data

Data Descriptor

SIMADL: Simulated Activities of Daily Living Dataset

Talal Alshammari [1,2,]*[iD]**, Nasser Alshammari** [1,3] [iD]**, Mohamed Sedky** [1] [iD] **and Chris Howard** [1]

1   Staffordshire University, College Road, ST4 2DE Stoke-on-Trent, UK;
     nasser.alshammari@research.staffs.ac.uk or nashamri@ju.edu.sa (N.A.); m.h.sedky@staffs.ac.uk (M.S.);
     c.howard@staffs.ac.uk (C.H.)
2   College of Computer Science and Engineering, University of Hail, Hail 81481, Saudi Arabia
3   College of Information and Computer Science, Al Jouf University, Sakaka 72388, Saudi Arabia
*   Correspondence: talal.alshammari@research.staffs.ac.uk or talal.alshammari@uoh.edu.sa;
     Tel.: +44-751-494-0500

Abstract: With the realisation of the Internet of Things (IoT) paradigm, the analysis of the Activities of Daily Living (ADLs), in a smart home environment, is becoming an active research domain. The existence of representative datasets is a key requirement to advance the research in smart home design. Such datasets are an integral part of the visualisation of new smart home concepts as well as the validation and evaluation of emerging machine learning models. Machine learning techniques that can learn ADLs from sensor readings are used to classify, predict and detect anomalous patterns. Such techniques require data that represent relevant smart home scenarios, for training, testing and validation. However, the development of such machine learning techniques is limited by the lack of real smart home datasets, due to the excessive cost of building real smart homes. This paper provides two datasets for classification and anomaly detection. The datasets are generated using OpenSHS, (Open Smart Home Simulator), which is a simulation software for dataset generation. OpenSHS records the daily activities of a participant within a virtual environment. Seven participants simulated their ADLs for different contexts, e.g., weekdays, weekends, mornings and evenings. Eighty-four files in total were generated, representing approximately 63 days worth of activities. Forty-two files of classification of ADLs were simulated in the classification dataset and the other forty-two files are for anomaly detection problems in which anomalous patterns were simulated and injected into the anomaly detection dataset.

Dataset: 10.5281/zenodo.1185172

Dataset License: CC-BY

Keywords: smart home; simulation; dataset; internet of things; machine learning; classification; anomaly detection

---

1. Introduction

Recent developments in technology have increased the adoption of smart devices and sensors in smart homes. With the realisation of the Internet of Things paradigm (IoT), the number of these internet-connected devices is likely to grow. In a study conducted by Gartner [1], the number of connected "Things" is 8.4 billion devices in 2017. This number grew by 31% from 2016 and the study predicts that the number will continue to grow and will reach 20.4 billion connected devices by 2020. Moreover, the spending on IoT services that provide design, development, and implementation of IoT solutions was estimated to reach $273 billion by the end of 2017.

With the widespread usage of smart devices in smart homes, these environments will generate an enormous amount of streaming data. These generated data have the potential to provide novel services to the smart home inhabitants to improve their standards of living. These services can benefit from the analysis of this generated data.

Machine learning has been widely applied to develop probabilistic and statistical methods and sequence-learning algorithms to classify and predict ADLs of inhabitants. Nowadays, machine learning models and their contribution to the Internet of Things (IoT) applications are becoming one of the most active and interesting research areas [2]. The smart home is one of the most prominent applications of the IoT paradigm. There are many advantages for adopting smart home technologies such as monitoring energy consumption, security, automation, entertainment, eldercare, etc. To implement machine learning techniques in any of the previous applications, a representative dataset for that application is required. The dataset will be used to train and test the machine learning models to evaluate and validate their performance.

There are real smart home datasets available in the literature (e.g., [3–5]), however, they lack the flexibility to cope up with the recent advancements in sensor techniques, and they are costly to build and construct. Up to the knowledge of the authors, there is no real-world dataset targeted at anomaly detection in the context of smart homes.

Smart home simulation tools are an alternative to constructing real smart homes. These tools allow the researcher to design a smart home suitable to the problem that they are investigating and generate a representative dataset. There is less cost and effort involved in the process, and they can cope with new emerging techniques. However, many of these simulation tools are not available in the public domain as an open-source project, or they lack the flexibility and accessibility for both the researchers and the participants.

The simulation tools regarding dataset generation approaches can be categorised into two approaches, model-based and interactive approaches. The model-based approaches generate datasets using pre-defined scripts that generate events, probability of the occurrence of events, and their duration. On the other hand, the interactive approaches capture the sensor activities and log them to the dataset in real-time. Examples of model-based approaches include [6–8]. Examples of interactive approaches include [9–11].

The two approaches suffer from disadvantages for the researchers and the participants alike. Model-based approaches allow the researcher to generate big datasets in short periods of time. However, the generated datasets do not capture realistic and fine-grained interactions that happen in real smart homes. The interactive approaches can capture these fine-grained interactions because they capture the output of the sensors directly to the dataset. However, the interactive approaches produce smaller datasets and take more time for the participants to perform their habits. Most of the interactive tools focus on context-awareness applications and not on generating datasets.

OpenSHS is an open-source, 3D, cross-platform simulation tool that follows a hybrid approach that combines the advantages of both approaches. It allows the researcher to design a smart home specific to their research problem and generate a sufficiently large dataset in reasonable time while retaining the fine-grained interactions that the participants are performing.

This paper presents two datasets generated by OpenSHS for classification and anomaly detection problems. The remainder of this paper is structured as follows: Section 2 presents the related work in the literature. Section 3 explains OpenSHS architecture and how we use it to generate the datasets. Section 4 presents our methodology to generate the two datasets. Section 5 provides a description of the datasets.

2. Related Work

In this section, we review some of the available real datasets in the literature and the simulation tools that allow the researchers to generate synthetic datasets.

2.1. Real Datasets

Alemdar et al. [3] published the ARAS (Activity Recognition with Ambient Sensing) dataset which is a real dataset for complex scenarios of multi-residents. The dataset was captured for the duration of two months for two different houses and each house had two inhabitants. ARAS dataset was used to assess ADLs classification algorithms.

The Centre for Advanced Studies in Adaptive Systems (CASAS) is a project for creating real smart homes for the researchers in this field. Cook et al. [12] designed a simple and lightweight toolkit called "smart home in a box". The components of this toolkit are assembled in a single small box and easily installed in a home to be able to provide smart tasks. They have installed the toolkit in 32 smart homes and generated several datasets. The datasets are publicly available online [4].

The TigerPlace [13] conducted a study on the ageing population. They used passive sensor networks that were installed in 17 flats within eldercare facility. They used many kinds of sensors such as motion sensors, pressure sensors, etc. In some of the flats, the collection of data took more than two years.

Some datasets focus on wearable technologies to monitor and acquire the activities performed by the participants. The Smartphone-Based Human Activities Recognition (SBHAR) dataset [14] is one example of such datasets. The authors collected the accelerometer and gyroscope data of 30 participants who performed several ADLs using a smartphone. Casale et al. [15] and Bruno et al. [16] are other examples of similar datasets.

The Intelligent System Laboratory (ISL) [17] generated a dataset from three smart homes in which a single participant was performing his ADLs. The dataset represent around two months worth of data. The first smart home had 14 sensors, the second had 23 sensors, and the third had 21 sensors.

Using a camera feed to capture a participant activities is another approach to recognise ADLs. Pirsiavash and Ramanan [18] presented a dataset of one million frames captured from a wearable camera that represent a first-person view. The data were gathered from 20 participants who performed unscripted ADLs in their homes.

The ContextAct@A4H dataset [19] is an example of recent datasets that focus on ADLs. The dataset was generated using a real flat equipped with many sensors of different types. The dataset consists of one week worth of captured data during the summer season and three weeks of the fall season. The authors proposed a new annotation method using temporal logic.

2.2. Simulation Tool

Synnott et al. [20] conducted a survey of existing simulation tools for generating datasets in a smart home environment. They found that, due to the sensors technology cost, availability limitations, time considerations and finding the optimal sensors configurations, simulation tools are valuable assets to have for smart home research. The authors also identified that most of the available simulation tools focus on context-awareness applications and not on generating representative datasets. Moreover, supporting multiple inhabitants was one of the features lacking in current simulation tools.

Cook et al. [21] presented some challenges facing the evaluation of machine learning performance and pervasive computing techniques. The authors identified the need to have real datasets and there is a lack of real datasets in the literature.

Bouchard et al. [8] designed a 3D smart home simulator for activity recognition to overcome the limitations of creating real datasets in a smart home. Many pre-recorded scenarios were captured from clinical experiments and used to generate datasets.

To evaluate activity recognition algorithms, researchers require good representative datasets. Due to the high cost of building real smart homes and due to privacy and ethical issues with the human subjects, Helal et al. [22] developed an event driven simulation tool for researchers in the smart home domain. The developed simulator is called "Persim" and it can generate realistic datasets for complex scenarios of the occupant's activities.

An improved version of Persim was developed by Helal et al. [23] called PerSim 3D. This tool helps to generate realistic datasets from the inhabitants activities in a smart home scenario. The major improvement was adding 3D simulations of the inhabitant, sensors and actuators. In addition, the tool supports the researcher by a Graphical User Interface (GUI) to envisage the activities in 3D.

The intelligent environment simulation (IE Sim) developed by Synnott et al. [9] to generate synthetic datasets that capture ADL of smart home users. IE Sim provides the researcher with a 2D graphical interface of the floor plan to design the smart homes. The researcher can add different types of sensors such as temperature sensors, pressure sensors, etc. Then, using an avatar, the simulation can be carried out to capture ADLs. The output of the simulation dataset is in the homeML [24] file format.

3. OpenSHS

Most of the available simulation tools follow two approaches to generate synthetic datasets, model-based and interactive approaches [20].

The model-based approach relies on already defined statistical models of activities to generate synthetic data. The statistical model determines the order of events, the probability of occurrence, and the duration of activities. The model-based approach makes it easy to generate large datasets in a short period of time. The disadvantage of this approach is the lack of capturing fine-grained interactions and/or unexpected accidents that are common in real activities.

The interactive approach, on the other hand, can capture more interesting interactions and fine-grained details. This approach uses a virtual avatar controlled by a researcher, a human participant or a simulated participant. The avatar moves and interacts with the virtual environment equipped with virtual sensors and/or actuators. These interactions can be passive or active.

An example of active interactions is opening a door or turning the light on or off. Another example of passive interactions is having a pressure sensor installed on the floor that detects the movements of the avatar without the avatar explicitly activating the sensor. The disadvantage of the interactive approaches is how long it takes to generate enough data: because of the nature of the approach, the interactions must be captured in real-time.

3.1. OpenSHS Advantages

Most of the simulation tools in the literature are not open-source, except for [8], which makes it harder for the researcher to acquire the software and modify it to the experiment's need. In addition, having a 3D simulation adds to the realism of the conducted experiment.

OpenSHS is an open-source smart home simulator that allows the participants to simulate their ADLs in a 3D virtual environment. OpenSHS is developed with open-source and cross-platform techniques that makes it easy for the researcher to modify the tool and extend it according to their needs.

The approach that OpenSHS uses to generate datasets can be thought of as a hybrid approach of the model-based and interactive approaches. OpenSHS offers a replication mechanism of the recorded ADLs which allows for a quick and large dataset generation, similar to the model-based approaches. The replications have fine-grained details as the activities are captured in real-time, similar to the interactive approaches.

OpenSHS has the flexibility to add different activity labels that can be customised by the researcher and tailored to their needs. It also has a fast-forwarding feature which facilitates the simulation of long inactivity periods.

We use OpenSHS to generate the two datasets. One is for classification and prediction of ADLs problems and the other is for anomaly detection problems.

4. Methodology

In this section, we present the design of the smart home and the contexts to be performed by the participants, followed by the aggregation and generation of the datasets.

4.1. Smart Home Design

We designed a smart home consisting of a bedroom, living room, bathroom, kitchen, and home office, as shown in Figure 1. Each room has several types of sensors.

Figure 1. The design of the smart home.

The smart home is equipped with twenty-nine binary sensors, as shown in Table 1. The binary sensor has two states, on (1) and off (0). The sensors can be divided into two groups, passive and active. The passive sensors do not explicitly require the participant to interact with them. Instead, they react to the participant movements and position. An example of this type is the carpet sensors. The carpet sensors turn on when the participant walks over them.

The other type of sensors are the active sensors. This type requires explicit action from the participant to change their state, for example, when opening a door or when turning on the light.

The activities labels that we decided to include in this dataset are: *sleep, eat, personal, work, leisure, and other*. The anomaly detection dataset includes an additional label *anomaly*.

The participant controls a 3D avatar in first-person view and navigates and performs his/her ADLs in the virtual smart home environment. Throughout the simulation period, OpenSHS will capture and record the state of all the smart devices and sensors every second. Some activities take a long time, such as staying at the office for studying. OpenSHS provides a solution for this problem by implementing a fast-forwarding mechanism which enables the participants to quickly perform the long constant activities.

During the simulation, when a participant wants to change his/her activity, they can do that by using the dialogue shown in Figure 2. It is worth noting that, when the participants change their activity label, it does not immediately apply the change in the dataset. The activity label changes when one of the sensor's state has changed. This approach ensures a clean separation when the participant transits from one activity to another.

OpenSHS uses the concept of a context which is a specific time-frame of interest to the researcher to be simulated [25]. In this work, we have chosen to simulate the interactions of the participants in different contexts. On the weekdays, we have two contexts, one in the morning and the other in the evening. On the weekends, we have the same contexts during the day. Thus, there are four different

contexts per participant. The day contexts are "morning" and "evening" contexts. The week contexts are "weekday" and "weekend".

Table 1. All smart home sensors.

| #	Name	Type	Description	Active/Passive
1	bathroomCarp	binary	Bathroom carpet sensor	Passive
2	bathroomDoor	binary	Bathroom door sensor	Active
3	bathroomDoorLock	binary	Bathroom door lock sensor	Active
4	bathroomLight	binary	Bathroom ceiling light	Active
5	bed	binary	Bed contact sensor	Passive
6	bedTableLamp	binary	Bedroom table lamp	Active
7	bedroomCarp	binary	Bedroom carpet sensor	Passive
8	bedroomDoor	binary	Bedroom door sensor	Active
9	bedroomDoorLock	binary	Bedroom door lock sensor	Active
10	bedroomLight	binary	Bedroom ceiling light	Active
11	couch	binary	Living room couch	Passive
12	fridge	binary	Kitchen fridge	Active
13	hallwayLight	binary	Hallway ceiling light	Active
14	kitchenCarp	binary	Kitchen carpet sensor	Passive
15	kitchenDoor	binary	Kitchen door sensor	Active
16	kitchenDoorLock	binary	Kitchen door lock sensor	Active
17	kitchenLight	binary	Kitchen ceiling light	Active
18	livingCarp	binary	Living room carpet sensor	Passive
19	livingLight	binary	Living room ceiling light	Active
20	mainDoor	binary	Main door sensor	Active
21	mainDoorLock	binary	Main door lock sensor	Active
22	office	binary	Office room desk sensor	Passive
23	officeCarp	binary	Office room carpet sensor	Passive
24	officeDoor	binary	Office door sensor	Active
25	officeDoorLock	binary	Office door lock sensor	Active
26	officeLight	binary	Office ceiling light	Active
27	oven	binary	Kitchen oven sensor	Active
28	tv	binary	Living room TV sensor	Active
29	wardrobe	binary	Bedroom wardrobe sensor	Active
30	Activity	String	The current participant activity	
31	timestamp	String	The timestamp every second	
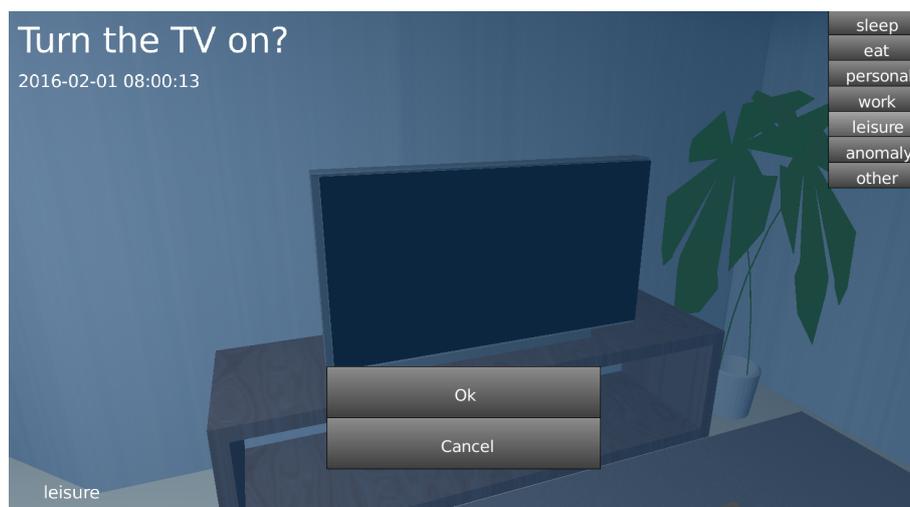
Figure 2. The activities selection dialogue.

4.2. The Participants

The participants in this work were chosen randomly but all of them have jobs. They also have experience with first-person games which will facilitate the learning curve of the tool.

The number of participants was 7, and the average time it took to conduct the simulation was 50 min ($min_{time} = 30$, $max_{time} = 75$, $std_{time} = 14.43$).

For each participant, we followed the following procedures:

1. The researcher guides the participant and shows him/her the virtual smart home.
2. The participant is asked to play with the virtual smart home to get familiar with it.
3. The participant's familiarity with the virtual smart home is tested by asking them to perform specific tasks.
4. The actual simulation takes place, and the participant is asked to give us their actual starting times for each context.
5. The participant is asked to complete the usability questionnaire.

4.3. The Anomalies

In some contexts, the definition of an anomaly is clear and can be quantified, for example, the heart rate for a patient. A heart rate that ranges from 60 to 100 beats per minutes is considered a normal resting heart rate for an adult. However, in the context of an inhabitant's behaviour in their smart home environment, the definition of what an anomalous behaviour is can be difficult and hard to quantify. Anomalous behaviour becomes much more subjective and varies from one inhabitant to another. Thus, anomalies in the datasets were not injected after the simulations were conducted, based on the researcher's idea of what an anomaly is.

To overcome the issues with defining what is an anomaly for an inhabitant, the researcher left this definition to the persons capable of defining these anomalies, the participants themselves.

Each participant performed an additional simulation that is intended to represent an anomaly from the point of view of the participant. All the anomalies are defined by the participants and no restrictions were imposed by the researcher. Table 2 shows each participant's anomaly that he/she simulated. Although there are seven anomalies in total, each anomaly is injected into six different contexts based on the user's behaviour.

Table 2. The anomalies defined by the participants.

Participants	Anomaly Definition
participant 1	leaving the fridge door open.
participant 2	leaving the oven on for long time.
participant 3	leaving the main door open.
participant 4	leaving the fridge door open.
participant 5	leaving the bathroom light on.
participant 6	leaving tv on.
participant 7	leaving light bedroom and wardrobe open.

4.4. Dataset Aggregation

To accelerate the process of generating the dataset, the participants are asked to perform several simulations of the same context. Since we record the activities of the participants in real-time, every simulation will be different and will contain unique information. OpenSHS provides an aggregation algorithm that uses all the real-time recorded simulations to generate a new and random dataset but in a controlled manner [25].

For each participant, we have generated six datasets with unique parameters. The parameters used to generate each dataset are as follows:

1. Days: We chose 30 and 60.
2. Start-date: We chose 1 February 2016.
3. Time-margin: We chose the values 0, 5, and 10.

The above parameters generated one month and two months worth of data. For the one-month set, we have three variants with 0, 5, and 10 time-margins. The same goes for the two-month set. This ensures that the generated datasets are different in the time dimension. Table 3 shows a sample of the final dataset.

Table 3. A sample of the final dataset output.

Timestamp	Bed Table Lamp	Bed	Bathroom Light	Bathroom Door	...	Activity
2016-04-01 08:00:00	0	1	0	0	...	sleep
2016-04-01 08:00:01	0	1	0	0	...	sleep
2016-04-01 08:00:02	0	1	0	0	...	sleep
2016-04-01 08:00:03	0	1	0	0	...	sleep
2016-04-01 08:00:04	1	1	0	0	...	sleep
2016-04-01 08:00:05	1	0	0	0	...	sleep
2016-04-01 08:00:06	1	0	0	1	...	personal
2016-04-01 08:00:07	1	0	0	1	...	personal
2016-04-01 08:00:08	1	0	1	1	...	personal
2016-04-01 08:00:09	1	0	1	1	...	personal
2016-04-01 08:00:10	1	0	1	1	...	personal
⋮	⋮	⋮	⋮	⋮	⋮	

5. Dataset Description

We generated a dataset for classification problems, and a dataset for anomaly detection problems. Each dataset consists of forty-two files, thus totalling eighty-four files. The naming convention used for the datasets files is d{x}-{y}m-{z}tm where:

- x is an index number to uniquely identify a dataset;
- y is the number of months generated; and
- z is the time-margin value.

The classification dataset has a target column of the previously mentioned labels of the activities, while the anomaly detection dataset has an additional label for the anomalous activity. In addition to the twenty-nine binary sensor readings, both datasets have a timestamp column.

Table 4 shows a listing of the number of records for both datasets excluding the header record. It is worth noting that, for each file in the classification dataset, OpenSHS generated the final output randomly from the record samples. The same procedure was used for the anomaly dataset, with the exception that the anomalous activity was injected in the last quarter of the file. This decision of injecting the anomalous activity towards the end of the file was made to allow the model to learn the normal patterns before detecting the anomalous ones in anomaly detection problems.

Figure 3 shows seven bar charts of the classification files. Each bar chart shows the proportions of the training records (the first 60%) and the testing records (the last 40%). Some files do not have all the labels included because the participants did not perform that activity, for instance, as shown in the dataset d1_2m_0tm where the participant did not perform the "work" activity.

Figures 4 and 5 show the frequency of the active sensor readings that are associated with the "leisure" label in the training and testing samples which shows that there are slight differences between the two. The remaining labels, figures, and dataset files are available online at http://datasets.openshs.org.

Table 4. The number of records for the forty-two files for both datasets.

Name	Classification Dataset	Anomaly Dataset
d1-1m-0tm	18,800	18,120
d1-1m-5tm	18,966	18,096
d1-1m-10tm	18,828	18,044
d1-2m-0tm	38,204	35,033
d1-2m-5tm	37,532	34,967
d1-2m-10tm	38,012	35,065
d2-1m-0tm	37,332	35,358
d2-1m-5tm	36,261	35,679
d2-1m-10tm	35,687	35,541
d2-2m-0tm	75,183	74,171
d2-2m-5tm	72,302	72,163
d2-2m-10tm	73,526	72,751
d3-1m-0tm	39,832	40,603
d3-1m-5tm	42,526	40,064
d3-1m-10tm	40,730	41,681
d3-2m-0tm	77,328	88,091
d3-2m-5tm	83,346	88,091
d3-2m-10tm	79,933	87,552
d4-1m-0tm	40,232	30,031
d4-1m-5tm	40,015	30,923
d4-1m-10tm	38,629	29,645
d4-2m-0tm	80,033	61,114
d4-2m-5tm	79,171	59,444
d4-2m-10tm	79,176	56,829
d5-1m-0tm	27,762	41,343
d5-1m-5tm	28,008	39,724
d5-1m-10tm	28,450	40,817
d5-2m-0tm	55,577	78,267
d5-2m-5tm	56,200	79,048
d5-2m-10tm	56,919	78,627
d6-1m-0tm	81,859	88,883
d6-1m-5tm	85,763	90,434
d6-1m-10tm	84,672	88,942
d6-2m-0tm	165,596	174,809
d6-2m-5tm	165,038	174,189
d6-2m-10tm	167,282	169,654
d7-1m-0tm	49,282	53,321
d7-1m-5tm	49,605	51,972
d7-1m-10tm	49,769	52,262
d7-2m-0tm	100,544	99,193
d7-2m-5tm	100,498	102,340
d7-2m-10tm	100,502	100,974
Total	2,674,910	2,743,855
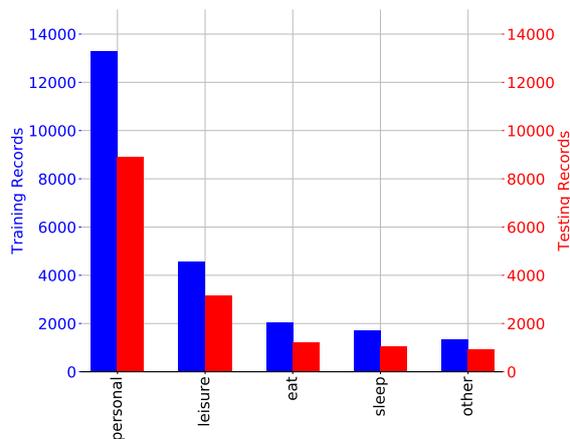
(**a**) d1_2m_0tm dataset
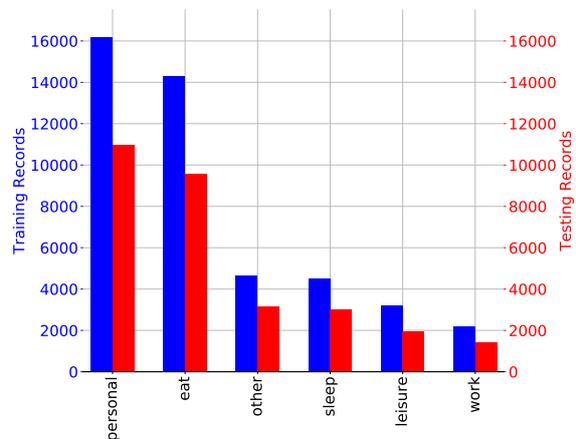
(**b**) d2_2m_0tm dataset

Figure 3. *Cont.*

(**b**) d3_2m_0tm dataset

(**c**) d4_2m_0tm dataset

(**d**) d5_2m_0tm dataset

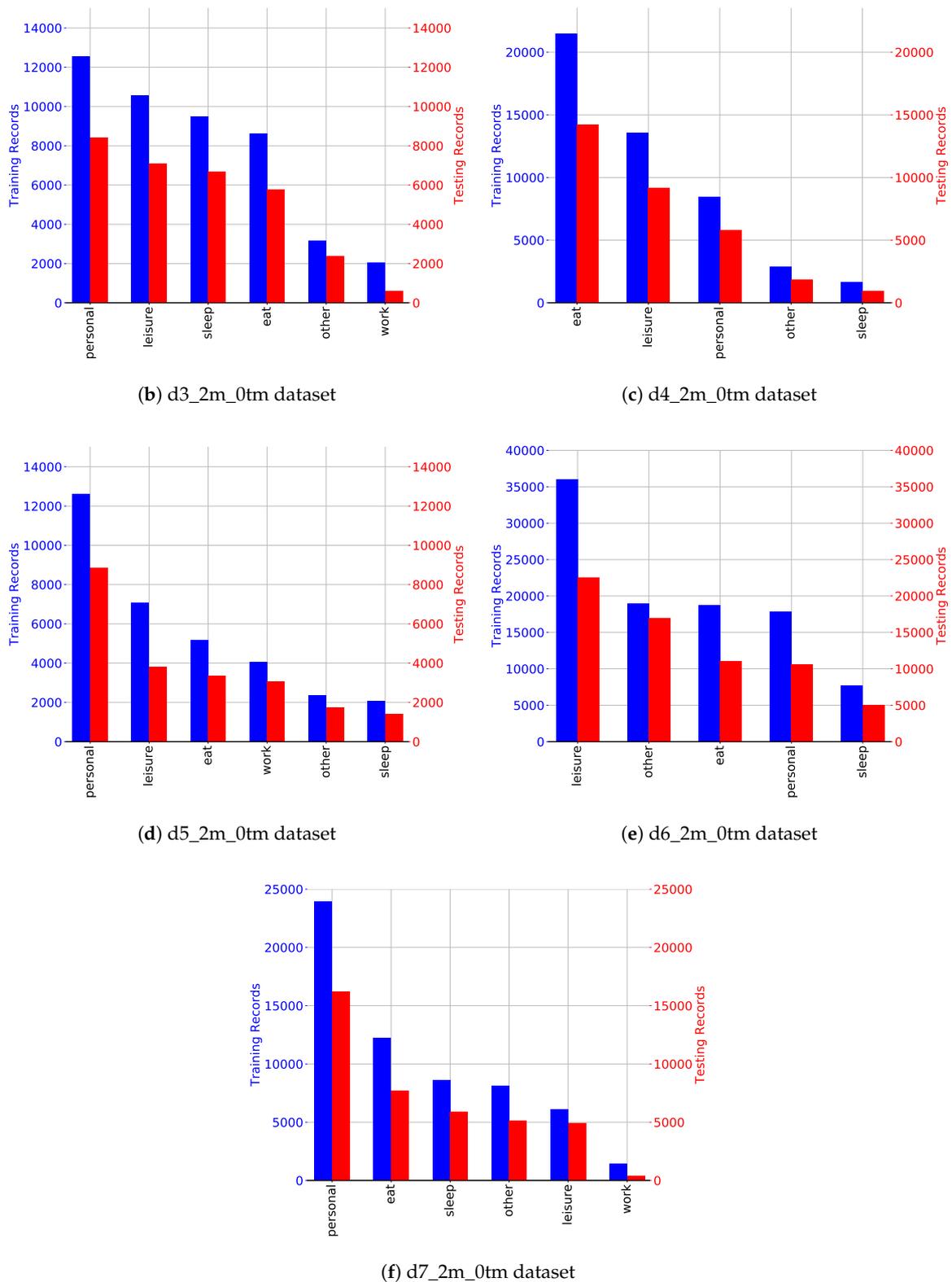(**e**) d6_2m_0tm dataset

(**f**) d7_2m_0tm dataset

Figure 3. Seven files from the classification dataset with a 60%/40% split for training and testing records.
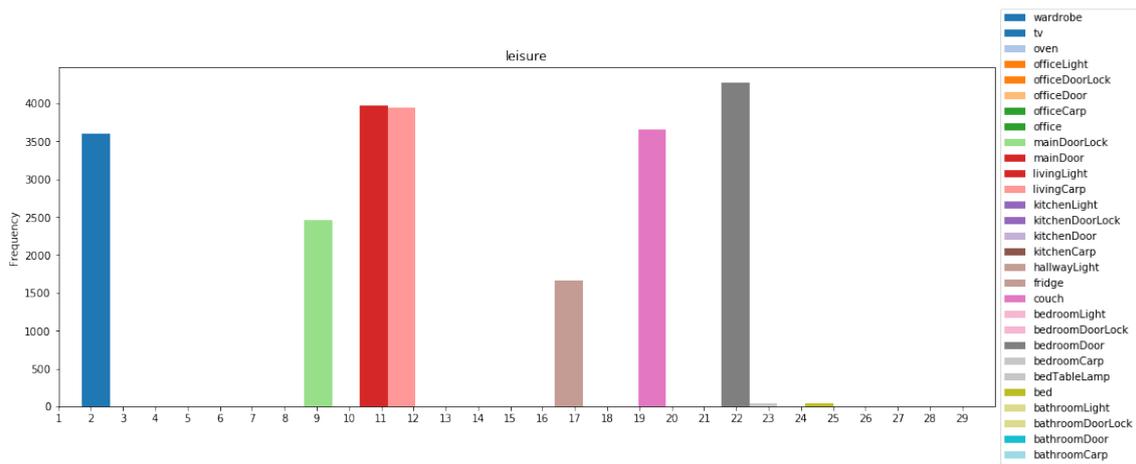
Figure 4. The sensor readings for the leisure activity in the training sample.
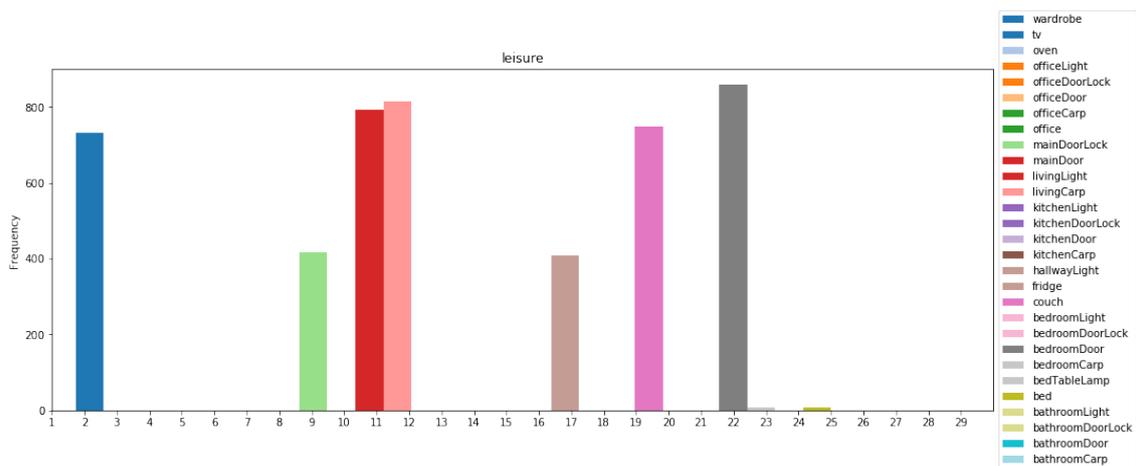

Figure 5. The sensor readings for the leisure activity in the testing sample.

6. Conclusions

This paper introduces two datasets for the smart home research community, one for classification and the other for anomaly detection. The two datasets are generated using a simulation tool (OpenSHS), and seven participants simulated their ADLs. The collection of the generated date accumulates to 63 days worth of patterns for both datasets.

Representative smart home datasets, such as the ones presented in this paper, have direct machine learning applications, mainly for the training, testing and validation of new models. Different datasets are needed depending on the machine learning target application, i.e., classification, clustering, prediction or anomaly detection. The contributed datasets can be used to validate machine learning models that perform classification tasks and/or anomaly detection tasks in the smart home domain. Classification and anomaly detection tasks are applicable to many use cases such as automation, eldercare, healthcare, entertainment, security, etc.

For future work, we will use the developed datasets to visualise smart home designs. This visualisation would allow researchers to identify drawbacks in a smart home environment. This will help and accelerate the development and proposition of new effective designs. Moreover, within the IoT paradigm, the contributed datasets will be used to test and validate IoT frameworks.

Author Contributions: Talal Alshammari contributed to the data collection and analysis and managed the participants during the experiments, the usability study and completed a review of existing datasets. Nasser Alshammari contributed to the data collection and analysis and managed the participants during the experiments. Mohamed Sedky and Christopher Howard provided guidance and direction for the implementation, development and evaluation of the research. All authors significantly contributed to the writing and review of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ADL　　　　　Activities of Daily Living
IoT　　　　　Internet of Things
OpenSHS　　Open Smart Home Simulator

References

1. 8.4 Billion Connected Things Will Be in Use in 2017, Up 31 Percent from 2016. 2017. Available online: http://www.gartner.com/newsroom/id/3598917 (accessed on 31 March 2018).
2. Alshammari, T.; Alshammari, N.; Sedky, M.; Howard, C. Evaluating Machine Learning Techniques for Activity Classification in Smart Home Environments. *Int. J. Comput. Electr. Autom. Control Inf. Eng.* **2018**, *12*, 48–54.
3. Alemdar, H.; Ertan, H.; Incel, O.D.; Ersoy, C. ARAS human activity datasets in multiple homes with multiple residents. In Proceedings of the 2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops, Venice, Italy, 5–8 May 2013; pp. 232–235.
4. WSU CASAS Datasets. Available online: http://ailab.wsu.edu/casas/datasets/ (accessed on 31 March 2018).
5. PlaceLab Datasets. Available online: http://web.mit.edu/cron/group/house_n/data/PlaceLab/PlaceLab.htm (accessed on 31 March 2018).
6. Lee, J.W.; Cho, S.; Liu, S.; Cho, K.; Helal, S. Persim 3D: Context-Driven Simulation and Modeling of Human Activities in Smart Spaces. *IEEE Trans. Autom. Sci. Eng.* **2015**, *12*, 1243–1256.
7. Kormányos, B.; Pataki, B. Multilevel simulation of daily activities: Why and how? In Proceedings of the 2013 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), Milan, Italy, 15–17 July 2013; pp. 1–6.
8. Bouchard, K.; Ajroud, A.; Bouchard, B.; Bouzouane, A. SIMACT: A 3D Open Source Smart Home Simulator for Activity Recognition. In *Advances in Computer Science and Information Technology*; Kim, T.H., Adeli, H., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 524–533.
9. Synnott, J.; Chen, L.; Nugent, C.; Moore, G. The creation of simulated activity datasets using a graphical intelligent environment simulation tool. In Proceedings of the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 26–30 August 2014; pp. 4143–4146.
10. Ariani, A.; Redmond, S.J.; Chang, D.; Lovell, N.H. Simulation of a Smart Home Environment. In Proceedings of the 2013 3rd International Conference on Instrumentation, Communications, Information Technology and Biomedical Engineering (ICICI-BME), Bandung, Indonesia, 7–8 November 2013; pp. 27–32.
11. Fu, Q.; Li, P.; Chen, C.; Qi, L.; Lu, Y.; Yu, C. A configurable context-aware simulator for smart home systems. In Proceedings of the 6th International Conference on Pervasive Computing and Applications (ICPCA), Port Elizabeth, South Africa, 26–28 October 2011; pp. 39–44.
12. Cook, D.J.; Crandall, A.S.; Thomas, B.L.; Krishnan, N.C. CASAS: A smart home in a box. *Computer* **2013**, *46*, 62–69.
13. Skubic, M.; Alexander, G.; Popescu, M.; Rantz, M.; Keller, J. A smart home application to eldercare: Current status and lessons learned. *Technol. Health Care* **2009**, *17*, 183–201.
14. Anguita, D.; Ghio, A.; Oneto, L.; Parra, X.; Reyes-Ortiz, J.L. A Public Domain Dataset for Human Activity Recognition using Smartphones. In Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Bruges, Belgium, 24–26 April 2013.
15. Casale, P.; Pujol, O.; Radeva, P. Human activity recognition from accelerometer data using a wearable device. In *Iberian Conference on Pattern Recognition and Image Analysis*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 289–296.

16. Bruno, B.; Mastrogiovanni, F.; Sgorbissa, A.; Vernazza, T.; Zaccaria, R. Analysis of human behavior recognition algorithms based on acceleration data. In Proceedings of the 2013 IEEE International Conference on Robotics and Automation, Karlsruhe, Germany, 6–10 May 2013; pp. 1602–1607.

17. Van Kasteren, T.; Englebienne, G.; Kröse, B.J. Transferring knowledge of activity recognition across sensor networks. In *International Conference on Pervasive Computing*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 283–300.

18. Pirsiavash, H.; Ramanan, D. Detecting activities of daily living in first-person camera views. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 2847–2854.

19. Lago, P.; Lang, F.; Roncancio, C.; Jiménez-Guarín, C.; Mateescu, R.; Bonnefond, N. The ContextAct@A4H real-life dataset of daily-living activities. In *International and Interdisciplinary Conference on Modeling and Using Context*; Springer: Cham, Switzerland, 2017; pp. 175–188.

20. Synnott, J.; Nugent, C.; Jeffers, P. Simulation of Smart Home Activity Datasets. *Sensors* **2015**, *15*, 14162.

21. Cook, D.; Schmitter-Edgecombe, M.; Crandall, A.; Sanders, C.; Thomas, B. Collecting and disseminating smart home sensor data in the CASAS project. In Proceedings of the CHI Workshop on Developing Shared Home Behavior Datasets to Advance HCI and Ubiquitous Computing Research, Boston, MA, USA, 4–9 April 2009; pp. 1–7.

22. Helal, S.; Lee, J.W.; Hossain, S.; Kim, E.; Hagras, H.; Cook, D. Persim-Simulator for human activities in pervasive spaces. In Proceedings of the 7th International Conference on Intelligent Environments (IE), Nottingham, UK, 25–28 July 2011; pp. 192–199.

23. Helal, A.; Cho, K.; Lee, W.; Sung, Y.; Lee, J.; Kim, E. 3D modeling and simulation of human activities in smart spaces. In Proceedings of the 2012 9th International Conference on Ubiquitous Intelligence and Computing and 9th International Conference on Autonomic and Trusted Computing, Fukuoka, Japan, 4–7 September 2012; pp. 112–119.

24. McDonald, H.; Nugent, C.; Hallberg, J.; Finlay, D.; Moore, G.; Synnes, K. The homeML suite: Shareable datasets for smart home environments. *Health Technol.* **2013**, *3*, 177–193.

25. Alshammari, N.; Alshammari, T.; Sedky, M.; Champion, J.; Bauer, C. OpenSHS: Open Smart Home Simulator. *Sensors* **2017**, *17*, 1003.