



# Contextualized behavior patterns for change reasoning in Ambient Assisted Living: A formal model

Paula Lago, Claudia Jiménez-Guarín, Claudia Roncancio

## ► To cite this version:

Paula Lago, Claudia Jiménez-Guarín, Claudia Roncancio. Contextualized behavior patterns for change reasoning in Ambient Assisted Living: A formal model. *Expert Systems*, 2017, 34 (2), 10.1111/exsy.12189 . hal-01659743

**HAL Id: hal-01659743**

**<https://hal.science/hal-01659743>**

Submitted on 23 Jan 2018

**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.

# Contextualized behavior patterns for change reasoning in Ambient Assisted Living: a formal model

Paula Lago<sup>1</sup>, Claudia Jiménez-Guarín<sup>1</sup>, and Claudia Roncancio<sup>2</sup>

<sup>1</sup>*Systems and Computing Engineering Department, School of Engineering, Universidad de los Andes, Colombia*

*Email: pa.lago52@uniandes.edu.co, cjimenez@uniandes.edu.co*

<sup>2</sup>*Univ. Grenoble Alpes, LIG, France*

*Email: Claudia.Roncancio@imag.fr*

**Abstract:** Ambient Assisted Living, smart environments and technology used for elder care, increases independent living time and cuts long-term care costs. An important requirement for these systems is detecting and informing about *abnormal* behavior in users' routines. Still, routines may change when certain context conditions hold, which does not mean these changes should concern caregivers. However, changes due to context conditions are not often considered in current proposals. In this paper, we formalize the main concepts related to activities in AAL and we introduce contextualized behavior patterns, a long-term behavior model that considers variability due to context. We adopt a semantic similarity that allows to better detect behavior changes and to understand what makes an observation of daily living different from *expected* patterns. The results of the experience with three public data sets are promising.

**Keywords:** behavior analysis · behavioral patterns · activities of daily living · elder care · context awareness · long term behavior modeling

## 1 Introduction

The aging process of the last decades is unprecedented, pervasive and enduring (DESA. Population Division. UN, 2001). With an increasingly older population, medical costs and demands for health services and for long-term care will also increase. Long-term care refers to assistance with everyday life. Ambient Assisted Living (AAL) systems emerge as an alternative and as a complement to traditional assisted living. They contribute to reduce costs of care at home, they assist caregivers and, globally, they help to enable aging at home.

One common scenario for AAL systems is to detect behavior changes. Commonly, behavior changes are regarded as warning signs of health deterioration, dementia or elder abuse. Being able to properly identify such changes (abrupt or subtle) is an important research direction. Some works have proposed to detect these changes by modeling behavior with logic models (propositions, ontologies) or probabilistic models (clusters, bayesian networks, statistical distributions, etc). Probabilistic models can be learned from data and are better at handling the inherent randomness of human behavior (Monekosso and Remagnino, 2010). On the other hand, logic-based models are better for human-understanding and thus, for explaining why an observation is considered as an anomaly. This *intelligibility* increases adoption and trust in the underlying system (Dey, 2009). To take advantage of the strengths of both kind of models, some authors use hybrid models (Chen et al., 2014; Díaz-Rodríguez et al., 2014) that combine logical modeling with probabilistic learning.

Providing accurate human behavior modelling and anal-

ysis is still challenging because it involves many activities performed in a personal manner. Each person has some kind of routines which may, in reality, be flexible. Such aspects motivate our work. Current models either do not consider routine variations or have them tightly coupled with activity recognition models. This hinders the understanding of personal frequent behaviors, called *behavior patterns* in this paper, and their possible variations due to context (i.e. weekend day, rainy weather). Current proposals either create different models, either introduce new activities for each variation, increasing the complexity of the recognition phase and reuse.

Previously, we demonstrated by experimentation (Lago et al., 2014) that behavior patterns change according to context features. In addition, we exposed the need of representing behavior patterns with its variations to differentiate real anomalies from normal changes. In this paper, we propose contextualized behavior patterns to close this gap. Our contribution is a model to represent personal behavior patterns independent from activity recognition models. Temporally and context-based variations are both explicitly modeled so that pattern understanding and anomaly explaining are possible. The reasoning behind this model is that activities have inherent characteristics that enable their detection through sensor events across multiple users and that personal variations (time, duration, variations due to context attributes) should be modeled in higher semantic level. Our model separates activity recognition from behavior modeling and understanding. This facilitates the reuse of activity models<sup>1</sup> and simplifies behavior models by modeling only personal characteristics at higher

---

<sup>1</sup>Activity recognition methods are out of the scope of this paper.

semantic levels instead of at raw sensor data level.

To illustrate our proposal, consider Alice who usually takes breakfast right after waking up around 8:30am and then takes a shower. Some Sundays, however, she does not take breakfast home but with her friends after church, especially if it is sunny. This is a contextualized behavior pattern, with possible variation due to some context features (day of the week and weather). A system that learns patterns without context, would only learn the first pattern (breakfast around 8:30am) and variations due to context, that are not as frequent, will not be learned. Therefore, sunny Sundays could be misleadingly identified as an "anomaly". With the contextualized behavior patterns we propose, all this can be handled correctly and it is possible to provide explanations about observed behaviours which are not so usual (e.g. not going out in a snowy Sunday).

This paper is organized as follows. In Section 2, we discuss related work in behavior modeling and point out the interest of an appropriate context management in AAL. In Sections 3 and 4 we present our contribution. We first present a model facilitating the description of behavior patterns and validity context. And then introduce the use of a similarity measure allowing capturing semantic similarities to improve reasoning about the behaviors. In Section 5 we report the experimentation of our proposal on three real datasets. In Section 6 we present our conclusions and some research directions to learn contextualized behavior patterns.

b

## 2 Related work on activity and behavior modeling

We first present the basic concepts of behavior analysis in AAL research in Section 2.1. We argue that there is no formal and standard definition of two main concepts: activity and behavior. These restrains the comparison of models and results. Next, in Section 2.2 we review behavior models used for three different goals in AAL: activity discovery, activity recognition and anomaly detection. We finish this review with a discussion to present the needs addressed by our model (Section 2.2.4).

### 2.1 Basic concepts disagreement

In Ambient Assisted Living literature, we often find concepts such as action, activity, behavior and situation used with different meanings. The distinctions are made based on sensor capabilities (what is measurable by a sensor), temporal characteristics (short duration or long duration), context features considered and regularity of the pattern. Even though authors agree that the semantic level of abstraction each term conveys is different, there is no agreement on neither where the difference lies nor what the different levels are.

Some works need no concept distinctions for their goals and use the terms activity and behavior interchangeably. For example, for Rashidi et al. (2011) a recognizable pattern in sensor events is an activity. In contrast, for Monkosso and Remagnino (2010) these patterns are referred to as a behavior. This latter work, however, concludes that creating a hierarchy of models facilitates noise handling. This is one reason to create a hierarchy of semantic concepts.

Other works discern semantic levels, each one inferred by combining one or more elements in the previous level (Aztiria and Augusto, 2013; Chaaraoui et al., 2012). Sensor events are the most basic level and higher semantic levels are defined based on them. For example, an action is defined by a combination of sensor measurements and an activity by different actions. This implies that higher semantic levels have longer durations. A behavior may be a composition of activities with order restrictions. Even though the limits between each concept are not clear, these works provide a semantic differentiation of the terms.

Yet, most disagreements emerge when modeling personal routines. For Aztiria and Augusto (2013) a frequent behavior is a network of actions with context variations and constraints. Conversely, for Chaaraoui et al. (2012) a behavior is a recurring sequence of activities, but context features are not considered. In addition, new concepts come up to describe frequent activities and patterns. For example, Chen et al. (2014) introduce the concepts of regular activities and patterns.

There is no clear distinctions among terms. In this work, activities describe quite general characteristics whereas behaviors involve personal routines. This distinction is based on the perception that activities have recognizable traits despite who is performing them. When someone is cooking, it is possible to know it regardless of the personal nuances of cooking preferences. The main advantage of this distinction is to separate activity recognition models from behavior pattern models. This facilitates the reuse of activity models and simplifies behavior models that describe personal patterns.

### 2.2 Behavior models

Requirements for modeling behavior patterns depend on the objective of the model. Although our goal is not activity recognition nor discovery it is important to discuss the behavior models used for these goals along with those used for anomaly detection.

#### 2.2.1 Behavior models for Activity Discovery

Activity discovery refers to finding "activity routines that are temporally coarse and recurring" (Seiter et al., 2014). For works dealing with activity recognition, an activity is a frequent pattern in sensor data and no interpretation is needed. For this, the term behavior is sometimes used to refer to the same patterns.

To discover activities to track in smart homes, unsupervised learning techniques such as topic models (Rieping et al., 2014; Huynh et al., 2008; Seiter et al., 2014) and clustering methods (Rashidi et al., 2011) are used. Sensor information is encoded in a vector which sometimes includes context information. Usually time and location are incorporated. A behavior is modeled with either a cluster or the probability distribution of clusters in topics. One strength of topic models is that an activity vocabulary may be created with clustering or sequential pattern mining techniques before topic modeling, creating a hierarchical model. Nonetheless, adding context features may reduce the accuracy of the model, which is why some authors create different models to deal with context conditions (Rieping et al., 2014).

Activity discovery methods use distance measures to group similar vectors as a single activity. These measures are usually a value based distance (such as euclidean distance) which for some context variables may not be suitable. Some context values may be semantically similar but numerically different. For example, the living room and the bedroom may be considered similar for a person that takes naps in both locations.

Finally, behavior models used for activity discovery are not intelligible because it is not an important requirement for their goal.

### 2.2.2 Behavior models for activity recognition

Activity recognition refers to inferring activity labels from sensor observations (Kim et al., 2011). Two main approaches are used: specification-based and data-based (Chen et al., 2012a).

Specification based methods, such as ontologies, describe properties of the activities (for example, start time, duration, location, object interactions made among others) in a formal manner. Specification-based models are naturally intelligible. They support personalization by instantiation of the ontology for each user. This can be done either using prior knowledge of his habits (Chen et al., 2012b) or using a learning algorithm (Chen et al., 2014). Most of these models do not support uncertainty and variability, but Díaz-Rodríguez et al. (2014) use a fuzzy ontology for this.

On the other hand, data-based models learn behavior patterns from observations collected from sensors. The most common models are Hidden Markov Models (Ordóñez et al., 2014), Conditional Random Fields (Nazerfard et al., 2010) and Bayesian Networks (Nazerfard and Cook, 2012). These models represent sequences of activities with uncertainty. They can learn personal patterns if the data is available. Their main drawback is their need of labeled data, which is often difficult to obtain. For this reason, they mainly use crowd learned models for application at different homes and do not provide insight in personal patterns. In addition, these models are black-box models that are not intelligible.

The main difference of the models used for activity recognition with our work is that they have personal patterns tightly coupled with activity models to increase accuracy of recognition.

### 2.2.3 Behavior models for anomaly detection

Anomaly detection means identifying observations that do not conform to expected behaviors (Kandhari et al., 2009). In AAL, an abnormal behavior can signal health or well-being deterioration and unsafe situations. Therefore, identifying anomalies has been a topic of research inside the community. Anomalies can be detected with statistical methods, distance based methods or rule based methods. In rule based methods anomalies are defined by a caregiver or a health expert. This is the most common method in commercial applications (Canary Care Limited, 2015; Lively Inc, 2014). Their main drawback is that it is difficult to consider variability and all cases and to adapt to changes. Both statistical and distance based methods are based on the principle that most observations constitute normal behavior and that anomalies are deviations from frequent patterns.

In statistical methods (Monekosso and Remagnino, 2010; Moshtaghi et al., 2015; Ordóñez et al., 2015; Forkan et al., 2014) anomalies are highly unlikely observations. Observations can be either sensor events or higher semantic interpretations. Monekosso and Remagnino (2010) use a Hidden Markov Model to model sequences of sensor events, and although they don't use context features as part of the model they recognize that creating different models for different day periods increases the precision of anomaly detection. Moshtaghi et al. (2015) use different models to consider both spatial and temporal features but their observations are just inactivity periods. Using long-tail statistical distributions (Pareto,  $\alpha$ ) they recognize an anomaly as a highly unlikely inactivity period for the spatio-temporal segment. Ordóñez et al. (2015) consider also sensor events as observations. They use model for sensor activation, sequence and sensor duration likelihood. They also recognize that considering periodic variations (routines for weekdays and weekends) can improve the precision of anomaly detection.

Forkan et al. (2014) model high level interpretation of sensor events as activities and use three different models to represent various features of behavior: sequences are modeled with a HMM, temporal features (start time and duration) are modeled with a normal distribution as are physiological features. Anomaly detection is made by first analyzing each variable separately and then combining information with fuzzy rules. We find two main differences with our model. First, to represent different occurrences of the same activity in a day they create different activity interpretations (breakfast, lunch, dinner to represent different occurrences of eating). This increases the complexity of the activity recognition model. Second, they do not consider periodic variations of behavior or context

features such as weather.

In distance based methods (Aztiria et al., 2013; Lundström et al., 2016; Park et al., 2010; Yuan and Herbert, 2014), anomalies are observations that are very different from learned behaviors based on a distance or similarity measure. Park et al. (2010) use low-level sensor events as observations and consider sequences of events as behaviors. Their similarity measure is based on sequence distance, start time and duration of the sequence. They do not consider any other context features or possible variations in the behavior. The behavior model in Aztiria and Augusto (2013) considers a network of sensor events (not high level interpretations) and possible variations due to different context conditions. However, their similarity measure (Aztiria et al., 2013) only considers the sequential elements and not the similarity between context conditions. Lundström et al. (2016) considers interpretations (by clustering sensor events) but do not consider sequential information. They consider temporal and spatial characteristics but use value-based similarity to detect anomalies. They do not consider periodic variations (day of the week, month) or other contextual features. Yuan and Herbert (2014) use high level interpretations of sensor events and different context conditions (environmental, physiological, day of the week, location, time, duration) but they do not have a model normal behavior. Instead they save all past observations and use a case based reasoning system to classify them into anomalies or not. This implies that initially, some cases must have been classified previously which is usually very difficult. It also implies that behavior patterns are unknown and thus, no understanding is provided of neither patterns or why an observation was considered abnormal.

#### 2.2.4 Summary and discussion

One important need when using behavior models for anomaly detection and notification is that they should be intelligible. The behavior model should allow to explain why there is an anomaly. This is helpful for caregivers in decision taking and increases adoption of AAL systems. Intelligibility can be provided by logic-based models (logic propositions, ontologies, calculus). Nonetheless, to be personal and adaptable, models should be combined with an efficient learning method. This is why hybrid models have been used lately.

Working with high semantic interpretations (activities) instead of raw sensor data or low level information has two main advantages. First, noise can be handled in event processing logic. Second, "intrinsic" properties of an activity (object interactions, for example) can be modeled apart from personal manners. This simplifies both the activity recognition and the behavior patterns models. Additionally, activity recognition models can be reused in different homes and behavior patterns can enable understanding of routines.

This work focuses on single-inhabitant scenarios. Con-

sidering that many elders live alone, this is a common scenario in AAL. The proposed model can be extended to multi-inhabitant scenarios as far as the person doing an activity can be identified. This requires the sensing infrastructure and the recognition process to handle it.

Modeling human behavior is complex. There are many sources of variability. Temporal variability in the activities plays an important role because an activity is done approximately at the same time or with a periodic pattern but not necessarily with a strict schedule. Context variations are also meaningful because human behavior may change depending on the conditions where he evolves (i.e. a person doesn't perform the same activities when alone as when there are visitors). Current models don't support these variations which would improve the quality of anomaly detection. They create different models, introduce new activities for each possible variation, or create false alarms. Our proposal allows to represent and to analyse such context data and variations and facilitates the reuse of the knowledge.

Finally, reasoning about anomalies may benefit from the semantics of the data. We propose to use a similarity function which takes advantage of such semantics to better analyse observed behaviors. Current models use value based distances which don't profit from the semantics.

In Table 1 we present a summary of the reviewed models for anomaly detection in AAL compared to our proposal.

	Monekosso and Remagnino (2010)	Moshtaghi et al. (2015)	Ordóñez et al. (2015)	Forkan et al. (2014)	Yuan and Herbert (2014)	Aztiria et al. (2013); Aztiria and Augusto (2013)	Park et al. (2010)	Lundström et al. (2016)	<i>Contextualized behavior patterns</i>
<b>An observation is...</b>	Sensor data	Inactivity period	Sensor data	Activity interpretations	Activity interpretations	Sensor data	Sensor data	Sensor data	Activity interpretations
<b>Context features considered</b>	Time of the day is considered by creating a different model for each period (morning, afternoon, evening)	Time of the day and location	Duration	Physiological data, start time and duration	Temporal, spatial, physiological and environmental data	Temporal and environmental data	Temporal, spatial, physiological and environmental data	Time of the day and spatial	Start time, duration, location, day of the week, weather. Can include others
<b>Is sequential order considered?</b>	Yes. With HMM	No	Yes. Sequential patterns	Yes. With HMM.	No. There is no model of normal behavior.	Yes. Network of events.	Yes. Sequential patterns (Episodes)	No. Uses clusters to represent behaviors	Yes, network of activities
<b>Method for detecting anomalies</b>	Statistical. An anomaly is a highly unlikely observation	Statistical. Distance from the mean inactivity period for the region and time is greater than a threshold.	Statistical. Distance from the mean duration is greater than a threshold. For sequential patterns, edit distance is calculated	Statistical. An anomaly is a highly unlikely observation in the sequence	Distance based between observations. An anomaly is a case that has not been seen before or has been rarely seen.	Distance based comparing observations to possible sequences in the behavior network	Distance based. Value based distance.	Distance based. An anomaly is an observation far from all clusters.	Distance based. Compares observations to patterns. An anomaly is a very distinct observations.
<b>Is context considered when detecting anomalies?</b>	Using a different model each time	Yes, by applying different threshold for each context	Yes. Considers significant deviations in duration.	Yes, detects significant deviations in each context feature and then combines information	Yes. Each context feature has a different weight	No. Just sequence with edit distance.	Yes. Value-based distance	Yes. Value-based distance	Yes, similarity based on semantic characteristics of each context attribute

Table 1: Review of behavior models used for anomaly detection in AAL

### 3 Representing contextualized behavior patterns

In this Section we introduce our model to represent contextualized behavior patterns. We first introduce, in Section 3.1, notions related to activities and behaviors and then present contextualized behavior patterns in Section 3.2.

#### 3.1 Activity and Behavior patterns

Uncertainty is an important factor when reasoning about human behavior patterns. To represent uncertain properties in a flexible way, our proposal relies on fuzzy intervals as defined in (Zadeh, 1965; Ohlbach, 2004) and recalled hereafter.

**Definition 1.** *Fuzzy Interval (Zadeh, 1965; Ohlbach, 2004)*

A fuzzy interval  $I$  is defined through a membership function  $fI(x)$  associating to each point  $x$  in a base set  $\mathbb{X} = \{x\}$  a real number in the interval  $[0, 1]$ . The values of  $fI(x)$  represent the "grade of membership" of  $x$  in  $I$ .

$$f : \mathbb{X} \rightarrow [0, 1]$$

Fuzzy intervals may be defined by a number of functions. Trapezoid functions, used in this work, can be specified algebraically by defining the four points  $(t_1, t_2, t_3, t_4)$  of the trapezoid, where the following always apply:

$$fI(x|x \leq t_1) = 0$$

$$fI(x|x \geq t_2 \leq x < t_3) = 1$$

$$fI(x|x \geq t_4) = 0$$

Fuzzy intervals will be used to capture uncertainty and variability, for example in the start time of an activity and its possible duration. They will also be used to represent ranges of other data values that have no strict definition. For example, "cold" temperatures can be considered as the temperatures in a fuzzy interval proper to each person.

**Example 1.** *Let's suppose Alice usually goes to bed between half past nine and half past ten at night. To introduce flexibility on her usual bedtime, we consider that she may break her schedule occasionally, by either advancing or delaying her bedtime at most half an hour. The fuzzy time interval (1), also showed by Figure 1, represents her bedtime routine.*

$$f(t) = \{21 : 00, 21 : 30, 22 : 30, 23 : 00\} \quad (1)$$

In the AAL context targeted by this work, interesting human behaviors involve mainly daily living activities such as sleeping, eating or cooking. Other more particular activities can be included as far as they can be recognized.

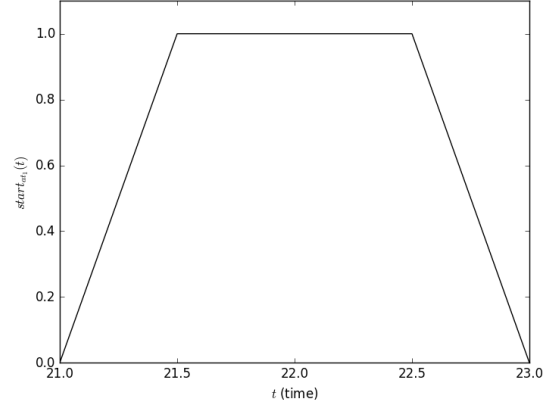


Figure 1: Fuzzy time interval  $f(t)$

**Definition 2.** *Activity*

An activity is an action performed by a person. Let  $\mathcal{A} = \{a_1, a_2, \dots, a_n\}$  be the set of activity labels.

**Example 2.** *The set of Alice's activities is  $\mathcal{A} = \{\text{SLEEP, SHOWER, EAT, COOK, GO\_OUT}\}$*

In practice, in our work, activities are semantic interpretation of sensor measures. Nevertheless, the model makes no assumption on the activity recognition method. Activities are usually performed at specific times and locations. Let  $\mathcal{L} = \{l_1, l_2, \dots, l_n\}$  be the set of locations at home.

**Definition 3.** *Activity Pattern*

An *activity pattern* represents when and where an activity usually occurs. It is defined as a tuple:

$$p = \{a_i, l, s_p(t), d_p(t)\} \quad (2)$$

where:

- $a_i \in \mathcal{A}$  is an activity label
- $l \subseteq \mathcal{L}$  is one or more usual locations where activity  $a_i$  takes place.
- $s_p(t)$  is a fuzzy time interval representing the usual start time of activity  $a_i$
- $d_p(t)$  is a fuzzy time interval representing the usual duration of activity  $a_i$

The temporal characteristics of an activity pattern are not deterministic but stochastic. There is variability in the start time of activities and their duration. Even though we do the same activity at approximately the same time every day, this time is not exact but rather a range. For this, we use fuzzy time intervals to represent start time and duration in the activity pattern. The width of the time interval depends on the person's routines. A person with strict routines will have narrow start intervals, but

a person with more relaxed routines has wider intervals. Several activity patterns may be defined for the same activity  $a_i$ , specifying different temporal characteristics for the activity.

**Example 3.** *Alice's bedtime routine takes place at her bedroom and she sleeps for 6 to 8 hours more or less. Alice also usually takes a one hour nap. This is at around 14:00 either in her bedroom or in the living room. These patterns represent typical occurrences of the activity sleep and they are specified in Eq. 3 and Eq. 4 respectively.*

$$sleep_1 = \{SLEEP, BEDROOM, s_{sleep_1}(t), d_{sleep_1}(t)\} \quad (3)$$

With

$$s_{sleep_1}(t) = \{21 : 00, 21 : 30, 22 : 30, 23 : 00\}$$

$s_{sleep_1}(t)$  is illustrated in Fig. 1.

$$d_{sleep_1}(t) = \{5, 6, 8, 9\}$$

Values for  $d_{sleep_1}(t)$  are in hours.

$$sleep_2 = \{SLEEP, \{BEDROOM, LIVING\}, s_{sleep_2}(t), d_{sleep_2}(t)\} \quad (4)$$

With

$$s_{sleep_2}(t) = \{13 : 45, 14 : 00, 14 : 15, 14 : 30\}$$

$$d_{sleep_2}(t) = \{0.8, 1, 1.2, 1.4\}$$

Notice that there is no need to introduce a new label (such as nap) to represent different occurrences of the same activity SLEEP. Activity patterns allow expressing typical times and locations for single activities. Let us now introduce *behavior patterns* which involve several activity patterns.

**Definition 4.** *Behavior pattern*

A *behavior pattern* expresses typical relationships among *activity patterns*. It defines order constraints on them and eventual temporal delays. A *behavior pattern* is a pair

$$b = (P, R) \quad (5)$$

Where

- $P$  is a set of activity patterns
- $R$  is a set of tuples relating activity patterns in  $P$

Each element  $r_i$  in  $R$  is a quadruple:

$$r_i = (p_j, p_k, r_{jk}, gap_{jk})$$

Where

- $p_j, p_k \in P$

- $r_{jk}$  is a relation between the time span of the activities referenced by  $p_j$ , and  $p_k$ . Allen's relations (Allen, 1983) are considered.
- $gap_{jk}$  is an optional parameter that can be used with the before relation. It allows defining a fuzzy time interval representing the time gap between the end of the activity of  $p_j$  and the beginning of the activity of  $p_k$ .

**Example 4.** *Let us consider that Alice takes a shower for a quarter to half an hour after waking up. To represent this behavior pattern, let us first define the activity pattern for the shower:*

$$shower_1 = \{SHOWER, BATHROOM, s_{shower_1}(t), \{0.2, 0.25, 0.5, 0.55\}\} \quad (6)$$

Equation 7 shows the behavior representing the sequence. No delay between the activities is provided.

$$(\{sleep_1, shower_1\}, \{(sleep_1, shower_1, \prec)\}) \quad (7)$$

A behavior pattern can specify a precedence and a gap between two occurrences of a same activity.

**Example 5.** *Let's consider Alice does HOUSEKEEPING daily. HOUSEKEEPING is done at home but with no specific start time.*

The activity pattern for housekeeping is defined in Eq. 8.

$$hk = \{HOUSEKEEPING, HOME, undefined, d_{hk}(t)\} \quad (8)$$

And the behavior is defined in Eq. 9:

$$(P = \{hk\}, R = \{(hk, hk, \prec, \{12, 18, 24, 30\})\}) \quad (9)$$

Having defined patterns of behavior, i.e., what usually happens, we now define observations, this means, what is happening.

**Definition 5.** *Event*

An *event* is an observation of a given property. It represents measures of environmental or physiological conditions or interactions of the user with the environment:

$$event = (n, m, l, s_t, f_t) \quad (10)$$

Where

- $n$  is the name of the observed property
- $m$  is the observed value for the property
- $l$  is the location where the measure was taken
- $s_t$  and  $f_t$  are timestamps defining a time interval



For raw sensor data,  $s_t = f_t$  provide the timestamp when the measure was taken whereas other events have a duration interval. For events concerning the recognition of an activity, the interval  $[s_t, f_t]$  is the time span of the **observed activity**,  $n = \text{ACTIVITY}$  and  $m$  is the label of the activity carried out ( $m \in \mathcal{A}$ ).

**Example 6.** The event in Eq. 11 represents a temperature measured at the kitchen.<sup>2</sup>

$$( \text{TEMPERATURE}, 32, \text{KITCHEN}, \\ 2013 - 02 - 25 12 : 15 : 00, 2013 - 02 - 25 12 : 15 : 00 ) \quad (11)$$

**Example 7.** The observation of Alice eating from 8:20am to 8:45am in the kitchen is represented by the following event:

$$( \text{ACTIVITY}, \text{EAT}, \text{KITCHEN}, 08 : 20 : 00, 08 : 45 : 00 ) \quad (12)$$

**Definition 6.** *Event Stream*

An *event stream* is an infinite sequence of events. When entering the stream, events get two additional attributes, a physical identifier (position order) and the current timestamp.

Streams can be separated according to the property name of the events.

$$\text{stream}(n_i) = \{e : \text{event} | e.n = n_i\}$$

The activity stream,  $\text{stream}(\text{ACTIVITY})$ , will be used to analyse human behavior. A sub-sequence of this stream is called a **behavior observation**.

### 3.2 Enriching behavior patterns with context data

Since context conditions may affect human behaviors, a better understanding and handling of context data is helpful in AAL. The model proposed in this Section allows to capture general context data such as weather, noise levels, visitor presence, or temporal precisions like day of the week, month or season. Context data will be combined with behavior patterns to better represent the reality.

**Definition 7.** *Context Attribute*

A context attribute is a property characterizing entities, the environment or situations. Let  $\mathcal{C} = \{c_1, c_2, \dots, c_n\}$  be the set of considered context attributes. An attribute  $c_i$  is a pair

$$c_i = [name, dom] \quad (13)$$

<sup>2</sup>For the sake of readability we omit date from the timestamps in the following examples

including the *name* of the attribute and its domain *dom*. The domain defines the set of possible values for the attribute. An IS-A relation can be defined for domains with particular semantics so as to introduce a taxonomy.

**Example 8.** One may introduce the domains ALLDAYS, WEEKENDDAY, WORKDAY with IS-A(WORKDAY, DAY) and IS-A(WEEKENDDAY, DAY).

**Definition 8.** *Context data record*

A *context data record* ( $cr$ ) represents the conditions of a particular context by providing values for a set of context attributes. It is a set of pairs:

$$cr = \{cr_1, cr_2 \dots, cr_n\} \\ \text{where} \\ cr_i = (k, v) \wedge \exists c \in \mathcal{C} : k = c.name \wedge v \subseteq c.dom \quad (14)$$

**Example 9.** A context data record representing rainy weekdays is represented by the set in Eq. 15

$$\{ \{ \text{DAY}, \text{WEEKDAY} \}, \{ \text{WEATHER\_DESC}, \text{RAINY} \} \} \quad (15)$$

**Definition 9.** *Contextualized behavior pattern*

A *contextualized behavior pattern* represents a *behavior pattern* that occurs *only* under certain context conditions given by a *context data record*. It is represented by the expression

$$b|cr \quad (16)$$

where  $b$  is a behavior pattern and  $cr$  is a context data record.

**Example 10.** Let us add the information that when it rains, Alice may take a nap after lunch (instead of going out).

The activity patterns are represented in Eq. 17.

$$\text{lunch} = \{ \text{EAT}, \text{DINING}, s_{\text{lunch}}(t), d_{\text{lunch}}(t) \} \\ \text{nap} = \{ \text{SLEEP}, \text{BEDROOM}, s_{\text{nap}}(t), d_{\text{nap}}(t) \} \quad (17)$$

The contextualized behavior is represented in Eq. 18.

$$(\{ \text{lunch}, \text{nap} \}, (\text{lunch}, \text{nap}, \prec, \text{undefined})) | \\ \{ \{ \text{WEATHER\_DESC}, \text{RAINY} \} \} \quad (18)$$

Let us now consider how observations are contextualized. The context conditions under which an *activity* or *behavior observation* took place are retrieved by querying the event streams.

**Definition 10.** *Context of an activity or behavior observation*

The context of an *activity or behavior observation* is the set of most relevant values of the context attributes in  $\mathcal{C}$  over the time interval when it took place. Each context attribute has a function to provide the most relevant value based on the set of values available during the time interval. For example the last seen value or the maximum value.

## 4 Similarity model

Having defined a model to represent contextualized behavior patterns and observations, we now define how to detect deviations from expected patterns. The comparison between them depends on both, the order of activities and, context data. As small differences are to be expected, we adopt similarity and not equality. In Section 4.1 we introduce a similarity measure for context attributes considering their semantics. Then in Section 4.2 we present the similarity measure to compare contextualized observations with contextualized behavior patterns.

### 4.1 Comparing context attribute values

The similarity between two values of a context attribute depends on the attribute they measure and its domain.

The context data record of an *activity or behavior observation* has a single value for each of the context attribute it uses. In contrast, the context data record of a contextualized behavior pattern can provide a set of values for a single attribute. These values are all considered as "valid" values.

**Definition 11.** *Similarity between context attribute values*

Let  $v_1$  be a value for a given context attribute and  $V_2$  a set of values for the same attribute. The similarity between  $v_1$  and  $V_2$  is defined as follows:

$$sim(v_1, V_2) = \begin{cases} 1 & \text{if } v_1 \in V_2 \\ semsim(v_1, V_2) & \text{otherwise} \end{cases} \quad (19)$$

Where  $semsim(v_1, V_2)$  is a customized function considering the semantics of the attribute and its domain. In the following we discuss three implementations.

#### *Comparing context attributes with fuzzy domains*

When fuzzy intervals are used to define a domain, the similarity of a measured value (which are crisp values) to the domain is done by using function  $fI(v)$ . A particular example are start times. As they are represented with fuzzy time intervals, the semantic similarity between a timestamp  $t_i$  of an observed event and a fuzzy time interval  $s(t)$  is defined by  $semsim(t_i, s(t)) = s(t_i)$ .

#### *Comparing locations*

Although we can compare locations based on their physical

distance, semantic properties of locations give better insight. In fact, the bathroom may be close to the bedroom but sleeping in a bathroom is not expected and should alert. Comparing locations based on the activities that are expected to occur in them offers better reasoning. Two locations where similar activities (for example, the main and auxiliary bedrooms) occur are similar even though they may be far apart.

Let  $\mathcal{A}(l_1)$  and  $\mathcal{A}(l_2)$  be the sets of activities that occur in location  $l_1$  and  $l_2$  respectively. The similarity between these two locations is calculated as the Jaccard similarity of both sets.

$$semsim(l_1, l_2) = \frac{\mathcal{A}(l_1) \cap \mathcal{A}(l_2)}{\mathcal{A}(l_1) \cup \mathcal{A}(l_2)} \quad (20)$$

The activities that may occur at each location depends on the inhabitant. For instance, let  $\mathcal{A}(\text{LIVING}) = \{\text{SLEEP, WATCH\_TV}\}$ ,  $\mathcal{A}(\text{BEDROOM}) = \{\text{SLEEP, DRESS}\}$  and  $\mathcal{A}(\text{STUDY\_ROOM}) = \{\text{WATCH\_TV, WORK}\}$

Comparing STUDY\_ROOM to the other locations, the function finds some similarity with the living but not at all with the bedroom.

$$semsim(\text{LIVING}, \text{STUDY\_ROOM}) = \frac{|\{\text{WATCH\_TV}\}|}{|\{\text{SLEEP, WORK, WATCH\_TV}\}|} \\ semsim(\text{LIVING}, \text{STUDY\_ROOM}) = 0.33 \quad (21)$$

$$semsim(\text{BEDROOM}, \text{STUDY\_ROOM}) = 0 \quad (22)$$

#### *Comparing domains with IS-A relationship*

When an IS-A relationship is defined between domains, it can be used to assess the similarity between values. The actual similarity function depends on the particular taxonomy created by the relation.

### 4.2 Comparing behavior patterns to behavior observations

Behavior patterns involve several activities. We first introduce the similarity between activity patterns and activity observations.

#### 4.2.1 Comparing activity patterns to activity observations

The objective is to find out if an observed activity "respects" a pattern defined (and expected) for this type of activity. Intuitively, an observed activity is similar to a pattern if its start time, duration and location are similar to the ones defined by the pattern. Since the pattern may define several possible location for the activity, the similarity function takes the most similar to the one of the observed location.

Let  $ap = \{ap.n, ap.l, ap.s(t), ap.d(t)\}$  be an activity pattern and  $ae = \{\text{ACTIVITY}, ae.label, ae.l, ae.start, ae.end\}$  be an activity observation.

$$\begin{aligned}
sim(ap, ae) = & 1_{ap.n=ae.label*} \\
& (w_1 * max_i(sim(ap.l_i, ae.l)) \\
& + w_2 * ap.s(ae.start) \\
& + w_3 * ap.d(ae.end - ae.start))
\end{aligned} \tag{23}$$

where

- $w_1 + w_2 + w_3 = 1$
- 1 is the indicator function.

The weights  $w_i$  can be equal or be adapted depending on the activity and user requirements. Changing their values allow to enforce the importance of one particular aspect, for example the location of the activity.

**Example 11.** Suppose an observation of Alice sleeping at her study room from 14:00 until 15:00 as shown in Eq. 24.

$$e_s = \{ \text{ACTIVITY, SLEEP, STUDY\_ROOM, 14 : 00, 15, 00} \} \tag{24}$$

It is compared to the activity pattern  $sleep_2$  (in Eq. 4) as shown in Eq. 25.

$$\begin{aligned}
sim(p, e_s) = & 1_{\text{SLEEP}=\text{SLEEP}*} \\
& (w_1 * max(sim(\text{BEDROOM, STUDY\_ROOM}), \\
& sim(\text{LIVING, STUDY\_ROOM})) \\
& + w_2 * s_{sleep_2}(14 : 00) \\
& + w_3 * d_{sleep_2}(1)) \\
sim(p, e_s) = & 1 * (\frac{1}{3} * max(0, 0.33) + \frac{1}{3} * 1 + \frac{1}{3} * 1) \\
sim(p, e_s) = & 0.77
\end{aligned} \tag{25}$$

This similarity value points out that the observed activity "matches" most aspects of the pattern but not all of them. An explanation tool may provide the information about the "unexpected" location where Alice slept.

Having introduced the similarity measure for activities, let's determine how to compare behavior patterns to behavior observations that involve one or more activities.

#### 4.2.2 Comparing behavior patterns to behavior observations

Consistent with its definition, a behavior observation  $bo$  is similar to a behavior pattern  $bp$  if the activities involved are similar to the activity patterns expressed in the pattern and are performed in the habitual order. Since a behavior pattern represents a partial order among the activities, there are various possible sequences ( $\mathcal{T}_{bp}$ ). The observation is compared to all possible sequences.

$$sim(bp, bo) = max_{(b \in \mathcal{T}_{bp})} \sum_i w_i * sim(b_i, bo_i) \tag{26}$$

with:

$$\bullet \sum_i w_i = 1$$

The weights  $w_i$  allow giving more importance to particular activities in a sequence.

#### 4.2.3 Comparing contextualized behaviors

We now introduce context information in the similarity calculation. We consider contextualized behavior patterns  $b|cr$  and a behavior observation  $bo$  together with the context data under which it was performed ( $cr_{bo}$ ). The similarity is measured by the product of the similarity between  $b$  and  $bo$  as defined previously, and the similarity of the context records as in Eq. 27.

$$sim(b|cr, bo|cr_{bo}) = sim(b, bo) * sim(cr, cr_{bo}) \tag{27}$$

The similarity between context records relies on the similarity of their attributes. When several values are defined for an attribute, the highest similarity of pairwise comparison between possible values is retained. Recall that a context record is a set of (name, value) pairs  $cr = \{(cr_1.name, cr_1.v), (cr_2.name, cr_2.v), \dots, (cr_n.name, cr_n.v)\}$

$$\begin{aligned}
sim(cr, cr_{bo}) = & \sum_{i,j} (1_{cr_i.name=cr_{bo_j}.name} \\
& w_r * max_{k,l} (sim(cr_i.v_k, cr_{bo_j}.v_l))
\end{aligned} \tag{28}$$

As in the case for activities, the weights  $w_r$  allows reinforcing the importance of particular attributes in the pattern.

Let us illustrate how two context data records are compared. Let

$$\begin{aligned}
cr_1 = & \{ \{ day, WEEKDAY \}, \\
& \{ NOISE, \{ low\_noise, medium\_noise \} \} \} \\
and \\
cr_2 = & \{ \{ day, MONDAY \}, \{ NOISE, \{ 22.5 \} \} \}
\end{aligned} \tag{29}$$

The similarity between them is calculated by comparing values of the same attribute. When there are several values for one attribute (as there is for noise), we take the maximum of the similarities of each possible pair. In this example, no weights are defined so we take  $w_1 = w_2 = 0.5$

$$\begin{aligned}
sim(cr_1, cr_2) = & 0.5 * max(sim(WEEKDAY, MONDAY) \\
& + 0.5 * max(sim(LOW\_NOISE, 22.5), \\
& sim(MEDIUM\_NOISE, 22.5))
\end{aligned} \tag{30}$$

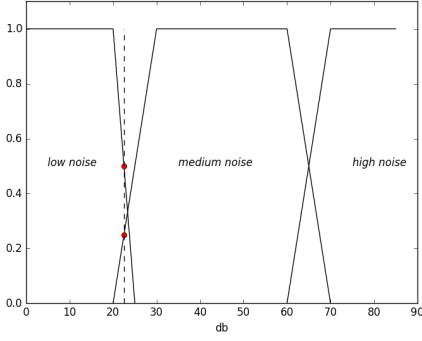


Figure 2: Fuzzy interval defining noise levels. A measurement of 22.5db (vertical line) belongs to both low and medium level with different degrees

Since  $\text{MONDAY} \in \text{WEEKDAY}$  and using noise intervals illustrated in Fig. 2:

$$\begin{aligned} \text{sim}(cr_1, cr_2) &= 0.5 * 1 + 0.5 * \max(0.5, 0.25) \\ \text{sim}(cr_1, cr_2) &= 0.5 * 1 + 0.5 * 0.5 \\ \text{sim}(cr_1, cr_2) &= 0.75 \end{aligned} \quad (31)$$

The similarity measure allows providing information on the unexpected characteristics of a behavior.

## 5 Experimental Validation

In this Section we present the experiments carried out to verify our proposal. We mined activity and behavior patterns from three datasets (Cook et al., 2013) (namely, HH120, HH122 and HH123) using an extension of the algorithm presented in (Lago et al., 2015). The algorithm is a frequent sequence mining algorithm that considers context attributes. It is based on prefix trees. Each node in the tree represent an activity sequence and has a matrix keeping the count of how many times a given sequence has occurred in a given context attribute value and time. A forest allows us to consider more attributes (one tree per context attribute). We refer the reader to the article for details of the algorithm, which is out of the scope of this work.

All datasets are annotated with the same activities, which makes them comparable, and offer a real world and long-term scenario to test our proposal. Each dataset has seventeen annotated activities. First, we present a summary of the dataset and the pre-processing undertaken (Section 5.1 and then present activity patterns found (Section 5.2), deviating observations (Section 5.3) and contextualized behavior patterns found the datasets (Section 5.4).

### 5.1 Dataset description and preprocessing

The datasets contain information about the daily life of people in three different apartments for a time span of 2

<i>Dataset</i>	<i># Days</i>	<i># Records</i>	<i># Locations</i>
HH120	63	2114	6
HH122	28	1103	8
HH123	28	999	6

Table 2: Summary of the datasets used for experimentation

```

Sleep,2,2012-01-28T00:04:14Z,2012-01-28T01:29:07Z
Bed_Toilet_Transition,1,2012-01-28T02:11:23Z,2012-01-28T02:14:41Z
Sleep,2,2012-01-28T02:14:58Z,2012-01-28T05:00:40Z
Bed_Toilet_Transition,1,2012-01-28T05:01:13Z,2012-01-28T05:03:56Z
Sleep,2,2012-01-28T05:04:06Z,2012-01-28T07:15:43Z
Toilet,1,2012-01-28T07:16:26Z,2012-01-28T07:19:20Z
Watch_TV,2,2012-01-28T07:19:40Z,2012-01-28T07:30:51Z
Dress,2,2012-01-28T07:30:53Z,2012-01-28T07:32:37Z
Watch_TV,2,2012-01-28T07:32:39Z,2012-01-28T07:37:57Z

```

Figure 3: Fragment of the activity stream used in validation

months, 1 month and 1 month. We only keep information about the activities performed, the location in which they were done and their start and end time. The location of each activity is taken as the room in which the most sensors were activated during the time the activity was performed. The sets  $\mathcal{L}$  for each one of these environments are shown in Eq. 32.

$$\begin{aligned} \mathcal{L}_{hh120} &= \{ \text{BATHROOM, BEDROOM, KITCHEN,} \\ &\quad \text{HALL, DINING} \} \\ \mathcal{L}_{hh122} &= \{ \text{BATHROOM1, BEDROOM, KITCHEN,} \\ &\quad \text{HALL, LIVING, ROOM2, BATHROOM2} \} \\ \mathcal{L}_{hh123} &= \{ \text{BATHROOM, BEDROOM, KITCHEN,} \\ &\quad \text{HALL, DINING} \} \end{aligned} \quad (32)$$

The transformed datasets are the activity streams for the experimentation, of which a fragment is shown in Figure 3.

Activity labels are coarse-grained definitions of the annotated activities in the dataset. For instance, instead of having COOK\_BREAKFAST and COOK\_LUNCH and COOK\_DINNER, we use the activity label COOK. The final set  $\mathcal{A}$  for this experiment is shown in Eq. 33

$$\begin{aligned} \mathcal{A} = \{ &\text{HYGIENE, RELAX, DRESS, BED\_TOILET, COOK,} \\ &\text{WATCH, TOILET, ENTER\_HOME, LEAVE\_HOME,} \\ &\text{READ, SLEEP, WASH\_DISHES, PHONE, WORK, EAT,} \\ &\text{TAKE\_MEDICINE, GROOM, BATH} \} \end{aligned} \quad (33)$$

Some examples from the dataset were removed to retain the same number of examples for each weekday. A summary of the number of records, locations and days in each of the final datasets is shown in Table 2. The total number of records per activity in each dataset is shown in Table 3.

Activity	# Records in HH120	# Records in HH122	# Records in HH123
Personal Hygiene	156	88	45
Relax	25	17	15
Dress	150	36	75
Bed Toilet Transition	115	62	23
Cook	57	54	82
Toilet	294	161	89
Watch TV	180	76	124
Enter home	203	70	92
Read	27	10	3
Leave Home	196	70	92
Sleep	229	104	79
Wash Dishes	43	65	93
Phone	5	4	39
Work	181	133	1
Eat	48	43	54
Take medicine	162	39	51
Groom	7	51	32
Bath	36	20	10

Table 3: Number of records per activity in each dataset

Context attributes (Eq. 34) used for these experiments correspond to the day of the week. Its domain defines the IS-A relationship as expressed in Example 8. We consider that two days in the same domain have similarity equal to 0.5.

$$\mathcal{C} = \{ (\text{DAY}, \{\text{MONDAY}, \text{TUESDAY}, \text{WEDNESDAY}, \text{THURSDAY}, \text{FRIDAY}, \text{SATURDAY}, \text{SUNDAY}\}) \} \quad (34)$$

Activity observations in the dataset are rewritten as events. An example is shown in Eq. 35. Activity observations are then rewritten to include context information as in Eq. 36.

$$e_{h_{120}} = \{ \text{ACTIVITY}, \text{LEAVE\_HOME}, \text{HALL}, \\ 2013 - 04 - 22T19 : 12 : 20, \\ 2013 - 04 - 22T19 : 12 : 29 \} \quad (35)$$

$$e_{h_{120}} = \{ \text{ACTIVITY}, \text{LEAVE\_HOME}, \text{HALL}, \\ 19 : 12 : 20, 19 : 12 : 29 \} \mid \{ \{ \text{DAY}, \text{MONDAY} \} \} \quad (36)$$

## 5.2 Mining activity patterns

For each dataset we mined activity patterns for every annotated activity. Frequent locations and days of the week

for each activity in the HH120 dataset are shown in Table 4 and the start time intervals are illustrated in Fig. 4. These results for the other two datasets are shown in Table 6, Table 7, Fig. 7 and Fig. 8 at the end of this article.

Eq. 37 is the formal representation of activity patterns for the LEAVE\_HOME activity in the HH120 dataset. Notice that these patterns are conditioned on the day of the week and that there are several patterns, each expressing a different start time interval (expressed in set form). The duration of each pattern is undefined.

$$\begin{aligned} leave_{hh120_1} &= \{ \text{LEAVE\_HOME}, \text{HALL}, \\ &\quad \{ 8 : 11, 8 : 15, 9 : 00, 9 : 04 \}, und \} \\ &\quad \mid \{ \text{DAY} \}, \{ \text{Tuesday}, \text{Wednesday}, \\ &\quad \text{Thursday}, \text{Saturday}, \text{Sunday} \} \} \\ leave_{hh120_2} &= \{ \text{LEAVE\_HOME}, \text{HALL}, \\ &\quad \{ 9 : 06, 9 : 15, 10 : 45, 10 : 54 \}, und \} \\ &\quad \mid \{ \text{DAY} \}, \{ \text{Tuesday}, \text{Wednesday}, \\ &\quad \text{Thursday}, \text{Saturday}, \text{Sunday} \} \} \\ leave_{hh120_3} &= \{ \text{LEAVE\_HOME}, \text{HALL}, \\ &\quad \{ 12 : 02, 12 : 15, 14 : 30, 14 : 43 \}, und \} \\ &\quad \mid \{ \text{DAY} \}, \{ \text{Tuesday}, \text{Wednesday}, \\ &\quad \text{Thursday}, \text{Saturday}, \text{Sunday} \} \} \\ leave_{hh120_4} &= \{ \text{LEAVE\_HOME}, \text{HALL}, \\ &\quad \{ 15 : 17, 15 : 30, 17 : 45, 17 : 58 \}, und \} \\ &\quad \mid \{ \text{DAY} \}, \{ \text{Tuesday}, \text{Wednesday}, \\ &\quad \text{Thursday}, \text{Saturday}, \text{Sunday} \} \} \end{aligned} \quad (37)$$

Activity patterns allow us to understand each inhabitant's personal routines. Each activity is described flexibly and we can observe differences among the inhabitants and variability for each activity. In terms of location, for example, the activity EAT in the first two datasets occurs in two different locations. However, for the third dataset it occurs in just one.

Start time intervals also have wide-ranging durations depending on the activity and the person. A larger interval means that the start time for the activity is less strict. On the other hand, a smaller interval means that the activity starts at a stricter time around the interval. People with less strict routines (or activities that follow less strict routines) will show larger intervals for the start times. For example, the intervals for the activity eat in the HH120 dataset are much stricter than the intervals for the same activity in the HH122 dataset. Similarly, the habits for the HH123 dataset are more irregular than those of the other two datasets in general. The ability to represent these intervals depending on the person and the activity instead of predefined time intervals makes the model flexible for a variety of users and habits.

<i>Activity</i>	<i>Location</i>	<i>Day of the week</i>
Personal Hygiene	Bathroom	All days
Relax	Bedroom	All days except Monday and Wednesday
Dress	Bedroom	All days
Bed Toilet Transition	Bathroom	All days
Cook	Kitchen	All days
Toilet	Bathroom	All days
Watch TV	Bedroom	All days except Tuesday and Sunday
Enter Home	Hall	All days
Read	Bedroom	Tuesday, Thursday, Friday, Saturday
Leave Home	Hall	All days
Sleep	Bedroom	All days
Wash Dishes	Kitchen	All days except Thursday
Phone	Bedroom	Tuesday, Thursday and Friday
Work	Dining	All days
Eat	Bedroom	All days
Take medicine	Kitchen	All days
Groom	Bathroom	All days except Tuesday and Sunday

Table 4: Activity patterns found in the HH120 dataset

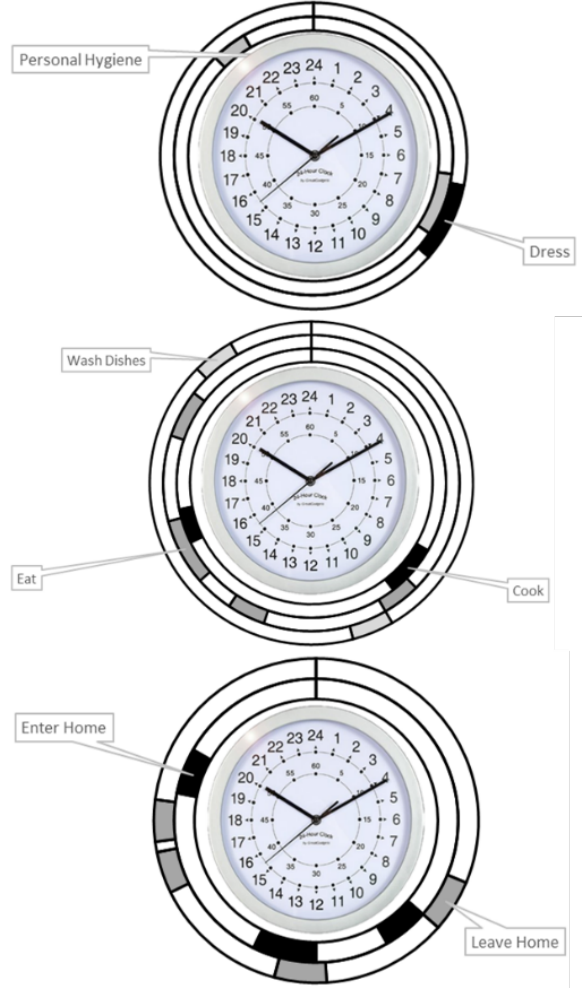


Figure 4: Frequent time intervals for the start time of different activities in the HH120 dataset

### 5.3 Assessing similarity of observations to patterns

Activity observations in each dataset were compared to the activity patterns using the similarity measure described in Section 4. Some observations from each dataset with similarity measures less than 0.5 are presented in Table 5. The unusual context attributes of each observation are underlined.

Observations with similarities close to 0.5 correspond to observations mostly made on a not so frequent but similar day to those of the pattern (also a weekday). In this experimentation, we gave each attribute the same relative importance so these changes do not have a strong impact. More context information could help better evaluate the deviation. Eating at home on Tuesdays (HH122), if Tuesdays is a day of usually eating out can be alarming. The person might have forgotten a compromise, or losing interest in social activity. Or it could be that it was just raining. Nonetheless, being able to point out the discrepancies can rise conversations from caregivers that can help evaluate the situation.

We would also like to remark the observation from dataset HH122 of sleeping in Room2. Even though it is not usual because sleeping is done at the bedroom, the inhabitant relaxes at both the bedroom and room2. This makes the observation a little more similar from the pattern than sleeping in a completely unrelated location.

### 5.4 Mining contextualized behaviors

To assess how contextualized behavior patterns add insight to the inhabitants routine (besides her activity patterns), we show some contextualized behavior patterns using day periods (morning, afternoon, night) as context attribute from the HH120 dataset in Eq. 38 and 39. The morning and afternoon behavior patterns are illustrated in Fig. 5 and Fig. 6, respectively.

$$\begin{aligned}
 (P = \{ & \text{SLEEP, TOILET, DRESS, BATH,} \\
 & \text{WORK, PHONE, WATCH\_TV} \}, \\
 R = \{ & \{ \text{SLEEP, TOILET, } \prec \}, \{ \text{SLEEP, DRESS, } \prec \}, \\
 & \{ \text{TOILET, BATH, } \prec \}, \{ \text{DRESS, TOILET, } \prec \}, \\
 & \{ \text{DRESS, WORK, } \prec \}, \{ \text{DRESS, PHONE, } \prec \}, \\
 & \{ \text{DRESS, WATCH\_TV, } \prec \}, \} \mid \{ \text{TIME, MORNING} \} \\
 & \quad \quad \quad (38)
 \end{aligned}$$

$$\begin{aligned}
 (P = \{ & \text{SLEEP, TOILET, RELAX, COOK,} \\
 & \text{LEAVE\_HOME, WATCH\_TV} \}, \\
 R = \{ & \{ \text{SLEEP, TOILET, } \prec \}, \{ \text{TOILET, DRESS, } \prec \}, \\
 & \{ \text{TOILET, COOK, } \prec \}, \{ \text{TOILET, LEAVE\_HOME, } \prec \}, \\
 & \{ \text{TOILET, RELAX, } \prec \}, \{ \text{TOILET, WATCH\_TV, } \prec \}, \} \\
 & \mid \{ \text{TIME, AFTERNOON} \} \\
 & \quad \quad \quad (39)
 \end{aligned}$$

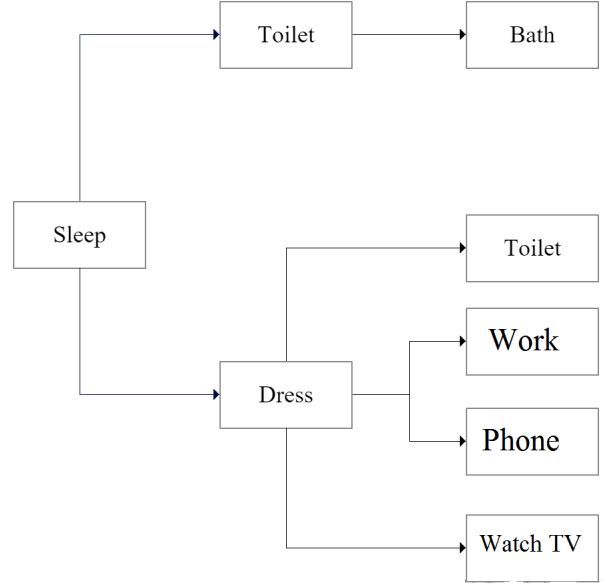


Figure 5: Morning routine for HH120 dataset

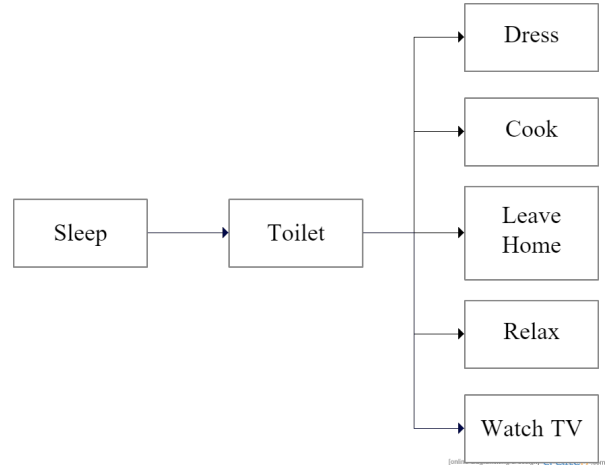


Figure 6: Afternoon routine for HH120 dataset

<i>Dataset</i>	<i>Activity</i>	<i>Date</i>	<i>Start time</i>	<i>Location</i>	<i>sim</i>
HH120	Eat	Wednesday	07:52:36	<u>kitchen</u>	0.33
HH120	Work	Monday	<u>21:10:28</u>	<u>kitchen</u>	0.33
HH120	Sleep	Sunday	<u>21:04:26</u>	<u>living</u>	0.33
HH122	Dress	<u>Sunday</u>	<u>21:56:51</u>	bedroom	0.33
HH122	Sleep	Sunday	<u>08:52:23</u>	<u>Room2</u>	0.385
HH123	Eat	Tuesday	<u>06:33:12</u>	bedroom	0.495
HH123	Phone	<u>Tuesday</u>	<u>06:29:14</u>	bedroom	0.495

Table 5: Deviating behaviors found in the dataset. Unusual context attributes for the activity are underlined

Contextualized behavior patterns improve understanding of the inhabitant’s routines. For example, LEAVE\_HOME after SLEEP in the morning is unusual, although the same sequence in the afternoon is expected. This also shows how the same activity can follow different patterns.

A behavior observation is shown in Eq. 40.

$$\begin{aligned}
&\{\text{ACTIVITY, SLEEP, BEDROOM,} \\
&\quad 2012 - 03 - 02T14 : 59 : 40Z, \\
&\quad 2012 - 03 - 02T16 : 17 : 36Z\}, \\
&\{\text{ACTIVITY, TOILET, BATHROOM,} \\
&\quad 2012 - 03 - 02T16 : 17 : 48Z, \\
&\quad 2012 - 03 - 02T16 : 20 : 51Z\}, \\
&\{\text{ACTIVITY, LEAVE\_HOME, HALL,} \\
&\quad 2012 - 03 - 02T16 : 30 : 08Z, \\
&\quad 2012 - 03 - 02T16 : 30 : 27Z\}
\end{aligned} \tag{40}$$

We remark that when context was not considered, afternoon patterns were not discovered since they are not as frequent in the complete database. Since the sequence SLEEP-TOILET-LEAVE\_HOME is not usual overall, the behavior observation in Eq. 40 always has low similarity measures when comparing the order. When contextualized behavior patterns are mined, the observation’s similarity to the pattern is increased significantly.

## 6 Conclusions and Future Work

In this paper we proposed contextualized behavior patterns as a long-term behavior model for AAL. The proposed model handles both temporal and context-based variations of human behavior patterns. Variations are expressed in a meaningful manner improving the understanding of long-term behavior patterns and reasoning about rare behaviors. Furthermore, the model facilitates the reuse of knowledge.

To find behavior deviations we propose a similarity measure that considers the semantic meaning of context attributes. This allows to reason about rarely seen or never before seen behaviors in the light of known patterns. The result of the comparison between observed behaviors and

known ”normal” patterns can be used by notification tools to explicitly indicate the differences and explain possible abnormal happenings.

The model was assessed using three public data sets. We used a sequence mining algorithm we proposed in previous work to mine activity patterns and contextualized behavior patterns. The experimentation showed that the model is flexible enough to represent activity and behavior patterns for different users and to capture context variations properly. The experience also confirmed the explanation capacity when behavior deviations were recognized.

Our current work focuses on extending the experiences by using of a larger variety of context attributes and improving the mining algorithm with more flexible time intervals handling. We are working in obtaining a long-term dataset with rich context data. This is done with the Amiqua4Home Intelligent Apartment founded by the French Research Agency (<https://amiqua4home.inria.fr/>).

Future research concerns a more complete solution providing a smart notification interface allowing care givers to fully take advantage of deviation information. We are also interested in providing efficient real time behavior pattern mining where one of the main issues is avoiding the exponential growth of pattern candidates.

## Acknowledgements

The authors would like to thank Nicolás Cardozo, Paola Gómez and the reviewers of this work for their valuable insights and suggestions.

## References

- Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843.
- Aztiria, A. and Augusto, J. C. (2013). Context-Aware Discovery of Human Frequent Behaviours Through Sensor Information Interpretation. In Guesgen, H. and Marsland, S., editors, *Human Behavior Recognition Technologies: Intelligent Applications for Monitoring and Security*, volume 2, chapter 2, pages 14–32. IGI Global, Hershey.



- Aztiria, A., Farhadi, G., and Aghajan, H. (2013). User Behavior Shift Detection in Ambient Assisted Living Environments. *JMIR mhealth and uhealth*, 1(1):e6.
- Canary Care Limited (2015). CanaryCare.
- Chaaaraoui, A. A., Climent-Pérez, P., and Flórez-Revuelta, F. (2012). A review on vision techniques applied to Human Behaviour Analysis for Ambient-Assisted Living. *Expert Systems with Applications*, 39(12):10873–10888.
- Chen, L., Hoey, J., Nugent, C. D., Cook, D. J., Yu, Z., and Member, S. (2012a). Sensor-Based Activity Recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 42(6):790–808.
- Chen, L., Nugent, C., and Okeyo, G. (2014). An Ontology-Based Hybrid Approach to Activity Modeling for Smart Homes. *IEEE Transactions on Human-Machine Systems*, 44(1):92–105.
- Chen, L., Nugent, C. D., and Wang, H. (2012b). A Knowledge-Driven Approach to Activity Recognition in Smart Homes. *IEEE Transactions on Knowledge and Data Engineering*, 24(6):961–974.
- Cook, D. J., Crandall, A. S., Thomas, B. L., and Krishnan, N. C. (2013). CASAS : A Smart Home in a Box. *IEEE Computer*, 46(7):62–69.
- DESA. Population Division. UN (2001). World population aging, 1950 - 2050.
- Dey, A. K. (2009). Modeling and intelligibility in ambient environments. *Journal of Ambient Intelligence and Smart Environments*, 1(1):57–62.
- Díaz-Rodríguez, N., Cadahía, O., Cuéllar, M., Lilius, J., and Calvo-Flores, M. (2014). Handling Real-World Context Awareness, Uncertainty and Vagueness in Real-Time Human Activity Tracking and Recognition with a Fuzzy Ontology-Based Hybrid Method. *Sensors*, 14(10):18131–18171.
- Forkan, A. R. M., Khalil, I., Tari, Z., Fofou, S., and Bouras, A. (2014). A context-aware approach for long-term behavioural change detection and abnormality prediction in ambient assisted living. *Pattern Recognition*, 48(3):628–641.
- Huynh, T., Huynh, T., Fritz, M., Fritz, M., Schiele, B., and Schiele, B. (2008). Discovery of activity patterns using topic models. *Proceedings of the 10th international conference on Ubiquitous computing - UbiComp '08*, page 10.
- Kandhari, R., Chandola, V., Banerjee, A., Kumar, V., and Kandhari, R. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3):15–58.
- Kim, E., Helal, S., and Cook, D. (2011). Human Activity Recognition and Pattern Discovery. *IEEE Pervasive Computing*, 9(1):48–53.
- Lago, P., Jiménez-guarín, C., and Roncancio, C. (2014). A Case Study on the Analysis of Behavior Patterns and Pattern Changes in Smart Environments. In Pecchia, L., Chen, L. L., Nugent, C., and Bravo, J., editors, *Ambient Assisted Living and Daily Activities*, volume 8277, pages 296–303. Springer International Publishing, Belfast, UK.
- Lago, P., Jiménez-Guarín, C., and Roncancio, C. (2015). Contextualized Behavior Patterns for Ambient Assisted Living. *Lecture Notes in Computer Science*, 9277:129–142.
- Lively Inc (2014). MyLively.
- Lundström, J., Järpe, E., and Verikas, A. (2016). Detecting and exploring deviating behaviour of smart home residents. *Expert Systems with Applications*, 55:429–440.
- Monekosso, D. N. and Remagnino, P. (2010). Behavior Analysis for Assisted Living. *IEEE Transactions on Automation Science and Engineering*, 7(4):879–886.
- Moshtaghi, M., Zukerman, I., and Russell, R. A. (2015). Statistical models for unobtrusively detecting abnormal periods of inactivity in older adults. *User Modeling and User-Adapted Interaction*, 25:231–265.
- Nazerfard, E. and Cook, D. J. (2012). Bayesian Networks Structure Learning for Activity Prediction in Smart Homes. *2012 Eighth International Conference on Intelligent Environments*, pages 50–56.
- Nazerfard, E., Das, B., Holder, L. B., and Cook, D. J. (2010). Conditional random fields for activity recognition in smart environments. *Proceedings of the ACM international conference on Health informatics - IHI '10*, page 282.
- Ohlbach, H. (2004). Relations between fuzzy time intervals. *Proceedings. 11th International Symposium on Temporal Representation and Reasoning, 2004. TIME 2004*.
- Ordóñez, F. J., de Toledo, P., and Sanchis, A. (2015). Sensor-based Bayesian detection of anomalous living patterns in a home setting. *Personal and Ubiquitous Computing*, 19(2):259–270.
- Ordóñez, F. J., Englebiene, G., de Toledo, P., van Kasteren, T., Sanchis, A., and Krose, B. (2014). In-Home Activity Recognition: Bayesian Inference for Hidden Markov Models. *IEEE Pervasive Computing*, 13(3):67–75.
- Park, K., Lin, Y., Metsis, V., Le, Z., and Makedon, F. (2010). Abnormal human behavioral pattern detection in assisted living environments. *Proceedings of the 3rd*

*International Conference on Pervasive Technologies Related to Assistive Environments - PETRA '10*, (March 2016):1.

Rashidi, P., Cook, D. J., Holder, L. B., and Schmitter-Edgecombe, M. (2011). Discovering activities to recognize and track in a smart environment. *IEEE Transactions on Knowledge and Data Engineering*, 23(4):527–539.

Rieping, K., Englebienne, G., and Kröse, B. (2014). Behavior analysis of elderly using topic models. In *Pervasive and Mobile Computing*, volume 15, pages 181–199.

Seiter, J., Amft, O., Rossi, M., and Tröster, G. (2014). Discovery of activity composites using topic models: An analysis of unsupervised methods. *Pervasive and Mobile Computing*, 15:215–227.

Yuan, B. and Herbert, J. (2014). Context-aware hybrid reasoning framework for pervasive healthcare. *Personal and Ubiquitous Computing*, 18(4):865–881.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3):338–353.

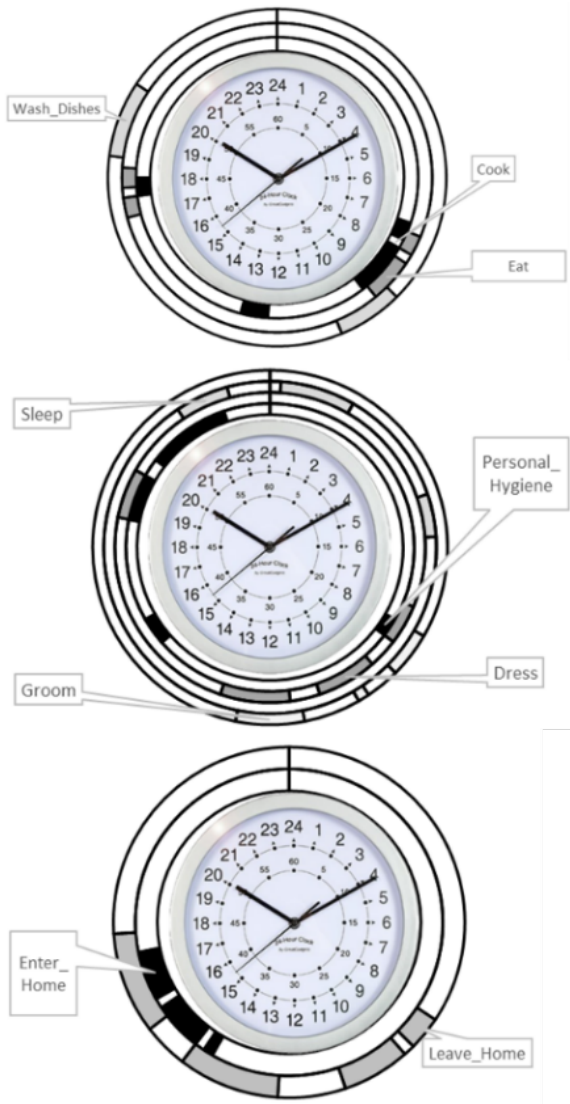


Figure 7: Frequent time intervals for the start time of different activities in the HH122 dataset

<i>Activity</i>	<i>Location</i>	<i>Day of the week</i>
Personal Hygiene	Bathroom1	All days but Saturday
Relax	Bedroom and Room 2	Monday, Tuesday, Wednesday, Friday
Dress	Bedroom	All weekdays
Bed Toilet Transition	Bathroom1	All days but Sunday
Cook	Kitchen	All days but Saturday
Toilet	Bathroom 1 and Bathroom 2	All days but Sunday
Watch TV	Room2	All days but Sunday
Enter Home	Hall	All days but Monday
Read	Bedroom	Monday, Wednesday and Saturday
Leave Home	Hall	All days but Monday
Sleep	Bedroom	All days
Wash Dishes	Kitchen	All days but Saturday
Phone	Living	Monday and Thursday
Work	Room2	All days but Sunday
Eat	Kitchen and Room2	All days but Saturday
Take medicine	Bathroom1	All days
Groom	Bathroom1	All days but Saturday
Bathe	Bathroom1	All days but Monday and Friday

Table 6: Activity patterns found in the HH122 dataset

<i>Activity</i>	<i>Location</i>	<i>Day of the week</i>
Personal Hygiene	Bathroom	All days but Thursday
Relax	Bedroom	All days but Friday
Dress	Bedroom	All days
Bed Toilet Transition	Bathroom	All days but Tuesdays
Cook	Kitchen	All days
Toilet	Bathroom	All days but Sunday
Watch TV	Bedroom	All days but Tuesdays
Enter Home	Hall	All days but Wednesday and Friday
Read	Bedroom	Sundays
Leave Home	Hall	All days
Sleep	Bedroom	All days
Wash Dishes	Kitchen	All days but Tuesdays and Fridays
Phone	Bedroom	All days but Tuesdays and Saturdays
Work	Dining	Sunday
Eat	Bedroom	All days but Tuesdays
Take medicine	Kitchen	All days
Groom	Bathroom	All days but Tuesdays
Bathe	Bathroom	Monday, Tuesday, Thursday

Table 7: Activity patterns found in the HH123 dataset

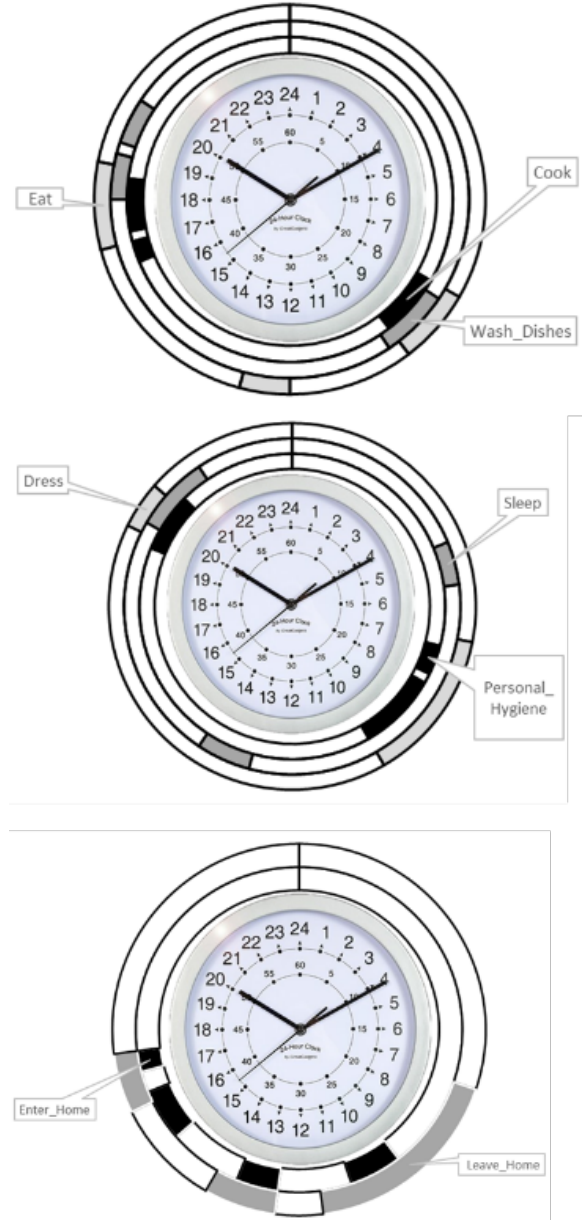


Figure 8: Frequent time intervals for the start time of different activities in the HH123 dataset