

Life Habits Modeling with Stochastic Timed Automata in Ambient Assisted Living

Kevin Fouquet, Gregory Faraut, Jean-Jacques Lesage

► **To cite this version:**

Kevin Fouquet, Gregory Faraut, Jean-Jacques Lesage. Life Habits Modeling with Stochastic Timed Automata in Ambient Assisted Living. 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Oct 2020, Toronto, Canada. hal-03039270

HAL Id: hal-03039270

<https://hal.archives-ouvertes.fr/hal-03039270>

Submitted on 3 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Life Habits Modeling with Stochastic Timed Automata in Ambient Assisted Living

Kevin Fouquet, Gregory Faraut and Jean-Jacques Lesage

Abstract—Recent improvements in connected tools and learning algorithms allow new opportunities in the field of Ambient Assisted Living (AAL). However, smart home inhabitant’s life habits are often required to obtain adequate results for energy management, security, Health at Home (HaH), and numerous other applications. In this paper, a model for life routines representation and algorithms for its generation is introduced. Study on the state of the art exposes that activity ordering and duration are key features of human behavior. Consequently, the presented approach focuses on a higher level of semantic by observing activities performed by the inhabitant rather than the sensor logs, which allow for better understanding of his comportment and universality of the model for multiple aims. Stochastic Time Automata (STA) is proposed as it adequately models activity ordering with probability associated to edges and activity duration through probability distribution associated to location delay. Presented approach does not require specific equipment besides sensors required for activity recognition and is versatile enough to be used in various applications. A case study highlights the relevancy of the chosen features and demonstrates that the proposed model is efficient to depict and understand inhabitants’ life habits.

Index Terms—Ambient Assisted Living, Machine Learning, Pattern Recognition, Discrete Event Systems

I. INTRODUCTION

Development of new connected tools such as wireless sensors, internet of things (IoT) and smart home coupled with development of new approaches in the field of machine learning offer new opportunities for improving quality of life using smart home technology. This field of research is often called Ambient Assisted Living (AAL) and clusters many fields such as home automation, activity of daily living or health at home (HaH).

Yet, numerous applications require precise information on the inhabitant’s habits to perform adequately. Knowing life habits allows for better energy management, medical home care [1]–[3], artificial database generation [4], and is often necessary to obtain convincing results in many AAL applications.

To fulfill this requirement, existing works such as [2]–[6] aim at extracting the life patterns of a smart home inhabitant by observing the sensor logs he generates during his everyday life. These contributions are often data-based and focus on a specific purpose which means they cannot be used for other projects that the one they were designed for. Moreover, most of them rely on sensor logs and offer a hard to read outlook of the inhabitant behavior.

*This work was not supported by any organization

Kevin Fouquet, Gregory Faraut and Jean-Jacques Lesage are with Université Paris-Saclay, ENS Paris-Saclay, LURPA, 94235 Cachan, France {first name.last name}@ens-paris-saclay.fr

In this paper, a model based approach for life habits representation is presented. This model aims at addressing most of AAL needs by proposing a versatile and easily readable model. To this end, it focuses on a higher semantic level thanks to activity recognition algorithms, and directly consider discovered activities rather than sensor events. Working on this level of semantic is relevant for many applications as that provides insightful and easily readable information about the inhabitant behavior. Any activity recognition process is suitable for the presented model, which signifies the approach proposed in this paper does not rely on specific technology and can be applied to most of already equipped smart home. Moreover, it does not necessitate additional equipment or information than the one needed by the chosen activity recognition algorithm. The features that the presented model has to depict are motivated by a state of the art of existing methods and needs in AAL.

Paper is structured as follows. Section 2 investigates existing solutions and their limits to define which behavioral indicators have to be deduced (extracted) from the model. Section 3 presents the global methodology and positioning of the model, as well as its definition, and section 4 introduces algorithms for its generation. Finally, a case study is proposed in section 5 including a discussion on the results and last section gives conclusions and perspectives.

II. STATE OF THE ART

As activity recognition has been more and more investigated in the past years, a few works are now considering the list of activities carried out by an inhabitant to model his life habits. In particular, most of the works using explicit model consider activity ordering and duration, as they are key features for behavior modeling.

Activity ordering is an interesting feature for home automation and HaH. It allows knowing which activity is the most likely to be executed in the near future, and to automatize some tasks accordingly. It is also a powerful tool for health evaluation, as routine alteration can be caused by new medical conditions or the evolution of already existing ones [5], [7].

In particular, [3] proposes a model to represent the usual way to carry out morning routine. The model uses operators representing activity order and dependency, including scenario choice and interruption. It also proposes a decomposition of activities into elementary tasks. Nevertheless, the model has to be built by an expert, which greatly reduces its scalability. It also makes this approach invalid when the life habits of the inhabitant are unknown, and can be used only to

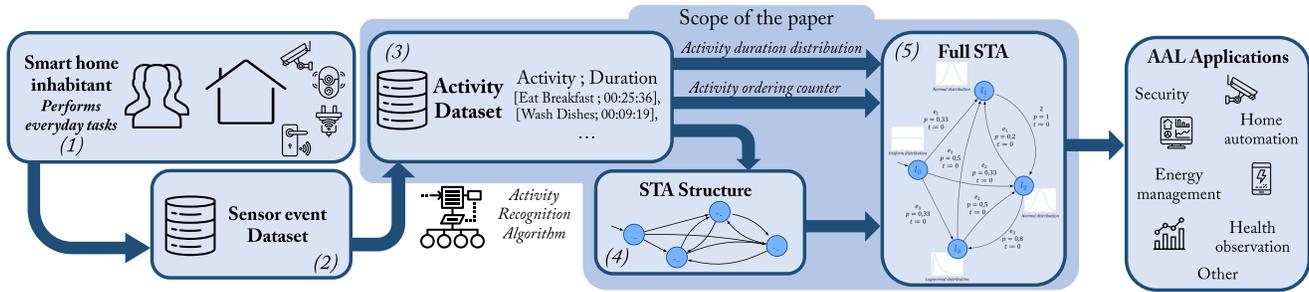


Fig. 1. Presented methodology and positioning within AAL applications

detect abnormal behavior. Other works such as [6] offer data mining approaches to build a model depicting life habits of a smart home inhabitant, but also require an expert to interpret it to know which activity is carried out. Many other works such as [2] considers Hidden Markov Model (HMM). The structure of these models is unobservable and is built only for the sake of probability distribution. Therefore they are harder to read than models with explicit structure which limits the ease of use for various applications. These approaches still testify the needs for a model able to express activity ordering for a convincing representation of life routines.

Additionally to activity ordering, activity duration gives insightful information about human behavior. In particular, such information is key features for energy management or security systems. Additionally, observing time needed for a person to carry out a given activity is a relevant information to evaluate mobility loss, loss of appetite and changes in sleep patterns which are common symptoms to numerous diseases [7].

Papers considering activity duration as an important feature to observe often consider probabilistic distribution for their approach, as it efficiently represents life habits trends despite the sometimes erratic human behavior. To model human activities, normal distribution is often used for its ease of use and readability. For instance, [2] proposes Gaussian distribution for activity time and duration modeling. Additionally, thresholds depending on the standard deviation are computed to detect outliers in order to evaluate the inhabitant behavior and provide medical staff with relevant information. [1] also uses normal distribution as an efficient tool to model activity time and duration and for artificial database generations. Yet, these works do not discuss on the relevancy of the normal distribution, and proposed methodologies are adapted only for a specific problematic. They can often not be adapted for other AAL related issues where other probabilistic distributions are more adapted. This is why this paper proposes a methodology which allows for other distribution more representative of reality and more suitable for various applications.

To the best of our knowledge, no method proposes to represent activity ordering and duration through the same model. Therefore, in this paper is proposed a model based approach for life habits representation focusing on activity duration and ordering without requiring expert knowledge.

III. PROPOSED APPROACH

A. Methodology description and positioning

As highlighted by the state of the art, activity ordering and duration are key features of routine behavior. In order to include this information in the model, observing the activities carried out by the inhabitant is necessary. Hopefully, activity recognition algorithms have been subject to many works in the past years.

Most of these works rely on observing logs from different types of sensors, generated by the inhabitant while he is carrying out some activities, such as [8]. In this paper, it is assumed the activities performed by the inhabitant are known thanks to such algorithms.

Proposed methodology and its positioning is depicted in figure 1. By performing his every day life (1), smart home inhabitant feeds a sensor event database (2) which is used by existing activity recognition algorithm to build an activity database (3). Sensors and algorithms used to achieve this task will not be considered. Identified activities are stored in the activity database in which each activity event is defined by a label (activity name) and duration (in seconds). This paper focus on the definition and generation of a model able to express life routines of the inhabitants of a smart home directly from this activity database. Next subsection III-B defines the model while subsection IV-A considers structure generation (4) and subsections IV-B and IV-C present the full model generation methodology (5). Applications of the proposed model can be manifold and include various AAL applications such as home automation, energy management or health at home applications.

B. Model definition

Human behavior can sometimes be very indeterministic. Even if most people have routine they follow, consciously or not, external disturbances may occasionally deviate the inhabitant behavior from its standard. Therefore, this paper assumes that the observed human behavior used to build the model is unsupervised as defined in [9]: an assumption is made that most of the observed behavior is consistent with life habits but some anomalies may appear. Consequently, proposed model must be robust to such anomalies. Moreover, life habits might be multi-class, i.e. a unique inhabitant may have several life habits. Therefore, the model must be able to represent all of them.

Taking into account these assumptions, probabilistic automaton is a powerful tool to model activity ordering as anomalies within data does not impact significantly probability if they are scarce. This kind of model is relevant for multi-class behavior modeling as well, as several paths may possess consequent associated probabilities depending on the automaton structure.

On the other hand, by observing existing labeled public databases such as CASAS database [10] or DOMUS database [11], [12], it can be highlighted that activity duration tends to follow specific probabilistic distributions. Some examples are displayed in figure 2. They are extracted from the CASAS database, and represents duration histograms of 4 different activities over a period of 30 days for home hh103.

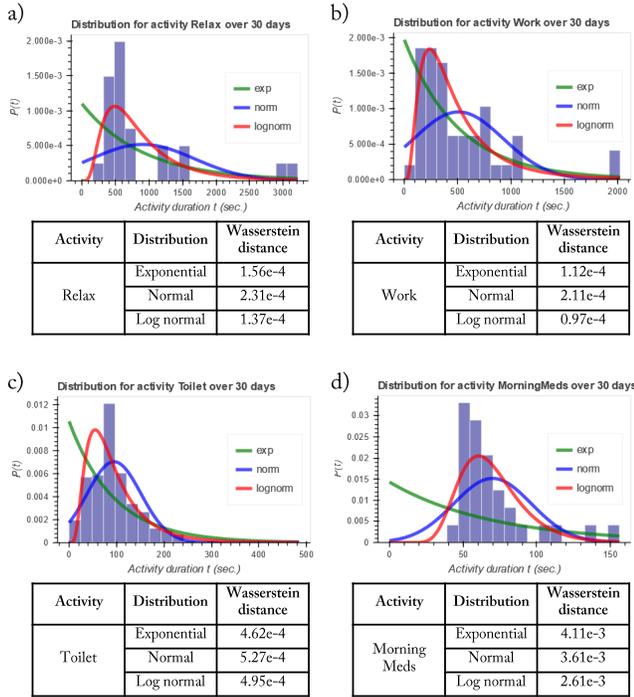


Fig. 2. Histograms of activity duration for some activities from CASAS hh103 database and associated Wasserstein distance

Some existing works already propose to model activity duration by normal distribution [1], [2] but they do not allow other distribution. Figure 2 evidences the need for more versatility. For instance, figure 2.b highlights that activity "Work" is better modeled by distribution exponential or log normal than normal, while activity "MorningMeds" on figure 2.d is better fitted by distribution log normal and normal. At this end, this paper proposes to model activity duration by any probabilistic distribution (such as exponential or log normal) depending on the activity observed.

Stochastic Timed Automaton (STA) [13] fulfill these requirements, as it is a probabilistic automaton with probabilistic distribution associated to location delay.

A Stochastic Timed Automaton is a 3-tuple $\langle A, \mu, P \rangle$ where A is the STA structure such as $A = \langle L, l_0, X, \Sigma, E \rangle$

with:

- L is a finite set of locations;
- $l_0 \in L$ is the initial location;
- X is a set of clocks;
- Σ is the alphabet;
- $E \subseteq L \times \sigma \times 2^X \times L$ is a set of edges composed of a departure location, an event, the list of clocks to reset and an arrival location.

Additionally, in compliance with definition from [14], the STA structure is said to be deterministic if from any location, no outgoing edges are labeled with the same event:

$$\forall l \in L, \forall \sigma \in \Sigma, |\{l' : (l, \sigma, l') \in E\}| \leq 1$$

μ is the probability measures over location delays $\mu = (\mu_s)_{s \in L \times R_+^X}$ and P is the probability defined on each edge $P = (p_e)_{e \in E} \in [0, 1]$. It should be noted that for each location, the sum of all probabilities from outgoing edges is always equal to 1.

$$\forall l \in L, \sum_{l' \in L, \sigma \in \Sigma} P(l, \sigma, l') = 1$$

An example of STA is given in figure 3. It possesses four locations, $L = \{l_0, l_1, l_2, l_3\}$ with l_0 the initial state, an alphabet $\Sigma = \{e_1, e_2, e_3\}$ and a unique clock t . A uniform distribution is associated to location l_0 delay, while normal distributions are associated to locations l_1 and l_2 and a log normal distribution to l_3 . Parameters of these distributions are not depicted for the sake of readability.

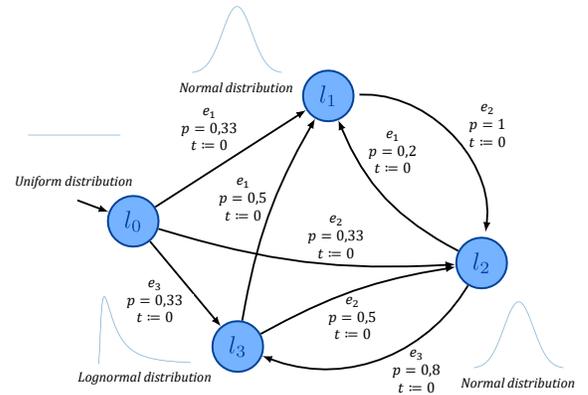


Fig. 3. Graphical representation of a STA

IV. MODEL GENERATION

In this section is proposed a methodology for life habits model generation. To this aim, structure of the model is generated first. Then edge probabilities and probabilistic distribution over location delay are computed from the activity database.

A. Structure generation

Generating STA structure can be seen as a similar problem of Probabilistic Finite Automaton (PFA) structure generation, which has already been explored [14]. Several paradigms

exist: A first possibility consists in generating the structure thanks to prior knowledge. Set of locations, events and edges are already known due to existing knowledge in the application domain. Second solution is based on learning approach. State-merging strategy is an example of learning method adapted to PFA. Last type of solution considers structure identification and probabilities altogether similarly to hidden models.

Hidden model structure is generated for the sake of probability generation, meaning it might not be possible to associate a given activity to each state. This is not consistent regarding the class of proposed model, as it will be necessary to associate a probabilistic distribution to each location. Moreover, learning approach requires important quantity of data to produce convincing results. As the inhabitant may carry out a specific activity only once per day, it requires to observe his behavior for many weeks or months before having satisfying amount of data, which is not always possible.

Using prior knowledge to produce model structure is the most adequate solution for life habits modeling. As the set of observed activities is already known thanks to existing activity recognition algorithms [8], [11], [12], structure building process is automatic and thus scalable. Moreover, structure produced by automatic approach can be deterministic by construction (in the sense of events), which significantly ease probability computation and further interpretation of the obtained model [14].

Proposed pseudo-code for structure generation is given in algorithm 1. It consists in creating a location for each activity, plus an initial location l_0 . Then, edges are created from every location to each one with adequate associated event.

Algorithm 1 STA Structure generation

Require: List of observed activities $l_{Activities}$

Ensure: STA Structure $STA_{Structure}$

Global initialization

1: $L = \{l_0\}, X = \{X_{STA}\}, \Sigma = \emptyset, E = \emptyset$

Locations creation

2: **for** $Activity \in l_{Activities}$ **do**

3: append $Activity$ to L

4: **end for**

Edges creation

5: **for** $e_{Activity} \in l_{Activities}$ **do**

6: **for** $location \in L$ **do**

7: append $e_{Activity}$ to Σ

8: $tr = (location, e_{Activity}, X_{STA} := 0, Activity)$

9: append tr to Σ

10: **end for**

11: **end for**

B. Probability computation

To model inhabitant life habits regarding activity ordering, probability defined on each edge P has to be computed. The objective is to estimate P so that the activity distribution expressed by the proposed model fit the most closely the real

distribution. To achieve this objective, counting the number of times the edge is fired produce the most satisfying result when the structure is given and deterministic [14], which is our case.

Consequently, considering:

- $P(l_i, e_j, -, l_j)$ probability associated to edge from location l_i to l_j annotated e_j (where l_j and e_j refer to the same activity);
- $N(e_i \rightarrow e_j)$ is the number of occurrences of the succession of activities $[e_i, e_j]$ in the training dataset;

Probability is defined as:

$$P(l_i, e_j, -, l_j) = \frac{N(e_i \rightarrow e_j)}{\sum_{l_k \in L} N(e_i \rightarrow e_k)}$$

Considering the initial location, as the very first activity in the dataset cannot be determined from the activity the user performed before, all edges from the initial location have equal associated probability $P(l_0, e_i, -, l_i) = \frac{1}{\|L\|-1}$, with $\|L\|$ number of locations.

C. Probabilistic distribution identification

Activity duration tends to follow specific probabilistic distribution. In particular, in [1], [2], normal law was used to model activity time and duration. Depending on the database used or the activity considered, other distribution might be better adapted, such as log-normal or exponential distribution.

Consequently, this paper proposes to model activity duration by probabilistic distribution associated to location delay. Each location possesses a stochastic distribution μ_s with its own parameters. Required parameters, for instance mean and standard deviation for normal distribution, are computed from the training database.

If a given location has unknown distribution, a uniform law can be used instead. In particular, this is the case for the initial location.

V. TEST ON REAL DATABASE

To highlight the relevance of the indicators and of the model chosen, a case study is presented. To this end, a public database available online allows to highlight the versatility and the scalability of our approach. DOMUS database [11], [12] is proposed for this case study as it focus on the morning routine of inhabitants which is often well defined. It also consider a limited amount of activities to observe making our model easy to read. Moreover, as it studies several participants, a different model can be generate for each of them and compared to conclude on the relevancy of the proposed approach.

A. DOMUS database description

DOMUS database [11], [12] proposes to observe the daily routines of 6 participants. Each experimenter is asked to carry his morning routine every day in the equipped smart home. As the experimenter is free to execute task as he desires, the obtain dataset is unsupervised: punctual anomaly may occur, but they are assume to be scarce. DOMUS smart

home is equipped with infrared movement detectors, pressure detectors, lamp light switches, door and switch contacts and flow meter. By triggering these sensors, the inhabitant generates sensor events labeled with the occurrence time, the name, the type and the localization of the sensor, and a value depending on the sensor's type. These events are then used by activity recognition algorithm as those presented in [11], [12] to generate the activity dataset. Smart home plan, sensors location and examples of sensor event and activity log are depicted in figure 4.

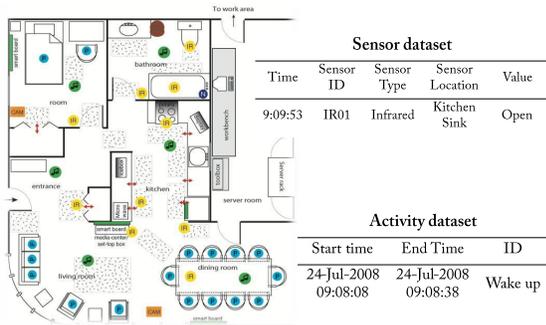


Fig. 4. DOMUS smart home plan [11], [12] with sensor and activity datasets

Observed activities are Waking up, Bathing, Preparing breakfast, Having breakfast, Washing dishes, and Other activities. Two series were carried out. In the first series, participants have to perform these 6 activities in whatever order they desire. In the second one, an additional activity (preparing tea) was imposed. Only first series of activity was used for model generation.

B. Model generation

Each user performs his morning routine 10 days. For each user, a model depicting his life habits is generated according to methods presented in section IV.

Concerning activity duration, normal distribution does not appear satisfying. Normality tests are not satisfying for most activities. Additionally, normal distribution is defined on R while activity duration is by definition defined on R^+ . It is generally not a problem if the probability of having a negative duration is low. If $\mu - 3 \times \sigma > 0$ with μ and σ respectively mean and standard deviation of the data, then it is considered as acceptable to use normal distribution even if the data to model are strictly positive. As depicted in figure 2, it is often not the case.

Log normal distribution is a powerful tool when considering duration of activities performed by a smart home inhabitant. Figure 2 depicts activity duration of various activities from public database. Wasserstein distance between histograms and most used probabilistic distribution tends to establish that log normal distribution fit the best most of activities duration. Additionally, log normal law is defined on R^+ and is still easily computable. Therefore, log normal distribution has been chosen to model activity duration in this case study.

Log-normal distribution can be modeled by the mean and standard deviation of the log of duration. In particular, if μ and σ are respectively estimations of the expectation and standard deviation of the log of duration:

$$\mu = \frac{1}{n} \sum_{i=1}^n d_i, \sigma = \frac{1}{n-1} \sum_{i=1}^n (d_i - \mu)^2 \text{ (Bessel's correction)}$$

Then e^μ the median and e^σ the shape parameters will be used as log-normal distribution parameters.

C. Results

Results are displayed in figure 5. Models have been generated using Python 3.6 with GraphViz library. Structure generation algorithm is $O(N^2)$ where N is the number of activities to observe (6 in this case study), probability and distribution computation are $O(n)$ where n is the number of activity events in the database. Median of activity duration e^μ is in seconds. These models depict what is the morning routine of each of the experimenters.

Concerning activity ordering, likelihood [14] is a powerful tool to compute probability for each sequence of event to occur and therefore know life habits pattern of the inhabitant. For user 1, the word of length 6 the most probable to occur is [Other activities, Wake up, Use toilet, Preparing breakfast, Having breakfast, Washing dishes] with a likelihood of 11.16%. His morning routine is very likely to correspond to this order. As stated in the state of the art section, authors in [6] also build a life habits model by observing sensor events and manually labeling activities with an expert. They applied their methodology to the same user 1 and obtained a similar morning routine, which tends to testify the correctness of the proposed approach. User 2 has the same pattern, but with a much lower likelihood (3.97%) and some other patterns such as [Wake up, Use toilet, Preparing breakfast, Having breakfast, Washing dishes, Use toilet] appear with a similar likelihood (2.20%). It means user 2 behavior is more diverse than user 1.

The models also allow for activity duration observation. In the presented example, User 1 activities related to breakfast are longer than User 2 (median time of 199s and 579s for preparing and having breakfast for user 1, while user 2 requires 118s and 311s with similar shape parameter), while other activities have roughly the same duration. It reflects different eating patterns and highlights the need for a specific model for each user instead of generic model.

VI. CONCLUSION

In this paper, a model for life habits representation has been presented. The model required to know which activities were performed during the observation phase. It is not dependent on the technology used for activity recognition and does not necessitate specific sensors or material except the one for activity recognition. Considering the needs express in the state of the art, the proposed model focus on activity ordering and activity duration. Stochastic Time Automata (STA) was introduced as it is a powerful tool to model activity ordering through probability associated to edges, and

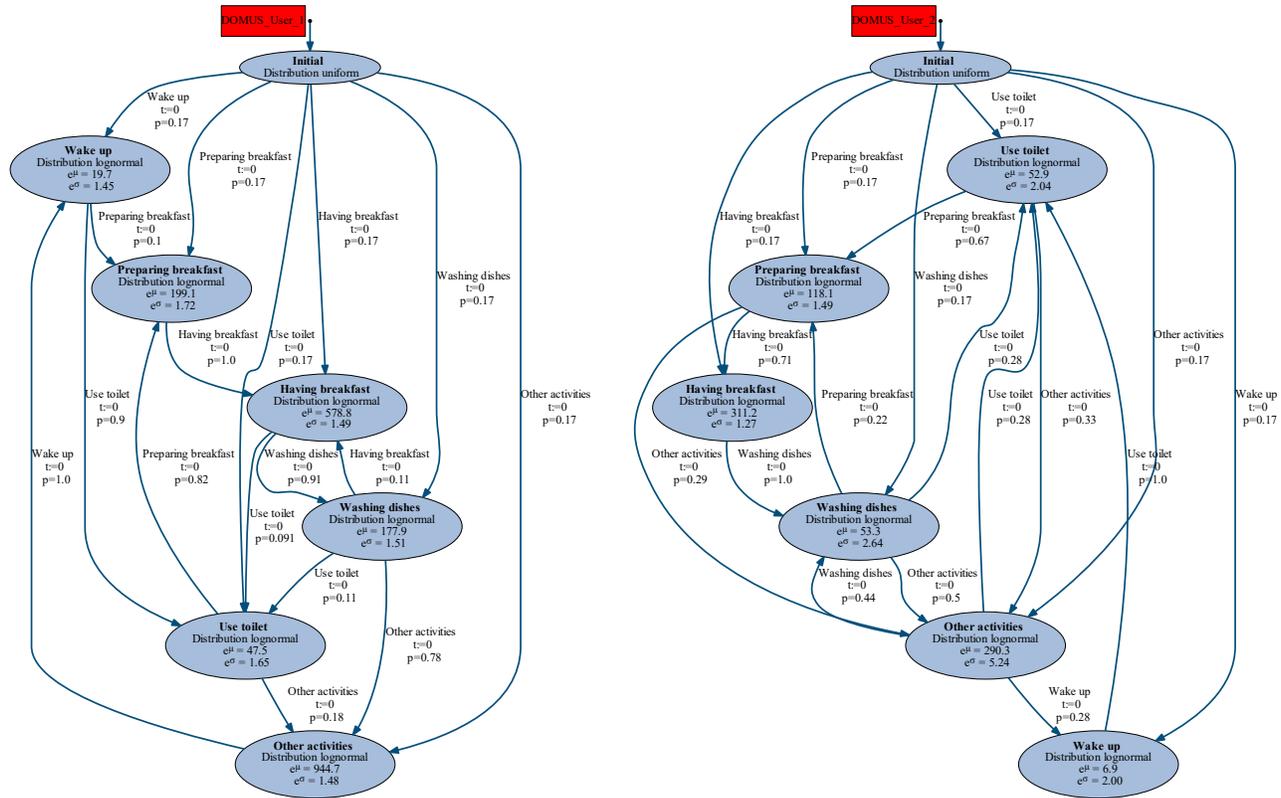


Fig. 5. Life habits model for User 1 and 2 from DOMUS database.

activity duration through probabilistic distribution associated to location delay. Case study highlighted the relevancy of the proposed approach as life habits appeared clearly and allowed to see the difference between several users.

This model allows numerous uses such as user identification, home automation or energy management in AAL fields and anomaly detection or behavior clustering for health at home. Further work could consider other behavior indicators such as activity time and number.

REFERENCES

- [1] C. Azefack, R. Phan, V. Augusto, G. Gardin, C. M. Coquard, R. Bouvier, and X. Xie, "An approach for behavioral drift detection in a smart home," in *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*. IEEE, aug 2019.
- [2] A. R. M. Forkan, I. Khalil, Z. Tari, S. Foufou, and A. Bouras, "A context-aware approach for long-term behavioural change detection and abnormality prediction in ambient assisted living," *Pattern Recognition*, vol. 48, no. 3, pp. 628–641, mar 2015.
- [3] P. Parvin, F. Paterno, and S. Chessa, "Anomaly detection in the elderly daily behavior," in *2018 14th International Conference on Intelligent Environments (IE)*. IEEE, jun 2018.
- [4] B. Ho, D. Vogts, and J. Wesson, "A smart home simulation tool to support the recognition of activities of daily living," in *Proceedings of the South African Institute of Computer Scientists and Information Technologists 2019 on ZZZ - SAICSIT '19*. ACM Press, 2019.
- [5] F. Kaddachi, H. Aloulou, B. Abdulrazak, P. Fraisse, and M. Mokhtari, "Long-term behavior change detection approach through objective technological observations toward better adaptation of services for elderly people," *Health and Technology*, vol. 8, no. 5, pp. 329–340, sep 2018.
- [6] J. Saives and G. Faraut, "Automated generation of models of activities of daily living," *IFAC Proceedings Volumes*, vol. 47, no. 2, pp. 13 – 20, 2014, 12th IFAC International Workshop on Discrete Event Systems (2014). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1474667015373742>
- [7] C. Chalmers, W. Hurst, M. Mackay, and P. Fergus, "Identifying behavioural changes for health monitoring applications using the advanced metering infrastructure," *Behaviour & Information Technology*, vol. 38, no. 11, pp. 1154–1166, feb 2019.
- [8] K. Viard, M. P. Fanti, G. Faraut, and J.-J. Lesage, "An event-based approach for discovering activities of daily living by hidden markov models," in *2016 15th International Conference on Ubiquitous Computing and Communications and 2016 International Symposium on Cyberspace and Security (IUCC-CSS)*. IEEE, dec 2016.
- [9] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection," *ACM Computing Surveys*, vol. 41, no. 3, pp. 1–58, jul 2009.
- [10] 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, jul 2013.
- [11] R. Kadouche, H. Pigot, B. Abdulrazaka, and S. Giroux, "Support vector machines for inhabitant identification in smart houses," in *Ubiquitous Intelligence and Computing*. Springer Berlin Heidelberg, 2010, pp. 83–95.
- [12] B. Chikhaoui, B. Chikhaoui, S. Wang, S. Wang, H. Pigot, and H. Pigot, "A new algorithm based on sequential pattern mining for person identification in ubiquitous environments," *Knowledge Discovery from Sensor Data (SensorKDD'10)*, p. 19, 2010.
- [13] N. Bertrand, P. Bouyer, T. Brihaye, G. Menet, C. Baier, M. Groesser, and M. Jurdzinski, "Stochastic timed automata," *Logical Methods in Computer Science*, vol. 10, no. 4, dec 2014.
- [14] E. Vidal, F. Thollard, C. de la Higuera, F. Casacuberta, and R. C. Carrasco, "Probabilistic finite-state machines – part i," *IEEE trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 7, p. 2005, 2005.