

# Detection of User Activities in Intelligent Environments

Agnese Augello and Salvatore Gaglio

**Abstract** Research on Ambient Intelligence (AmI) focuses on the development of smart environments adaptable to the needs and preferences of their inhabitants. For this reason it is important to understand and model user preferences. In this chapter we describe a system to detect user behavior patterns in an intelligent workplace. The system is designed for a workplace equipped in the context of *Sensor9k*, a project carried out at the Department of Computer Science at the University of Palermo (Italy).

## 1 Introduction

Research in Ambient Intelligence (AmI) focuses on the development of smart environments, generally equipped with wireless sensor networks, allowing the gathering of data about the environment state [1]; such data needs to be processed and analyzed in order to deduce useful information. Ambient Intelligence brings intelligence to our everyday environments making those environments sensitive, and adaptive to us [2]. The definition of appropriate user profiles can allow an AmI system to anticipate their needs, and adapt the environment settings to their preferences [3, 4, 8]. User profiling can also be used to detect significant changes in resident behaviors [2, 5], to customize building energy and comfort management systems [6], or to allow automatic setting of system parameters in order to optimize energy consumption [7].

Most systems perform an explicit profiling or derive users presence and activity by analyzing the sensor data and the use of actuators. In many projects, data mining

---

A. Augello (✉)

ICAR-CNR, Viale delle Scienze, Edificio 11, Palermo, Italy  
e-mail: augello@pa.icar.cnr.it

S. Gaglio

DICGIM-UNIPA, Viale delle Scienze, Edificio 6, Palermo, Italy  
e-mail: salvatore.gaglio@unipa.it

methodologies are used to detect recurrent behaviors in time sequences of events, to find the relationship between variables of interest, or to prefigure the future behavior of some entities. As an example, in MavHome [9] hierarchical models of inhabitant behaviors are learned by means of data-mining techniques aimed to discover periodic and frequent episodes of activity patterns, while in [2] they are used to predict the location, routes and activities of the residents in order to adaptively control home environments .

An important aspect to be considered is the choice of the models used to represent the information acquired in an AmI system. As an example, the authors of [10] propose an ontology to model the domain knowledge of a smart building system. In [11] the problem of modelling interaction events of smart objects in a environment is discussed; ontologies can help developers to infer the possible connections between these objects, enable device interoperability on a semantic level. In [12] there is a discussion on the importance of common sense ontology in a real AmI system; a reasoning on the acquired information can be performed to deal with unforeseen requirements or needs, making decisions out of the prefixed behavioral patterns.

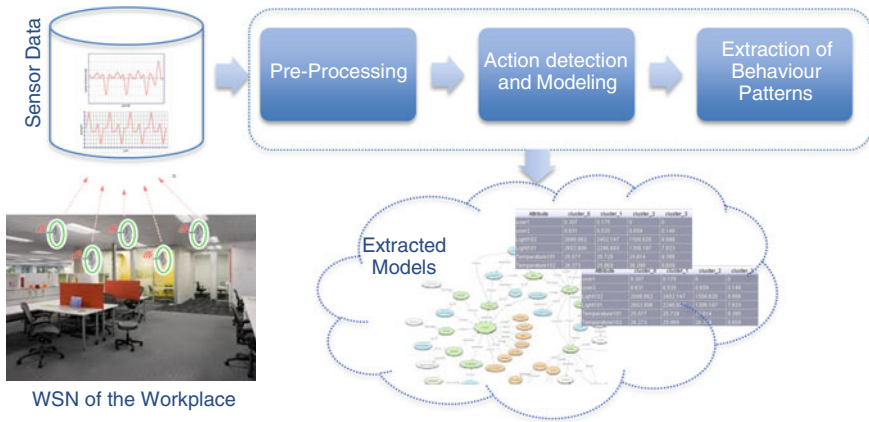
The architecture proposed in this chapter arises from the need to implicitly analyze and profile the users of an intelligent workplace equipped in the context of Sensor9k project [13]. In this work we consider a workplace equipped with a set of sensors aimed at the detection of environmental variables, such as light, temperature and humidity, while there are no sensors providing direct information about the user except for their presence (which is detected by RFID sensors). In our opinion user profiles are implicitly described by the data collected by the sensors located in the workplace rooms, data which explicitly show the consequences of users' actions over the environment state.

The approach proposed here exploits different methodologies of analysis and reasoning on data sensors in order to detect user actions on the environment. In particular the analysis is aimed at the detection of meaningful changes which can be considered as consequences of user actions. The sensory data and the recognized events are arranged in appropriate models in order to highlight the existence of relationships among environmental data or events and the users' presence in the office room. Moreover an analysis and reasoning on the information described in such models can be performed in order to infer users behaviors and preferences.

## 2 Proposed Approach

The proposed approach consists essentially of three different phases of analysis of the data gathered by the wireless sensor network, regarding the environment state and the user presence (see Fig. 1).

The *pre-processing* analyzes data in order to detect anomalies, remove outliers and replace missing data, the *action detection and modelling* phase analyzes sensor data trends to infer changes which can be ascribed to human actions; appropriate models allow for a better understanding of the extracted information. Finally the *Extraction*



**Fig. 1** The proposed approach

of *behavior Patterns* phase is accomplished to find relationships of interest and to detect similar clusters in the Aml data.

## 2.1 Pre-Processing

At this phase, the observed variables trends are analyzed to recognize those events that can be ascribed to human intervention. First of all the data collected by the sensors are preprocessed in order to detect anomalies, remove invalid values and estimate missing values. For example the assumption of a time correlation in the data can be exploited for the estimation of missing values by means of a linear interpolation between preceding and subsequent observations, while the assumption of spatial correlation in the readings of sensors located in small indoor environments can be profitably employed in order to detect outliers, and to replace them with the combination of neighbour sensors readings.

However, for our aim, it is important to analyze differences between sensors belonging to different, but close areas. We assume that variations in inter-areas sensors readings can be due to the use of actuators from users, such as turning on/off the light or changing the settings for the temperature and humidity control systems.

## 2.2 Action Detection and Modelling

In this phase of analysis we consider the placement of the sensors in areas within the environment and analyze the dynamics of the observed series, i.e., the mechanism by which they evolve over time. In particular, the time series obtained from sensor

readings can be decomposed into a set of components: a *trend*, a *seasonal* and a *remainder* component [14]. The trend component  $T_t$  defines the long-term trend of the variable and can be considered as the tendency to increase, decrease or remain constant over a long period of time. The seasonal or periodic component  $S_t$  is given by one or more periodic components, taking the same or similar values at a fixed distance in time. Finally, the remainder component  $R_t$  determines short-term fluctuations in the series. This decomposition is important in order to estimate and remove a regular and predictable component which could hide useful information, and to consider only meaningful changes in the data trend. In particular, in our context, changes in data can be due to regular and repetitive factors, for example to natural light and temperature changes during the day, while other changes can be due to human actions on actuators.

Let  $R_t$  the remainder of the analyzed series, and  $R'_t$  the corresponding derivative, representing its time variation. The function  $R'_t$  shows changes in  $R_t$  function. Every change is characterized by a strength and a direction. If the detected change is attributed to a user action, the direction will allow the interpretation of the type of action. For example a positive direction will be considered as a sign of the light being turned on, or in the case we are analyzing temperature data could be interpreted as an increase of temperature settings. Given an experimentally defined threshold, called  $\vartheta$ , only changes with strength greater than  $\vartheta$  will be considered.

After the preprocessing of the data we suggest the use of a probabilist model to estimate if the detected changes in sensor measurements can be attributed to users' actions. The choice of using such a model depends on the fact that we are bound to reason about uncertain knowledge. The interest variables and their possible states are modelled as nodes of a dynamic Bayesian network. The nodes are connected with directed links representing the influence among the nodes where the influence of parent nodes on a variable  $X_i$  is quantified by means of conditional probabilities  $P(X_i | ParentsValues)$  represented in opportune tables associated to each node. We model two kinds of variables, *UA* variables, representing possible user actions and *SO* variables, representing sensory observations. The model thus built is used to estimate the probability that the event we want to investigate did occur, based on the sensor readings. As an example, Fig. 2 shows the Bayesian network used to estimate the user actions controlling ambient light; the possible action are modelled by the *UserAction*  $t$  variable which can take three different states: *on*, corresponding to turning on action, *off*, corresponding to turning off action, and *none* if no action is carried out. We also have four variables of type *SO*: *UserPresence*, *OutdoorLight*  $t + 1$ , *LightTrend*  $t$  and *LightTrend*  $t + 1$ . The first can assume the states *in* and *out* depending on users presence in the room. Variable *OutdoorLight*  $t + 1$  represents the external light at a subsequent time instant and can be *high*, *medium*, or *low*; it is used to better understand how much the light change in the room at a subsequent instant may be due to the external light status or to an user action. The variables *LightTrend*  $t$  and *LightTrend*  $t + 1$  represent the evolution of light in two successive time instants, and can assume the states *increasing*, *decreasing* and *stable*. The actions are modelled as states of the *UserAction*  $t$ , which depends on the state of the light

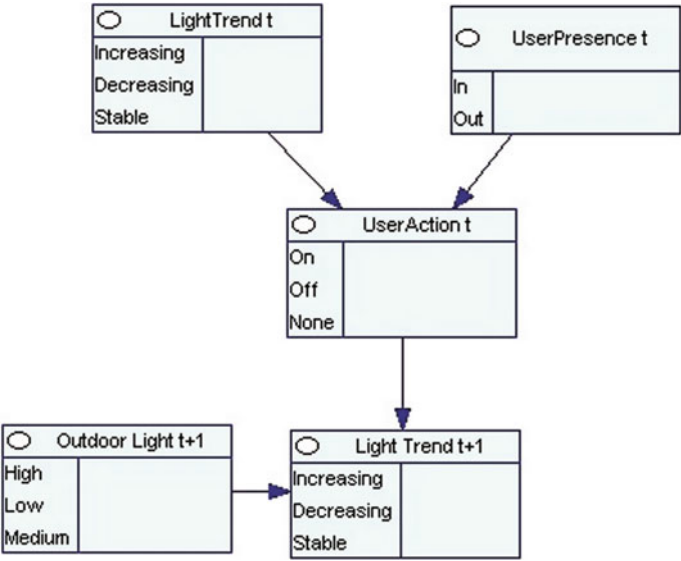
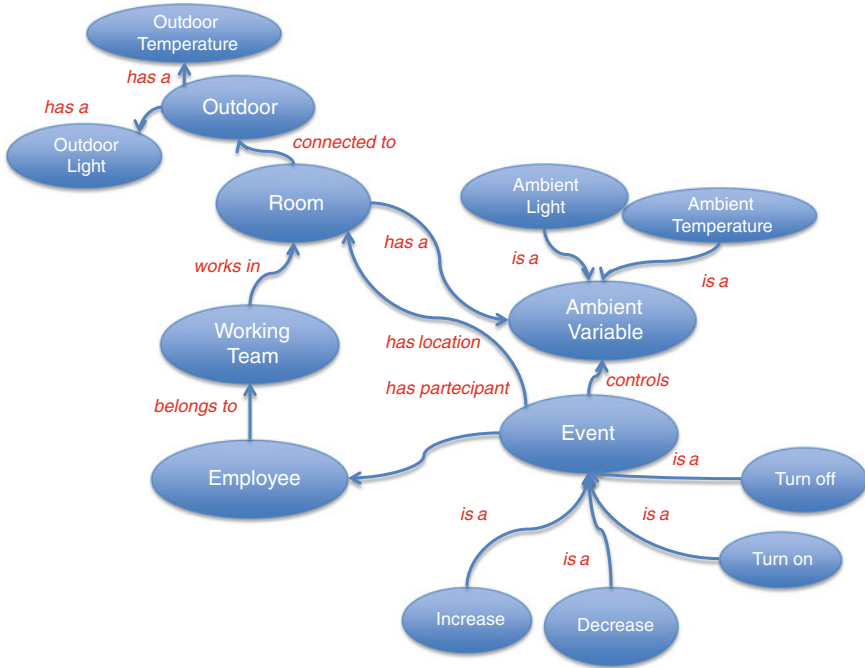


Fig. 2 A probabilistic model to estimate user actions controlling ambient light

( *LightTrend t* variable) and by information on the users presence (*UserPresence t* variable) in workplace. The state of *UserAction t* variable and the state of the outdoor light *OutdoorLight t +1* influence in their turn the state of the light at the next instant (*LightTrend t + 1* variable).

The detected events, which in the particular case are possible user actions on the actuators, can be modelled in an ontology (a portion is shown in Fig. 3).

Each action modelled in the ontology is associated to the room in which the event took place, to the employee who may have executed that action (we have information only about the presence of people in a particular room but we do not know who actually have done that action), to the time in which the action was accomplished, and to other correlated concepts. Into the ontology, other information is also represented, related to the domain knowledge; for example the variables that can be controlled by the user, such as the temperature and light, and the environmental status of the outdoor. The ontology can be used to reason about the acquired information in order to deduce new knowledge. For example, analyzing either the information regarding the state of the temperature inside and outside a room, and the actions performed on the thermostat when in that room is present a specific user, we can acquire the preferences of that user regarding the setting of the temperature.



**Fig. 3** An ontological formalization of the actions on the intelligent workplace

### 2.3 Extraction of Behavior Patterns

The sensors data are analyzed to identify relationships among the variables of interest and the users. The measurements recorded by the different sensors for each physical variable in a specific period (such as for instance an entire working day) and the occurrences of events such as user actions in specific instants are represented in a matrix. In particular the rows of the matrix represent the different observations in a given period, while the columns the sample values detected from each sensor during the observations. The number of sensors inside the room can be large, so as to generate several columns. Therefore, we perform a dimensionality reduction to evaluate only the most meaningful informative content. Let  $\mathbf{X}$  indicate the matrix representing a dataset composed by a set of  $m$  vectors of length  $n$ , each one representing the set of measurements obtained by the  $n$  sensor at a specific observation for a specific variable  $x$   $X = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ .

As an example, let  $\mathbf{T} = [\mathbf{T}_1, \dots, \mathbf{T}_n]$  be a  $m \times n$  matrix composed of a set of column vectors, each one representing the set of observations regarding the temperature measured by  $n$  sensors in an office room.

A dimensionality reduction process, by means of PCA [15] is performed on such set of observations allowing for the projection of original dataset space into a smaller space. PCA extracts a set of orthogonal vectors, called principal components, by

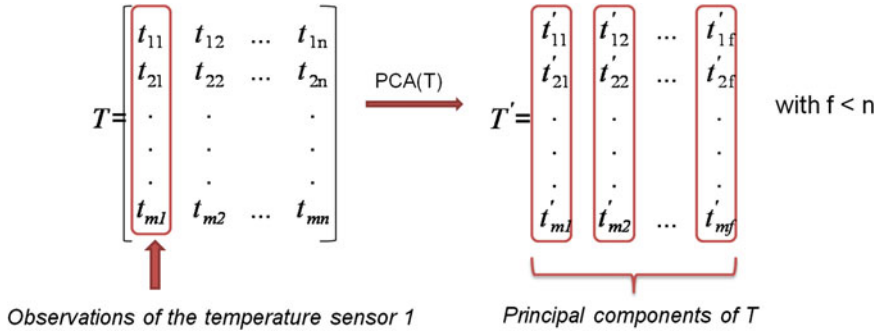


Fig. 4 Principal component analysis performed on temperature dataset

a linear transformation of the original variables, and arranges them according to decreasing variance values. This transformation has the effect to capture the major associational structure in the dataset, removing information which contribute less to the variance of data, and are thus less relevant. It should be highlighted that PCA performs the dimensionality reduction process by a combination of original vectors, while other methods merely select a subset of items from the original dataset [16].

After performing PCA we obtain a  $m \times f$  matrix, with  $f \leq n$ ,  $\mathbf{T}' = [\mathbf{T}'_1, \dots, \mathbf{T}'_f]$  (see Fig. 4). The same procedure can be performed over the entire set of observed variables.

The whole set of variables and events observations is then modelled as a matrix  $\mathbf{X}(m \times nv)$ , where a row  $\mathbf{X}_i$  represents an observations at a specific time  $i$  and a specific column  $\mathbf{X}_j$  represents the entire sample of observations of the  $j$ -th variable in the considered period.

In our specific case, matrix  $\mathbf{X}$  is given by:

$$\mathbf{X} = [\mathbf{U}_1, \dots, \mathbf{U}_d, \mathbf{T}_1, \dots, \mathbf{T}_f, \mathbf{L}_1, \dots, \mathbf{L}_g]$$

composed of a set of vectors, each one representing the set of observations of a specific variable: the set  $\mathbf{U} = \{\mathbf{U}_j\}_{j=1\dots d}$  represents observations about the presence of  $d$  users in the considered period, the sets  $\mathbf{T} = \{\mathbf{T}_j\}_{j=1\dots f}$  and  $\mathbf{L} = \{\mathbf{L}_j\}_{j=1\dots g}$  represent observations about temperature and light exposure respectively, related to the  $f$  and  $g$  variables obtained after the application of PCA on temperature and light matrices as described in the previous section.

We therefore compute a correlation matrix  $\mathbf{C}(nv \times nv)$  in order to highlight the relationships among the variables, where the  $i, j$ -th element of  $\mathbf{C}$  is given by the correlation coefficient  $c_{ij}$  between the  $i$ -th and the  $j$ -th variable, as given by:

$$c_{ij} = \text{Corr}(\mathbf{X}_i, \mathbf{X}_j) = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$$

In the previous formula,  $\sigma_{ij}$  represents the covariance between  $\mathbf{X}_i$  and  $\mathbf{X}_j$  and  $\sigma_i$  and  $\sigma_j$  respectively the standard deviation of  $\mathbf{X}_i$  and  $\mathbf{X}_j$ .

In this way it is possible extract sub-matrices, representing correlation patterns between the observations related to the presence of users in office rooms and values representative of specific environment variables. It is thus possible to obtain a characterization of users with respect to values of the observed variables in a specific period.

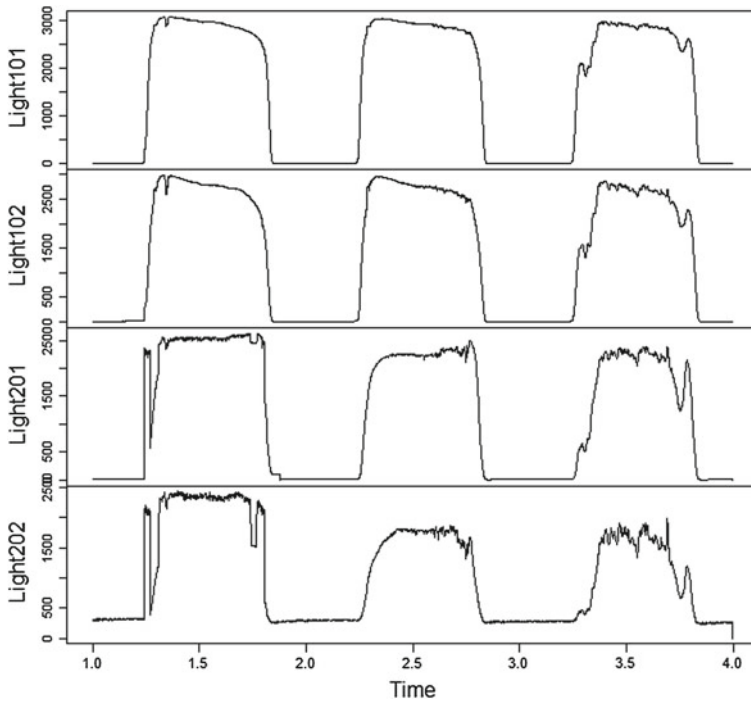
The extracted patterns can be clustered in order to identify group of users with similar preferences about variables setting, or users performing the same actions in similar environment conditions. In this way it is possible obtain a subdivision of users' behavior patterns into a set of similar profiles. In particular the data extracted from the correlation matrix  $\mathbf{C}$  are classified by means of a K-means [17] algorithm, setting experimentally the number of clusters to be obtained and the function to evaluate distances between data points and cluster centers. An iterative process is then performed, during which  $k$  data in the dataset are randomly chosen to constitute the first centroids of the clusters. The metric distance allows to assign the remaining data to the cluster on the strength of their closeness with the centers of clusters, then, new centers are detected evaluating the average of each cluster. The process ends when the obtained result satisfies a predetermined criterion of termination.

### 3 Experimental Results

In this section we report some experimental results obtained testing the proposed approach on a dataset of the *Sensor9k* project [18, 19]. The dataset contains a set of measurements obtained from a testbed for WSN-based Ambient Intelligence applications [20–22], built in a workplace of the University of Palermo. The office rooms of the workplace have been equipped with sensor nodes monitoring indoor and outdoor physical quantities such as relative humidity, temperature, and light exposure; additionally, RFID sensors allow for detecting the employees' presence in the workplace through the use of personal badges. To test the proposed approach we have considered a subset of the sensors used in [18], in particular the experiments have been conducted analyzing data measured from MTS300 sensor nodes, where the analyzed variables are light, and temperature, the access of the employees in the workplace and outdoor measurements of light and temperature. In particular we have data regarding two office rooms, *Room1* and *Room2*. The former is an office room, used by two employees *User1* and *User2*, whereas the latter in a common area. The two rooms share similar exposition (thus similar trends for the considered variables), and are connected by a door.

We have analyzed data measured from two sensors per room. We will indicate light and temperature measurements collected by the two sensors in *Room1* as *Light101*, *Light102*, *Temperature101* and *Temperature102*, respectively; analogously, *Light201*, *Light202*, *Temperature201* and *Temperature202* will be the





**Fig. 5** Time series for *Light101*, *Light102*, *Light201* and *Light202*

measurements related to *Room2*. Figure 5 shows the time series of the all four light measurements in a period of three days.

The two topmost plots clearly show the similarity in the trends of sensors located within the same room, and the same consideration holds for the two plots at the bottom; if we consider the two central plots, we can also identify significant similarities, although not as striking as in the previous cases; we argue that the differences in measurements for sensors in different rooms are partially due to different placements, and mainly to the effects caused by a different use of the actuators by the users.

Figures 6 and 7 show the decomposition of two light time series belonging to the two rooms, i.e. *Light101* and *Light201*.

Each figure shows the original series, and its seasonal, trend and remainder components. The plots show how the seasonal component is related to day-night cycles, while the trend is related to level of brightness of days.

Figure 8 shows the analysis of the remainder component of the *Light101* series and the corresponding relevant variations computed as derivative of the function, which presumably correspond to actions on part of some user. The plot shows only instantaneous changes since we are looking for actions on artificial light settings (a different analysis would be performed for the detection of actions on temperature settings, because temperature takes longer to stabilize).

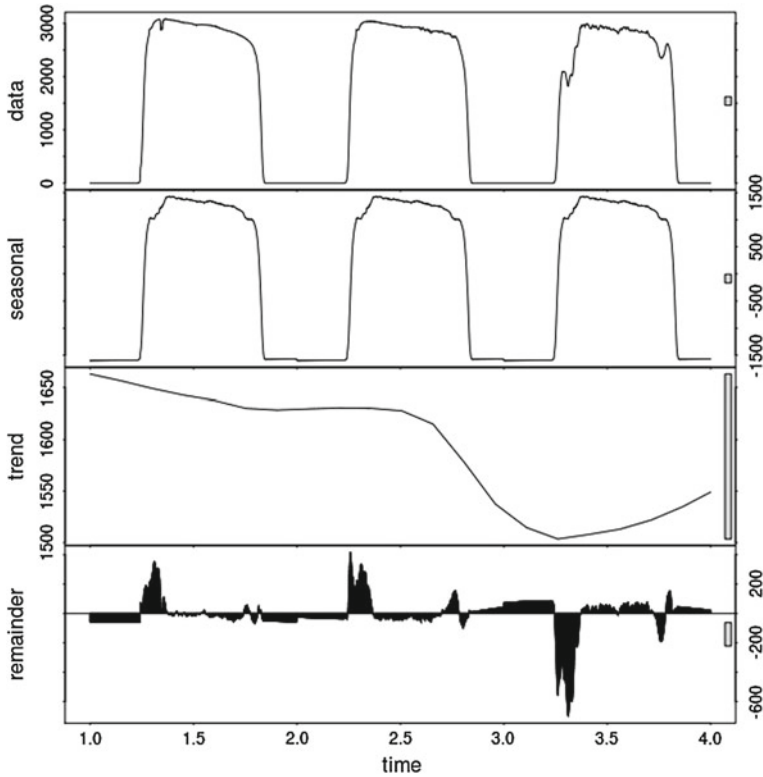
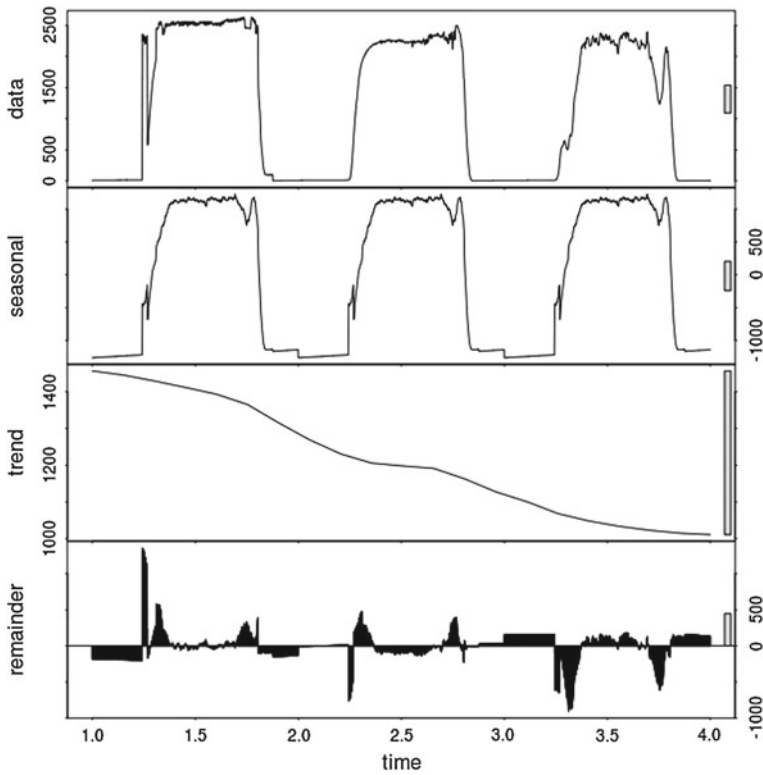


Fig. 6 Decomposition of Light101 temporal series

Figure 9 shows a reasoning process to disambiguate one of the detected variations. In the example, evidence coming from sensor observations is set, and the probability of user action states is evaluated. In particular, the user was in the room, the external light was high and the internal light had an increasing change in two subsequent instants. The result of the reasoning process is that the user action could be a *turning on* action, with a probability of 0.32, a *turning off* action with a probability of 0.11, and with a probability of 0.57 the increasing will be due to the increasing of the external light (independent of the user action).

The dataset used to validate our approach so far contains information about the presence of only two users. For this reason, we conducted a proof-of-concept experiment involving a set of time slots in a working day, also considering information about the users' presence and environmental conditions. Figure 10 shows the results obtained by a k-means clustering on one-day matrix observations, related to the presence of the two users *User1* and *User2*, and the measurements of light and temperature.

The Table shows that *User2* had more significant influence on the measured quantities as it was more present; this is especially evident for the last two clusters,

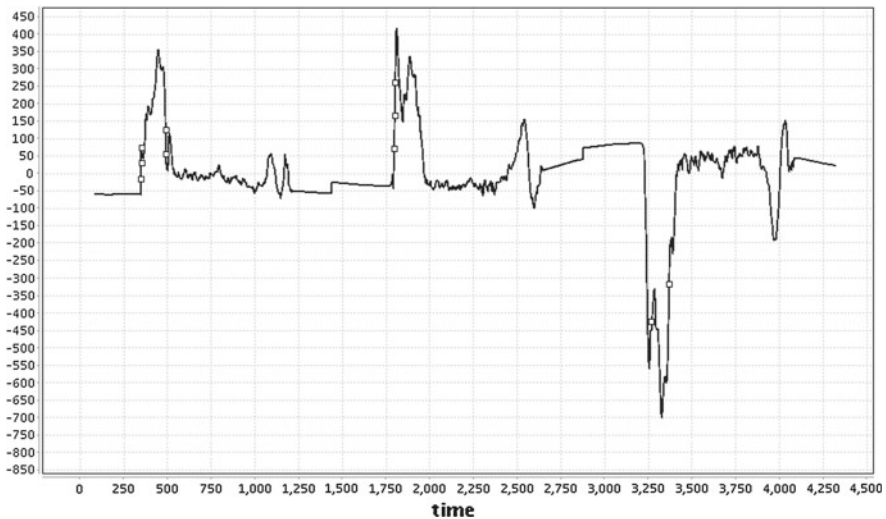


**Fig. 7** Decomposition of Light201 temporal series

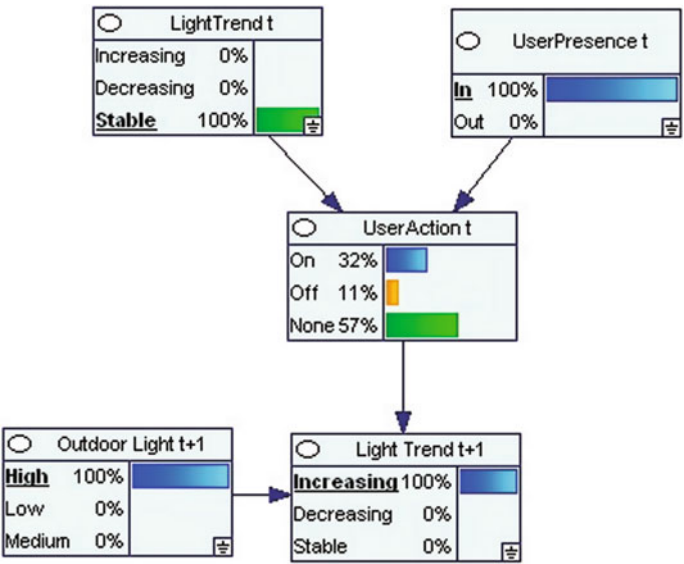
regarding the later time of the day, where the influence of *User2* may be easily singled out; the significant differences in the numerical values captured by cluster 3 are easily explainable by considering that this cluster contained data measured during nighttime.

## 4 Conclusion

In this chapter we have described an approach aimed at implicitly detecting behavior patterns of users working in an intelligent environment, equipped in the context of *Sensor9k* project. The approach consists in different modules for the extraction of useful information regarding user actions and habits. Such information can be used for different purposes, for example to adapt the environment settings to users preferences, or to monitor the energy consumption in the workplace.



**Fig. 8** Light101 relevant events: the curve represents the trend of light, while squares represent the recognized events



**Fig. 9** Probabilist reasoning on a recognized event

**Acknowledgments** This work has been partially supported by the PO FESR 2007/2013 grant G73F11000130004 funding the SmartBuildings project.

Attribute	cluster_0	cluster_1	cluster_2	cluster_3
user1	0.307	0.175	0	0
user2	0.631	0.535	0.659	0.148
Light102	2686.862	2402.147	1508.620	9.898
Light101	2603.906	2246.683	1306.187	7.833
Temperature101	25.577	25.729	25.614	9.395
Temperature102	26.273	25.969	26.399	9.609

**Fig. 10** Relation between users presence and temperatures and lights values in a day

## References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: A survey on sensor networks. *IEEE Commun. Mag.* **40**(8), 102–114 (2002)
2. Cook, D.J., Augusto, J.C., Jakkula, V.R.: Ambient intelligence: Technologies, applications, and opportunities. *Pervasive Mob. Comput.* **5**(4), 277–298, (2009). ISSN 1574–1192, doi:[10.1016/j.pmcj.2009.04.001](https://doi.org/10.1016/j.pmcj.2009.04.001)
3. Khalili, A., Wu, C., Aghajan, H.: Autonomous learning of user's preference of music and light services in smart home applications. In: Behavior Monitoring and Interpretation Workshop at German AI Conference, September (2009)
4. Mozer, M.C.: Lessons from an adaptive home. In Cook, D.J., Das, S.K. (eds.) *Smart Environments: Technology, Protocols, and Applications*, pp. 273–298. Wiley, UK (2004)
5. Akhlaghinia, M.J., Lotfi, A., Langensiepen, C., Sherkat, N.: Occupant behaviour prediction in ambient intelligence computing environment. In: Special Issue on Uncertainty-Based Technologies for Ambient Intelligence Systems, vol. 2, Issue number 2, May (2008)
6. Dong, B., Andrew, B.: Sensor-based occupancy behavioral pattern recognition for energy and comfort management in intelligent buildings. In: *Proceedings of Building Simulation '2009, an IBPSA Conference*. Glasgow, U.K. (2009)
7. Barbato, A., Borsani, L., Capone, A., Melzi, S.: Saving, home energy, through a user profiling system based on wireless sensors. In: *ACM Buildsys, : (in conjunction with SenSys 2009)*, Berkeley, CA. Nov. 3, (2009)
8. Doctor, F., Hagrais, H., Callaghan, V.: A fuzzy embedded agent-based approach for realizing ambient intelligence in intelligent inhabited environments. *IEEE Trans. Syst. Man Cybern. Part A* **35**(1), 55–65 (2005)
9. Youngblood, G.M.: Automating inhabitant interactions in home and workplace environments through data-driven generation of hierarchical partially-observable Markov decision processes. Ph.D. thesis, The University of Texas at Arlington, (2005)
10. Stavropoulos, T.G., Vrakas, D., Vlachava, D., Bassiliades, N.: BOnSAI: a smart building ontology for ambient intelligence. In: *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics (WIMS '12)*. ACM, NY, USA (2012)
11. Niezen, G.: Ontologies for interaction: enabling serendipitous interoperability in smart environments. *JAISE* **5**(1), 135–137 (2013)
12. Santofimia, M.J., Moya, F., Villanueva, F.J., Villa, D., Lpez, J. C.: How intelligent are ambient intelligence systems? *Int. J. Ambient Comput. Intell. (IJACI)* **2**(1), 66–72 (2010). doi:[10.4018/jaci.2010010106](https://doi.org/10.4018/jaci.2010010106)
13. De Paola, A., Farruggia, A., Gaglio, S., Lo Re, G., Ortolani, M.: Exploiting the human factor in a WSN-based system for ambient intelligence. In: *CISIS* pp. 748–753 (2009)
14. Brockwell, P.J., Davis, R.A.: *Time Series: Theory and Methods*. Springer, NY (1998). ISBN: 038797429
15. Jolliffe, I.T.: *Principal Component Analysis*, p. 487, Springer, NY (1986). doi:[10.1007/b98835](https://doi.org/10.1007/b98835) ISBN 978-0-387-95442-4

16. Han, J., Kamber, M.: Data mining: concepts and techniques, 2nd edn. In: Jim G. Series (ed.) The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann Publishers, San Francisco (March 2006). ISBN 1-55860-901-6
17. MacQueen, J.B.: Some methods for classification and analysis of multivariate observations. In: Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability. vol. 1, pp. 281–29. University of California Press, Berkeley (1967)
18. De Paola, A., Gaglio, S., Lo Re, G., Ortolani, M.: Sensor9k : a testbed for designing and experimenting with WSN-based ambient intelligence applications. *Pervasive Mob. Comput.* **8**(3), 448–466 (2012)
19. Augello, A., Ortolani, M., Lo Re, G., Gaglio, S.: Sensor mining for user behavior profiling in intelligent environments. *Advances in Distributed Agent-Based Retrieval Tools* 143–158 (2011)
20. De Paola, A., La Cascia, M., Lo Re, G., Morana, M., Ortolani, M.: User detection through multi-sensor fusion in an Am I scenario. In: Proceedings of the 15th International Conference on, Information Fusion, pp. 2502–2509 (2012)
21. De Paola, A., La Cascia, M., Lo Re, G., Morana, M., Ortolani, M.: Mimicking biological mechanisms for sensory information fusion. *J. Biologically Inspired Cogn. Architectures* **3**, pp. 27–38 (2013)
22. Cottone, P., Lo Re, G., Maida, G., Morana, M.: Motion sensors for activity recognition in an ambient-intelligence scenario. In: Proceedings of the 5th International Workshop on Smart Environments and Ambient, Intelligence, pp. 646–651 (2013)

Advances onto the Internet of Things

How Ontologies Make the Internet of Things Meaningful

Gaglio, S.; Lo Re, G. (Eds.)

2014, IX, 352 p. 125 illus., Softcover

ISBN: 978-3-319-03991-6