VPM: Analyzing Human Daily Habits through Process Discovery

Francesco Leotta[a], Silvestro V. Veneruso[a]

[a]*Dipartimento di Ingegneria Informatica Automatica e Gestionale "A. Ruberti", Sapienza Università di Roma, via Ariosto 25, Roma, Italy*

Abstract

Models usually employed for Ambient Intelligence (AmI) in smart homes are usually obtained directly from sensor logs composed by timestamped sequences of sensor measurements. Such approaches, still effective at different tasks, have the drawback of producing representations difficult to read and validate. In this paper we propose a tool, called Visual Process Maps (VPM), intended to allow the analysis of human routines at the human action level thanks to log preprocessing and the application of process discovery.

1. Introduction

The term *smart space* refers to an environment enriched with a set of devices (e.g., sensors and actuators) which aims at providing intelligent services to the human user, realizing the paradigm known as *ambient intelligence* (AmI) [1]. In order to provide people with these automatic or semi-automatic utilities, an AmI system acquires data from the environment in the form of a *sensor log*, i.e., a sequence of measurement values acquired from sensors.

The literature in the area proposes several solutions to analyze sensor logs and to automatically perform actions on the basis of the current context, user preferences and habits. All of these approaches are based on *models* which describe the relationship between *(i)* the behavior of a smart space's inhabitant(s), *(ii)* the specific contextual state of the environment, and *(iii)* a portion of sensor measurements from the log. Most of these techniques anyway, and especially those based on machine learning techniques (the so called *learning-based methods*), only rely on raw sensor measurements, whereas *human actions* are usually neglected, making it difficult to visually inspect daily routines and habits [2].

Process Mining [3] combines data mining methods with techniques from the Business Process Management (BPM) area [4]. In particular, it aims at extracting meaningful information from *event logs*, i.e., sequences of readable human actions performed by user(s).

Therefore, applying techniques from the BPM area can be a good compromise to overcome the gap between raw sensor measurements and human actions. Due to the BPM's input limitation, a log-preprocessing step is required to infer human actions from the sensor log.

Proceedings of the Demonstration & Resources Track, Best BPM Dissertation Award, and Doctoral Consortium at BPM 2021 co-located with the 19th International Conference on Business Process Management, BPM 2021, Rome, Italy, September 6-10, 2021

leotta@diag.uniroma1.it (F. Leotta); veneruso@diag.uniroma1.it (S. V. Veneruso)

In this paper we propose the *Visual Process Maps* (VPM) system, consisting of a complete pipeline formed by *(i)* a tool for the visual analysis of sensor logs, *(ii)* a method to transform raw movement measurements into actions, and *(iii)* a method to identify and visually analyze precedence relationships between human actions. The tool and related resource can be downloaded from https://www.diag.uniroma1.it/leotta/demos/vpm_bpm2021.html.

2. Relationship with the State of the Art

VPM is a tool that allows to graphically represent log activations and then to extract habit models out of them. The intended user of such a tool is a domain expert, who analyses the activities in the environment (e.g., an house) and takes decisions about the design of the space itself.

The literature about representing models of human habits is wide. In this section, we will highlight the differences between our work and those representation tools available in literature, which produce a human readable output that is easy to validate (once the representation is known).

In [5], the Context Modeling Language (CML) is presented. CML represents an initial approach to provide a graphical representation to a formalism for situations and contexts. Derived from software and databases design techniques, it introduces constructs to model typical smart spaces issues, as missing and/or conflicting information, events dependencies and constraints.

A survey about context modeling and reasoning techniques over contexts and situations is provided in [6]. However, these works on context-awareness are focused on the representational and reasoning issues; conversely they do not address the issue of learning the model from the sensor logs. This is indeed the main contribution of our work, in which we revert to visual process maps as representation for human habits.

Recently, the BPM research community has applied process mining techniques to smart spaces [4]. In [7], authors propose both supervised and unsupervised techniques to fill the gap between raw sensor measurements and human readable models (as discussed in Section 1), to then apply an inductive miner and obtain a Petri-net, i.e., a graphical representation of an obtained model characterized by nodes and arcs.

These first attempts to apply techniques taken from the business process management were focused on workflow specifications to anticipate user actions. A workflow is composed by a set of tasks related by qualitative and/or quantitative time relationships. In [8, 9], authors implemented a system called "Sequential Patterns of User Behavior System" (SPUBS) to automatically retrieve these workflows from sensor data.

This category of workflow-based modeling techniques (like SPUBS) can, in the case of highly variable processes, produce unreadable models; furthermore this approach requires the sensor log to be already segmented before mining.

With VPM, indeed, we decided to apply a specific technique named *fuzzy mining* that extracts models, which are more suitable than workflows employed in SPUBS, to describe human habits that are flexible by nature. As an advantage, this representation naturally supports visual analytics.

Finally, we want to mention *Sitivus*, a tool adopting visual analytics in smart spaces [10]. In

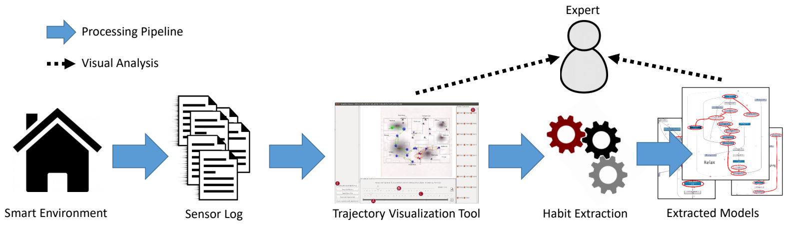

Figure 1: The proposed approach.

their work, they remark the importance of providing domain experts with a tool allowing to graphically inspect models of smart spaces. However, with respect to our work, *sitivus* is focused on modeling and representing situations more than habits. A *situation* is an interpretation of the context at a higher level and can comprehend the state of multiple users and consider the activities that are executed by each of them (e.g., “users are having lunch”), while a habit is focused on the routine of the user on specific time intervals (e.g., “what the user usually does after work?”).

3. Features Walkthrough

Figure 1 shows the data processing pipeline implemented through the tool. The sensor log generated by the smart home is converted to an event log (containing human actions instead of raw sensor measurements) by running a modified version of the *TRACLUS algorithm* (see Section 3.1). At this point, *human habit extraction* is performed thanks to fuzzy mining [11]. The proposed technique has been applied to the Aruba dataset, part of the CASAS project[1] freely available datasets [12]. The Aruba dataset contains sensor data that were collected for two years in the home of a volunteer adult.

3.1. Sensor Log conversion and Trajectory Visualization Tool

As discussed in Section 1, in order to apply process mining techniques, the sensor log $\mathcal{S}$ must be translated into a suitable event log $\mathcal{E}$. In this work, we adapted a portion of the TRACLUS algorithm by [13] to perform this task.

It considers the entire sensor data log as a single complex trajectory in which it can identify several sub-trajectories, each one describing a portion (a sub-log) of the original sensor log, which are related to the same action. This segmentation is based on finding the so called “characteristic points”: points in which the trajectory’s behaviour changes rapidly. Finally, it classifies each sub-trajectory by considering information about its movement and also which are the sensors involved and for how long, and labels them with an *action* (MOVEMENT, AREA or STAY), and their relative *location* inside the smart environment (e.g., <STAY Kitchen_table>). Now the log can be visually analysed by a human domain expert, who could be interested in replaying the log (or a portion of it) by employing the *trajectory visualization tool* (see Figure 2). Furthermore, it can now be processed through process mining techniques.

[1]http://casas.wsu.edu/datasets/

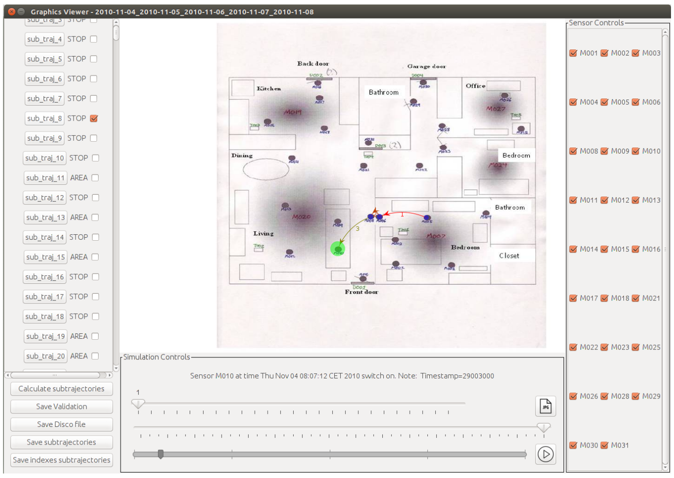

Figure 2: The trajectory visualization tool.

3.2. Habit Extraction

The event log, obtained from the previous step (Section 3.1), is fed into the *fuzzy miner*. The fuzzy miner produces a directed graph where each node represents a task (or in our case a human action) and each arc represents a precedence constraint between an action and its successor. Nodes and arcs are weighted, respectively with the number of times an action appears in the log and the number of times that an action follows another in the log.

These information are used to extract frequent sequences of human actions (i.e., paths inside the graph), by using an entire day as time reference.

Finally, we merge all these sequences into single activity models. As this approach was evaluated on the partially labeled Aruba dataset, we had at our disposal a *ground truth* consisting of sequences labeled with the name of the activity (e.g., *Relax*, *Eating*, *Sleeping*, and others). All frequent sequences extracted were compared with these labeled "ground truth" sequences, by computing the *Jaccard coefficient*: each sequence is then assigned to the activity label of the ground truth sequence obtaining the maximum score. As a final step, the sequences belonging to the same activity, are combined together: a graph for each activity class, that is also a subgraph of the original graph. The results obtained are very interesting, as the segments extracted are representative of the activity performed.

4. Discussion

This paper introduces VPM, a tool for the automated analysis of smart home sensor logs through process discovery.

At the current stage, the proposed graphical representation and, as consequence, models produced by the tool are limited by the type of sensors considered during the analysis. At the current stage, only PIR sensors have been considered; this does not allow to understand precisely the action performed by the user, but only the usage of a device. Supporting other types of sensors, would led to better classify actions and to more readable and accurate models, thus allowing a greater precision of the whole approach. Extending the analysis to other types of sensors will be an important starting point for our future work.

In addition, the employment of the tool in a real scenario would require also automatic segmentation techniques in order to describe human routines at a level of granularity finer than an entire day.

References

[1] M.-R. Tazari, F. Furfari, A. Fides-Valero, S. Hanke, O. Höftberger, D. D. Kehagias, M. Mosmondor, R. Wichert, P. Wolf, The universaal reference model for aal., Handbook of Ambient Assisted Living 11 (2012) 610–625.

[2] F. Leotta, M. Mecella, J. Mendling, Applying process mining to smart spaces: Perspectives and research challenges, in: International conference on advanced information systems engineering, Springer, 2015, pp. 298–304.

[3] W. M. P. van der Aalst, Process Mining: Data Science in Action, 2 ed., Springer, 2016.

[4] M. Dumas, M. La Rosa, J. Mendling, H. A. Reijers, Fundamentals of Business Process Management (2nd ed.), Springer, 2018.

[5] K. Henricksen, J. Indulska, Developing context-aware pervasive computing applications: Models and approach, Pervasive and mobile computing 2 (2006) 37–64.

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

[7] N. Tax, N. Sidorova, R. Haakma, W. M. van der Aalst, Event abstraction for process mining using supervised learning techniques, in: Proceedings of SAI Intelligent Systems Conference, Springer, 2016, pp. 251–269.

[8] A. Aztiria, A. Izaguirre, R. Basagoiti, J. C. Augusto, D. J. Cook, Discovering frequent sets of actions in intelligent environments., in: Intelligent Environments, iOS Press, 2009, pp. 153–160.

[9] A. Aztiria, A. Izaguirre, R. Basagoiti, J. C. Augusto, D. Cook, Automatic modeling of frequent user behaviours in intelligent environments, in: Intelligent Environments (IE), 2010 Sixth International Conference on, IEEE, 2010, pp. 7–12.

[10] A. K. Clear, T. Holland, S. Dobson, A. Quigley, R. Shannon, P. Nixon, Situvis: A sensor data analysis and abstraction tool for pervasive computing systems, Pervasive and Mobile Computing 6 (2010) 575–589.

[11] C. W. Günther, W. M. Van Der Aalst, Fuzzy mining–adaptive process simplification based on multi-perspective metrics, in: International conference on business process management, Springer, 2007, pp. 328–343.

[12] D. J. Cook, Learning setting-generalized activity models for smart spaces, IEEE intelligent systems 27 (2012) 32–38.

[13] J.-G. Lee, J. Han, K.-Y. Whang, Trajectory clustering: a partition-and-group framework, in: Proceedings of the 2007 ACM SIGMOD international conference on Management of data, 2007, pp. 593–604.