Energy Anomaly Detection and Modelling on Smart Premises using SDAR

Sachin Gupta[#1], Bhoomi Gupta[*2]

[#]*School of Engineering and Technology, MVN University*
Palwal, Haryana, India

[1]sachin.gupta@mvn.edu.in

[*]*Department of IT, Maharaja Agrasen Institute of Technology, GGSIP University*
Delhi, India

[2]bhoomigupta@mait.ac.in

***Abstract*— Energy as a commodity is facing a globally prevalent shortage with the ever increasing gap between the demand and supply. The strain on the non-renewable conventional energy resources is evident from the multitude of industries in the manufacturing hubs of China that have come to standstill in the face of non-availability of coal. The phenomenon is particularly acute in all developing countries with an unbalanced approach to the energy management vis-a-vis distribution losses. The Internet of Things (IoT) ecosystem enabled smart homes and industrial premises have shown promising application towards achieving energy efficiency via creation of dynamically adjusting demand based systems. It is possible to predict the energy consumption requirements based on application of Machine Learning (ML) based models on statistical data obtained from energy sensors. The data streams from IoT devices however, more often than not throw up surprises in the form of outliers and changes which can affect the Machine learning based time series forecasting. The problem is more accentuated in the case of non-stationary time series sources where it is imperative to ascertain whether an anomaly is momentarily affecting the time series as an outlier or it is a permanent change never returning to the original trend. This paper uses a sequentially discounting auto regression (SDAR) learning algorithm to detect and classify the anomalies in energy consumption usage for model accuracy. Specifically, we have applied the online SDAR algorithm on energy consumption IoT dataset from kaggle as a demonstration to distinguish between outliers and permanent changes over the time series which can be used for interpretability and increasing model accuracy while prediction of energy consumption. We were able to forecast sudden changes in the energy consumption requirements well in advance, based on the previous years' usage patterns as the results indicate.**

***Keywords*— IoT, Machine Learning, Anomaly Detection, Time Series, SDAR, Energy efficiency.**

I. Introduction

The global energy economy is plotting remarkably different charts along the renewables and the non renewable energy resources post the recession induced by Covid during the 2019-2020 period. While the renewables led by solar energy are making great strides making the energy cheaper, coal and natural gas on the other hand are rallying to steep price rise amidst global shortages. With a precariously biased ratio between sustainability goals with limiting energy consumption vs ever rising global emissions, heavily skewed towards the latter, energy efficiency is today being considered equally important to growth in production capacity. Considering 2019 as a benchmark for the energy requirements for a perspective, the year 2020 witnessed a 4% reduction due to covid, but the rebound is happening at a 4.6% increase in 2021 across the world [1]. The unfortunate statistic about coal demand being projected for a 60% rise leading to a minimum 5% rise in global emissions is alarming.

Energy efficiency is at the heart of all sustainability goals and while it is difficult to ascertain how far will the nations adhere to their sustainability goals commitment, the smart management of energy on industrial premises and in homes using IoT is a promising research and development thrust area with realistic projections. The IoT technology has made it possible for continual consumption data sensing, processing and sharing for the network integrated devices [2]. There are however several problems associated with non-stationary time series data reported through literature, as discussed in the next section. The large-scale data generation through multiple sensors is subject to processing using big data analytics and several representative cases of the big data analytics using ML have been considered in [3][4] and [5], while performance modelling of a large scale system has been discussed in [6].

The remainder of this paper is organized as follows: the related research and development work in the areas of IoT based energy management along with anomaly detection techniques has been discussed in Section II. The methodology of this study including training and testing data used for modelling has been introduced in Section III, with the detailed result description and discussion in Section IV. The future research directions have been presented in Section V with the conclusions of this study.

II. Related Work And Data Description

IoT has remained a prime research area during the last decade owing largely to its wide application scope with parallel strides happening in networking, security and cryptocurrencies. There are comprehensive surveys available on IoT applicability in diverse technological fields across

978-1-6654-3542-0/21/$31.00 ©2021 IEEE

industry as available in [7-9]. There has been a growing awareness about energy efficiency being researched both in the perspective of conservation as well as renewables to save the globe from toxicity associated with fossil fuel based energy resources [10][11]. The potential energy management systems based on IoT have been proposed from the perspective of high level grid based solutions [12-14] down to residential buildings and office premises at the individual building level, which is primarily computed on consumption by hot water, ventilation, and air conditioning (HVAC) systems [15]. The scope of IoT in smart energy management is wide, but this study considers only the on-premise smart home / office systems using IoT smart meters to record the time series data of consumption across multiple devices during different seasons of the year.

Time series data is known to have demonstrated several kinds of anomalies, and in this paper we have applied the SDAR algorithm to classify and analyse anomalies present in IoT sensor based dataset. Any time series may be categorized as a stationary time series if the basic statistical parameters like mean and variance along with high order moments and autocorrelation functions do not change over time. If however, even a single parameter from the above mentioned is violated, the series is called a non-stationary time series[17]. System behaviors being modelled through any time series data sequence is subject to gradual or abrupt changes over time due to intrinsic or extrinsic factors [18,19], and the change points can be tracked with the help of change point detection algorithms [20]. We have used the changepoint detection method based on [21] in this study.

The dataset used for the purpose of demonstration has been taken from Kaggle [16] and the details presented below include both the device information and the prevalent weather conditions when the sensor data was captured. Table 1 shows the description of the attributes of interest related to energy usage and generation along with a brief description.

TABLE I
ENERGY USAGE / GENERATION ATTRIBUTES

S.No	Attribute	
	Energy Usage (kW)	**Description**
1	Use	Total Consumption
2	Gen	Total generated by non renewable resources
3	Premise Overall	Total consumption by building
4	Dishwasher	Energy consumed by device
5	Furnace 1	Energy consumed by device
6	Furnace 2	Energy consumed by device
7	Home Office	Energy consumed by device
8	Refrigerator	Energy consumed by device
9	Wine Cellar	Energy consumed by device
10	Garage Door	Energy consumed by device
11	Kitchen 12	Energy consumed by device
12	Kitchen 14	Energy consumed by device
13	Kitchen 38	Energy consumed by device
14	Barn	Energy consumed by device
15	Well	Energy consumed by device
16	Microwave	Energy consumed by device
17	Living Room	Total consumption by living room
18	Solar	Generation by solar

The duration of data capture is close to a full year spread across one minute readings making it a perfect choice for time series assessment. The dataset has 32 fields each for 503910 rows indexed by timestamps and can be fairly considered big data at a size of 130 MB.

The weather information has also been captured by the sensors contributing to the dataset and the details have been summarized in Table 2 with attribute information.

TABLE 2
WEATHER ATTRIBUTES

S.No	Attribute	
	Weather Parameter	**Description**
1	Temperature	Measured in degrees centigrade
2	Humidity	Atmospheric Vapour
3	Visibility	Distance by 5% reduction in Lux
4	Apparent Temperature	Perceived outdoor temp in presence of wind, humidity
5	Pressure	Air Pressure
6	Windspeed	Speed of wind flowing
7	CloudCover	Okta measure of cloudiness
9	DewPoint	Varies on pressure and humidity
10	PrecipProbability	Probability of rain
11	PrecipIntensity	Intensity of Rain

A. Exploratory Time Series Analysis

The dataset description mentions that the sensor readings from smart meters have been recorded one minute apart along with weather related data. To assess the readiness of

the dataset for a time series evaluation, a basic plot of temperature values recorded is plotted and shown in Figure 1.

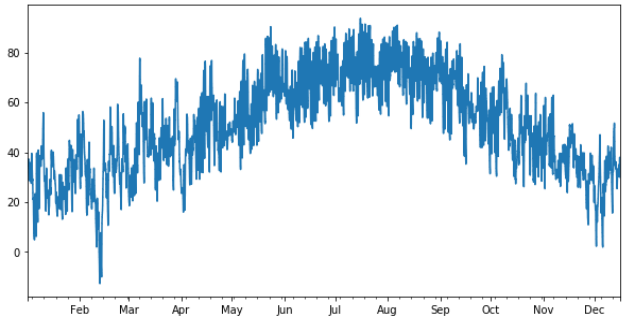

Fig. 1 Exploratory Time Series Analysis for temperature attribute (Temperature on x axis and Time on y axis).

The energy consumption trends were also plotted as a time series over daily averages to aggregate the data points and are shown in Figure 2. It is interesting to notice the usage spike in the middle of the year around July to September, as visible from the trend analysis of usage and the gradual upsurge from January to July and a decline thereof.

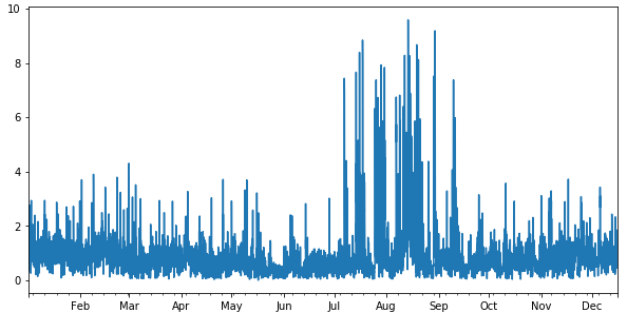

Fig. 2 Exploratory Time Series Analysis for energy usage (Energy consumption on x axis and Time on y axis)

III. Methodology

Analysing changes happening in a long running time series data has traditionally been either offline or online, with the basic difference being that offline algorithms consider the complete time series data for the assessment of the change point in the series and usually processes a batch mode operation. The online change detection methods however consider a dynamically changing subset of the complete dataset, drawing in from the **"n"** recent most data points and forgetting the prior data in order to keep an *almost* concurrent information about the changes *as they happen* [22].

If we take a large number of data points before assessing change, we are moving towards the offline model, and if the data points being considered moves close to 1, the model is truly concurrent but the precision and utility becomes lesser. Intuitively, it can be deduced that the offline algorithms work better in most stationary time series assessments, but to gain meaningful insights from non-stationary time series datasets, the online change detection algorithms shall produce better results. The sequentially discounting auto regression algorithm, SDAR, which has been used in the underlying ChangeFinder [21] algorithm works on the principle of online change detection, and works on the below mentioned machine learning methodology.

The SDAR algorithm uses an optimally chosen discount parameter for progressively forgetting past data records, to ensure good results despite using non-stationary data in time series. The ChangeFinder algorithm first trains the time series prediction model on each point using the SDAR approach and then using this model, predicts the likelihood of the next point plot. It then calculates a log based loss function score for **outlier calculation** as per equation 1.

$$Score\,(x_t) \;=\; -\,log\,P_{t-1}\,(x_t \mid x_1, x_2, \ldots, x_{t-1}) \quad \ldots\ldots (1)$$

The outlier score is then processed using a smoothing function as shown in equation 2, with a variable window to ascertain whether the change is just transient or has been there for a long time to classify it as an outlier or a change point.

$$ScoreSmoothed_{(x_t)} = \frac{1}{W} \sum_{t=t-W+1}^{t} Score(x_t) \qquad \ldots\ldots (2)$$

The model is again trained using the score obtained by SDAR algorithm and the logarithmic loss is recalculated to finally compute the **change score.** The changefinder library available in Python has been used in this study, and the hyperparameter tuning for the above models can be achieved through adjusting the value or r between 0 to 1, with smaller values affecting greater influence of past data points, and the window smoothing parameter which allows assessment of essential changes for small values.

We have used manual hyperparameter tuning by hit and trial for a quick assessment of the outlier scores and change scores. The details have been presented in the results section.

IV. Results

The Auto-regression model used allows for a specific order of how far the past points may be included in the model. For the purpose of standardization, we have kept the order parameter fixed, varying the r values between 0.01 and 0.02, in combination with the value of smooth parameter at 3 and 8. The values are chosen because out of these ranges, the time series dataset stopped giving any change/outlier values or the threshold went out of focus.

The resultant graphs from all 4 combinations of the above two variables are presented in Figure 3 to 6 below, corresponding to the cases of a) r = 0.01, smooth =3; b); r

=0.01 smooth =8; c) r = 0.02 smooth = 3 and d) r =0.02 and smooth =8.

It can be observed that increasing the value of r increases the threshold for a value to be termed an outlier thereby missing several values as can be seen from figure 3, where r has been chosen as 0.02 and plotted against a smooth value of 8. The change detection points are missed across major data point clusters. The same may be compared with figure 4 for reference with a smaller value of r at 0.001 with higher resultant change scores and thus more change detection points.

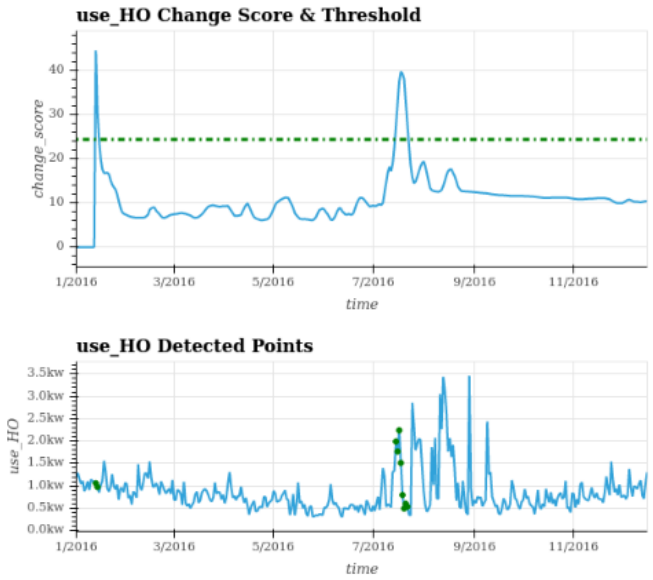

Fig. 3 Overall Energy Use Time series plots with r=0.02 , smooth = 8

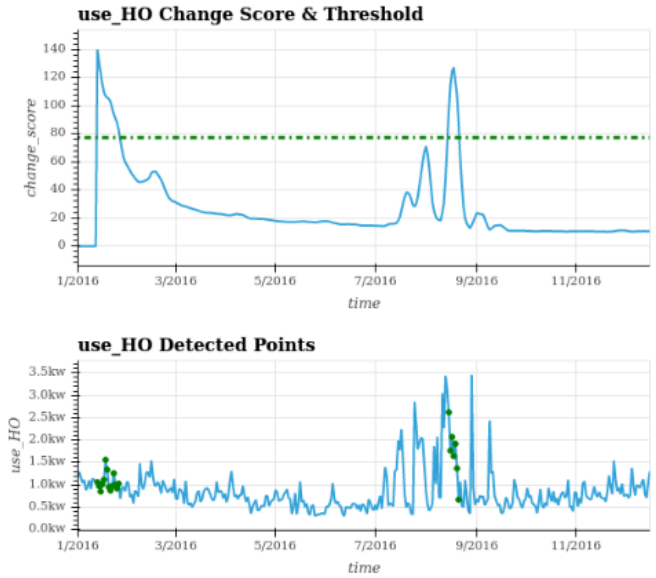

Fig. 4 Overall Energy Use Time series plots with r=0.001 , smooth = 8

In the other case of varying the smooth parameter, it was observed that a lower smooth trained the model better with lower threshold of outliers and identification of many more number of change points for further studies. The same has been shown in figure 5, with r =0.001 and smooth = 3 presented below:

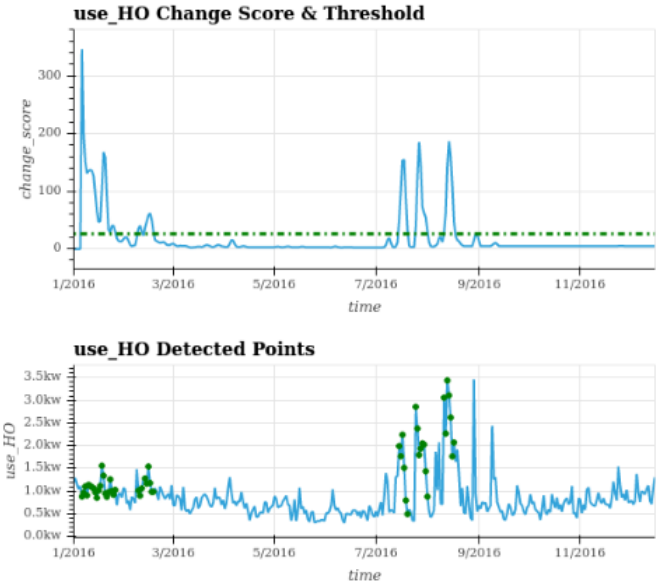

Fig. 5 Overall Energy Use Time series plots with r=0.001 , smooth = 3

V. Conclusions

The results allowed us to observe that the anomaly in usage tendency can be assessed from the past usage patterns. The results also indicate that all change points happening between July, with a rapid increase in consumption are captured along with the rapid decrease happening in September as well, visible on the plots as green dots. These points may be considered as anomaly points, and an extensive domain study may help to forecast a sudden spurt in the energy consumption requirements well in advance, based on the previous years' usage patterns. Provisions for surplus energy in the peak requirement months can be planned across the entire city accordingly and a reduced supply in the short use months may allow for a judicious energy efficient plan of action. The study has not factored in the external environmental data, which could yield useful insights with future research in the domain.

References

[1] IEA (2021), World Energy Outlook 2021, IEA, Paris https://www.iea.org/reports/world-energy-outlook-2021

[2] Affum, E. A., Agyekum, K. A. P., Gyampomah, C. A., Ntiamoah-Sarpong, K., & Gadze, J. D. (2021). Smart Home Energy Management System based on the Internet of Things (IoT). International Journal of Advanced Computer Science and Applications, 12(2). https://doi.org/10.14569/IJACSA.2021.0120290

[3] Kalra S., Gupta S., Prasad J.S. (2020) Predicting Trends of Stock Market Using SVM: A Big Data Analytics Approach. In: Batra U., Roy N., Panda B. (eds) Data Science and Analytics. REDSET 2019. Communications in Computer and Information Science, vol 1229. Springer, Singapore.https://doi.org/10.1007/978-981-15-5827-6_4

[4] Sneh Kalra and Dr. Sachin Gupta, "Performance Evaluation of Machine Learning Classifiers for Stock Market Prediction in Big Data Environment" Published - Journal of Mechanics of Continua and Mathematical Sciences - Web of Science - for Vol – 14, No – 5,

September-October,2019,http://doi.org/10.26782/jmcms.2019.10.00022

[5] V. Sachdeva and S. Gupta, "Basic NOSQL Injection Analysis And Detection On MongoDB," 2018 International Conference on Advanced Computation and Telecommunication (ICACAT), Bhopal, India, 2018, pp. 1-5, doi: 10.1109/ICACAT.2018.8933707

[6] Sachin Gupta & Bhoomi Gupta (2019) "Performance modeling and evaluation of transportation systems using analytical recursive decomposition algorithm for cyclone mitigation", Journal of Information and Optimization Sciences, 40:5, 1131-1141,DOI:10.1080/02522667.2019. 1638003

[7] Da Xu, L.; He, W.; Li, S. Internet of Things in Industries: A Survey. IEEE Trans. Ind. Inform. 2014, 10, 2233–2243.

[8] Talari, S.; Shafie-Khah, M.; Siano, P.; Loia, V.; Tommasetti, A.; Catalão, J. A review of smart cities based on the internet of things concept. Energies 2017, 10, 421.

[9] Ibarra-Esquer, J.; González-Navarro, F.; Flores-Rios, B.; Burtseva, L.; Astorga-Vargas, M. Tracking the evolution of the internet of things concept across different application domains. Sensors 2017, 17, 1379.

[10] Connolly, D.; Lund, H.; Mathiesen, B. Smart Energy Europe: The technical and economic impact of one potential 100% renewable energy scenario for the European Union. Renew. Sustain. Energy Rev. 2016, 60, 1634–1653.

[11] 13. Grubler, A.; Wilson, C.; Bento, N.; Boza-Kiss, B.; Krey, V.; McCollum, D.L.; Rao, N.D.; Riahi, K.; Rogelj, J.; De Stercke, S.; et al. A low energy demand scenario for meeting the 1.5 C target and sustainable development goals without negative emission technologies. Nat. Energy 2018, 3, 515–527.

[12] Hossain, M.; Madlool, N.; Rahim, N.; Selvaraj, J.; Pandey, A.; Khan, A.F. Role of smart grid in renewable energy: An overview. Renew. Sustain. Energy Rev. 2016, 60, 1168–1184.

[13] Karnouskos, S.; Colombo, A.W.; Lastra, J.L.M.; Popescu, C. Towards the energy efficient future factory. In Proceedings of the 2009 7th IEEE International Conference on Industrial Informatics, Cardiff, UK, 23–26 June 2009; pp. 367–371.

[14] M. Avci, M.E.; Asfour, S. Residential HVAC load control strategy in real-time electricity pricing environment.In Proceedings of the 2012 IEEE Conference on Energytech, Cleveland, OH, USA, 29–31 May 2012; pp. 1–6

[15] Vakiloroaya, V.; Samali, B.; Fakhar, A.; Pishghadam, K. A review of different strategies for HVAC energy saving. Energy Convers. Manag. 2014, 77, 738–754.

[16] T. S. Anttal, "Smart home dataset with weather information," Kaggle, 06-Apr-2019.[Online]. Available: https://www.kaggle.com/taranvee/smart-home-dataset-with-weather-information. [Accessed: 09-Nov-2021].

[17] Brown, R.G.; Hwang, P.Y.C. Introduction to Random Signals and Applied Kalman Filtering; John Wiley and Sons, Inc.: Hoboken, NJ, USA, 2012.

[18] Kawahara Y, Sugiyama M (2009) Sequential change-point detection based on direct density-ratio estimation. In: SIAM international conference on data mining, pp 389–400

[19] Montanez GD, Amizadeh S, Laptev N (2015) Inertial hidden Markov models: modeling change in multivariate time series. In: AAAI conference on artificial intelligence. pp 1819–1825

[20] Aminikhanghahi, S., & Cook, D. J. (2017). A survey of methods for time series change point detection. Knowledge and Information Systems, 51(2). https://doi.org/10.1007/s10115-016-0987-z

[21] Argmax.jp. 2021. changefinder - Argmax.jp. [online] Available at: <https://argmax.jp/index.php?changefinder> [Accessed 11 November 2021].

[22] Downey AB (2008) A novel changepoint detection algorithm. arXiv:0812.1237. Accessed 9 Nov 2021