A context-aware hybrid framework for human behavior analysis

Roghayeh Mojarad[1] , **Abdelghani Chibani**[1] , **Ferhat Attal**[1] , **Yacine Amirat**[1]
[1]Univ Paris Est Creteil, LISSI, F-94400 Vitry, France
{roghayeh.mojarad, abdelghani.chibani, ferhat.attal, amirat}@u-pec.fr

Abstract

Human behavior analysis is a significant component of Ambient Assisted Living (AAL) systems and personal assistive robots. It allows discovering people's preferences, activities, and habits to provide them intelligent services intended to improve their quality of lives in terms of autonomy, well-being, and safety. In this paper, a framework is proposed to better characterize the human context by inferring new knowledge about his/her behaviors using commonsense reasoning. The proposed framework is composed of four main components. In the first component, human activities are recognized using a CNN-LSTM model. In the second component, the different contexts of human activities, including location, object, frequency, duration, and sequences of frequent activities, are extracted to analyze human behaviors. The obtained activity contexts are mapped to an ontology, called Human AcTivity (HAT) ontology conceptualizing the human activities and their contexts. In the last component, Answer Set Programming (ASP), a high-level expressive logic-based formalism, is used to represent human behaviors and carry out commonsense reasoning to infer new knowledge about these behaviors. The proposed framework was evaluated using the *Orange4Home dataset*, which consists of routine daily activities. Moreover, two quantitative experiments were carried out to demonstrate the ability of the proposed framework to better characterize human behaviors.

1 Introduction

Having a healthy life style can help people in preventing chronic disease and long time illnesses (Mai). Human lifestyle depends on human daily living behaviors; therefore, human behavior analysis is an important component of Ambient Assisted Living (AAL) systems and personal assistive robots to improve people's quality of lives in terms of autonomy, well-being, and safety (Mojarad et al. 2018), (Ayari, Chibani, and Amirat 2013), (Chibani et al. 2013). These assistive systems must be endowed with sufficient reasoning and data analytics capabilities to analyze sensor data, and therefore recognize the human contexts (Chibani et al. 2015), (Ayari et al. 2015), (Bettini et al. 2010). Dey (2001) defined context as "any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves." One of the main challenges of the aforementioned systems is to provide a consistent description of the human contexts from sensor data (Mojarad et al. 2018). In this study, the main focus is the recognition of human contexts from the human behavior point of view.

Since human behavior is complex, designing AAL systems and personal assistive robots with the ability to analyze human behaviors poses several challenges. One of them is to provide a standard and machine-understandable definition of human behavior (Salah et al. 2012). In the literature, human activity and human behavior are usually used interchangeably (Rodríguez et al. 2014) while there is a distinction between them. Only a few studies (Lago, Jiménez-Guarín, and Roncancio 2015) distinguished these two terms such that human behavior is usually defined as frequent activities that the user performs in different situations (Banovic et al. 2016). This definition is not comprehensive due to not taking into account the different contexts of human activity such as location, object, duration, daypart, and etc. In this paper, human behavior is defined comprehensively; it is defined as a structure with six components: (i) frequent activities in specific locations, e.g., *eating in the kitchen*, (ii) frequent activities with specific objects, e.g., *eating with a fork and a knife*, (iii) frequent activities in particular dayparts, e.g., *eating in the morning*, (iv) frequent activities within particular ranges of duration, e.g., *eating takes between d_{min} to d_{max} minutes, where d_{min} and d_{max} represent the minimum and maximum duration of eating activity, respectively*, (v) recurrent activities with particular frequencies per day, e.g., *frequency of eating per day is between f_{min} and f_{max}, where f_{min} and f_{max} represent the minimum and maximum frequency of eating, respectively*, and (vi) frequent sequences of activities, e.g., *the activity sequence eating-cleaning*.

Most of the existing approaches analyzing human behaviors are data-driven approaches (Nigam, Singh, and Misra 2019). For example, machine-learning models such as Support Vector Machine (SVM) and Random Forest (RF) rely on datasets, which usually include limited information and may not include the different contexts of human activities. Commonsense reasoning can be considered as an appropriate approach to take into account different human activity contexts and deal with the limitations of data-driven approaches.

In this study, a framework is proposed to better characterize the human contexts by inferring new knowledge about

his/her behaviors using commonsense reasoning. Mueller (Mueller 2006) defines commonsense reasoning as a process that involves taking information about specific aspects of a scenario and infers new knowledge about other aspects of this scenario based on commonsense knowledge, e.g., *cooking takes place in the kitchen*. The language used to represent commonsense knowledge should be non-monotonic to model changes in knowledge. Commonsense reasoning can handle concurrent events, indirect effects, preconditions, triggered events, default reasoning, temporal projection, abduction, and postdiction (Mueller 2006). The proposed framework is composed of four main components: (i) human activity recognition, (ii) extracting human activity contexts, (iii) mapping to ontology, and (iv) human behavior analysis. In the first component, human activities are recognized using a CNN-LSTM model. The human activity contexts in terms of locations, object, duration, frequency, and activity sequence are then extracted. An ontology, called Human AcTivity (HAT) ontology, is used to provide a formal specification of a shared conceptualization to describe human activity. In the third component, the extracted contexts are mapped into the proposed HAT ontology to conceptualize human activity and its contexts. In the last component, ASP, a high-level expressive logic-based formalism, is used to represent human behaviors and carry out commonsense reasoning to infer new knowledge about human behaviors. We choose ASP because it is a rich KR formalism with roots in logic programming and non-monotonic reasoning. It allows handling partial and incomplete context information. Among the features of ASP that we exploited is the use of the Negation As Failure (NAF) in the rules, which allows default reasoning, i.e, when a given activity or context (location, object, etc.) becomes true at a time-point but next time-point it becomes no more true (user moved away, changed activity, etc.), the conclusion can be retracted. Activities and their contexts are defined in the ontology as general concepts and properties; the latter are mapped into the ASP program to define general knowledge about how these activities happen in a normal context, i.e. default context. By exploiting NAF, we introduced exception rules that allow inferring, for instance, when a given mandatory activity is not happening or an activity happens in the wrong context, an abnormal human behavior is inferred. In NAF-based reasoning, if there are no facts matching with exception rules, the default context remains valid. However, strong negation is used to represent false facts; e.g., reading-a-book can happen in two locations: living-room and bedroom; strong negation is used to represent that bed is not recognized; i.e., the location is not bedroom; therefore, ASP reasoner allows inferring the location of current activity is living-room. Postdiction is used to obtain information about the previous states of user based on his/her activities; e.g., the user first sleeps in the bedroom, and then the user takes a shower in bathroom, from the fact "living-room is in the way from bedroom to bathroom", ASP reasoner allows inferring that the user was in living-room.

The main contribution of this paper is a hybrid framework that combines a CNN-LSTM model with ASP to better characterize human behaviors. To the best of our knowledge, this study is the first attempt to use ASP for this purpose. The proposed framework was evaluated using the *Orange4Home dataset*, which consists of routine daily activities. Moreover, two quantitative experiments were carried out to demonstrate the performance of the proposed framework to infer new knowledge.

This paper is organized as follows: section II is dedicated to related works. The details of the proposed framework are presented in section III. The experimental results are provided and discussed in section IV. Finally, section V provides a summary of the proposed framework and research perspectives.

2 Related Works

Researchers have developed many approaches for HAR, drawing on ideas from two very different categories, data-driven and knowledge-driven approaches. The data-driven approaches such as machine-learning models strongly rely on data while knowledge-driven approaches such as ontological-based models rely on priori contextual information, such as the temporal and spatial relationships between activities and the possible objects involved in the activities (Azkune and Almeida 2018), (Ferrari et al. 2020). Deep learning models, built using neural networks, are one type of machine learning model that have recently received remarkable attention from researchers. In (Agarwal and Alam 2020), a Lightweight Deep Learning Model combining Shallow Recurrent Neural Network (RNN) combined with Long Short Term Memory (LSTM) for HAR requiring less computational power is proposed. In (Ronao and Cho 2016), a deep Convolutional Neural Network (RNN) for HAR is proposed. The model derives relevant and more complex features in order to increase the performance of HAR. In (Ordonez and Roggen 2016), a deep framework using convolutional and LSTM recurrent units is proposed for HAR. On the other hand, knowledge-driven approaches are used for HAR using knowledge representation and different types of reasoning. The basis of several studies in this domain in action theory that is an area in philosophy concerned with the processes causing intentional human movement. In (Kautz 1991), a formal theory for plan recognition, which can be used for HAR, is proposed. This study is limited to recognize instances of plans whose types appear in the hierarchy. It means that their theory can not be used to recognize new plans created by chaining together the preconditions and effects of other plans. In (Gabaldon 2009), a framework is proposed to recognize activities using reasoning about actions, based on action theory, that consist of a notion of intended actions. This study considers the intentions of the users based on his/her activities and properties of the world, all at various points of time. In (Blount and Gelfond 2012), the behavior of a user intending to execute a sequence of actions is formalized. The obtained axioms are represented using knowledge representation language Answer Set Prolog. In (Oetsch and Nieves 2018), a logic-based framework specifying complex activities using activity theory is proposed. In this study, complex activity is defined as a process that mediates the relationship between an individual and certain motivating objects, generating a hier-

archy of objectives that direct actions. Several researchers extend HAR to analyze human behaviors due to its importance in the safety of people. Although there are differences between two terms of *human behavior* and *human activity*, in the literature, these two terms are often considered interchangeable (Riboni et al. 2016), (Baxter, Robertson, and Lane 2015), (Batchuluun et al. 2017), (Bruno et al. 2013), (Makantasis et al. 2016), (Rodríguez et al. 2014). In (Bruno et al. 2013), a framework is proposed to recognize human behaviors using Gaussian mixture modeling and Gaussian mixture regression. In (Makantasis et al. 2016), a CNN model is used to automatically construct high-level features capturing the spatio-temporal information on human behaviors. Multi-Layer Perceptron (MLP) is then used to classify human behaviors. In (Baxter, Robertson, and Lane 2015), a framework is proposed to recognize human behaviors using Dynamic Bayesian Networks (DBNs). In this study, human behaviors are classified into two categories: high-level and low-level behaviors. A low-level behavior is isolated and does not include long-term dependencies, while a high-level behavior is composed of sequences of low-level behaviors.

In (Batchuluun et al. 2017), a fuzzy inference system is proposed to recognize and predict human behaviors. The proposed system consists of two main fuzzy classifiers; the first one is used for behavior recognition, and the second one is applied for behavior prediction using the fuzzy probabilities obtained with the first classifier.

In (Phan et al. 2017), an ontology-based deep learning model (ORBM+) is proposed to predict human behaviors using undirected and nodes-attributed graphs. The authors propose a bottom-up algorithm to learn user representation from ontologies. Then the learned representation is used to integrate social influences and environmental events in a model of human behavior prediction based on the Restricted Boltzmann Machine.

However, in few studies, human behavior and human activity are defined differently; *human behavior* is defined as human activity routines (Yürüten, Zhang, and Pu 2014), (Banovic et al. 2016), (Lago, Jiménez-Guarín, and Roncancio 2015). The approach proposed in (Yürüten, Zhang, and Pu 2014) combines low-rank matrix decomposition and time-warping techniques to analyze human activities. The routines and deviations are separated into different clusters using Dynamic Time Warping (DTW). The Silhouette index (Kaufman and Rousseeuw 2009) is then used to specify the optimal number of clusters. Afterward, a cross-product between routine-clusters and deviation clusters is performed to find the final memberships of clusters for each day.

In (Banovic et al. 2016), the proposed approach allows extracting routines from human behavior logs automatically. The authors define human behavior as frequent actions performed in different situations. A Markov Decision Processes (MDP) framework is used to demonstrate routine behavior, and the MaxCausalEnt algorithm is then applied to predict human behavior from the logs.

Most of the existing studies on human behavior recognition are data-driven approaches. The latter does not consider the context of human behaviors. Moreover, the data-driven approaches rely on datasets that usually include a lot of sensors, which takes a lot of effort to collect the data and assign suitable labels to them. The current datasets are typically limited to the few numbers of activities and their contexts. Therefore, inferring new contexts by reasoning on the existing information in these datasets can be an appropriate solution to deal with these limitations. In this paper, a hybrid framework overcoming the mentioned limitations is proposed to analyze human behaviors. In the proposed framework, a data-driven approach is combined with commonsense reasoning.

3 Proposed Framework

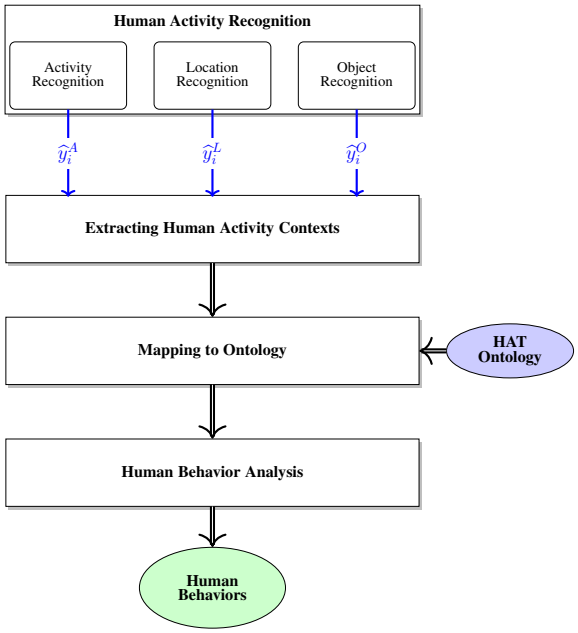

Figure 1: Architecture of the proposed framework.

3.1 Human activity recognition

In the human activity recognition component, three machine-learning models are independently used to classify sensor data into human activities by predicting a set of labels, namely: human activity, location, and object, see Fig. 1. Assigning one machine-learning model to each label allows the proposed framework to analyze human behaviors even without specific information about location or object labels. This component can be formalized as follows:

$$\begin{aligned} \widehat{Y}_i^A &= f_A(X_i) \\ \widehat{Y}_i^L &= f_L(X_i) \\ \widehat{Y}_i^O &= f_O(X_i) \end{aligned} \tag{1}$$

where X_i represents the i^{th} data sample. $\widehat{Y}_i^A$, $\widehat{Y}_i^L$, and $\widehat{Y}_i^O$ represent the labels predicted by models dedicated to activity, location, and objects recognition, respectively. The model used in functions f_A, f_L, and f_O is CNN-LSTM model. This model uses CNN layers for feature extraction

on input data and LSTM layers to support sequence prediction. LSTM is a specific kind of Recurrent Neural Networks (RNN) with the capability of learning long-term dependencies. LSTM is explicitly designed to avoid the long-term dependency problem and is particularly well suited for modeling problems with temporal relations (Gers, Schmidhuber, and Cummins 1999). Since human daily living activities usually include sequential data structures, LSTM is suitable to model human activities. The model is trained in a fully-supervised way and consists of 11 layers. The first three layers are *1D convolution* layer, which uses a convolution kernel that is convolved with the input layer over a single temporal or spatial dimension to produce a tensor of outputs. The kernel size of these layers equals three. The fourth layer is *dropout layer* with a rate of 0.5 to prevent overfitting. The fifth layer is a *max-pooling* layer with a max-pooling window size equals to two to reduce the number of parameters and computation time. The sixth layer is *flatten* layer that is used to convert the pooled feature map into a single column. The seventh and eighth layers are LSTM layers with 200 and 100 units, respectively. The ninth layer is *dropout* layer with a rate of 0.5. The two last layers are *dense* layers with *relu* and *softmax* activation functions, respectively. The parameters of models related to the functions f_A, f_L, and f_O are optimized using the *adam* optimization function and *categorical-crossentropy* loss function. The hyper-parameters of the CNN-LSTM model are estimated using a grid search (parameter sweep) method, see Table 1.

Table 1: .

Hyperparameters	Testing Values	Optimal Value
layers	[4-15]	11
filters for each convolution layer	[8,16,32,64,128]	[64],[32],[16]

3.2 Extracting human activity contexts

Human behaviors depend on different human activity contexts. Extracting these contexts is required to analyze human behaviors since these contexts have significant roles in characterizing human behaviors better. In this study, an algorithm is developed to extract five contexts, namely: (i) frequent activities in particular locations, (ii) frequent activities with particular objects, (iii) frequent activities within particular ranges of duration, (iv) recurrent activities with particular frequencies per day, and (v) frequent sequences of activities. This component is formalized as a function g:

$$R_i^{loc}, R_i^{obj}, R_i^{dur}, R_i^{freq}, R_i^{seq} = g(\widehat{Y}_i^A, \widehat{Y}_i^L, \widehat{Y}_i^O) \quad (2)$$

where R_i^{loc} represents the list of frequent activities in specific locations; R_i^{obj} represents the list of frequent activities with specific objects. R_i^{dur} represents the list of frequent activities within specific ranges of duration. R_i^{freq} represents the list of recurrent activities with specific frequencies. R_i^{seq} represents the list of frequent sequence of activities. In the developed algorithm to extract these five lists, seven hash tables are generated; each table is dedicated to one of the human activity contexts, namely: locations, objects, minimum duration, maximum duration, minimum frequency, maximum frequency, and previous activity. Each hash table maps activities to one list of human activity contexts.

3.3 Ontology

An ontology provides a formal specification of a shared conceptualization that can be used to describe human activities and context using classes, individuals, and relations (Guarino, Oberle, and Staab 2009). To model human activity based on the defined specification, the HAT ontology, inspired by *ConceptNet* semantic network (Speer, Chin, and Havasi 2017), is proposed; *ConceptNet* is a knowledge graph that link two terms, such as words and phrases of natural language, with labeled edges (Speer, Chin, and Havasi 2017), e.g., two terms *an oven* and *cooking* are linked using the *is used for* labeled edge. Fig 2 shows the overview of the HAT ontology modeled using the Semantic Web Ontology Language (OWL) (McGuinness, Van Harmelen, and others 2004). The HAT ontology is built based on two upper-level concepts: (i) *Event* and (ii) *Object*. The knowledge about human activity and its contexts can be asserted using six concepts, namely: *Activity*, *Location*, *Time*, *Physical Object*, *Duration*, and *Frequency*, which are derived from the two aforementioned concepts. The concepts are connected using six relationships, namely: *has place*, *has frequency*, *has duration*, *has time*, *is used for*, and *is a*. Table 2 represents the formalized relationships between the main concepts in the HAT ontology. The latter is used to conceptualize the outputs of extracting human activity contexts component, e.g., the activities *preparing food*, *making food*, and *cooking* are conceptualized to the *cooking* activity concept based on HAT ontology. The conceptualized information are given as input to the human behavior analysis component. In the latter, the HAT ontology is also used in defining the ASP rules, i.e, concepts and relationships among concepts defined in the HAT ontology are exploited in defining ASP rules.

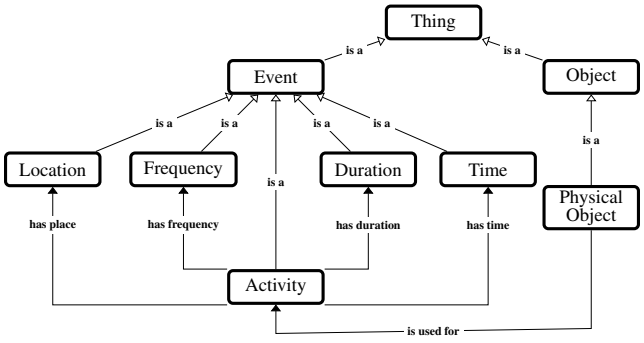

Figure 2: Overview of the HAT ontology.

Table 2: Formalized relationships between the main concepts in the HAT ontology.

$\mathsf{Object}(x) \vee \mathsf{Event}(x) \rightarrow \mathsf{Thing}(x)$.
$\mathsf{Location}(x) \vee \mathsf{Activity}(x) \rightarrow \mathsf{Event}(x)$.
$\mathsf{Activity}(a) \rightarrow \exists o\, \mathsf{Object}(o) \wedge \mathsf{isusedfor}(o, a)$.
$\mathsf{Activity}(a) \rightarrow \exists l\, \mathsf{Location}(l) \wedge \mathsf{hasplace}(a, l)$.
$\mathsf{Activity}(a) \rightarrow \exists t\, \mathsf{Time}(t) \wedge \mathsf{hastime}(a, t)$.
$\mathsf{Activity}(a) \rightarrow \exists d\, \mathsf{Duration}(d) \wedge \mathsf{hasduration}(a, d)$.
$\mathsf{Activity}(a) \rightarrow \exists f\, \mathsf{Frequency}(f) \wedge \mathsf{hasfrequency}(a, f)$.

3.4 Human Behavior Analysis

The ASP syntax is derived from the Prolog language, and its semantics is defined using the stable model semantics introduced by Gelfond et al. (1988). Formally, a standard ASP rule $Head \Leftarrow Body$ states that if the body of a rule is true, the reasoner can infer that the head of the rule is also true:

$$l : -b_1, ..., b_n, not\, c_1, ..., not\, c_n \ \ (n \geq 0).$$

where $b_1, ..., b_n, not\, c_1, ..., not\, c_n$ is the body of the rule and l is its *head*. A program Π in ASP is composed of a collection of standard ASP rules. The program Π can be seen as a specification for answer sets, knowledge obtained using the reasoner regarding the program Π. Each rule in the program Π can be seen as a constraint on the answer set; in the mentioned rule if $b_1, ..., b_n$ are in the answer set and none of $c_1, ..., c_n$ is in the answer set, then l must be included in the answer set. The latter, also called stable model, should be minimal and justified. Answer set is composed of ground atoms defined as atoms (symbols) in which there are no variables. An answer set is formalized as follows:

Definition 1.
Suppose a program Π *composed of rules:*

$$l : -b_1, ..., b_m, not\, c_1, ..., not\, c_n \ \ (n, m \geq 0).$$

Let S be a set of ground atoms,
A Reduct Π^S *is obtained from* Π *by deleting:*
- each rule having $not\, c_i$ *in its body with* $c_i \in S, i \geq 0$
- all negative atoms of the form $not\, c_i$ *from all other rules*
Answer set S is a minimal model of the reduct (Π^S*)*

Consider the following illustrative example with the following facts and rules:

Facts:
activity(drinking, 4, 7).
activity(washing, 8, 15).
Rules:
before(Act_1, Act_2):- activity(Act_1, T_{1s}, T_{1e}), activity(Act_2, T_{2s}, T_{2e}), $T_{1e} < T_{2s}$.
meet(Act_1, Act_2):- activity(Act_1, T_{1s}, T_{1e}), activity(Act_2, T_{2s}, T_{2e}), $T_{1s} = T_{2e}$.

where the fluent *activity(Act,* T_{1s}*,* T_{1e}*)* denotes that the user performs the activity *Act* from the timestamp T_{1s} to the timestamp T_{1e}. The *before(*Act_1*,* Act_2*)* show the temporal relationship, *before*, among the activities Act_1 and Act_2, i.e., the activity Act_1 happens before the activity Act_2 when the ending point of the activity Act_1 is less than the starting point of the activity Act_2. The *meet(*Act_1*,* Act_2*)* show another temporal relationship, *meet*, among the activities Act_1 and Act_2, i.e., the ending point of the activity Act_1 is equal with staring point of the activity Act_2. The Based on the existing facts and rules, the answer set for this model is as follows:

activity(drinking, 4, 7), activity(washing, 8, 15), before(drinking, washing)

The new knowledge *before(drinking, washing)* is inferred using ASP; this inferred knowledge represents that the user performs the activity *drinking* before the activity *washing*. In this study, human behaviors are modeled using several axioms. In order to represent human behavior based on activity, location, involved object, consider Axiom 1 :

Axiom 1. *behaviorActivityLocationObject(Act, Loc, Obj) :-activity(Act,* T_{1s}*,* T_{1e}*), location(Loc,* T_{2s}*,* T_{2e}*), object(Obj,* T_{3s}*,* T_{3e}*),* $T_{2s} > T_{1s}$, $T_{1e} > T_{2e}$, $T_{3s} > T_{2s}$, $T_{2e} > T_{3e}$.

where the fluent *activity(Act,* T_{1s}*,* T_{1e}*)* denotes the human activity. *Act*, T_{1s}, and T_{1e} respectively represent the name of activity, starting time, and ending time of that activity. The fluent *location(Loc,* T_{2s}*,* T_{2e}*)* denotes the human location. *Loc*, T_{2s}, and T_{2e} respectively represent the name of location, starting time, and ending time of human presence in that location. The fluent *object(Obj,* T_{3s}*,* T_{3e}*)* represents the human object. *Obj*, T_{3s}, T_{3e} denote respectively the object name, starting time, and ending time of using that object by the user. Axiom 1 is used to represent human behaviors when the user performs a specific activity in a specific location with a specific object. For example, a user may have the behavior "reading a book in the bedroom".

The human behavior without location and object can be respectively extracted from the Axiom 2 and Axiom 3 derived by the Axiom 1.

Axiom 2. *behaviorActivityLocation(Act, Loc) :- activity(Act,* T_{1s}*,* T_{1e}*), location(Loc,* T_{2s}*,* T_{2e}*),* $T_{2s} > T_{1s}$, $T_{1e} > T_{2e}$.

Axiom 3. *behaviorActivityObject(Act, Obj) :- activity(Act,* T_{1s}*,* T_{1e}*), object(Obj,* T_{2s}*,* T_{2e}*),* $T_{2s} > T_{1s}$, $T_{1e} > T_{2e}$.

Relationships between time intervals of human activities are necessary to better characterize human behaviors. Allen's interval algebra (Nebel and Bürckert 1995) is exploited to model temporal relationships of human activities using ASP through the Axioms 4-10; where Act_m represents the m^th activity. T_{ms} and T_{me} represent the starting and ending time of the activity Act_m.

Axiom 4. *before(*Act_1*,* Act_2*):- activity(*Act_1*,* T_{1s}*,* T_{1e}*), activity(*Act_2*,* T_{2s}*,* T_{2e}*),* $T_{1e} < T_{2s}$.

Axiom 4 represents the fact that an activity happens before another one if the ending time of the first activity is before starting time of the second one.

Axiom 5. *meet(*Act_1*,* Act_2*):-*
activity(Act_1*,* T_{1s}*,* T_{1e}*), activity(*Act_2*,* T_{2s}*,* T_{2e}*),*
$T_{1s} = T_{2e}$.

Axiom 5 defines the fact that an activity meets another one if the ending time of the first activity is the same as the starting time of the second one.

Axiom 6. *overlap(*Act_1*,* Act_2*):-*
activity(Act_1*,* T_{1s}*,* T_{1e}*), activity(*Act_2*,* T_{2s}*,* T_{2e}*),*
$T_{2s} < T_{1e}$, $T_{1s} < T_{2s}$, $T_{1e} < T_{2e}$.

Axiom 6 states the fact that when an activity overlaps another one, the staring and ending time of the first activity is before the starting and ending time of the second one meanwhile the ending time of the first activity is after the starting time of the second activity.

Axiom 7. *starts(*Act_1*,* Act_2*):-*
activity(Act_1*,* T_{1s}*,* T_{1e}*), activity(*Act_2*,* T_{2s}*,* T_{2e}*),*
$T_{1s} = T_{2s}$.

Axiom 7 represents the fact that an activity starts with another one if the staring time of both activities is the same.

Axiom 8. *during(*Act_1*,* Act_2*):-*
activity(Act_1*,* T_{1s}*,* T_{1e}*), activity(*Act_2*,* T_{2s}*,* T_{2e}*),*
$T_{1s} > T_{2s}$, $T_{2e} > T_{1e}$.

Axiom 8 defines the fact that an activity is during another one if the starting time of the first activity is after the starting time of the second one while the ending time of the first activity is before the ending time of the second one.

Axiom 9. *finishes(*Act_1*,* Act_2*):-*
activity(Act_1*,* T_{1s}*,* T_{1e}*), activity(*Act_2*,* T_{2s}*,* T_{2e}*),*
$T_{1e} = T_{2e}$.

Axiom 9 represents the fact that an activity is finished with another one if the ending time of both activities is the same.

Axiom 10. *isEqualTo(*Act_1*,* Act_2*):-*
activity(Act_1*,* T_{1s}*,* T_{1e}*), activity(*Act_2*,* T_{2s}*,* T_{2e}*),*
$T_{1e} = T_{2e}$, $T_{1s} = T_{2s}$, $Act_2 \; ! = Act_1$.

Axiom 10 states the fact that when an activity is temporally equal to another one, the starting and ending time of both activities are the same.

The sequences of human activities have a significant role in the representation and analysis of human behavior; axiom 11 formalizes these sequences:

Axiom 11. *SequentialActivity(*Act_1*,* Act_2*, ...* Act_n*):-*
activity(Act_1*,* T_{1s}*,* T_{1e}*), activity(*Act_2*,* T_{2s}*,* T_{2e}*), ...*
activity(Act_n*,* T_{ns}*,* T_{ne}*), meet(*Act_1*,* Act_2*), ...*
meet(Act_{n-1}*,* Act_n*).*

Axiom 11 defines the fact that a set of activities have a sequential relationship with each other, if each activity in the set has temporal relationship *meet* with the next activity from the set. It is worth mentioning, activity sequences inferred by ASP is more comprehensive than the ones generated by the component of extracting human activity contexts as the latter is the simplest case of the former with $n = 2$.

Human behaviors can be represented using daypart; for example, a behavior can be "reading a book at noon". Therefore, ten axioms are defined to represent different temporal parts of the day. The Axiom 12 represents the structure of these ten axioms, where *daypart* represents the temporal parts of a day. TH_s, and TH_e respectively represent starting time, and ending time of the *daypart* which are defined in HAT ontology. Table 3 shows the required conditions for each specific daypart; for instance, an activity happens in the specific daypart *morning* when this activity happens in the period from 8:00 to 11:00.

Axiom 12.
daypart(Act):- $activity(Act, T_s, T_e)$, $T_s < TH_e$, $T_e \geq TH_s$.
daypart(Act):- $activity(Act, T_s, T_e)$, $T_s > T_e$, $T_e \geq TH_s$.

To consider the period which exceeds the midnight, formalized as ($T_s > T_e$) when the time changes from 23:59 to 00:00, the second rule is defined (daypart(Act):- $activity(Act, T_s, T_e)$, $T_s > T_e$, $T_e \geq TH_s$).

Table 3: Definition of different dayparts for Axiom 12.

Temporal parts of a day (daypart)	Threshold of starting time (TH_s)	Threshold of ending time (TH_e)
nightAfterMidnight	0	5
earlyMorning	5	8
morning	8	11
lateMorning	11	12
noon	12	13
earlyAfternoon	13	14
afternoon	14	16
lateAfternoon	16	18
evening	18	21
nightBeforeMidnight	21	24

The new knowledge obtained from a set of axioms, including Axioms 1-12 that are defined in ASP makes better characterize human behaviors.

4 Performance Evaluation

In this section, the performances of the human activity recognition component are evaluated in terms of *precision*, *recall*, *F-measure*, and *accuracy* on the *Orange4Home dataset* (Cumin et al. 2017). Since the objective of the proposed framework is characterizing the human behaviors, the framework is evaluated in order to show its effectiveness to infer new knowledge about human behaviors. Moreover, another experiment is conducted to evaluate the inference of missing knowledge. For the implementation and evaluation, a computer equipped with an Intel i7-8650U 2.11GHz CPU with 32GB RAM is used.

4.1 Description of the dataset

Orange4Home dataset (Cumin et al. 2017) is collected from 236 sensors that capture information about the operation of doors, the use of electrical equipment, water consumption, etc. The sensors are located in the different places of an instrumented home. In this dataset, seventeen daily living

activities are performed by one occupant during four consecutive weeks of working days. The *Orange4Home* dataset includes three main human activity contexts, namely: time-of-day, place, and activity. Time-of-day considers temporal information of activities such as date and time. Place takes into account a geographical location in the home, such as *kitchen*. Activity considers the activity performed by the occupant. Table 4 shows the list of places in the *Orange4Home* dataset and activity performing in those places.

Table 4: List of activities grouped by places in the *Orange4Home* dataset.

Place	Activities
Entrance	Entering, Leaving
Kitchen	Preparing, Cooking, Washing the dishes
Living Room	Eating, Watching TV, Computing
Toilet	Using the toilet
Staircase	Going up, Going down
Bathroom	Using the sink, Using the toilet, Showering
Office	Computing, Watching TV
Bedroom	Dressing, Reading, Napping
Common to all places	Cleaning

4.2 Performances of the human activity recognition component

The *Orange4Home* dataset consists of two labels: activity and location; therefore, in the human activity recognition component, two CNN-LSTM models are independently used to classify input data into activity and location labels.

Table 7: Performance obtained using the CNN-LSTM model and baseline models in the case of activity recognition.

Models	Precision	Recall	F-measure	Accuracy
RF	63.60	69.34	57.80	79.50
KNN	77.12	47.30	57.76	73.30
CNN	48.45	69.60	57.13	77.52
CNN-LSTM	76.75	77.64	75.35	84.70

Table 8: Performance obtained using the CNN-LSTM model and baseline models in the case of location recognition.

Models	Precision	Recall	F-measure	Accuracy
RF	73.33	69.90	57.63	76.58
KNN	78.96	76.67	77.53	84.62
CNN	46.12	67.91	54.93	75.43
CNN-LSTM	81.00	81.60	80.63	87.70

These CNN-LSTM models are evaluated in terms of precision, recall, F-measure, and accuracy. With regard to the *Orange4Home* dataset, three baseline models, namely Random Forest (RF), K-Nearest Neighbors (KNN), Convolutional Neural Network (CNN), are compared with the CNN-LSTM models.

Table 7 shows the performance results in the case of activity recognition. The CNN-LSTM model yields the highest average F-measure, 84.70%, while for RF, the average F-measure is 57.80%. CNN-LSTM obtains sensibly better performances in comparison with the other machine-learning models.

Table 8 shows the performance results in the case of location recognition. The obtained results show that the CNN-LSTM model provides the best results in terms of precision, recall, F-measure, and accuracy, followed by KNN, RF, and CNN, respectively.

The results show that the CNN-LSTM model provides the best performance results in both cases, activity and location recognition. This can be explained by the fact that in this model, the CNN layers extract suitable features, which help the model to be trained well with the useful features. Moreover, in this model, LSTM layers take into account the temporal information among the activities, which help the model to perform well with predicting correct labels since the structure of human activity are time-series.

The performance results obtained by machine-learning models in the case of location recognition are better in comparison with those in the case of activity recognition. This can be explained by the fact that the number of samples in the case of location is higher than in the case of activity; hence, these models can be trained better in the case of location recognition. Moreover, the classes in the case of location are more distinguishable compared with those in the case of activities.

4.3 Performances of the human behavior analysis component

In this study, ASP has been implemented using CLINGO (Gebser et al. 2014), an ASP tool to ground and solve logic programs. The human behavior analysis component is evaluated in terms of new inferred knowledge about human behaviors. Moreover, another evaluation is conducted to show the performance of the human behavior analysis component in missing data compensation since one objective of the proposed framework is dealing with the problem of missing and limited information in datasets.

Table 5 shows the performance obtained using the human behavior analysis component in terms of inferring new knowledge about human behaviors and time consumption to obtain that inferred knowledge. In the evaluation, the number of initial facts varied from 50 to 700. The initial facts are based on activity and location contexts, while the inferred knowledge is based on temporal, dayparts, and activity sequences contexts. The results show that the behavior analysis component remarkably increases the contextual knowledge about human behaviors. The most inferred knowledge is related to temporal information, i.e., the temporal relationships among different activities. In the *Orange4Home* dataset, the user's activities are similarly repeated in each day. When the number of initial facts is between 50 and 100,

Table 5: Performance of the human behavior analysis component in terms of new inferred knowledge about human activity contexts.

Number of facts before reasoning	50	100	200	300	400	500	600	700
Activity-based context	25	50	100	150	200	250	300	350
Location-based context	25	50	100	150	200	250	300	350
Number of facts after reasoning	661	1145	1304	1474	1628	1785	1945	2103
Temporal context	545	926	932	935	936	937	938	939
Daypart context	25	50	100	150	200	250	300	350
Sequence-based context	46	69	72	89	92	98	107	114
Time consumption (ms)	13	22	43	53	80	128	182	236

Table 6: Performance of the human behavior analysis component when the location-based context is missing.

Number of activity-based facts before reasoning	25	50	100	150	200	250	300	350
Number of inferred location-based knowledge	25	50	100	150	200	250	300	350
Accuracy of inferred location-based knowledge	72.00	80.00	83.00	84.00	83.5	84.0	84.66	83.43

the temporal knowledge-base enhance remarkably however after these numbers of initial facts, the temporal knowledge-base is saturated, since the inferred temporal knowledge already exists in the knowledge-base. Consequently, the size of temporal knowledge-base stays similar when increasing the number of initial facts is increased.

For each activity, contextual knowledge related to daypart is inferred since each activity has a context of daypart. The sequences of activities are another important context to characterize human behaviors. These activity sequences are obtained based on the specific ASP rules, Axiom 11. The consuming time for reasoning is directly relevant to the number of initial facts; i.e., when the number of initial facts is increasing, the ASP rules should process these facts which need more time.

In order to show the ability of the human behavior analysis component to deal with the limitations of datasets, another evaluation is performed. In this evaluation, the location-contexts are removed from the initial facts, and then the location knowledge is inferred using ASP. Because of the types of activities and related locations are in accordance with commonsense, ASP has inferred them with high accuracy. However, the accuracy of the inferred knowledge is not complete, 100%, because some activities of the dataset, e.g., *cleaning*, happen in different locations in the dataset. Table 6 show the performance results obtained using the human behavior analysis component in terms of accuracy of inferred location-based knowledge. The results show that more than 70% of the obtained knowledge is correct, which shows the effectiveness of the framework to obtain new knowledge in the case of missing knowledge to better characterize human behaviors. Moreover, the results show that the accuracy of the inferred location-based knowledge stays similar after the first 50 activity-based facts; because in the dataset, the activities and their locations are repetitive; the first 50 activity-based facts are arguably enough to represent the whole dataset in terms of activities and locations.

5 Conclusion

In this paper, a framework is proposed to better characterize human behaviors. It allows providing a comprehensive description of human behavior. The proposed framework consists of four main components: (i) human activity recognition, (ii) extracting human activity contexts, (iii) mapping to ontology, (iv) human behavior analysis. In the first component, a CNN-LSTM model is used to recognize human activities. In the second component, the different contexts of human activities are extracted. The HAT ontology is proposed to provide a formal specification of a shared conceptualization that is used to describe human activities. In the third component, the extracted contexts are mapped into specific concepts of the HAT ontology. The last component is used to analyze human behaviors; ASP is exploited to represent human behaviors and carry out commonsense reasoning to infer new knowledge. Two quantitative experiments were carried out to show the ability of the proposed framework to infer new knowledge about human behaviors. In terms of research perspectives to this study, an interesting topic is to enhance the framework with a recommendation system based on healthy life guidelines. Another interesting research direction to explore is to handle the uncertainty of data and rules in the human behavior analysis component.

References

Agarwal, P., and Alam, M. 2020. A lightweight deep learning model for human activity recognition on edge devices. *Procedia Computer Science* 167:2364 – 2373. International Conference on Computational Intelligence and Data Science.

Ayari, N.; Chibani, A.; Amirat, Y.; and Matson, E. T. 2015. A novel approach based on commonsense knowledge representation and reasoning in open world for intelligent ambient assisted living services. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 6007–6013.

Ayari, N.; Chibani, A.; and Amirat, Y. 2013. Semantic management of human-robot interaction in ambient intelligence environments using n-ary ontologies. In *IEEE International Conference on Robotics and Automation*, 1172–1179.

Azkune, G., and Almeida, A. 2018. A scalable hybrid activ-

ity recognition approach for intelligent environments. *IEEE Access* 6:41745–41759.

Banovic, N.; Buzali, T.; Chevalier, F.; Mankoff, J.; and Dey, A. K. 2016. Modeling and understanding human routine behavior. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '16, 248–260. New York, NY, USA: ACM.

Batchuluun, G.; Kim, J. H.; Hong, H. G.; Kang, J. K.; and Park, K. R. 2017. Fuzzy system based human behavior recognition by combining behavior prediction and recognition. *Expert Systems with Applications* 81:108–133.

Baxter, R. H.; Robertson, N. M.; and Lane, D. M. 2015. Human behaviour recognition in data-scarce domains. *Pattern Recognition* 48(8):2377 – 2393.

Bettini, C.; Brdiczka, O.; Henricksen, K.; Indulska, J.; Nicklas, D.; Ranganathan, A.; and Riboni, D. 2010. A survey of context modelling and reasoning techniques. *Pervasive and Mobile Computing* 6(2):161 – 180.

Blount, J., and Gelfond, M. 2012. Reasoning about the intentions of agents. In *Logic programs, norms and action*. Springer. 147–171.

Bruno, B.; Mastrogiovanni, F.; Sgorbissa, A.; Vernazza, T.; and Zaccaria, R. 2013. Analysis of human behavior recognition algorithms based on acceleration data. In *IEEE International Conference on Robotics and Automation*, 1602–1607.

Chibani, A.; Amirat, Y.; Mohammed, S.; Matson, E.; Hagita, N.; and Barreto, M. 2013. Ubiquitous robotics: Recent challenges and future trends. *Robotics and Autonomous Systems* 61(11):1162 – 1172.

Chibani, A.; Bikakis, A.; Patkos, T.; Amirat, Y.; Bouznad, S.; Ayari, N.; and Sabri, L. 2015. *Using Cognitive Ubiquitous Robots for Assisting Dependent People in Smart Spaces*. Cham: Springer International Publishing. 297–316.

Cumin, J.; Lefebvre, G.; Ramparany, F.; and Crowley, J. L. 2017. A dataset of routine daily activities in an instrumented home. In Ochoa, S. F.; Singh, P.; and Bravo, J., eds., *Ubiquitous Computing and Ambient Intelligence*, 413–425. Cham: Springer International Publishing.

Dey, A. K. 2001. Understanding and using context. *Personal Ubiquitous Comput.* 5(1):4–7.

Ferrari, A.; Micucci, D.; Mobilio, M.; and Napoletano, P. 2020. On the personalization of classification models for human activity recognition. *IEEE Access* 8:32066–32079.

Gabaldon, A. 2009. Activity recognition with intended actions. In *Twenty-First International Joint Conference on Artificial Intelligence*.

Gebser, M.; Kaminski, R.; Kaufmann, B.; and Schaub, T. 2014. Clingo = ASP + Control: Preliminary Report. *arXiv:1405.3694 [cs]*.

Gelfond, M., and Lifschitz, V. 1988. The stable model semantics for logic programming. 1070–1080. MIT Press.

Gers, F. A.; Schmidhuber, J.; and Cummins, F. 1999. Learning to forget: continual prediction with lstm. In *1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470)*, volume 2, 850–855 vol.2.

Guarino, N.; Oberle, D.; and Staab, S. 2009. What is an ontology? In *Handbook on ontologies*. Springer. 1–17.

Kaufman, L., and Rousseeuw, P. J. 2009. *Finding groups in data: an introduction to cluster analysis*, volume 344. John Wiley & Sons.

Kautz, H. A. 1991. A formal theory of plan recognition and its implementation. *Reasoning about plans* 69–125.

Lago, P.; Jiménez-Guarín, C.; and Roncancio, C. 2015. Contextualized Behavior Patterns for Ambient Assisted Living. In Salah, A. A.; Kröse, B. J.; and Cook, D. J., eds., *Human Behavior Understanding*, volume 9277. Cham: Springer International Publishing. 132–145.

Maintaining a healthy lifestyle. https://www.foundationforpn.org/living-well/lifestyle/. Accessed: 2020-02-00.

Makantasis, K.; Doulamis, A.; Doulamis, N.; and Psychas, K. 2016. Deep learning based human behavior recognition in industrial workflows. In *IEEE International Conference on Image Processing (ICIP)*, 1609–1613.

McGuinness, D. L.; Van Harmelen, F.; et al. 2004. Owl web ontology language overview. *W3C recommendation* 10(10):2004.

Mojarad, R.; Attal, F.; Chibani, A.; Fiorini, S. R.; and Amirat, Y. 2018. Hybrid approach for human activity recognition by ubiquitous robots. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 5660–5665.

Mueller, E. T. 2006. Chapter 1 - introduction. In Mueller, E. T., ed., *Commonsense Reasoning*. San Francisco: Morgan Kaufmann. 1 – 16.

Nebel, B., and Bürckert, H.-J. 1995. Reasoning about temporal relations: a maximal tractable subclass of allen's interval algebra. *Journal of the ACM (JACM)* 42(1):43–66.

Nigam, S.; Singh, R.; and Misra, A. K. 2019. A review of computational approaches for human behavior detection. *Archives of Computational Methods in Engineering* 26(4):831–863.

Oetsch, J., and Nieves, J.-C. 2018. A knowledge representation perspective on activity theory. *arXiv preprint arXiv:1811.05815*.

Ordonez, F., and Roggen, D. 2016. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors* 16(1):115.

Phan, N.; Dou, D.; Wang, H.; Kil, D.; and Piniewski, B. 2017. Ontology-based deep learning for human behavior prediction with explanations in health social networks. *Information Sciences* 384:298 – 313.

Riboni, D.; Bettini, C.; Civitarese, G.; Janjua, Z. H.; and Helaoui, R. 2016. Smartfaber: Recognizing fine-grained abnormal behaviors for early detection of mild cognitive impairment. *Artificial Intelligence in Medicine* 67:57 – 74.

Rodríguez, N. D.; Cuéllar, M. P.; Lilius, J.; and Calvo-Flores, M. D. 2014. A survey on ontologies for human behavior recognition. *ACM Comput. Surv.* 46(4):43:1–43:33.

Ronao, C. A., and Cho, S.-B. 2016. Human activity recog-

nition with smartphone sensors using deep learning neural networks. *Expert Systems with Applications* 59:235 – 244.

Salah, A. A.; Ruiz-del Solar, J.; Meriçli, Ç.; and Oudeyer, P.-Y. 2012. Human behavior understanding for robotics. In Salah, A. A.; Ruiz-del Solar, J.; Meriçli, Ç.; and Oudeyer, P.-Y., eds., *Human Behavior Understanding*, 1–16. Berlin, Heidelberg: Springer Berlin Heidelberg.

Speer, R.; Chin, J.; and Havasi, C. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Thirty-First AAAI Conference on Artificial Intelligence*.

Yürüten, O.; Zhang, J.; and Pu, P. 2014. Decomposing activities of daily living to discover routine clusters. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, AAAI'14, 1348–1354. AAAI Press.