International Journal of Engineering & Technology, 7 (3.34) (2018) 931-934

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

IoT Data based Predictive Modeling for Energy Usage of Appliances in Smart Home

[1]Mrs. Archana Shirke, [2]Dr. M.M. Chandane

[1,2]Veermata Jijabai Technological Institute, Mumbai
**Corresponding Author E-mail: [1]archanashirke25@gmail.com , [2]mmchandane@vjti.org.in*

Abstract

Internet of things (IoT) has emerged as the new trends in the wireless technology in last few years. This new area has greatly influenced the working of humans. IoT has applications in every domains of human being. The growth rate of IoT devices is exponential due to its wide applicability. Therefore, the data generated by these devices is huge and contains high variability. Such a huge amount of data needs to be modeled precisely. The effectiveness of the IoT applications lies in the preciseness of the data represented by the models. Predictive analysis helps business analysts to build models to predict trends, make tradeoff decisions, and model the real world for decision making support system. This paper presents the study on various models used for IoT data analytics. Various predictive models such as Multiple Linear Regression (LR), Support Vector Machine for regression (SVR), Random Forest (RF), Gradient Boosting Machine (GBM) and extreme Gradient Boosting Machine (XGBM) are applied on the sensor data collected from Smart Home. The comparative results produced by these models have been analyzed with reference to energy consumption and prediction. The implementation of the models is carried out on R language. The results show that XGBM model perform better based on RMSE, R-squared and MAE for given data set. It has less RMSE and high R-squared which indicates it has captured high variability in the data.

Keywords: *Predictive Modeling, IoT, Regression, GBM, XGBM*

1. Introduction

The popularity of IoT is increased rapidly in last few years because it has created opportunities for research, innovation and business transformations. Gartner predicts that, by 2020, more than half of major new business processes and systems will incorporate some element of the IoT. The data generated by diverse IoT devices has opened new opportunities for data analytics. IoT applications need optimized, real-time analytics techniques to infer events for automatic decision making.

Smart Sensors, machine-to-machine, and network data plays major role in the success of IoT. The Analytics of IoT data is possible with increasedprocessing power and platform with high speed.IoT data analytics includes revealing patterns, unseen events and correlations which helps understanding business quickly [1] [2] [3]. One of the most important questions that arise now days is that how to convert the data generated or captured by IoT into knowledge to provide a more convenient environment. It enables researchers to analyze large volume of unstructured and stream data precisely [4] [5].

With the evolution of wireless sensor network and communication technology,it becomes feasible for IoT environment to integrate sensor devicesto communicate seamlessly for information sharing from different platforms conveniently. IoT analytics make the things smarter and intelligent. Intelligent thing evolves as a collaborative model and communicate with one another to accomplish tasks. Existing things such as in the home, office, factory floor, and medical facility becomes intelligent with the use of IoT devices and delivers the power of AI.IoT is also widely adopted in smart cities by developing intelligent systems, such as smart office, smart transportation, smart healthcare, and smart energy. In smart city, smart devices including mobile devices, public transport and daily home appliances are used for data collection.

Many real life things like home appliances, emergency alarm and vending machines can be connected through wireless sensor network and controlled remotely. Data acquisition devices sense and transmit data using WSN solutions, such as Bluetooth, RFID, WiFi and GSM to systems for further analysis in order to improve standard of living [6][7].

IoT network can be divided into three components: Sensors Network, Communication Network and Analysis Network. The recent development of different sensors and communication technologies put IoT in front line.IoT Analytics network which does Data Analytics is also important in the fame of IoT systems because they involved in the decision-making process to bring transformation in business and individual. IoT analytics use different algorithms, techniques and tools as any other analytics that can run in the cloud.IoT analytics has input from variety of sources like streams of sensor data, environmental data sensors and other entities. Accurate and timely processing this data requires cutting edge algorithms and processing platforms. It also involves attention to security, communication, data storage, application integration, governance and other considerations beyond analytics.

After introduction, the paper describes different types of data analytics algorithms along with evaluation measures. The next section shows exploratory data analysis experiment results for model selection. The paper ends with a comparison of models and conclusion.

2. Types of Data Analytics ALGORITHMS

Data analytics are divided into different types based on the requirements of applications. These types are presented in this subsection under following categories.

1. Historical Analytics is based on the lot of historical data analysis and takes time. It is used when quick response is not expected. We can useoff-line batch processing architecture to improve efficiency. Historical analytics can be of type Descriptive and/or Diagnostic used for different types of reporting [1] [8].
2. Predictive analytics includes various techniques from predictive modelling, machine learning, and data mining which analyses historical and present data to make predictions about unknown patterns. It is generally performed on sensor data. In sensor based system, data change frequently, and fast data analytics techniques are required to obtain result within a short period [1][8].

Real-time Analyticsprovide lighter-weight analytics very quickly. There are two types of real time analytics i.e. Predictive and Prescriptive Analytics. Predictive Analytics is used to make predictions about unknown future events where as Prescriptive analytics is used to find the best course of action for a given situation [1] [2] [8].

Stream IoT Analytics deals with batches or streams of IoT data with none or moderately high time constraints (e.g. Environment data Analytics) [1][2][8]

Different predictive analytics Algorithms are found in the literature and are broadly categorized based on designer's perspective of the algorithm and objective function used.

1. Multiple Linear Regression(LR) [13]–Multiple Linear Regression is a very simple approach for predicting a quantitative response Y on the basis of multiple predictor variable Xp.In general, suppose that we have p distinct predictors. Then the multiple linear regression model takes the form

$$Y = \beta 0 + \beta 1X1 + \beta 2X2 + \cdots + \beta pXp + \varepsilon$$

2. Random Forest(RF) [21]- Random forest is an ensemble learner used for classification, regression and other tasks, which build by constructing multiple of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

3. Support Vector Machine(SVR) [22]–It is a supervised learning model that analyzes data for classification and regression. In SVM model, each example is considered as points in space so that the examples of the separate categories are divided by a clear gap that is as wide as possible. It not only perform linear classification but also efficiently perform a non-linear classification usingkernel function, implicitly mapping their inputs into high-dimensional feature spaces.

4. Gradient Boosting Machine(GBM)[14] [15] [23]is a collection of weak prediction models which produces an output.This boosting model is built in steps and it generalizes them by allowing optimization of an arbitrary differentiableloss function.

5. Extreme Gradient Boosting Machine(XGBM) [24] -- XGBM is widely used by data scientists to achieve state-of-the-art results on many data mining challenges. It is proposed to build a novel sparsity-aware algorithm for sparse data and weighted Quantile sketch for approximate tree learning. It provides insights on cache access patterns, data compression and sharping to build a scalable tree boosting.

3. Implementation Details

3.1 Experimental Setup

An experiment was designed and implemented for studying and analyzing statistical models using R programming tool. It consists of the process of creating the design and collecting the data, shows how to perform the proper analysis of the data, and illustrates the interpretation of results. Different R packages including "caret" has been installed to complete the experiment. It is one of the most powerful packages for machine learning algorithms in R.

3.2 Case Study

Appliances energy prediction Data Set from UCI library is used for experiment [16]. The data set consist of data collected over period of 4.5 moths for every 10 minutes. The data collected from temperature and humidity sensor were monitored with a WSN. Each sensor node is configured to read and transmit temperature and humidity reading about every 3.3 minutes averaged over 10 minutes' duration. The energy data was logged every 10 minutes with m-bus energy meters. The nearest weather station data was also integrated from public data set from Reliable Prognosis and merged together with the experimental data sets using the date and time column.

3.3 Parameters Used for Analysis

Different parameters are used evaluation of models. ForThe precision of the model is very important from evaluation point of view. Different Evaluation Measuresare used to compare different algorithms [12].

1. RMSE(Root Mean Square Error)

$$RMSE = \sqrt{\frac{\sum_{n}^{i=1}(Y_i - \hat{Y}_i)^2}{n}}$$

2. R^2 (R squared)

$$R^2 = 1 - \frac{\sum_{n}^{i=1}(Y_i - \hat{Y}_i)^2}{\sum_{n}^{i=1}(Y_i - \bar{Y})^2}$$

3. MAP (Mean Absolut Error)

$$MAE = \frac{\sum_{n}^{i=1}|Y_i - \hat{Y}_i|}{n}$$

4. MAPE(Mean Absolute percentage error)

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i - \hat{Y}_i|}{Y_i}$$

Where Yi is the actual measurement (energy consumption), $\hat{Y}_i$is the predicted value and n is the number of measurements.

4. Performance Analysis

The experimental results of various models are listed below with their inferences.

4.1 Multiple Linear Regressions

Figure 1 shows residual plot for the linear regression model. It is obvious that the relationship between the variables and energy consumption of appliances is not well represented by linear model since the residuals are not normally distributed around the horizontal axis. The result analysis shows that this is not a good model for modelling of data

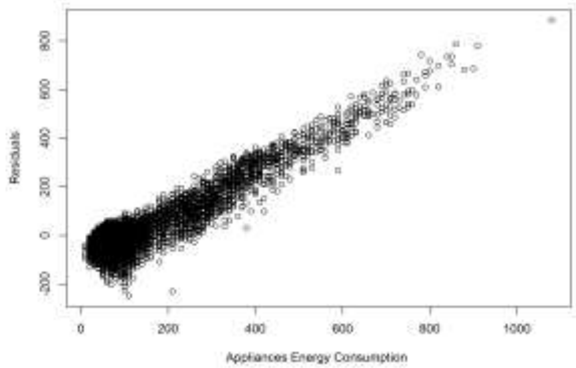

Figure 1: Linear Regression Model

4.2 Support Vector Regression

Figure 2 shows SVM with radial kernel used for regression (also called as SVR) is used. The SVR model requires two tuning parameters sigma and cost variables besides predictors. The result analysis shows that this is a good model for modelling of data.

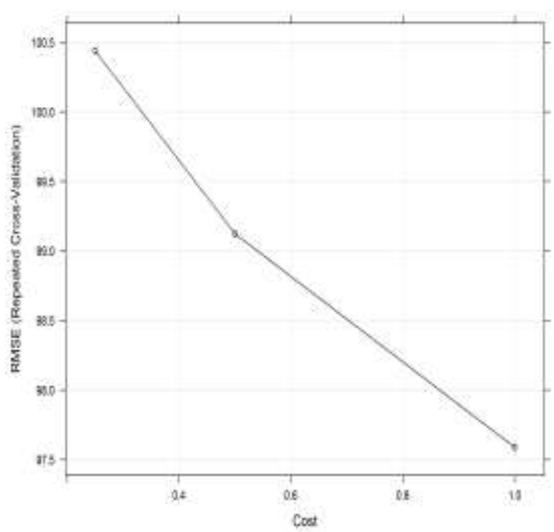

Figure 2: Support Vector Regression Model

4.3 Random Forest Regression

Figure 3 shows random forest model which is a Tree-based model used output of several regression trees. However, each tree is built with a random sample of selected predictors. The result analysis shows that this is not a good model for modelling of data.

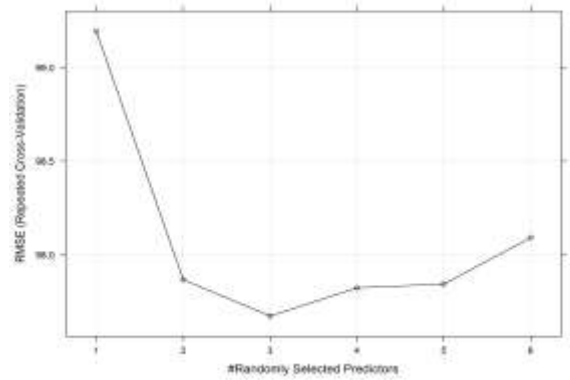

Figure 3: Random Forest Regression Model

4.4 Gradient Boosting Machine

Figure 4 shows GBM model which is based on Boosting try to improve the prediction by using information from the first trees and also require the selection of optimal parameters for the number of trees and maximum tree depth. The result analysis shows that this is a good model for modelling of data.

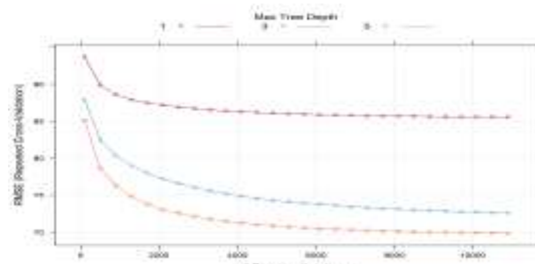

Figure 4: Gradient Boosting Machine Model

4.5Extreme Gradient Boosting Machine

Figure 5 shows extreme GBM model based on boosting which is very innovative and recent algorithm of prediction. It is found to be best for the given data set but training time is very high. The result analysis shows that this is a very good model for modelling of data.

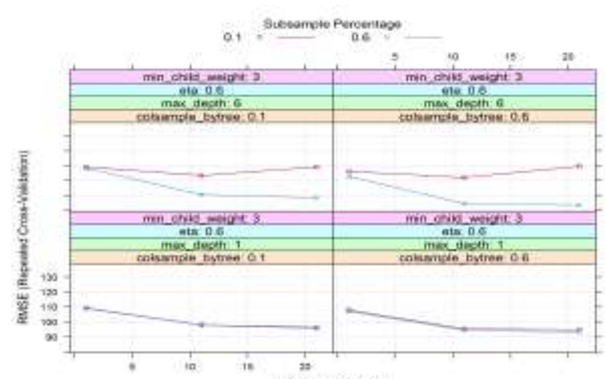

Figure 5: Extreme Gradient Boosting Machine Model

4.6 Comparison of All Models

The comparison of all the models test error is listed in the figure 6. The best model is selected based on lessRMSE and high R-squared value. It is found that XGBM has performed better than other models.

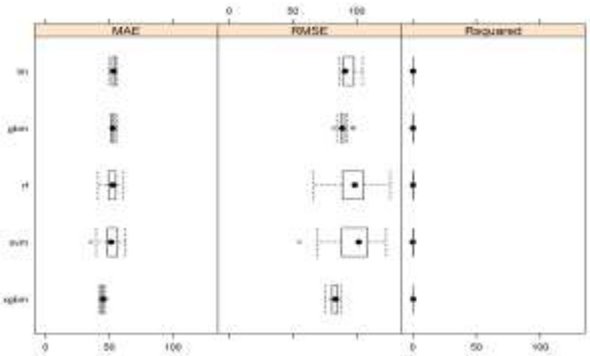

Figure 6: Comparison of all models error

5. Conclusion

The various predictive models (LR, SVR, RF, GBM and XGBM) have been implemented using R language. The comparative study of these models is carried out on energy prediction Data Set from UCI library with the goal to identify models that produces better estimates of energy prediction based on different parameters. The parameters such as RMSE, MAE, R-squared and MAPE are used for comparative statistics. The result shows curious relationship between variables.It has less RMSE and high R-squared value which indicates it has captured high variability in the data.

The study shows that models developed are not perfect and needs optimization for better prediction. The future work includes considering hybrid model which could be useful to improve the prediction and find relationship with different features. The existing IoT data analytics solutions are still in the early stage of development and hybrid data modelling may provide accurate insight which is swiftly becoming the mainstream.

References

[1] Mohsen Marjani, Fariza Nasaruddin2, Abdullah Gani1, (Senior Member, IEEE), Ahmad Karim, Ibrahim AbakerTargioHashem, Aisha Siddiqa, AndIbrarYaqoob "Big IoT Data Analytics: Architecture, Opportunities, and OpenResearch Challenges", IEEE Access, 29 March 2017, page 5247 - 5261.

[2] Chun-Wei Tsai, Chin-Feng Lai, Ming-Chao Chiang, and Laurence T. Yang "Data Mining for Internet of Things: A Survey" IEEE Communications Surveys & Tutorials, Vol. 16, No. 1, First Quarter 2014, page 77-97.

[3] HongmingCAI,BoyiXu, Lihong Jiang, and Athanasios V. Vasilakos, "IoT-Based Big Data Storage Systems in Cloud Computing: Perspectives and Challenges" IEEE Internet of Things Journal, Vol. 4, No. 1, February 2017 Page 75-87.

[4] Shanzhi Chen, HuiXu, Dake Liu, Bo Hu, and Hucheng Wang "A Vision of IoT: Applications, Challenges, and Opportunities with China Perspective", IEEE Internet of Things Journal, Vol. 1, No. 4, August 2014 Page 349-359.

[5] Ala Al-Fuqaha, Mohsen Guizani, Mehdi Mohammadi, Mohammed Aledhari, and MoussaAyyash, "Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications" IEEE Communication Surveys & Tutorials, Vol. 17, No.4, Fourth Quarter 2015,Page 2347-2376.

[6] Huadong Ma, Liang Liu, Anfu Zhou, and Dong Zhao "On Networking of Internet of Things: Explorations and Challenges" IEEE Internet of Things Journal, Vol. 3, No. 4, August 2016, Page 441-452.

[7] ShikharVerma, Yuichi Kawamoto, Zubair Md. Fadlullah, Hiroki Nishiyama, and Nei Kato,"A Survey on Network Methodologies for Real-Time Analytics of Massive IoT Data andOpen Research Issues", IEEE Communications Surveys & Tutorials, Vol: 19, Issue: 3, April 2017, Page 1457-1477.

[8] A. Akbar, A. khan, F. Carrez, and K. Moessner, "Predictive analytics for complex IoT data streams," IEEE Internet of Things Journal, vol. PP, no. 99, pp. 1–1, 2017.

[9] A. Akbar, F. Carrez, K. Moessner, and A. Zoha, "Predicting complex events for pro-active IoT applications," in Internet of Things (WF-IoT), 2015 IEEE 2nd World Forum on, Dec 2015, pp. 327–332.

[10] Paula Ta-Shma, A. Akbar, Guy Gerson-Golan, Guy Hadash, F. Carrez and K. Moessner, "An Ingestion and Analytics Architecture for IoT applied to smart City Use Cases," IEEE Internet of Things Journal, vol. PP, no. 99, pp. 1–1, 2017.

[11] Luis Candanedo, V. Feldheim, D. Deramaix, "Data driven prediction models of energy use of appliances in a low energy house" Energy and Buildings, Volume 140, 1 april 2017, pp. 327-341.

[12] Kecskemeti, G and Casale, G and Jha, DN and Lyon, J and Ranjan, R (2017), "Modelling and Simulation Challenges in Internet of Things", IEEE Cloud Computing, 4 (1). pp. 62-69. ISSN 2325-6095.

[13] Friedman J, Hastie T, Tibshirani R, et al. (2000). "Additive logistic regression: a statistical view of boosting," the annals of statistics, 28(2), 337–407.

[14] Friedman JH (2001). "Greedy function approximation: a gradient boosting machine." Annals of Statistics, pp. 1189–1232.

[15] Luis Candanedo,Appliances energy prediction Data Set, Online: http://archive.ics.uci.edu/ml/datasets/Appliances+energy+prediction,April 2017.

[16] Chris Raphael, "Why Edge Computing Is Crucial for the IoT",Online:https://www.rtinsights.com/why-edge-computing-and-analytics-is-crucial-for-the-iot/, 12 November 2015.

[17] Jennifer Sewell, 5 Myths of IoT analytics, Online:http://mnubo.com/5-myths-iot-analytics, 26 September 2016.

[18] The 4 stages of an IoT architecture, Online: https://techbeacon.com/4-stages-iot-architecture, Jan 2017.

[19] Jonathan Holdowsky, Monika Mahto, Michael E. Raynor, Mark Cotteleer, "Inside the Internet of Things",Online:https://dupress.deloitte.com/dup-us-en/focus/internet-of-things/iot-primer-iot-technologies-applications.html, 21 August, 2015.

[20] B. Dong, C. Cao, S.E. Lee, Applying support vector machines to predict buildingenergy consumption in tropical region, Energy Build. 37 (5) (2005) 545–553.

[21] G. James, D. Witten, T. Hastie, R. Tibshirani, an Introduction to Statistical Learning, Springer, 2013.

[22] Gradient Boosting Machine, M. Kuhn, caret: Classification and Regression Training, 2015.

[23] Tianqi Chen, Tong He, "boost: eXtreme Gradient Boosting", Package Version: 0.6-4, 4 January, 2017.

[24] Edwin de Jonge, Mark van der Loo, "An introduction to data cleaning with R", statistics Netherland, Discussion paper, 2013.

[25] Nerurkar, P., Shirke A., Chandane M., Bhirud S., "Empirical Analysis of Data Clustering Algorithms" in Procedia Computer Science 2018 Vol. 125 Pg. 770-779

[26] Nerurkar, P., Shirke A., Chandane M., Bhirud S., "ANovel Heuristic for Evolutionary Clustering" in Procedia Computer Science 2018 Vol. 125 Pg. 780-789

[27] G. Indrajith and K.Vijayakumar, "Automatic Mathematical and Chronological Prediction in Smartphone Keyboard" International Journal of Engineering and Computer Science ISSN: 2319-7242Volume 5 Issue 5 May 2016, Page No. 16714-16718.

[28] K. Vijayakumar and C. Arun, "Continuous Security Assessment of Applications in Cloud Environment", International Journal of Control Theory and Applications, ISSN: 0974-5645 volume No. 9(36), Sep 2016, Page No. 533-541.