

# Knowledge Inference Through Analysis of Human Activities

Leandro O. Freitas<sup>1</sup>, Pedro Rangel Henriques<sup>1</sup> and Paulo Novais<sup>1</sup>

<sup>1</sup>ALGORITMI Center, University of Minho, Braga, Portugal

leanfrts@gmail.com, {prh, pjon}@di.uminho.pt

**Keywords:** Activity analysis, Attribute grammar, Uncertainty handling.

**Abstract:** Monitoring human activities provides context data to be used by computational systems, aiming a better understanding of users and their surroundings. Uncertainty still is an obstacle to overcome when dealing with context-aware systems. The origin of it may be related to incomplete or outdated data. Attribute Grammars emerge as a consistent approach to deal with this problem due to their formal nature, allowing the definition of rules to validate context. In this paper, a model to validate human daily activities based on an Attribute Grammar is presented. Context data is analysed through the execution of rules that implement semantic statements. This processing, called semantic analysis, will highlight problems that can be raised up by uncertain situations. The main contribution of this paper is the proposal of a rigorous approach to deal with context-aware decisions (decisions that depend on the data collected from the sensors in the environment) in such a way that uncertainty can be detected and its harmful effects can be minimized.

## 1 Introduction

Activity recognition systems have to monitor user's movements aiming to understand his actions to build models capable of being processed. Attribute Grammars (AGs) emerges as a good approach to define models to validate context data related to human actions. Despite having their origin coined to programming languages specification (Knuth, 1968), they are already used in Intelligent Environments field to analyse the structure of human tasks (Freitas et al., 2020). AG provides a formal and strict structure, and specifies semantics through the definition of attributes for symbols and evaluation or validation rules associated with production rules.

Many times, uncertainty in context-aware systems is neglected. Explicit representation of it allows the quantification of its interference in the dataset (Camara et al., 2018). Different sources of data, with a wider diversity of sensors, require optimized reasoning, to help decrease the negative impact on users (Tian et al., 2018).

In this paper, an approach to perform knowledge inference over human daily activities is presented. The validation of activities was done through the definition of an Attribute Grammar, ensuring the correctness of data structure. The proposal was validated through two public datasets, presented as case studies.

The paper is organized as follows: Section 2 presents related works to this approach. The datasets and the attribute grammar developed to validate the context data are described in Section 3. Section 4 describes in details the approach proposed to deal with uncertainty in context-aware systems with its validation. Section 5 presents conclusions of the study and future research directions.

## 2 Related Work

Spacial-temporal aspects are considered in (Chen et al., 2018) to divide uncertainty into: *Point Of View*, including problems related to the conversion of sensor data into global position; *Temporal*, related to time lapses; *Transaction Attribute*, considering that attribute values may have different meanings depending on the source; *Location*, referring to problems with GPS logs with missing data, and; *Identity*, including problems with user's personal data. In (Chahuara et al., 2017), a framework for decision making in smart houses through voice recognition was presented. Hierarchic knowledge model for inference under situations with incomplete or outdated data was developed. Uncertainty is tackled through the learning process and the Markov model to the recognition of tasks execution.

These contributions differ from this paper once

here the formal model is based on an attribute grammar. It validates the context data and provides a set of semantic rules that help the identification of dubious situations. Other papers like (Bobek and Nalepa, 2017) and (Noor et al., 2016) can also be related to the proposal of this paper. However due to the sake of limited space, they will not be discussed in details.

### 3 Validation of human activities

This section firstly introduces the datasets that were used for the development of the work. In spite of using public datasets, the experiments were created based on an independent research problem, with independent goals. After that, the attribute grammar is described.

#### 3.1 Dataset 1: Human activities recognition dataset

This dataset of human activities in a smart house was created by (Cook and Schmitter-Edgecombe, 2009), considered a reference in the field of context awareness, at the Washington State University, USA. The activities monitored were *phone calling*, *washing hands*, *cooking*, *eating* and *cleaning*. The dataset has six thousand and four hundred and ninety-seven registers, acquired through motion, item, water and burner sensors.

During the *phone call* activity the volunteers to receive information about the tasks they should execute after finishing the call. Besides that, all the participants were asked to cook the same meal (oatmeal), following the same recipe. This is important because one of the parameters taken into consideration for the experiments' analysis is the time spent in each activity. And, specifically for these two, the time may naturally vary. For the development of this work, it was assumed that the information from the dataset refers to the same user, i.e., supposing that only one person was monitored. This makes possible to simulate different behaviours of the same user, taking into consideration the context, preferences and needs.

No other information about the work developed by (Cook and Schmitter-Edgecombe, 2009) was used in the research presented here.

#### 3.2 Dataset 2: Activity Daily Living and binary sensors

The dataset was developed by (Ordóñez et al., 2013) at the University of Madrid, Spain. It contains data

captured from two users during a total of 35 days. For the development of this work it was decided to use data from only one user, since one of the goals is to identify patterns of behaviours and uncertainties related to it. Thus, the chosen dataset contains two thousand and three hundred and thirty five registers, collected during 22 days. The dataset was properly normalized and prepared.

The house used in the experiment have five rooms: Entrance, Living, Bedroom, Bathroom and Kitchen. The monitored activities were *SpareTime/TV*, *Grooming*, preparing *Breakfast*, *Lunch*, *Dinner* and *Snack*, *Sleeping*, *Toileting*, *Showering* and *Leaving*. To recognize the activities binary sensors were used (PIR - Passive Infra-Red, magnetic, flush, pressure and electric).

The registers contain information about the *Start* and *End* time of activities, *Location* of the sensor (object in the house that had a sensor installed), *Type* of the sensor, *Place* (room of the house) and the *Activity's* name. No other information from the experiments developed by (Ordóñez et al., 2013) was used.

#### 3.3 Attribute grammar for context validation

The main advantage of attribute grammars is that they allow lexical, syntactic and semantic data verification. According to (Burger et al., 2010), in the first step (lexical analysis), the data is transformed into terminal symbols and will represent leaves (terminal nodes) in an Abstract Syntax Tree (AST). Then, non-terminal symbols are organized into a hierarchical structure, creating the AST, (syntactic analysis). At last (semantic analysis phase), the actual data associated with the symbols in the AST nodes via inherited or synthesized attributes is processed, considering pre-established semantic rules, to extract (infer) new contextual information and ensure the correctness of the input as required to produce the desired output.

Below, the some of the grammar production rules (the first ones), that stablish the input data structure (define the language syntax), are presented.

**p1:** *set\_of\_activities* → *activity*+

**p2:** *activity* → *activity\_name*,

*record*, *status*, *duration\_status*, *measurements*

**p3:** *record* → *activity\_time*, *sensor\_ID*+

**p4:** *activity\_time* → *date*, *hour*

**p5:** *sensor\_ID* → *location*, *type\**

The first production rule (**p1**) represents the axiom of the grammar. In this case, it refers to the definition of a set of activities. This symbol represents the root of the AST and comprises all the activities. This will

be the reference for the output, after validating the tuples from the dataset. At last, it can be composed of one or more *activity* symbols. The production **p2** represents the structure of each activity in the dataset.

Rule **p3** (*record*) defines the details about sensors used and data collected. This includes the time the activity was performed and the sensors used for it. Each sensor has an identification, which may have different structure depending on the dataset. For instance, it can be an alphanumeric string in one dataset and in other it can be the composition of two or more values from different columns (**p5**). The symbol *status* keeps the information about the beginning and ending time of an activity. During the execution of an activity, the system can receive several data from the sensors, which are analysed by the symbol *measurements*. The symbol *activity\_time*, according to **p4**, is composed of the activity's beginning date and hour. The symbols *activity\_name*, *date*, *hour*, *location* and *type* store terminal values.

## 4 Reasoning over context data

Inference rules are applied to datasets aiming deduce new information and identify patterns of behaviour. From that, it is possible to tackle uncertainties which may be caused by sensor failures or unexpected changes of behaviour. This section presents two scenarios to illustrate the approach. Based on the grammar productions (syntactic rules) and taking into account the attributes associated with the symbols, it is possible to create any set of attribute evaluation rules. The following rules (to evaluate attributes *duration*, *durationStatus*, *record*, *status*) were created to be applied in the case studies.

1. `activity.duration = getSystemTime(TEXT.value) - record.hour;`
2. `activity.durationStatus = durationAnalysis (activity.duration, expectedDuration[]);`
3. `activity.record = featureAnalysis(activity_time, activity.durationStatus);`
4. `activity.status = uncertaintyAnalysis(activity.record, expectedTotalAmount);`

Rule 1 computes in the total time spent by the user to finish the activity. Rule 2 analyses if this duration is within an expected time interval. The total amount of activities performed in a specific day is calculated in rule 3. The result of it, is used in the *uncertainty-Analysis* (rule 4), described in details in subsection 4.2.

## 4.1 Case study: abnormal amount of activities per day

The experiment described in this section was conducted in both of the datasets. Details are provided in the paragraphs below.

### 4.1.1 Dataset 1

In this experiment it was analysed the total of activities performed per day. A sudden change of these values along the days may represent change related to behaviour patterns or to the user's health. Figure 1 presents the values computed when analysing dataset 1.

The graph shows good consistency regarding the number of activities executed per day. The majority of the days, 84.61%, the user performed between five and ten activities. Despite having an increased number of activities in the last day of monitoring, it is reasonable to say that it does not affect his standard behaviour, once this represents an increasing of five tasks (16.67%). Semantically, this graph says that the user has a regular routine, probably using a constant time to perform his tasks.

The uncertain situation is evidenced in *day 4* where the number of activities was around 66.66% more than what can be defined as the normal amount in a regular day. There are several reasons for that, but considering only the information from the dataset, it is impossible to choose one with a high level of confidence. For instance, it might refer to a weekend day where the user usually performs more tasks (e.g. receive visitors or clean the house).

Regardless, as the number of activities in this day has a high discrepancy when compared to other days it is imperative for the system to be capable of identifying it. This is the first step in order to minimize the impact of uncertainty in the domain. By the identification of its existence, it will be possible to circumvent the situation with security measures.

### 4.1.2 Dataset 2

The second dataset differs from the first in the sense that for the identification of the days it was used date format, instead of sequential numbers. However, this does not interfere in the analysis in any way. Figure 2 present the values computed when analysing dataset 2.

The graph shows that in most of the days the user performs between 88 and 120 daily activities (72.72%). Considering that this number represents the great majority of the days, it is reasonable to say that his daily routine follow a pattern. Besides that,

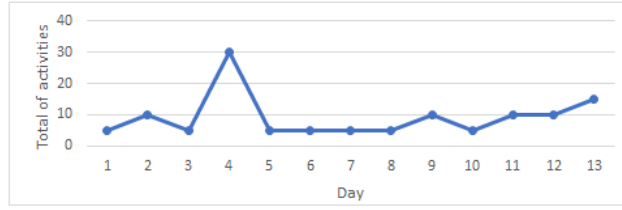


Figure 1: Total of activities per day

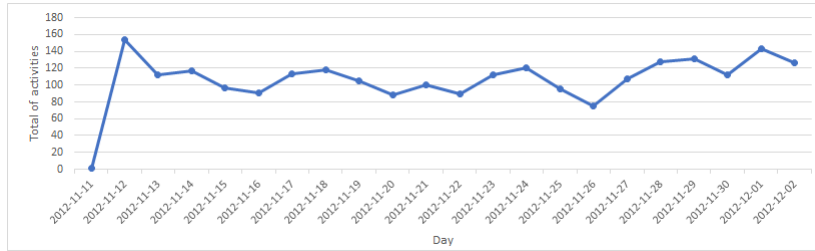


Figure 2: Total of activities per day

in the days that he performed a lower amount of activities (75), this number is still close to the majority. The context-aware system could analyse this situation as abnormal or not, depending on its objectives. The first day has only one activity performed because the monitoring started close to midnight.

Regarding the days where the user performed more activities than the average, those that present higher discrepancy were recorded in 2 days (2012-11-12, 154 activities and 2012-12-01, 143 activities). As aforementioned, to consider this difference as a problem, it is important to analyse the system's goals. Anyway, these two days correspond to the most critical situations.

Nevertheless, the graph shows that the user, for some reason, changed his routine. This can be related to a variety of reasons, including system failures. Cases like this generate uncertainty. The following section describes an approach to deal with uncertainty in Activity Daily Living environments.

## 4.2 Uncertainty handling approach

In spite of the two datasets used in this study have different structures, both of them represent the same type of domain: smart home. This means that they have similarities and these can be analysed following the same approach. This section tackles the problem of uncertainty, which was identified on the analysis of both datasets. Misinterpretation of sensor signs results in unreliable decision making. Hardware failures of sensors and communication problems may influence this. Thus, the goal is to analyse the total of activities performed in one day and answer questions like: why the user performed (or, why the system de-

tected) a different number of activities than the usual? Algorithm 1 describes how this uncertainty is tackled.

---

### Algorithm 1: uncertaintyAnalysis (activity.record)

---

```

1 for allDays do
2   | expectedTotalAmount[] =
   | countActivities(activity.record)
3 end
4 totalInCurrentDay ←
  countActivities(activity.record)
5 if !checkTotalActivities(totalInCurrentDay,
  expectedTotalAmount[]) then
6   | SD ← runSensorsDiagnosis(totalActivities)
7   | if SD then
8     | messageAlert(SD)
9   | end
10  | GD ← runGapsDiagnosis()
11  | if GD then
12    | userInteraction(GD)
13  | end
14 end

```

---

The method *countActivities* calculates the total amount of activities performed in one day. This process is done for all registered days (lines 1 and 2). The results are stored in a list and represent the *expectedTotalAmount* of activities. They represent the normal amount of activities of a day. After that, the same process is executed for one specific day, which is the object of analysis. The result is stored in the variable *totalInCurrentDay* (line 4). The method *checkTotalActivities* (line 5) analyses the *totalInCurrentDay* against the list of expected values. A return *false* indicates a discrepancy between the amount of activities performed by the user and the expected total for

him to perform. If a considerable difference between them is evidenced, some actions must be taken in order to prevent the system of using inconsistent or erroneous data. The first one is to run diagnosis tests aiming to find hardware failures (line 6). The method *runSensorDiagnosis* starts communications with the sensors to ensure they are working as expected. If problems are found, the system send warnings to the user (method *messageAlert* of lines 7 and 8) about it for him to solve the problem (e.g. repair or replace the sensor). Then, the system runs diagnosis tests (line 10) seeking information gaps and their influence in decision making. The method *runGapsDiagnosis* searches for any type of data that is necessary to build contexts for this activity. These gaps may influence the composition of contexts, once the system will lack context data to build situations. One approach to deal with this is asking the user for help, to acquire more data or to eliminate ambiguities (Lim and Dey, 2011). Thus, if the system detects gaps of information, it starts an interaction of such kind (method *userInteraction* of lines 11 and 12).

## 5 Conclusion

This paper presents a formalization of human daily activities through attribute grammars. Another contribution is related to how uncertain situations can be evidenced through the execution of inference rules. Uncertainty is naturally embedded in daily life. Thus, the challenge relies on finding ways of minimizing its negative impact on the system's behaviour and in the user's routine. Thus, the next steps of the research include the improvement of uncertainty analysis through the combination of machine learning algorithms and the validation provided by the grammar.

## Acknowledgements

"This work has been supported by national funds through FCT – Fundação para a Ciência e Tecnologia within the Project Scope: UID/CEC/00319/2019."

## REFERENCES

Bobek, S. and Nalepa, G. J. (2017). Uncertainty handling in rule-based mobile context-aware systems. *Pervasive and Mobile Computing*, 39:159 – 179.

Burger, C., Karol, S., and Wende, C. (2010). Applying attribute grammars for metamodel semantics. *ECOOP*

2010 Workshop Proceedings - International Workshop on Formalization of Modeling Languages, FML'10.

Camara, J., Peng, W., Garlan, D., and Schmerl, B. (2018). Reasoning about sensing uncertainty and its reduction in decision-making for self-adaptation. *Science of Computer Programming*, 167:51 – 69.

Chahuaara, P., Portet, F., and Vacher, M. (2017). Context-aware decision making under uncertainty for voice-based control of smart home. *Expert Systems with Applications*, 75:63 – 79.

Chen, S., Wang, Z., Liang, J., and Yuan, X. (2018). Uncertainty-aware visual analytics for exploring human behaviors from heterogeneous spatial temporal data. *Journal of Visual Languages and Computing*, 48:187 – 198.

Cook, D. J. and Schmitter-Edgecombe, M. (2009). Assessing the quality of activities in a smart environment. *Methods of information in medicine*, 48 5:480–5.

Freitas, L. O., Henriques, P. R., and Novais, P. (2020). Attribute grammar applied to human activities recognition in intelligent environments. In Novais, P., Lloret, J., Chamoso, P., Carneiro, D., Navarro, E., and Omatu, S., editors, *Ambient Intelligence – Software and Applications – 10th International Symposium on Ambient Intelligence*, pages 62–70, Cham. Springer International Publishing.

Knuth, D. E. (1968). Semantics of context-free languages. In *Mathematical Systems Theory*, pages 127–145.

Lim, B. Y. and Dey, A. K. (2011). Investigating intelligibility for uncertain context-aware applications. In *Proceedings of the 13th International Conference on Ubiquitous Computing*, UbiComp '11, pages 415–424, New York, NY, USA. ACM.

Noor, M. H. M., Salcic, Z., and Wang, K. I.-K. (2016). Enhancing ontological reasoning with uncertainty handling for activity recognition. *Know.-Based Syst.*, 114(C):47–60.

Ordóñez, F. J., De Toledo, P., and Sanchis, A. (2013). Activity recognition using hybrid generative/discriminative models on home environments using binary sensors. *Sensors*, 13(5):5460–5477.

Tian, W., Heo, Y., de Wilde, P., Li, Z., Yan, D., Park, C. S., Feng, X., and Augenbroe, G. (2018). A review of uncertainty analysis in building energy assessment. *Renewable and Sustainable Energy Reviews*, 93:285 – 301.