Conference on
Information Technology and
Data Science

Multi-Resident Location Detecting in Smart Home*

Anca Alexan, Alexandru Alexan, Ștefan Oniga

Department of Electric, Electronic and Computer Engineering,
Technical University of Cluj-Napoca, North University Center Baia Mare
anca.alexan@cunbm.utcluj.ro
alexanalexandru@gmail.com
stefan.oniga@cunbm.utcluj.ro

Proceedings of the 1st Conference on Information Technology and Data Science
Debrecen, Hungary, November 6–8, 2020
published at http://ceur-ws.org

CEUR-WS.org/Vol-2874/short3.pdf

Abstract

Our homes become smarter every day, as new devices become connected and provide internet of things functionality. Using sensors, our homes can detect our presence and even the activity executed. Multi-residence environments provide some difficulties in detecting the executed activity accurately, as it is often difficult to identify the residents and determine the correct activity. Our proposed system handles CASAS activity data and tries to improve the activity detection rate for multi-resident scenarios. The neural network implementation is done using Matlab.

Keywords: Smart home, CASAS, multi-resident tracking

1. Introduction

Smart homes equipped with activity detection sensors, which have a binary output, can be considered low-cost activity monitored locations. Although this type of activity detection system has a low-profile and is very affordable, it can successfully

Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

*This work was supported by the Intelligent Embedded Systems research laboratories at the Technical University of Cluj-Napoca, North University Centre of Baia Mare.

be used in multiple domains such as automation, health monitoring, security, and activity detection [8]. This system works very well if detecting simple actions or monitoring a single subject. However, if multiple residents live in the same smart home, correctly determining the activity type and pinpointing the correct subject can get very challenging [3, 6]. One option to improve the activity detection rate and accuracy is to use Artificial Intelligence to interpret the data gathered from the Smart House sensor networks. Since we are dealing with multiple resident scenarios, one of the most difficult roles of the artificial intelligence layer is to correctly match the sensor data to the appropriate resident [2] to correctly identify the performed activity. The amount of data and quality can greatly affect the chosen data analysis algorithm and predictions. Using machine learning algorithms (MLA) [7] can greatly increase the detection accuracy due to the multiple existing data processing functions [5]. Since MLA is a universal nonlinear tool for modeling and processing complex data [1], it is a great choice for complex activity recognition. In this article, we propose an activity detection and recognition system that is capable of operating in a multi-resident smart house. We propose to recognize complex activities and assign them to the correct resident.

2. Related Work

Smart environments are evolving continuously, and their purpose is to simplify everyday life by optimizing time, energy, and cost. Smart systems are formed by a series of devices that can communicate, transfer data, to optimize performance automatically without human intervention. The main components of this kind of system are sensors, processing networks, data analysis, and system monitoring [4]. One of the most recent innovations is activity detection in a multi-resident environment. Data processing and activity recognition from CASAS data-sets are shown [8]. One algorithm for solving this difficult recognition problem is sMRT, which handles multi-resident smart houses with binary output motion sensors. This algorithm is trying to learn the special-temporal sensor relation. To identify an input and output relation, a dynamic model was used to predict the state for each resident in the next timeframe, based on the current resident state. This dynamic model will become data input for the Bayes estimation process. The entire process is divided into two phases, phase one consists of a learning phase that constructs the dynamic model and a second phase that predicts the number of residents in the location.

3. Proposed System

3.1. Casas System Dataset

The CASAS system contains the following sensors types: PIR sensors for presence detection, temperature sensors, humidify sensors, light sensors, other misc. sensors that can vary. For testing the implemented location detecting system, the dataset used was one from the CASAS system, created at the Washington State University. The dataset has two users, and its name is Kyoto. The dataset is labeled with the activity type and the user that is executing it. The entire dataset is provided in a text format (.txt) and contains the activities in chronological order. Each record contains the date and time, the active sensor's name, the sensor's state (ON, OFF, OPEN, ABSENT). Besides this data that is obtained directly from the CASAS system, other labels are present with information regarding users and activities, as shown in Figure 1.

```
2008-11-10      14:28:20.615459 M01     ON 2 2
2008-11-10      14:28:21.00566  M17     ON 1 1
2008-11-10      14:28:22.025769 D07     OPEN 1 1
2008-11-10      14:28:22.23544  M19     OFF 1 1
2008-11-10      14:28:22.48164  M23     OFF 2 2
2008-11-10      14:28:22.81818  M21     OFF 2 2
2008-11-10      14:28:23.050209 M22     OFF 2 2
```

Figure 1. Kyoto dataset.

3.2. Activity Detection

A very important learning process aspect is data preprocessing. This can be implemented by using statistic data processing flows, numeric data scaling, primary component analysis, or selecting certain attributes that have a common property. All these preprocessing operations were applied to the Kyoto database provided by the WSU. From this dataset, Kyoto, details regarding four different activities were extracted. These activities were executed in three different scenarios with 2 different users. The dataset also provides the map for the used smart house, a floor plan containing each room, and a list of sensors alongside their placement on the map. Analyzing the dataset activity data, we have noticed that the succession of active sensors is different for the same activity. As we can observe in Figure 2, the number of active sensors is different according to the activity that is executed. One thing that we have in common for a certain activity is the same location where the activity takes place.

Since different activities share the same location, other characteristics were required to be extracted to aid the activity recognition process. The active sensors list alongside the required sensors for each activity were centralized. From this list, it's clear that for the same activity, the list of active sensors and the sensor's activation order were different. We can also notice that for each activity execution, the number of active sensors is different, and also the total activity execution time differs. From this centralized list, data was extracted and the active sensor enable probability was calculated for each activity. A hierarchy was created and the first three most probable enabled sensors for each activity were chosen. For example, for activity 1, the first three most probable sensors to be activated are M17, M15, and M16. For each characteristic an index was added. This index will be added together with the sensor rank from the previously generated hierarchy (top sensor 1, 2, or

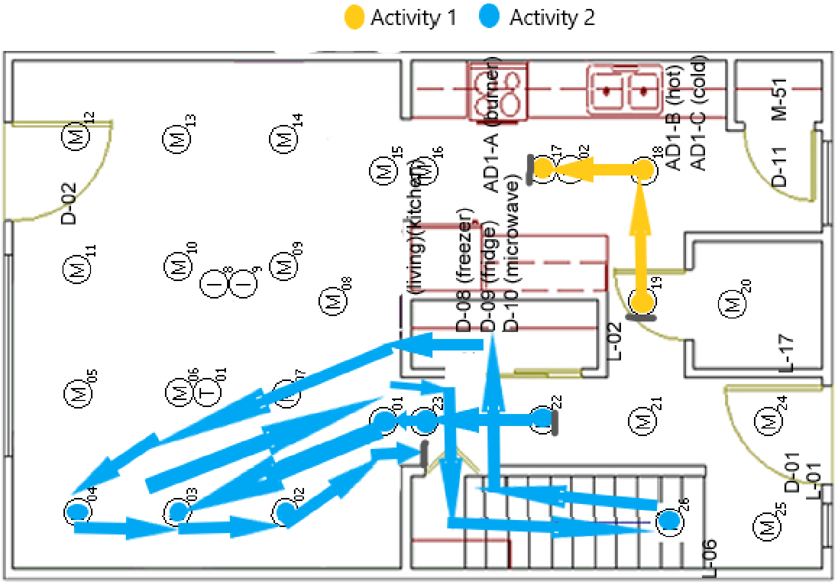

Figure 2. Activity active sensors.

3) to form a new characteristic that will be added for each dataset record(the text file). To further improve the activity recognition rate, the probability coefficient for a sensor to be enabled(ON state) reported to the total number of active sensors was calculated. Also, the probability coefficient for a sensor to be enabled(ON state) reported to the total number of active sensors for each user was calculated as well. All the generated coefficients were added in the dataset file, as shown in Table 1.

Table 1. Additional probability characteristics.

Time	Sensor	State	Person	Activity	PSA1	PSA2	PSA3	PPS1	PPS2	PS
14:28:17.986759	M22	ON	2	2	0	0	0	0	0.531914894	0.362318841
14:28:18.78605	M19	ON	1	1	0	0	0	1.136363636	0	0.362318841
14:28:19.551189	M23	ON	2	2	4	0	0	1.136363636	23.40425532	16.30434783
14:28:20.048559	M18	ON	1	1	0	0	0	2.272727273	0	0.724637681

Sensor data is sequential even though the activities are executed in parallel because the CASAS data recording precision is very high. In Figure 3 we can clearly notice the concurrent activities of the two users.

Each activity can be interpreted as an active sensor ordered list, a set. The smart house can be interpreted as a space S that is composed of multiple sub-spaces that represents the house's rooms. Each sub-space contains the list of sensors from the corresponding room, thus the house space is a reunion of all sub-spaces. An

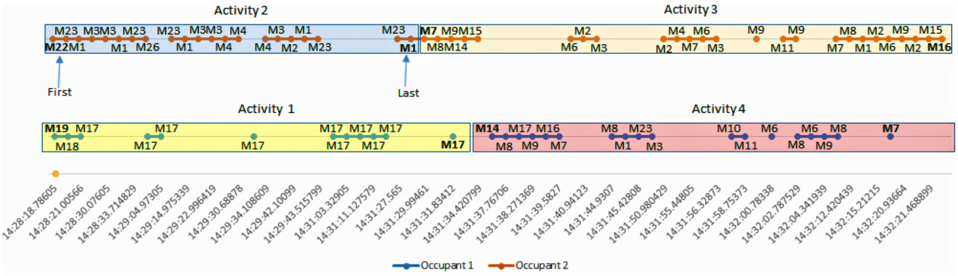

Figure 3. The two users concurrent activities diagram.

activity can be viewed as an intersection of certain sub-spaces. The data was imported in Matlab using the import tool, separating the input and output data. The columns that contain information on the sensors name an also the new characteristics that were computed in the preprocessing phase was used as input data (phase 2). The columns containing the person's information and the executed activity were used as output data and these were used as tags for the input data. The total number of records for the used dataset is 1100. For the first phase, phase 1, the data was used without the additional characteristics added. Using the Neural Net Fitting tool, a new neural network was created. The chosen to learn algorithm was Bayesian. The Bayesian reasoning offers a probabilistic approach for the inference. The decision logic takes into account the probabilistic distribution based on the interest quantities but also the observed training data. This approach is very important for the machine learning process because it provides a method of quantitative approximation that is necessary for justifying the alternative hypotheses evaluation.

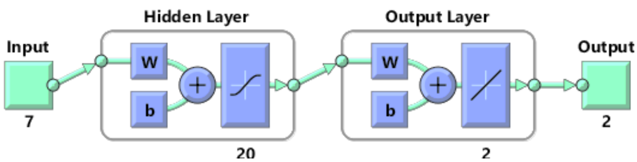

Figure 4. The neural network configuration.

As we can notice, the neural network chosen architecture uses 20 hidden layers, as is showed in Figure 4. The sampling type is Matrix rows. For the input data, each row contains 7 columns. The first column contains the active sensors name and it's followed by 3 columns displaying the hierarchy of the first 3 active sensors for each activity. The remaining columns contain the probability coefficient for a certain sensor to be enabled, ON state, reported to the total number of active sensors. As output data, each row contains two columns, one for the person that executes the activity and one for the activity that is being executed. For the learning process, 70% of total data were used. One part of the remaining data, 15%, was used for data validation and the remaining 15% were used for data verification. The algorithm learning process took 455 epochs, the best result being obtained in the 175 epoch with a value of 3.3183. The obtained result for this phase, the phase 1, was approximately 75 percent. For the next phase, the phase 2, the data used contained the additional computed characteristics. Using the same set-up, algorithm and hidden layer configuration, the recognition rate obtained was almost 88 percent.

4. Conclusions

This paper presents the implementation of a multi-resident activity detection system using data from the CASAS system. Since the sensors used are PIR based, having a binary output, activity detection in multiple resident locations is challenging due to the system's inability to identify residents. Our proposed method of increasing the neural network accuracy is computing additional characteristics based on the system behavior and adding them to the initial dataset data. The Matlab tool was used to create and train a neural network, comparing the results with and without the added computed characteristics. Thus, our proposed neural network and data preprocessing algorithm proved to significantly increase the activity recognition rate by 13 percent.

References

[1] A. Alexan, A. Alexan, S. Oniga, I. Pap: *Analysis of activity detection data pre-processing*, in: Oct. 2019, pp. 282–286,
DOI: 10.1109/SIITME47687.2019.8990804.

[2] P. Gupta, R. McClatchey, P. Caleb-Solly: *Tracking changes in user activity from unlabelled smart home sensor data using unsupervised learning methods* (Oct. 2020),
DOI: 10.1007/s00521-020-04737-6.

[3] M. Kanevski: *Advanced Mapping of Environmental Data: Introduction*, in: Feb. 2010, pp. 1–17, ISBN: 9781848210608,
DOI: 10.1002/9780470611463.ch1.

[4] *Machine learning for internet of things data analysis: a survey*, Digital Communications and Networks 4.3 (2018), pp. 161–175, ISSN: 2352-8648,
DOI: https://doi.org/10.1016/j.dcan.2017.10.002,
URL: http://www.sciencedirect.com/science/article/pii/S235286481730247X.

[5] T. Majoros, B. Ujvari, S. Oniga: *EEG data processing with neural network*, Carpathian Journal of Electronic and Computer Engineering 12 (Dec. 2019), pp. 33–36,
DOI: 10.2478/cjece-2019-0014.

[6] S. Ranasinghe, F. Al Machot, H. Mayr: *A Review on Applications of Activity Recognition Systems with Regard to Performance and Evaluation*, International Journal of Distributed Sensor Networks 12 (July 2016),
DOI: 10.1177/1550147716665520.

[7] J. Suto, S. Oniga, C. Lung, I. Orha: *Comparison of offline and real-time human activity recognition results using machine learning techniques*, Neural Computing and Applications 32 (Oct. 2020),
DOI: 10.1007/s00521-018-3437-x.

[8] T. Wang, D. J. Cook: *sMRT: Multi-Resident Tracking in Smart Homes with Sensor Vectorization*, IEEE Transactions on Pattern Analysis and Machine Intelligence (2020), pp. 1–1, doi: 10.1109/TPAMI.2020.2973571.