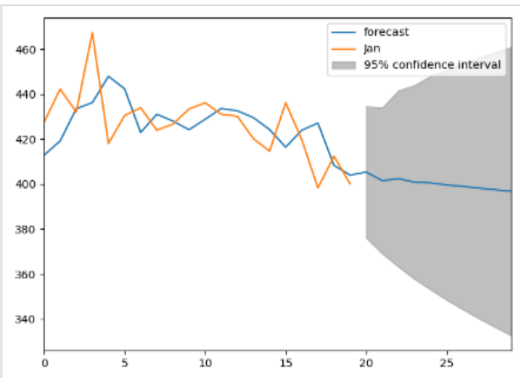
$$\hat{Y}_t = \mu + Y_{t-1}$$

The PACF of the UNITS series is telling us that, if we don't difference it, then we should fit an AR (1) model which will turn out to be equivalent to taking a first difference. In other words, it is telling us that UNITS really needs an order of differencing to be stationarized.

Power consumption dataset is evaluated and ARIMA model is applied on it. The accuracy of predicted data can be checked either using ACF and PACF or comparing the predicted values with the previous ones. If they are matched then the model is accurate otherwise not. Here predicted valued seems to be very close to the values in the dataset hence the model is accurate and gave better results.

Result

In this paper we evaluated power consumption of smart home and future power consumption of a particular smart home is predicted. Time series data is obtained from smart home and it is saved as dataset. Using this dataset, we predicted the future power consumption and the results shows that the predicted values are similar to the previous values and have confidence interval of 95%. The 95% confidence interval defines a range of values that you can be 95% certain contains the population mean. With large samples, you know that mean with much more precision than you do with a small sample, so the confidence interval is quite narrow when computed from a large sample.



Electronic copy available at: https://ssrn.com/abstract=3819512

Conclusion

Data science is one of the most developing areas of research and the system has a wide range of applicability in the fields of home automation, electrical energy management, *etc*. In this paper, power consumption analysis of a smart home is developed to provide an environment for recognizing the total power consumption of a smart home and to predict future consumptions which leads to reduce the wastage of electrical energy. Power consumption of a dataset is analysed and future consumption is predicted using ARIMA model, which is used to predict values in time series data. ARIMA gives the results which is similar to the previous values that shows the model gives accurate values. Prediction and analysis of power consumption of a system helps to know the total consumption of a smart home and the usage of electricity bill can be reduced. Also, ARIMA model gives much more accurate results when a huge dataset is used.

Future enhancement

As a future enhancement, understand the user behaviour and predicting the upcoming electricity bill of a smart home can be included in the system. Also, we can add tips to reduce the bill amount through an effective analysis. An intelligent chatbot can be used to control the devices in a home which makes the system more user friendly. Now the system focuses on simple hardware connections and analysed power consumption of a single smart home. Developing similar systems in a wider range to evaluate and analyze total power consumption of multiple houses, even in a city results in a very useful system for analysing the usage and wastage of electricity.

References

[1] JayavardhanaGubbi,RajkumarBuyya,SlavenMarusic,MarimuthuPalaniswami," ***Internet of Things (IoT): A vision, architectural elements, and future directions****", Sep-2013*

[2] Rajiv Ranjan, Omer Rana, Surya Nepal, Mazin Yousif, Philip James, Zhenyu Wen,StuartBarr, Paul Watson, Prem Prakash Jayaraman, Dimitrios Georgakopoulos, Massimo Villari, Maria Fazio, Saurabh Garg, Rajkumar Buyya, Lizhe Wang, Albert Y. Zomaya, Schahram Dustdar ***"The Next Grand Challenges Integrating the Internet of Things and Data Science"****, June 2018*

[3] G. Peter Zhang ***"Time series forecasting using a hybrid ARIMA and neural network model"****, 2001*

[4] Rizwan Mushtaq ***"Testing time series data for stationarity"***

[5] Fang-Mei Tseng, Hsiao-Cheng Yu, Gwo-Hsiung Tzeng ***"Combining neural network model with seasonal time series ARIMA model",*** *2002*

*[6] Kaibin Bao, Florian Allerding, Hartmut Schmeck****," User Behaviour Prediction for Energy Management in Smart"****, 2011*

[7] Gulnar Mehdi, Mikhal Roshchin ***" Electricity consumption constraints for smart-home automation: An overview of models and applications"****, 2015*

Electronic copy available at: https://ssrn.com/abstract=3819512

[8] https://www.wespeakiot.com/how-smart-homes-help-saving-energy/

[9]https://www.forbes.com/sites/louiscolumbus/2016/09/21/strong-analytics-skills-3x-more-likely-to-lead-to-success-with-internet-of-things-iot/

[10] https://www.skyhook.com/blog/iot/internet-of-things-statistics/

Electronic copy available at: https://ssrn.com/abstract=3819512