User Behavior Analysis Using IoT Data

Azade Fotouhi*, Marwa Oudi, Billel Gueni
and Mouna Ben Mabrouk

Telecom & Media Division, Altran Technologies, France
E-mail: azade.fotouhi@altran.com; marwa.oudi@altran.com;
billel.gueni@altran.com; mouna.benmabrouk@altran.com
*Corresponding Author

Abstract

While the number of connected objects in the world of Internet of Thing (IoT) is increasing, an efficient and intelligent solution to exploit the huge amount of generated data does not exist. Smart homes powered by IoT devises are able to automate and monitor the every day activities of home owners, and improve the life quality especially for elderly and disabled people. In this paper, we propose a machine learning based model in order to analyze the IoT data and to provide data-driven services. Hence, this will make it possible to extract meaningful information from data and make intelligent decisions in smart environments. Then, the proposed model is evaluated using collected data from IoT devices based on different communication protocols.

Keywords: IoT, smart homes, machine learning, smart wireless communication.

1 Introduction

With the development of computers and electronics, everything is miniaturized and connected to the Internet, such as phones, smart-watches, sensors, etc. Therefore, the number of devices are growing enormously every day. It

is estimated that 30 billion connected devices will be deployed by 2025 [1]. Consequently, new applications related to connected devices will revolutionize our daily life. Due to the multiple technologies and advanced devices, the concept of Internet of Things (IoT) has evolved. The IoT is the network of devices that contains electronics, software, actuators, and connectivity which allows these things to connect, interact and exchange data. The IoT ecosystem is indeed quite complex because it integrates several technologies and area of expertise. A large number of applications including smart cities [2], smart homes [3], health-care [4], industries [5], and more have been impacted by the recent advancement in the IoT world.

Despite the rapid growth in IoT technologies and devices, there still exist several problems in exploiting them efficiently.

First of all, the diversity of communication technologies and standards of devices puts up a technological barrier against interoperability. IoT devices need a common protocol to communicate with each other. Interoperability, therefore, is still one of the major issues in IoT world. In the literature, most of the IoT's applications are limited in heterogeneity and exchange of information terms [6–8].

Secondly, the recent progresses in computer science have led to the development of a high number of low-cost sensors. Their low power, low cost and high security characteristics attract more and more experts and amateurs to employ them for diverse reasons. Consequently, they are generating a huge amount of data. Yet, the applications and platforms that provide cohesive and simple data collection tools are not well developed [9].

And finally, processing and analyzing the collected data from a wide variety of IoT devices and sensors enables the proliferation of data-driven decision making systems [10]. To this end, we must rely on machine learning based solutions to extract the valuable information, learn from them, and create smart environments.

The goal of this paper is to present a testbed which is able to communicate with a set of sensors belonging to different communication protocols, collect their information and process them, and eventually extract useful patterns from the processed data. Here, we particularly put our focus on smart homes domain powered by IoT. A smart home, is a home equipped with connected devices to automate and monitor the every day activities of home owners. The connected devices generates data while residents are doing their daily tasks. With the recent advances in the domain of artificial intelligence (AI), data analysis, and sensors, the benefits of this application in improving the

quality of life of the household members are threefold: Automation, Security and Energy-Efficiency.

Due to the continuous monitoring of data, automation of several daily activities becomes feasible. Analyzing the data from sensors help us to find the pattern of different actions, and consequently predicting them. Automation will bring comfort and ease for the home residence, specially the elderly and disabled people [11]. Using the machine learning technique, it is possible to learn the behaviors, the preferences and the daily activities of the dwellers, perform them automatically, while minimizing the human interactions. Enhanced security is one the main benefits of IoT technology and smart homes. These systems enables the home to be controlled from anywhere and anytime remotely. Additionally, after learning the life style of inhabitants, the intelligent modules can detect the abnormal behaviors such as unexpected entrance in home, or the unusual data pattern from IoT devices.

In order to save energy and reduce the costs, it is essential to recognize the energy consumption patterns of the both home appliances and the users. According to an investigation conducted by PwC [12], the very primary motivation to buy smart home devices is to reduce energy bill and increase the energy efficiency. Converging all IoT devices and applying artificial intelligence techniques makes the energy consumption more efficient.

The rest of this paper is organized as follows: The related work will be presented in Section 2. In Section 3, the system model is described. The results are explained in Section 4. Finally, the conclusion is given in Section 5.

2 Related Work

In the following, we review the recent advances in using machine learning methods to provide data driven services for a smart home including numerous IoT devices.

Authors in [13] have installed multiple sensors at different places of a home, and processed the collected data. After manually labeling a part of the data, they have used support vector machine (SVM) to detect different activities by 84% accuracy.

The sensor data from home testbeds are used in [14] to predict next occurrence time of a set of activities. The authors have assumed that an activity recognition algorithm exists to label the sensor information with the activities, and the output can feed their proposed activity prediction algorithms.

(a) CASAS model

(b) ARAS model
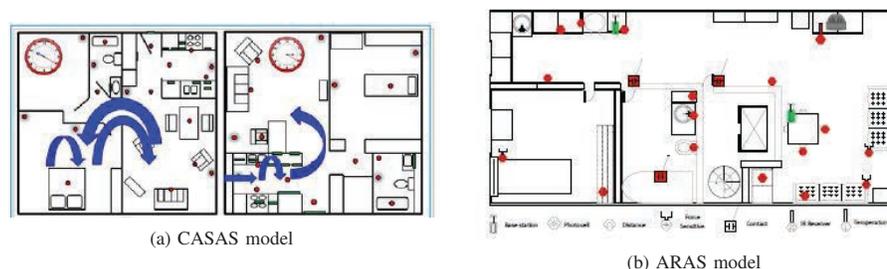
Figure 1   Sensor placement in two different smart home models.

MavHome (Managing an Adaptive Versatile Home) [15], is another examples of using machine learning algorithms to provide comfort for the inhabitants and reducing the operation cost. MavHome is more focused on mobility control for the users. By integrating machine learning, robotics, databases, mobile computing, and multimedia, MavHome tries to predict the mobility patterns of the inhabitants. To this end, the coverage area is divided to zones or sectors, and a corresponding graph is created to show the possible movements between zones. According to this work, location prediction is an essential task for the intelligent environments [15].

CASAS [13,16], a smart home in a Box, employs machine learning techniques to automate the actions. In this model, multiple sensors are installed at different places in a home, as shown in Figure 1, and the collected data is processed. Using the labeled data is can recognize the type of activity of a resident by 84% accuracy. The importance of energy saving in the building is also considered in CASCAS. By collecting more data and powering the box with more powerful algorithm, automation strategies for the residential is also possible as a future functionality for CASAS.

In order to collect data to analyze the human activities, authors in [17] developed a system composed on 20 binary sensors that communicate using zigbee. The data is collected in two different houses with multiple resident during two months.

3 System Model

To build a model representing a smart environment, we have used different sensors including proximity, luminosity and temperature to measure a variety of parameters in the environment. Raspberry Pi 3 B+ had been chosen as a gateway to communicate with the sensors and collect the received data and
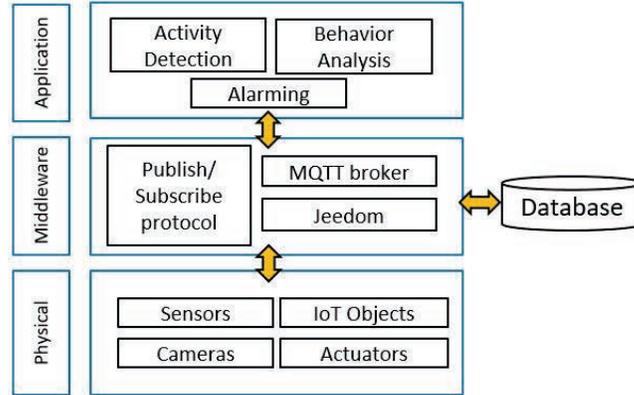
Table 1 List of devices

Name	Description
Raspberry Pi 3 model B+	as a gateway
Proximity sensor	zigbee based
Temperature sensor	zigbee based
Proximity sensor	two wired sensors
luminosity sensor	wired sensor

Figure 2 Demo with sensors and Raspberry pi 3.

store them on a database. Raspberry Pi has several advantages; first of all, it comes with multiple components, enabling it to communicate with Bluetooth, and Zigbee devices, and additionally it can be connected directly to a variety of the sensor and devices by cables. Moreover, it give us a huge processing unit in a small and compact board. Being low cost, and supporting python are among the other advantages of Raspberry Pi [18]. Table 1 lists the devices that have been used, and Figure 2 illustrates it.

To be able to communicate with Zigbee we have used zigbee2mqtt which is an Open Source project allows to build a Zigbee gateway. In order to run zigbee2mqtt on Raspberry Pi, CC debugger, CC2531 USB sniffer, Downloader cable CC2531 are needed [19]. Additionally, a MQTT broker is needed on the system. MQTT is a lightweight publish-subscribe network protocol that receives all messages from the clients and then routes the messages to the appropriate destination clients. Here, we used Mosquitto as the broker in the middle layer. The structure of our model is presented in Figure 3.

Figure 3   Architecture of our model to communicate with zigbee.

(a) Collected data from temperature sensor

Date	Illuminance (Lx)	Occupancy
2019-9-9 16:38:03	282	VRAI
2019-9-9 16:38:03	282	VRAI
2019-9-9 16:39:30	360	VRAI
2019-9-9 16:39:30	360	FAUX
2019-9-9 16:41:00	360	FAUX

(b) Collected data from proximity sensor

Figure 4   Collected data sample.

We then started collecting data from all sensors and storing them in csv files. The time with millisecond precision and measured values are registered for each sensor. The temperature sensor, is able to keep track of temperature, pressure and humidity. The data has been collected during three months continuously. Figure 4 displays an example of collected data from different sensors. The data of all sensors are merged in a chronological order, the noisy data are removed and the missing values are filled properly.

In order to have a better understanding on the data structure and the dependencies, we first performed different analysis and visualization models. According to our analysis, the distribution of data has a strong temporal characteristics. Therefore, the moving window model is used to extract the useful features from data. Each window can contain κ sequence of events. By processing each window several information such as starting time, duration, average value of each sensor, number of changing status are calculated. Then the results are used to build a clustering model. It must be mentioned that the the the optimal number κ should be computed in a way that maximize the performance of clustering method.

For detecting different actions in data and performing clustering Agglomerative and Kmean methods are used. The results shows that both are able to detect few actions and patterns in the data. To evaluate the result, a variety of scores such as silhouette, accuracy and NMI (Normalized Mutual Information) have been employed.

The results show that our model is capable of detecting different pattern and actions with considerable accuracy. These actions help in shaping the life style of each user and recognizing uncommon activities in future. Additionally, considering the potential of our model, adding more sensors and devices, even in different communication protocol, will enhance the behavior analysis and detection. The detail about the feature extraction model and the machine learning algorithms, along with the results will be provided in the next section.

4 Data Analysis

To find the best number of clusters or events in the data, Silhouette Coefficient is calculated using the scikit-learn library [20]. It is a metric to measure how
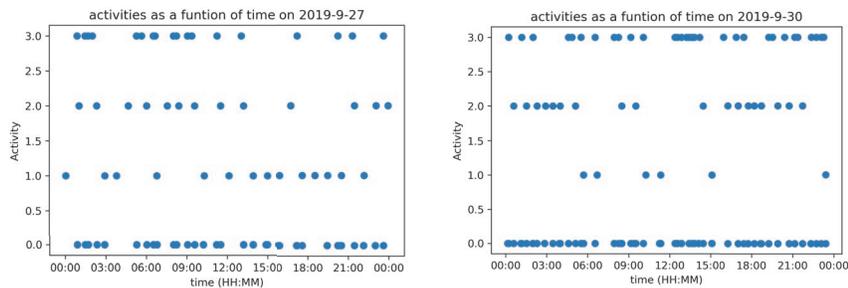
Figure 5 Activities as a function of time.

The silhouette plot for the various clusters.
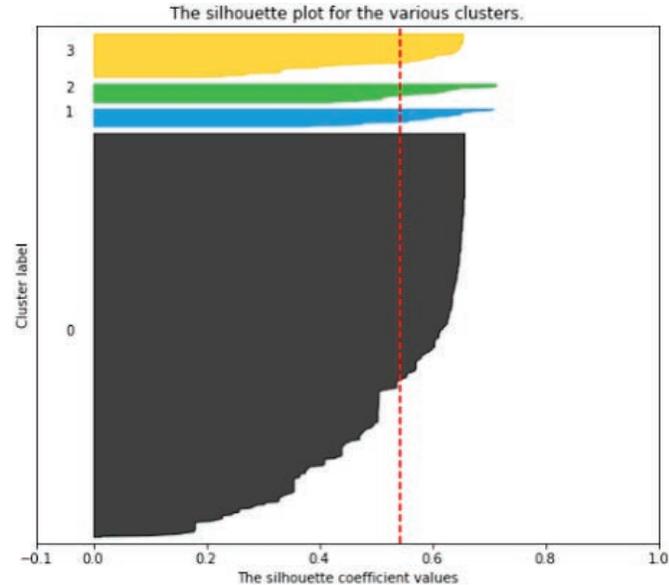
Figure 6   silhouette coefficient values for four clusters.

close each point in one cluster is to points in the neighboring clusters. The best value is 1 and the worst value is -1. Values near 0 indicate overlapping clusters. Negative values generally indicate that a sample has been assigned to the wrong cluster, as a different cluster is more similar.

We have implemented an algorithm to find the best number of clusters based on silhouette coefficient. As illustrated in Figure 6, the four actions will give us the best clustering for our data.

Then, considering four different actions, the data is presented as a function of time. Two different models are depicted in Figure 5. The results show that the clusters or the action are well detectable in this model. additionally, another important point is the strong temporal dependencies between actions and time. It will help in extracting more features based on time, to separates well the data for each action.

5 Conclusion

In this paper, we showed that IoT and machine learning have brought a huge potential in the domain of smart homes. The collected data from a wide category of devices can be used in analyzing, detecting, and predicting the

daily activities of home residents. The model that we developed is able to collect data from different protocol such as Bluetooth, and ZigBee. It is efficient, simple, and scalable. The data is analaysed in order to see how it can be used for more processes. In our future works, we will presents how collected data in long duration, and with a richer sensor category, can discover and predict the actions, that can be useful especially for elderly and disables people.

References

[1] Nokia, "LTE-M – Optimizing LTE for the Internet of Things," *Nokia white paper*, May 2015.

[2] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, Feb 2014.

[3] K. Agarwal, A. Agarwal, and G. Misra, "Review and performance analysis on wireless smart home and home automation using iot," in *2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 2019, pp. 629–633.

[4] S. M. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K. Kwak, "The internet of things for health care: A comprehensive survey," *IEEE Access*, vol. 3, pp. 678–708, 2015.

[5] A. Vakaloudis and C. O'Leary, "A framework for rapid integration of IoT Systems with industrial environments," in *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*, 2019, pp. 601–605.

[6] V. R. Konduru and M. R. Bharamagoudra, "Challenges and solutions of interoperability on iot: How far have we come in resolving the iot interoperability issues," in *2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon)*, 2017, pp. 572–576.

[7] T. Rahman and S. K. Chakraborty, "Provisioning technical interoperability within zigbee and ble in iot environment," in *2018 2nd International Conference on Electronics, Materials Engineering Nano-Technology (IEMENTech)*, 2018, pp. 1–4.

[8] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," vol. 17, no. 4, 2015, pp. 2347–2376.

[9] T. Munasinghe, E. W. Patton, and O. Seneviratne, "Iot application development using mit app inventor to collect and analyze sensor data,"

in *2019 IEEE International Conference on Big Data (Big Data)*, 2019, pp. 6157–6159.

[10] A. M. Ghosh and K. Grolinger, "Deep learning: Edge-cloud data analytics for iot," in *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, 2019, pp. 1–7.

[11] P. Mtshali and F. Khubisa, "A smart home appliance control system for physically disabled people," in *2019 Conference on Information Communications Technology and Society (ICTAS)*, 2019, pp. 1–5.

[12] D. Bothun and M. Lieberman, "Smart home, seamless life unlocking a culture of conveniences," in *Consumer Intelligence Series*.

[13] D. J. Cook, A. S. Crandall, B. L. Thomas, and N. C. Krishnan, "Casas: A smart home in a box," *Computer*, vol. 46, no. 7, pp. 62–69, July 2013.

[14] B. D. Minor, J. R. Doppa, and D. J. Cook, "Learning activity predictors from sensor data: Algorithms, evaluation, and applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 12, pp. 2744–2757, Dec 2017.

[15] S. K. Das, D. J. Cook, A. Battacharya, E. O. Heierman, and Tze-Yun Lin, "The role of prediction algorithms in the mavhome smart home architecture," vol. 9, no. 6, 2002, pp. 77–84.

[16] D. J. Cook, N. C. Krishnan, and P. Rashidi, "Activity discovery and activity recognition: A new partnership," vol. 43, no. 3, 2013, pp. 820–828.

[17] H. Alemdar, H. Ertan, O. D. Incel, and C. Ersoy, "Aras human activity datasets in multiple homes with multiple residents," in *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*, 2013, pp. 232–235.

[18] "Raspberry pi foundation," https://www.raspberrypi.org/, accessed: 2020-04-30.

[19] G. pages, "Zigbee2mqtt documentation," *url: https://www.zigbee2mqtt.io/*, 2019.

[20] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and Édouard Duchesnay, "Scikit-learn: Machine learning in python," vol. 12, no. 85, 2011, pp. 2825–2830. [Online]. Available: http://jmlr.org/papers/v12/pedregosa11a.html

Biography

Azade Fotouhi received her PhD from University of New South Wales (UNSW), Australia in the school of computer science and engineering. Following a postdoctoral fellowship at Data61|CSIRO, she joined Altran Research and Innovation, France, in 2018. Currently she is working as a Research engineer in the Machine learning and Telecom. She has authored several papers in IEEE journals and conferences, all in recognized venues, and was the recipient of the NASSCOM Student business Innovation Awards, Australia, in 2016. Her research interests include UAV Communication, Machine Learning, IoT, and Mobile Networks.