Streaming and Learning the Personal Context

Fausto Giunchiglia[1*] , **Marcelo Rodas Britez**[1] , **Andrea Bontempelli**[1] and **Xiaoyue Li**[1]

[1]University of Trento, Trento, Italy

{fausto.giunchiglia, marcelo.rodasbritez, andrea.bontempelli, xiaoyue.li}@unitn.it

Abstract

The representation of the personal context is complex and essential to improve the help machines can give to humans for making sense of the world, and the help humans can give to machines to improve their efficiency. We aim to design a novel model representation of the personal context and design a learning process for better integration with machine learning. We aim to implement these elements into a modern system architecture focus in real-life environments. Also, we show how our proposal can improve in specifically related work papers. Finally, we are moving forward with a better personal context representation with an improved model, the implementation of the learning process, and the architectural design of these components.

1 Introduction

Every person makes sense of their personal context differently because of their different sets of personal characteristics (intelligence) and behaviour (life choices). However, the machine's understanding of the personal context is radically different from the user's understanding. This limitation is due to the limited definition of the personal context, and the lack of tools to make sense of the personal context. For instance, while the person you are with now can be linked to a name, for people it has more meaning than just a name, e.g., friend and colleague. Additionally, these meanings are not fixed, they may change at any time, and every person can assign additional meaning using different criteria. Thus, effective context recognition requires a complex and dynamic representation of the personal context and the collaboration of the people to fill the cognitive gap of machines.

The addition of human collaboration into the context recognition learning of machines is an important part of supervised machine learning [Vapnik, 2013]. These interactions bring new challenges to the implementation of machine learning algorithms. For instance, humans can be defined as the expert of the supervised algorithms, interacting in an offline fashion by annotating sensor data [Webb, 2003], or the interaction can be directly online, as active learning [Settles, 2009;

---

[*]Contact Author

Hoque and Stankovic, 2012; Hossain *et al.*, 2017]. The human collaboration is important when we are moving into real-life scenarios [Kwapisz *et al.*, 2011].

Other challenges of this collaborative approach are the possibility of overwhelming the humans and the possible differences between the assignment of meaning between people [Chang *et al.*, 2017], thus, making the annotation a personal activity. Then, having humans as the input of the information opens the possibility of human error in the collaboration process [Tourangeau *et al.*, 2000], and this issue is well known in social sciences and psychology, because of response biases in answering self-reports [West and Sinibaldi, 2013], and more importantly, these biases are not well-understood [Freedman *et al.*, 2013].

We propose a novel context model based on the work from [Giunchiglia *et al.*, 2018]. That work focused on ensuring the reliability of annotations, whereas our focus is on improving personal context representations to get closer to work in real-life scenarios. So, we propose to add a more precise representation of personal context that can also work with machine learning algorithms. We formalize a context model based on ontology and use it with the streaming data to have a knowledge representation of context data. This formalization allows moving towards a generic definition of context that can work with existing multi-label machine learning approaches, using a conversion algorithm. Eventually, the last piece of the puzzle will be the design and development of the Streaming System to manage technically the dynamic context data, and it will be organized in the system architecture with modular components for independent development and easy deployment in current cloud environments.

Some examples of our model improvement can be seen compared with our main related work [Giunchiglia *et al.*, 2018; Bontempelli *et al.*, 2020; Zeni *et al.*, 2019]. All of them can take benefit from our novel personal context representation and can use our conversions algorithms to explicitly implement the transformations needed for the machine learning algorithms.

The paper is structured as follows. Section 2 introduces context modeling. Section 3 illustrates our representation of personal context, while we provide the formal representation in Section 4. Then, we show the learning process to transition from our formal representation to machine learning representation in Section 5 and how our formal representation

can be converted to a Direct Acyclic Graph (DAG). Finally, Section 6 describes works related to ours, and Section 7 concludes our paper.

2 The context in time

When we talk about the context, we concentrate on the context of a person called observer. The observer's context is the representation of a partial view of the world. We describe this context into three main dimensions: the viewpoint, the part-whole relation, and the endurant-perdurant.

Firstly, we have the viewpoint dimension divided as outside viewpoint and observer viewpoint. The outside viewpoint is the view of an ideal observer who can describe everything from a certain point of view. We distinguish this view from the world's static and dynamic properties. In the static property, there are whatever does not change in time, e.g., mountains, buildings, streets. In the dynamic property, there are not only the moving people, but also the moving animals, and facilities in their manifestations, like trains. Then, the observer viewpoint describes how the observer perceives what is around her or him. In this view, we also have the property of being static or dynamic, but it is relative to the movements of the observer.

Secondly, context is a part-whole relation. In our everyday life, when we do things, we are always embedded in the world. From an ontological view, we are part of the whole world. Thus, we call reference context as the element of the outside context with a volume and extension that is large enough to contain all our movements and changes. For instance, the reference context is the city of Trento when the user walks around, or the users' home when they are at home. In turn, our body as a whole has parts (e.g., arms, mind, legs) that are with us all the time, and they define the internal context of the user. The internal context identifies the elements of the user's body at different levels of abstraction. We usually distinguish between physical parts, such as arms, body, fingers, and mental parts, such as mind, memory, emotions. The context as a part-whole relation is divided into reference context and internal context, and both contexts with different dimensions play a role in our life.

Thirdly, context as endurant and perdurant refers to the bigger relation to changes in time or space. Events and actions are perdurants and elements, like me, are endurants.

2.1 The spatio-temporal context

The context, as viewpoints, defines the reference point from which we construct the context and the context, as part-whole, defines which parts we should consider. So, next, we need to define how we keep track of the context from a quantitative point of view, with a set of quantitative and qualitative measures. Therefore, based on these measures, we introduce the spatio-temporal context.

The spatio-temporal context consists of the temporal and spatial reference context. The former includes dates, times and all the additional notations like weekdays and seasons. The latter contains the world coordinate system. There are various reasons why context should be represented as a spatio-temporal context. First, this is a common representation when we think of the world. Second, any device today

can easily retrieve the time and time zones, and the space coordinates (e.g., via GPS). Third, time can be used to measure the changes in all the elements of the world, all evolving at different speeds, thus temporal context allows us to use them together based time. Finally, a lot of data about the spatial reference context and its sub-contexts are available from external sources (e.g., Google Maps, OpenStreetMaps).

The spatio-temporal context, also called objective context, at time t is defined as:

$$o_t = (D_t, T_t : L_t, \; me, \; \mathrm{coord}_t(me),$$
$$P_t^1 : \mathrm{coord}_t(P^1), \ldots, P_t^k : \mathrm{coord}_t(P^k),$$
$$O_t^1 : \mathrm{coord}_t(O^1), \ldots, O_t^m : \mathrm{coord}_t(O^m))$$

where D_n, T_n stands for date and time respectively, L_n is the location, namely the smallest possible spatial reference context that we can compute. Here me is the observer, P^i are persons and O^m are objects. The function $\mathrm{coord}(\ldots)$ computes the spatial coordinates of me, objects, and persons.

The number and type of persons and objects change over time. Hence, we will have a sequence of time-tagged states, namely $\mathcal{O} = \{o_1, \ldots, o_n\}$. We call the sequence $\mathcal{O}$ as the streaming context. In the streaming context, within the given reference location, it is easy to compute spatial relations (e.g., near, right, left, in front, far relative to the location) of the different elements among themselves. For instance, the system can compute that the smartphone is *in* the home building and the smartphone is *near* the computer.

2.2 The objective and subjective context

The spatio-temporal context is also called objective context, since all the relations are computed concerning what is objectively measured, in terms of spatial relations. However, notice that different observers will have different views of the world. For instance, the school building has the function of study-place from the point of view of a student and is the teacher's workplace. The word function here is used with the precise meaning defined in [Giunchiglia and Fumagalli, 2017; Giunchiglia *et al.*, 2018]. Hence, "*the function of an object formalizes the behavior that an object is expected to have*" [Giunchiglia and Fumagalli, 2017]. For instance, objects are trains and buildings. The expected behavior may be the purpose of the object (e.g., fridge) or the role of a person (e.g., friend).

The subjective context includes both the objective context elements and the function of persons and objects as seen by the user. Thus, the subjective context at time t is defined as:

$$s_t = (D_t, T_t : \; L_t, me, \mathrm{coord}_t(me), F_t(P^1), \ldots, F_t(P^k),$$
$$F_t(O^1), \ldots, F_t(O^m)),$$

where $F_n(P^k)$ and $F_n(O^m)$ are the functions with respect to a person P^k and an object O^m, respectively. The number and type of persons and objects, and their functions change over time. The sequence of subjective contexts over time is defined as the subjective streaming context $\mathcal{S} = \{s_1, \ldots, s_n\}$.

2.3 The endurant and perdurant context

In the endurant context [Giunchiglia *et al.*, 2017], its parts are endurants, essentially objects where the spatial extension

of their actions is contained by the space defined by the spatial (reference) context. The actions "*represents how objects change in time*" [Giunchiglia and Fumagalli, 2017]. For instance, running in a park performed by a runner. We also need to represent actions, in particular, the actions that are executed by the endurant me and also by any other elements of the outside dynamic context. Actions can be seen in two ways: *(i)* actions modeled as processes, namely as sequences of single micro-steps, each of length close to zero; *(ii)* actions modeled as events, which are often also called perdurants, namely as complete movements which last for a certain duration. Actions as events have key properties, similar to endurants. An event and an action can associate with a set of component sub-events and sub-actions.

Considering the mentioned concepts, the fundamentally different role of space and time should become clear. Whereas the parts of the space context are only used to limit the space where things happen, the parts of time have the main goal to detail how actions get executed. Things get complicated by adding objects, people, and functions as shown in the data representations shown in Table 1.

It is worth noting that each function of a person is associated with a limited set of actions, and the type of action that a person can perform can be considerably limited by knowing his or her function.

The actions apply to me and persons, whereas functions apply to me, other objects and persons. Notice also, that the stored location L is limited by the most specific location that we can compute. This is because the bigger locations are assumed to be static and stored in the system. For events, instead, we store the smallest possible most general event as well as those component actions which are done during a certain period. Thus, for instance, the action/event meeting can have sub-actions such as talking, walking, listening, typing. Table 2 reports the streaming context matrix of Example 1 and shows how the context changes over time.

3 The current context

The streaming context describes the contexts of an observer according to time. To represent each context occurrence, we define a set of notions to build a figure of the current context. We mainly divide the current context into four types of context according to how things compose in space and time. From top to bottom we have the following cases:

- 1L1E (One Location One Event), such as a lecture holds in a classroom;
- 1LME (One Location Multiple Events), such as a sequence of meetings hold in an office, or eat breakfast, lunch and dinner at home;
- 1E1L (One event One Location), this is the same as 1L1E. The difference is in the focus, which is now the location, e.g., the classroom where the lecture holds;
- 1EML (One Event Multiple Locations), such as a travel goes from many different places;

An 1EML example is shown in Figure 1, which describes the following travel scenario around observer me.

Example 1 *In the travel scenario, me is Xiaoyue, she has travel named Travel 1 in the Trentino of Italy from 12:00 to 13:00 of 2th, June 2021. From 12:00 to 12:30 on this day, she takes train 1 from Rovereto to Trento, she sits on seat 1 by herself. From 12:30 to 12:55 on the same day, she walks on Roads 2 from Trento Train Station to Xiaoyue's Home, together with her friend named Haonan. In addition, Xiaoyue talks to Haonan, and Haonan listens to Xiaoyue when they are walking. This scenario involves one event travel and multiple locations.*

In general, in the Figure 1, all elements are divided into Perdurants and Endurants. Perdurants are the Event and Action, and Endurants include Person, Object, and Location. An Event happens in a Location, a Person and Object appear in an Event and Action is in a Person. Most elements' inclusion relationships IN can be represented by the positions of those elements' internal boxes. For the top-level Location and Event, we can add an extra attribute box for them respectively as $InLocation(Location)$ and $InEvent(Event)$, to represent their belongs.

In the rest of this section, we list the attributes of Location, Event, Object, Person and Action, each kind of attributes is represented as a box in the figure. The **Location** has the following attributes:

- Spatial properties: Coordinates (x_i, y_i, z_i), Volume $(\Delta x_i, \Delta y_i, \Delta y_i)$, and InLocation(L_i) shows the super Location for the top level Location;
- Visual properties, namely some properties of Location that can be observed visually;
- Location' functions: $FunctionOf(U)$ with $U \in \{P^1, \ldots, P^k, O^1, \ldots, O^M\}$, which shows location's functions for persons and objects;
- An extra box: including the rest part of the context that happens in the Location.

The **Event** orders by time and is represented by a box with round corners. Events include actual events and virtual events. An Event can have Sub-Events, the Event and Sub-Event have the following attributes:

- Super Event for the top level event: $InEvent(E_i)$, which shows the super event E_i for the current Event;
- Temporal properties: Begin Time - End Time ($Date_i$, $Time_i$ - $Date_j$, $Time_j$), which shows the date and time of the start and end of the event;
- An extra box: including the rest part of the context that happens in the Event.

The **Object** appears in Event and has the following attributes:

- Spatial properties: Coordinates (x_i, y_i, z_i), In/Far/... $(P^i/O^m/...)$;
- Visual properties, namely some properties of Object that can be observed visually;
- Object's functions: $FunctionOf(U)$ with $U \in \{P^1, \ldots, P^k, O^1, \ldots, O^M\}$, which shows Object's functions for persons and other objects.

The **Person** appears in Event and has the following attributes:

$$\{$$

$$\left(D_1, T_1 : \text{super}(L_1), \text{super}(E_1), L_1, E_1, me, \text{coord}_1(me), A_1^{me}, F_1(P^1) : A_1^{P^1}, \dots, F_1(P^k) : A_1^{P^k}, F_1(O^1), \dots, F_1(O^m)\right),$$

$$\left(D_2, T_2 : \text{super}(L_2), \text{super}(E_2), L_2, E_2, me, \text{coord}_2(me), A_2^{me}, F_2(P^1) : A_2^{P^1}, \dots, F_2(P^k) : A_2^{P^k}, F_2(O^1), \dots, F_2(O^m)\right),$$

$$\dots,$$

$$\left(D_n, T_n : \text{super}(L_n), \text{super}(E_n), L_n, E_n, me, \text{coord}_n(me), A_n^{me}, F_n(P^1) : A_n^{P^1}, \dots, F_n(P^k) : A_n^{P^k}, F_n(O^1), \dots, F_n(O^m)\right)$$

$$\}$$

Table 1: The personal streaming context, where E_n is an event, $\text{super}(L_n)$ and $\text{super}(E_n)$ are the super-classes of L_n or E_n, respectively. The set of actions performed by *me* or by the persons based on their functions are denoted with $A_n^k = \{a_1, \dots, a_i\}$, with $k \in \{me, P^1, \dots, P^k\}$.

D_n	T_n	$\text{super}(L_n)$	$\text{super}(E_n)$	L_n	E_n	$\text{coord}_n(me)$	A_n^{me}	$F_n(P^1) : A_n^{P^1}$	$F_n(O^1)$
02/06/2021	12:15	Trentino	Travel 1	Train 1	Take Train	x41, y41, z41	Sitting	*NaN*	RestToolOf(Xiaoyue, Seat 1)
02/06/2021	12:30	Trentino	Travel 1	Roads 2	Walk	x43, y43, z43	Walking, Talking	FriendOf(Xiaoyue, Haonan): Walking, Listening	*NaN*

Table 2: A streaming context matrix representing the travel scenario of Example 1 from the point of view of Xiaoyue, i.e., the observer me. P^1 is Haonan and O^1 is the object "Seat 1". Each column is a property, and every row stands for the current context in a specific timestamp.

- Spatial properties: Coordinates (x_i, y_i, z_i), In/Far/...$(P^i/O^m/...)$;
- Visual properties, namely some properties of Person that can be observed visually;
- Person's functions: $FunctionOf(U)$ with $U \in \{P^1, \dots, P^k, O^1, \dots, O^M\}$, which shows Person's functions for other persons and objects;
- Internal states: Physical states $(InPain())$, Mental states $(InMood(), InStress())$;
- An extra box: including the Actions of Person.

The **Action** is similar with Event, it orders by time and is represented by a box with round corners. An Action can has many Sub-Action, the Action and Sub-Action have following attributes, each attribute has a box for itself.

- Temporal properties: Begin Time - End Time $(Date_i, Time_i - Date_j, Time_j)$, which shows the date and time of the start and end of the Action;
- Visual properties, namely some properties of Action that can be observed visually;
- Means of the Action: Means$(O^m/...)$;
- Sub-action: Action i;
- Action's functions: $FunctionOf(U)$ with $U \in \{P^1, \dots, P^k, O^1, \dots, O^M\}$, which shows Action's functions for persons and objects.

4 The current context as a Knowledge Graph

We use the *Entity Type Graph* (ETG) and the *Entity Graph* (EG) in ontology to represent the context. ETG is a knowledge graph where nodes are entity types, which are further decorated with data properties. The object properties are presented in the graph representing the relations among the entity types.

In Figure 2, white box nodes are entity types that include data properties with data types, the green box nodes enumerates the values of data property. The object properties are connecting all entity types, represented by diamond nodes with arrows. The inheritance relation in the ETG is represented by an arrow from the super-class to the sub-class, and the sub-class inherits all the data properties and object properties of its super-class. The EG populates the entity types and properties defined in the ETG with specific values. It is a data graph where nodes are entities that are connected by object property values representing the relations. Each entity further includes data property values. The streaming context can be viewed as a stream of EGs, in which each EG describes the context at a different time.

We design an EG example as Figure 3 according to the scenario in Example 1. The graph represents the context around "Me", this contains entities shown by nodes, e.g., "Smartphone", "Talking", "Walk". Also, we can see object property values.

5 Learning Context

AI applications like smart personal assistants provide a service to the users based on their context. The context information is usually not available to the machine, and hence it has to infer the location or the activity of the user from sensor data (e.g., GPS, accelerometer, nearby Bluetooth devices). In our scenario, Xiaoyue is carrying a smartphone that generates a stream of sensor readings, and she annotates the data by answering questions about her context, e.g., "Where are you"

Location (Trentino)
Coordinates: x40, y40, z40 Volume: Δx40, Δy40, Δz40 InLocation(Italy)
Area: 13607 square kilometers
LivingPlaceOf (People)

Event (Travel 1)
InEvent (Going Home)
from: 2th, June, 2021, 12:00; to: 2th, June, 2021, 13:00

Location (Train 1)	Location (Roads 2)
Coordinates: x47, y47, z47 Volume: Δx47, Δy47, Δz47 FromLocation: Rovereto ToLocation: Trento	Coordinates: x48, y48, z48 Volume: Δx48, Δy48, Δz48 FromLocation: Trento Train Station ToLocation: Xiaoyue's Home
Area: 100 square metres	Area: 1000 square metres
TransportationPlaceOf (People)	WallkingRoadOf (People)

Event1 (Take Train)
from: 2th, June, 2021, 12:00; to: 2th, June, 2021, 12:30

Object (Seat 1)
Coordinates: x42, y42, z42
Below(Xiaoyue)
Model: IVECO Seat
RestToolOf (Xiaoyue)

Me (Xiaoyue)
Coordinates: x41, y41, z41
On (A Seat 1 of Train 1)
Gender: Female
Occupation: Student
PassengerOf (Train 1)
InPain (None)
InMood (Good)
InStress (None)

Action (Sitting)
from: 2th, June, 2021, 12:00
to: 2th, June, 2021, 12:30
Informal Sitting
A Seat 1 of Train 1
None SubActions
Rest

Event2 (Walk)
from: 2th, June, 2021, 12:30; to: 2th, June, 2021, 12:55

Me (Xiaoyue)
Coordinates: x43, y43, z43
Left (Haonan)
Gender: Female
Occupation: Student
FriendOf (Haonan)
InPain (Head)
InMood (Good)
InStress (None)

Action (Walking)
from: 2th, June, 2021, 12:30
to: 2th, June, 2021, 12:40
Slow Walking
Foot
None SubActions
Moving

Action (Talking)
from: 2th, June, 2021, 12:35
to:2th, June, 2021, 12:40
Informal Talking
Mouse
None SubActions
Sharing ideas

Person (Haonan)
Coordinates: x44, y44, z44
Right (Xiaoyue)
Gender: Male
Occupation: Student
FriendOf (Xiaoyue)
InPain (None)
InMood (Good)
InStress (None)

Action (Walking)
from: 2th, June, 2021, 12:30
to: 2th, June, 2021, 12:40
Slow Walking
Foot
None SubActions
Moving

Action (Listening)
from: 2th, June, 2021, 12:35
to: 2th, June, 2021, 12:40
Informal Listening
Ears
None SubActions
Hearing ideas

Figure 1: One Event Multiple Locations: a travel around me. Representation of the Example 1.

and "What are you doing?". The sensor data are aggregated in time windows generating a stream of instances (e.g., the average number of nearby Bluetooth devices in the last 30 minutes). On each incoming instance, the machine decides whether to query the user to acquire the labels. The machine learning technique defines the query strategy, and in the simplest case, the labels are acquired on every instance.

The user's context recognition is a supervised learning problem in which an instance $\mathbf{x}$ is associated to multiple concepts $\mathbf{y}$ (aka classes in machine learning). The concepts are organized in a ground-truth hierarchy $\mathcal{H} = (C, I)$, which is a direct acyclic graph (DAG) where nodes $C = \{1, \ldots, c\}$ are the concepts and edges $I \subset C \times C$ are *is-a* relations, i.e., $I = \{(c_i, c_j) \mid c_i, c_j \in C \text{ and } c_i \text{ is a child of } c_j\}$ [Silla and Freitas, 2011]. The labels of the instances are indicator vectors $\mathbf{y} \in \{0, 1\}^c$, where the i-th elements is 1 if $\mathbf{x}$
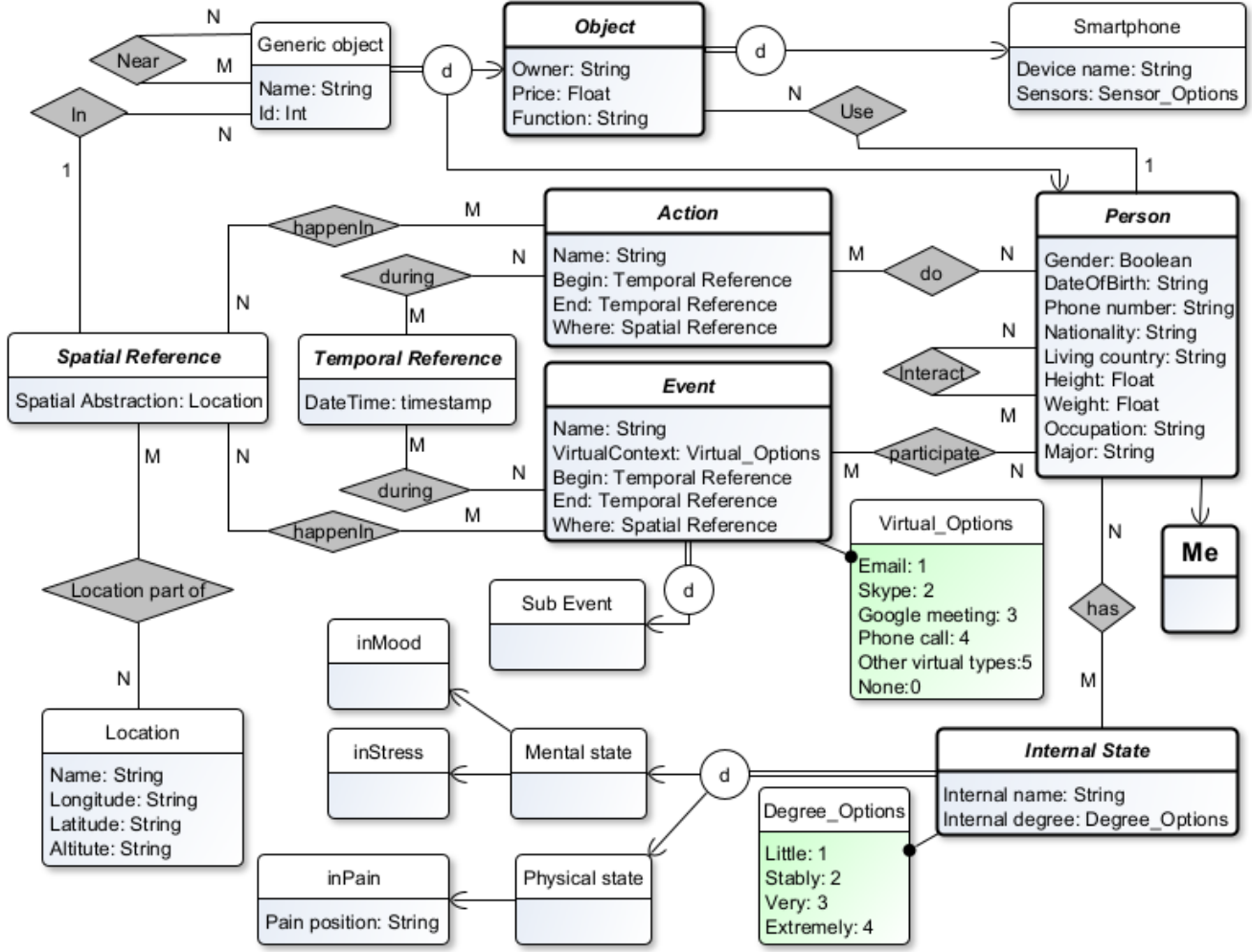
Figure 2: An ETG representing partially the personal context in our travel example.

belong to i-th concept in $\mathcal{H}$ and 0 otherwise. The machine is trained on a stream of examples $\mathbf{z}_t = (\mathbf{x}_t, \mathbf{y}_t)$ drawn from a ground-truth distribution $P(\mathbf{X}, \mathbf{Y})$ that is always consistent with a ground-truth hierarchy $\mathcal{H}$, i.e., if there is an edge from class c_i to class c_j, then $y^i = 1$ implies $y^j = 1$ and conversely $y^j = 0$ implies $y^i = 0$. The goal of this hierarchical classification tasks is to learn a classifier that recognize well the context on future sensor readings.

The ETG and EG introduced in Section 4 can be used as prior knowledge about the structure of the hierarchy. They encode the available information about the user and the world. Algorithm 1 shows the conversion from ETG and EG to a DAG $\mathcal{H}$. The first step is to convert each entity type (etype) in the ETG as a node in $\mathcal{H}$ (lines 3 - 5). Second, each entity in EG also becomes a node that is added as a child of the node referring to the etype of the entity (lines 6 - 10). The hierarchy encodes the information about the current user, so the *Me* etype and the corresponding entity (e.g., Xiaoyue entity in Figure 3) are not considered.

The properties of the ETG are grouped in properties that are context depends and properties that are static. The value of the former changes every time the users change their context and, in Figure 2 are $Q = \{near, use, interact, in, do, happenIn, during, participate\}$. For instance, if Xiaoyue travel from the city of Trento to Rovereto, the *in* property will change accordingly. In contrast, the fact that Trento is *partOf* Italy can be assumed to remain valid even if user's context is changed. This distinction is necessary since the value of context-dependent properties are derived from the output of the context recognition task (e.g., the machine recognizes that the user is in the city of Trento and thus updates the *in* property accordingly). The object properties that are not contextual are converted as follow. Given a object property $p \in \{isA, partOf, has\}$ that links the etype A to B, then the node referring to etype A becomes a child of the node of etype B. For the other object properties, a new node referring to the property is added as child of the codomain etype node (lines 13 - 19). For every object property value i of the prop-
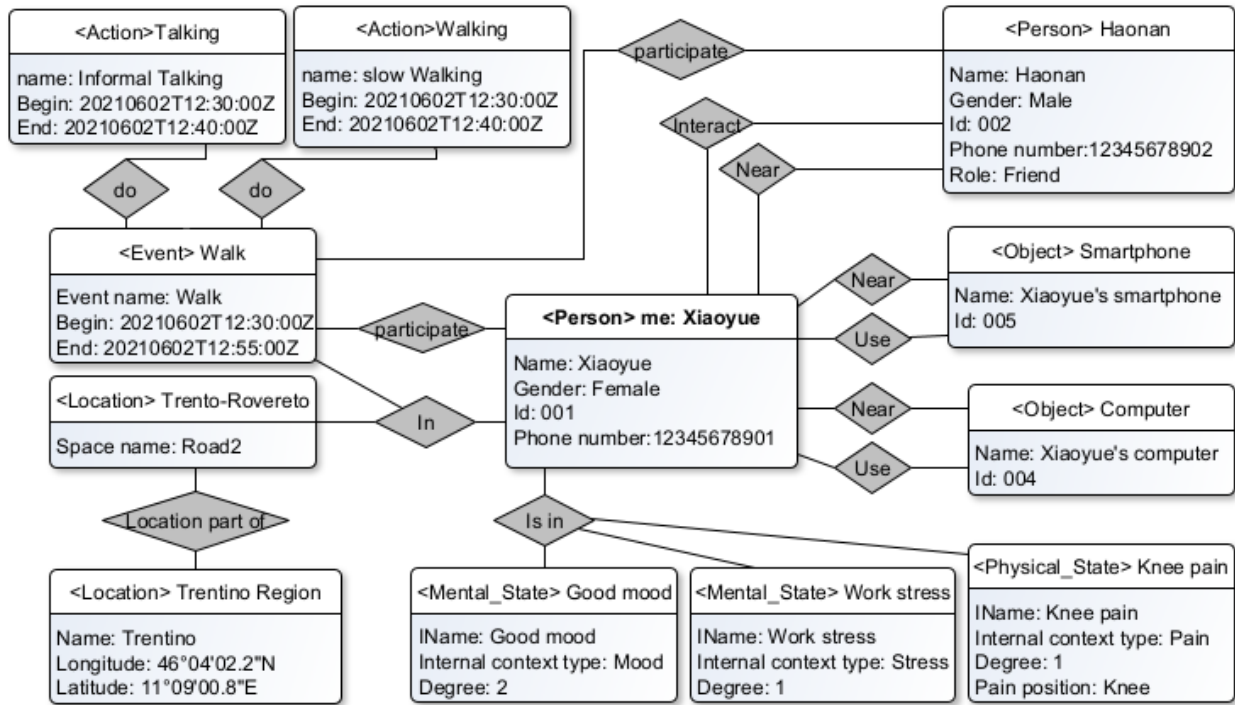
Figure 3: An EG representing partially the scenario about travelling of Example 1.

erty p linking the entities a and b, a new node c_i is added as child of c_p (i.e., the node referring to the property P), and as parent of c_b (i.e., the node pointing to the entity b) (lines 20 - 25). Finally, all nodes that does not have a parent are connect to the root node and the transitive reduction is applied (lines 27 - 30). Figure 4 shows an extract of the DAG resulting from applying Algorithm 1 on EG and ETG presented in Section 4.

Every node in $\mathcal{H}$ has a unique identifier that is used to reference back to the ETG and EG. The node name can be translated into a human-readable text that is used to interact with the user. This aspect is left as future work. The concept hierarchy available at the beginning can evolve over time and has to be continually updated. This aspect has been defined as knowledge drift and is addressed in [Bontempelli *et al.*, 2021].

6 Related Work

Considering our novel context modelling of the personal context, the context learning, and the architecture presented in this paper, we can show as use cases the papers that individually compartmentalizes examples for these improvements. Current works on context recognition focus on learning the relationship between the input data (sensor readings) and the target concepts (context). The structure of the context is implicitly learned by the implemented machine learning algorithm during the training phase. The new parts are described in Section 3 and compared with related work we can outline the following: 1) the connection between the learned context model; 2) The extension of new dimensions
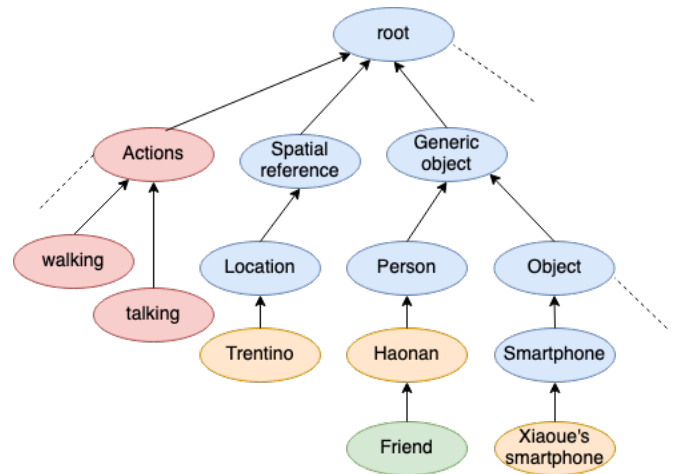

Figure 4: A partial representation of the DAG obtained from ETG and EG examples in Section 4. Orange nodes are derived from EG. Blue nodes are derived from ETG. Red nodes are actions and green nodes are the functions.

Algorithm 1 Convert ETG and EG in a DAG.

Inputs: ETG, EG, and the set of context dependent object properties Q
Outputs: $\mathcal{H} = (C, I)$

1: $C \leftarrow \emptyset$
2: $I \leftarrow \emptyset$
3: **for** every etype $A \neq Me$ in ETG **do**
4:     let c_A being the node of etype A
5:     $C = C \cup \{c_A\}$

6: **for** every entity a such that $\neg\, Me(a)$ in EG **do**
7:     let c_a being the node of the entity a
8:     let c_A being the node of the etype of a
9:     $C = C \cup \{c_a\}$
10:     $I = I \cup \{(c_a, c_A)\}$

11: **for** every object property p in ETG such that $p \notin \mathcal{Q}$ **do**
12:     let c_p being the node of p
13:     **for** every $p(A, B)$ in ETG **do**
14:         let c_A and c_B being the nodes of entity A and B respectively
15:         **if** $p \in \{isA, partOf, has\}$ **then**
16:             $I = I \cup \{(c_A, c_B)\}$
17:         **else**
18:             $C = C \cup \{c_p\}$
19:             $I = I \cup \{(c_p, c_B)\}$
20:     **for** every $p(a, b)$ in EG **do**
21:         let c_i being the node encoding $p(a, b)$
22:         let c_b being the node of entity b
23:         $C = C \cup \{c_i\}$
24:         $I = I \cup \{(c_i, c_p)\}$
25:         $I = I \cup \{(c_b, c_i)\}$

26: let c_0 being the root node
27: $C = C \cup \{c_0\}$
28: **for** every node $c \in C$ that $\mathrm{parent}(c) = \emptyset$ **do**
29:     $I = I \cup \{(c, c_0)\}$ add an edge from c to c_0
30: apply transitive reduction on $\mathcal{H} = (C, I)$

and classification of context, i.e. Internal State, Functions, Actions. The Algorithm 1 bridges the gap between the two modellings allowing context recognition to leverage both machine learning solutions and knowledge graph representation (ETG and EG). For instance, the ETG and EG can generate the questions needed by the machine to interact with the users. For instance, the personal context recognition model shown in [Giunchiglia *et al.*, 2018], showed to be a good approach to increase the accuracy of the context recognition algorithms. Our work described in Section 3 enhance the representation of the personal context with the hope to perform better in real-life scenarios. In practice, existing approaches for context recognition using batch or streaming sensor data [Vaizman *et al.*, 2018; Bontempelli *et al.*, 2020; Zeni *et al.*, 2019] do not leverage on an explicit context modeling. The context representation is implicit in the la-

bels used to train the machine learning model. The context formalization introduced in this work can be used to structure the output of these machine learning models according to our representation. Moreover, it can help machine learning approaches to interact with users. For instance, the machine can ask if Haonan is a friend of Xiaoyue since they are walking together. Approaches that use active learning strategies (e.g., [Settles, 2009; Hoque and Stankovic, 2012; Hossain *et al.*, 2017]) can benefit of our representation.

Also, existing frameworks for creating context-aware mobile applications, such as Ferreira *et al.*, do not consider the modelling of the context.

7 Conclusion and Future Work

In this paper, we moved forward with a better representation of personal context in real-life environments. We proposed an improved representation of the personal context, adding the internal state, functions, and actions. The learning aspect of our work is the formal definition of an algorithm to transform the streaming input data to ML algorithms. We will put all these components in the system architecture.

In comparison with the work on personal context recognition for human-machine collaboration [Giunchiglia *et al.*, 2018], we have shown an enhancement related to the model representation of the personal context.

Additionally, we have shown how our novel personal context representation can also be leveraged by machine learning algorithms to include prior knowledge about the structure of their output and can be used to drive the interaction with the user. Future work will focus on evaluating the impact of our formalization on an existing approach for fixing mislabeled data when learning the users' contexts [Zeni *et al.*, 2019].

The next step is to propose and implement a modern design of the services related to iLog [Zeni *et al.*, 2014] by a centralised streaming system and linking the personal context data collections with other distributed services of machine learning. This implementation will allow us to test and evaluate our novel context model in near real-life scenarios.

References

[Bontempelli *et al.*, 2020] Andrea Bontempelli, Stefano Teso, Fausto Giunchiglia, and Andrea Passerini. Learning in the wild with incremental skeptical gaussian processes. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 2886–2892, 7 2020.

[Bontempelli *et al.*, 2021] Andrea Bontempelli, Fausto Giunchiglia, Andrea Passerini, and Stefano Teso. Human-in-the-loop handling of knowledge drift. *arXiv preprint arXiv:2103.14874*, 2021.

[Chang *et al.*, 2017] Yung-Ju Chang, Gaurav Paruthi, Hsin-Ying Wu, Hsin-Yu Lin, and Mark W Newman. An investigation of using mobile and situated crowdsourcing to collect annotated travel activity data in real-word settings. *International Journal of Human-Computer Studies*, 102:81–102, 2017.

[Ferreira *et al.*, 2015] Denzil Ferreira, Vassilis Kostakos, and Anind K. Dey. Aware: mobile context instrumentation framework. *Frontiers in ICT*, 2:6, 2015.

[Freedman *et al.*, 2013] Vicki A. Freedman, Jessica Broome, Frederick Conrad, and Jennifer C. Cornman. Interviewer and respondent interactions and quality assessments in a time diary study. *Electronic international journal of time use research*, 10(1):55, 2013.

[Giunchiglia and Fumagalli, 2017] Fausto Giunchiglia and Mattia Fumagalli. Teleologies: Objects, actions and functions. In *International conference on conceptual modeling*, pages 520–534. Springer, 2017.

[Giunchiglia *et al.*, 2017] Fausto Giunchiglia, Enrico Bignotti, and Mattia Zeni. Personal context modelling and annotation. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 117–122. IEEE, 2017.

[Giunchiglia *et al.*, 2018] Fausto Giunchiglia, Mattia Zeni, and Enrico Big. Personal context recognition via reliable human-machine collaboration. In *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 379–384. IEEE, 2018.

[Hoque and Stankovic, 2012] Enamul Hoque and John Stankovic. Aalo: Activity recognition in smart homes using active learning in the presence of overlapped activities. In *2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, pages 139–146. IEEE, 2012.

[Hossain *et al.*, 2017] HM Sajjad Hossain, Md Abdullah Al Hafiz Khan, and Nirmalya Roy. Active learning enabled activity recognition. *Pervasive and Mobile Computing*, 38:312–330, 2017.

[Kwapisz *et al.*, 2011] Jennifer R. Kwapisz, Gary M. Weiss, and Samuel A Moore. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter*, 12(2):74–82, 2011.

[Settles, 2009] Burr Settles. Active learning literature survey. 2009.

[Silla and Freitas, 2011] Carlos N Silla and Alex A Freitas. A survey of hierarchical classification across different application domains. *Data Mining and Knowledge Discovery*, 2011.

[Tourangeau *et al.*, 2000] Roger Tourangeau, Lance J. Rips, and Kenneth Rasinski. The psychology of survey response. 2000.

[Vaizman *et al.*, 2018] Yonatan Vaizman, Nadir Weibel, and Gert Lanckriet. Context recognition in-the-wild: Unified model for multi-modal sensors and multi-label classification. 1(4), January 2018.

[Vapnik, 2013] Vladimir Vapnik. *The nature of statistical learning theory*. Springer science & business media, 2013.

[Webb, 2003] Andrew R. Webb. *Statistical pattern recognition*. John Wiley & Sons, 2003.

[West and Sinibaldi, 2013] Brady T. West and Jennifer Sinibaldi. The quality of paradata: A literature review. *Improving surveys with paradata*, pages 339–359, 2013.

[Zeni *et al.*, 2014] Mattia Zeni, Ilya Zaihrayeu, and Fausto Giunchiglia. Multi-device activity logging. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous Computing: Adjunct publication*, pages 299–302, 2014.

[Zeni *et al.*, 2019] Mattia Zeni, Wanyi Zhang, Enrico Bignotti, Andrea Passerini, and Fausto Giunchiglia. Fixing mislabeling by human annotators leveraging conflict resolution and prior knowledge. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(1):1–23, 2019.