Sensor mining for user behavior profiling in intelligent environments

Article

Accepted version

A. Augello, M. Ortolani, G. Lo Re, S. Gaglio

In Advances in Distributed Agent-Based Retrieval Tools, 2011, pp. 143-158

Publisher: Springer

http://link.springer.com/chapter/10.1007%2F978-3-642-21384-7_10

Sensor Mining for User Behavior Profiling in Intelligent Environments

A. Augello, M. Ortolani, G. Lo Re, and S. Gaglio

Abstract The proposed system exploits sensor mining methodologies to profile user behaviors patterns in an intelligent workplace. The work is based in the assumption that users' habit profiles are implicitly described by sensory data, which explicitly show the consequences of users' actions over the environment state. Sensor data are analyzed in order to infer relationships of interest between environmental variables and the user, detecting in this way behavior profiles. The system is designed for a workplace equipped in the context of Sensor9k, a project carried out at the Department of Computer Science of Palermo University.

1 Introduction

Research in Ambient Intelligence (AmI) focuses specifically on users and on how they relate to the surrounding environment; namely AmI systems attempt to sense the users' state, anticipate their needs and adapt the environment to their preferences [1]. For this reason, AmI systems can benefit from the latest research in users modeling and profiling. User profiling process consists of collecting, and analyzing users informations. The acquired user information can be used to build appropriate user models which can be considered as representation of the system's beliefs about the user [2]. User profiling can be explicit or implicit: an explicit profiling can be done through the formulation of questions about user's preferences, while implicit profiling methods construct user profiles by inferring user ratings from interest indicators by means of user interactions with the system [3]. AmI systems rely on specific hardware in order to gather information about the environment state and the user presence; such data may for instance be collected via pervasively deployed wireless sensor nodes [4], i.e. small devices equipped with sensors, a processor and

Agnese Augello, Marco Ortolani, Giuseppe Lo Re and Salvatore Gaglio
DINFO Dept. of Computer Engineering. University of Palermo, Viale delle Scienze, ed. 6 – Palermo, Italy, e-mail: (augello, ortolani)@dinfo.unipa.it, (lore, gaglio)@unipa.it

a transceiver unit. The idea presented here aims to investigate how such collected data might be profitably used to identify and implicitly profile user habits.

In this context, data mining techniques allow for an intelligent analysis of environmental data in order to detect behavior patterns and classify them into profiles. Careful processing of sensory data may be used to infer descriptive models showing the relationships of interest between environmental variables and the user, while predictive models may proved reliable inference on future behavior of users populating the considered environment [5].

In this work sensor mining methodologies are exploited to profile user behaviors in an intelligent workplace. The workplace has been equipped in the context of Sensor9k, a project carried out at our Department [15]. Our work is based in the assumption that users' habit profiles are implicitly described by sensory data, which explicitly show the consequences of users' actions over the environment state. The system analyzes time data collected by the sensors located in the workplace rooms, and through a data mining process tries to detect changes which can be considered as consequences of user actions. Moreover 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. The emerging behavioral patterns may finally be grouped based on their relative similarities by means of a clustering process in order to draw users profiles. The rest of the paper is organized as follows: after a discussion about the state of the art, the system architecture will be described, reporting some experimental results and discussing about future works. Moreover the advantages that can be achieved by extending the proposed approach to a distributed architecture will be discussed.

2 Ambient Intelligence for Behavior Profiling

Ambient Intelligence brings intelligence to our everyday environments making those environments sensitive, and adaptive to us [6]. Wireless sensor nodes allow the gathering of data about the environment state [4]. The huge amount of data, obtained from sensor measurements has to be processed and analyzed in order to deduce useful information.

In literature, several methodologies of reasoning on sensor data have been proposed to perform user modeling and profiling, activity recognition and also decision making processes [6]. User profiling applications in AmI have targeted environment personalization as for instance in [8], and [7], where reinforcement learning algorithms are used to learn preferred music and lighting settings, adaptable to preferences changes. User profiling can also be used to detect significant changes in resident's behavior preserving their safety [6][17]. Other applications regard personalization of building energy and comfort management systems [19]. In [9], data collected by wireless sensor are used to create profiles of the inhabitants, and a prediction algorithm allows the automatic setting of system parameters in order to optimize energy consumption. In many projects the aim is to anticipate the location, routes and activities of the residents in order to adaptively control home environments [6]. In MavHome [10], hierarchical models of inhabitant behaviours are learned by means of data-mining techniques aimed to discover periodic and frequent episodes of activity patterns. In the iDorm system [11], intelligent agents are embedded in the user environments to control them according to the needs and preferences of the user. In particular an unsupervised, data-driven, fuzzy technique is used by agents for analyze the actuator readings and sensor states and therefore extracting fuzzy rules representing users behaviors in the environment. A depth examination of AmI methodologies and applications can be found in [6]. Most systems analyze explicit feedback or implicit feedback deriving from the use of actuators, or data related to the presence of the user obtained for example with motion sensors, RFID sensors and so on. The main feature of the proposed systems is that they can only access information concerning such environmental features as light, temperature and humidity, and have no direct information about the user except for their presence detected by RFID sensors. Therefore, such systems exploit different methodologies of analysis and reasoning on these data in order to infer the possible user actions on the environment, namely a careful analysis of the data, discarding what can be attributed to noise or to the natural daily trend of the analyzed environmental variables. Our system, similarly to what was proposed by [19] uses probabilistic reasoning to interpret the possible events; moreover, similarly to [9], it extracts an appropriate representation in order to highlight correlations between events. It also uses clustering methodologies to detect similar user behavior patterns.

3 Sensor Data Mining

Data mining methodologies play a fundamental role in sensor data understanding. The main purpose of a data mining process is to find meaningful patterns, relationships and models in a huge amount of data. Pattern discovery methodologies are designed to automatically find new patterns in a predictive way, prefiguring the future behavior of some entities, or in a descriptive way, finding human-interpretable patterns [12]. The various data mining techniques can be classified according to the main categories of applications.

Among predictive methodologies, classification techniques are used to identify the features indicating that an entity belongs to a certain class, learning from a set of pre-classified examples, while prediction and regression techniques analyze known values to estimate unknown quantities. Among descriptive methodologies, clustering techniques are aimed to identify groups with homogeneous elements, association techniques are exploited to identify elements that often appear together during a specific event and sequential patterns are aimed to the detection of recurrent behavior in time sequences of events [13].

Predictive data mining algorithms are used in mining sensory data from an indoor environment to estimate physical at a location where no sensor is placed, to predict failures or user behaviors, while descriptive methodologies can be exploited to find

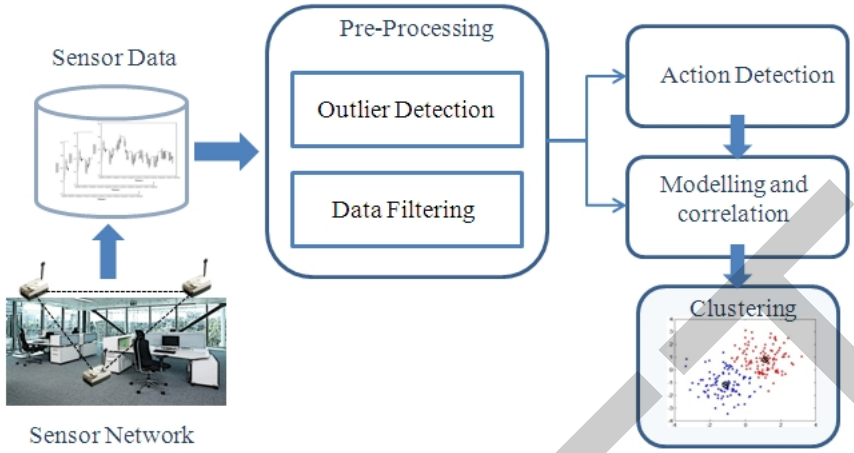

Fig. 1 System Architecture

the relationship between variables of interest, as an example to relate users behavior with energy consumption or system failures [5].

4 System Architecture

The proposed system aims to learn users' behavior profiles in the context of a smart workplace, such as that of the Sensor9k project [15]. Office rooms 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.

The overall system architecture, shown in Figure 1, has been designed according to a modular approach. A *preprocessing* module is used to improve the quality of sensory data while an *action detection* module analyzes data trends to infer changes which can be ascribed to human actions. The information extracted by sensors, and the recognized actions are arranged in appropriate models also keeping into account parameters of interest, such as timing information or any particular environmental conditions. A *correlation* module is devoted to find relationships among the information described in the models; finally a *clustering* module allows to classify the patterns extracted by the correlation module. The following sections will outline the most relevant features of each of the mentioned components.

4.1 Data Preprocessing

The data collected by the sensors are often affected by errors, due to imprecise measurements or to environmental noise. This module is devoted to the detection and removal of invalid values in raw sensory data and to possibly estimating missing data. Initial filtering can be performed assuming that there exist some admissible ranges for the values of the observed variables, and removing all values out of that range. Moreover, spacial and time redundancy can be exploited to detect anomalies in data, or to replace missing data. In particular time redundancy consists in the correlation between consecutive observations read from a sensor, while spatial correlation regards readings from neighboring sensors at a given time [5].

Time correlation can be exploited for the estimation of missing data by means of a linear interpolation between preceding and subsequent observations. Also spatial correlation may be exploited for sensors located in small indoor environments, such as office rooms; in this case, it is reasonable to assume that nearby sensors will retrieve similar measurements, due to the intrinsic nature of the considered physical quantities, so that analysis techniques pointing out significant differences in sensed data may be effectively identify potential outliers. On the other hand, measurements collected in such small indoor areas are typically subject to diverse influencing factors, not necessarily referable to natural phenomena; for instance, the influence of actuators (such as the air conditioning systems, as regards temperature and humidity, or artificial lights, as regards ambient light) cannot be disregarded altogether.

Taking into account the location of sensors within the environment it is thus possible to detect areas in which sensors readings should be correlated at a time point. However, for our aims, it is also important to analyze differences between sensors belonging to different, but close areas. In fact, in this case, variations in sensors readings can be due to the use of actuators from users. So at this stage of preprocessing, spatial correlation property among intra-area sensors can be profitably employed in order to detect outliers, and to replace them with the combination of neighbor sensors readings, while later, the action detection module will analyze in depth the inter-area differences.

4.2 Action Detection

This module aims to perform a deeper analysis of the observed variables trends ,and to recognize those events that can be ascribed to human intervention. We assume here that sudden changes in observed values can be consequence of users' actions, such as turning on/off the light or changing the settings for the temperature and humidity control systems.

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 in order to detect and remove repetitive and regular changes in data, so as to consider only meaningful changes in the data trend. Finally, a probabilistic inference process can be performed to associate the changes detected in the series to possible user actions on the environment.

4.2.1 Decomposition of sensory data time series

The analyzed time series obtained from sensor readings can be decomposed into a set of components: a *trend*, a *seasonal* and a *remainder* component [20].

The trend component T_t defines the long-term trend of the variable and can be defined as the tendency to increase, decrease or remain constant over a long period of time; it varies slowly over time and essentially determines the level of the series.

The seasonal or periodic component S_t is given by one or more periodic components, characterized from taking the same or similar values at a fixed distance in time. The remainder component R_t determines short-term fluctuations in the series. The three components can contribute to the composition of the observed series y_t in different ways, additive ($y_t = T_t + S_t + R_t$), multiplicative ($y_t = T_t S_t R_t$), or multiplicative with an additive remainder component ($y_t = T_t S_t + R_t$).

The techniques of decomposition of a time series into its components depend on its composition model. This decomposition is important in order to estimate and remove a regular and predictable component which could hide useful information. In fact, 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.

4.2.2 Analysis of changes in time series data

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 an turning on of the light, or in the case we are analyzing temperature data could be interpreted as an increase of temperature settings. Moreover the strength allows to analyze a subset of detected changes. In fact, given a threshold experimentally defined, called ϑ, only changes with strength greater than ϑ will be considered.

4.2.3 Probabilistic inference for action detection and interpretation

Probabilistic models, based on dynamic Bayesian networks, are then used to estimate which of those events may be in fact associated to human actions and to interpreting them. The choice of using these knowledge models depends on the fact that we are bound to reason about uncertain knowledge. We can model the interest variables and their possible states as nodes in the network, connected with directed links representing the influence among the nodes. 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.

In particular, two kinds of variables will be modeled, *UA* variables, representing possible user actions and *SO* variables, representing sensory observations. Therefore the built model is used to estimate the probability that the event we want to investigate did occur, based on the sensor readings. As an example, Figure 2 shows the Bayesian network used to estimate the user actions controlling ambient light.

This network represents the possible action on the light settings, by means of 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 modeled as states of the *UserAction t*, which depends on the state of the 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 turns the state of the light at the next istant (*LightTrend t + 1* variable).

4.3 Modeling and Correlation

The correlation module is devoted to identifying relationships among the extracted information. For example it is possible correlate users with sensors measurements within an environment. Matrix models are used to represent the values recorded by different sensors for what concerns a given physical variable during a specific period (such as for instance an entire working day) and to represent the occurrences of events such as user actions in specific instants. Therefore, for each analyzed variable we can build a matrix, whose rows 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 to generate several columns. Anyway, it is interesting to evaluate only the most meaningful informative content. We can therefore process data in order to perform a dimensionality reduction and then evaluate the correlation between variables of interest.

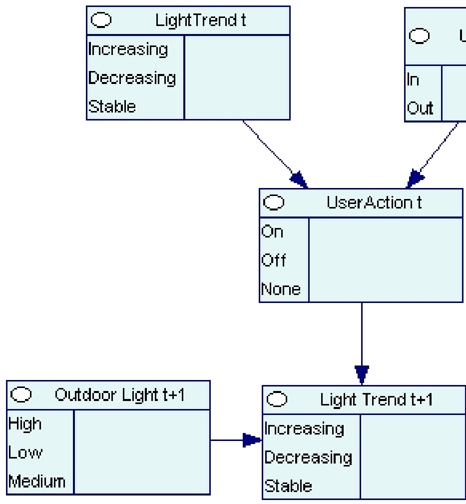

Fig. 2 A Probabilist model to estimate user actions controlling ambient light.

4.3.1 Dimensionality Reduction

The set of observations is to undergo a dimensionality reduction process, by means of PCA [21]. Principal components analysis is a methodology which allows the projection of original dataset space into a smaller space: by a linear transformation of the original variables, PCA extracts a set of orthogonal vectors, called principal components, 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 [22].

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 $X = [\mathbf{x}_1, \ldots \mathbf{x}_n]$.

As an example, let $\mathbf{T} = [\mathbf{T}_1, \ldots \mathbf{T}_n]$ 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. After performing PCA we obtain a $m \times f$ matrix, with $f \leq n$, $\mathbf{T'} = \left[\mathbf{T'}_1, \ldots \mathbf{T'}_f\right]$ (see Figure 3). The same procedure can be performed over the entire set of observed variables.

4.3.2 Correlation among variables of interest

The entire set of variables and events observations is then represented as a matrix $\mathbf{X}(m \times nv)$, where a row $\mathbf{X}_i$ represents an observations at a specific time i and a

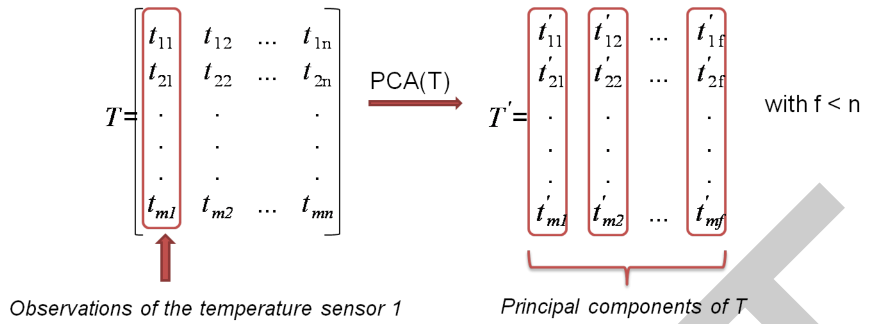

Fig. 3 Principal Component Analysis performed on temperature dataset

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:

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

composed of a set of vectors, each one represents the set of observations of a specific variable. In particular the set $\mathbf{U} = \{\mathbf{U}_j\}_{j=1\dots d}$ represents observations about the presence of d users in the considered period, while sets $\mathbf{T} = \{\mathbf{T}_j\}_{j=1\dots f}$ and $L = \{\mathbf{L}_j\}_{j=1\dots g}$ 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.

The correlation matrix $\mathbf{C}(nv \times nv)$ is then computed in order to highlight the relationships among the variables. 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:

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

where σ_{ij} is the covariance between $\mathbf{X}_i$ and $\mathbf{X}_j$ and σ_i and σ_j are respectively the standard deviation of $\mathbf{X}_i$ and $\mathbf{X}_j$.

In this way 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.

4.4 Clustering

The clustering module allows to classify the pattern extracted by the correlation module. The clustering leads to the subdivision of users' behavior patterns into a set of profiles based on their similarities. In this way we can group users with similar preferences about variables setting, or users performing the same actions in similar environment conditions. In particular the K-means [23] algorithm can be used to classify data from the sub-matrices extracted from the correlation matrix **C**.

The algorithm requires the number of clusters to be obtained and a distance function to evaluate distances between data points and cluster centers. These parameters can be experimentally defined. An iterative process is then performed, during which k data in the dataset are randomly chosen to constitute the first centroids of the clusters. Then 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.

5 Analysis of Sensor9k Dataset

In this section we report first evaluations of the proposed system on Sensor9k dataset. A set of experiments have been conducted analyzing data measured from MTS300 sensor nodes, where the analyzed variables are light, and temperature. Additional information has been collected to examine users presence, namely the status of the door of the office room, and outdoor light measurements.

In particular we have analyzed light and temperature 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 measurements related to *Room2*.

5.1 Analysis of light time series for the analyzed rooms

Figure 4 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 dif-

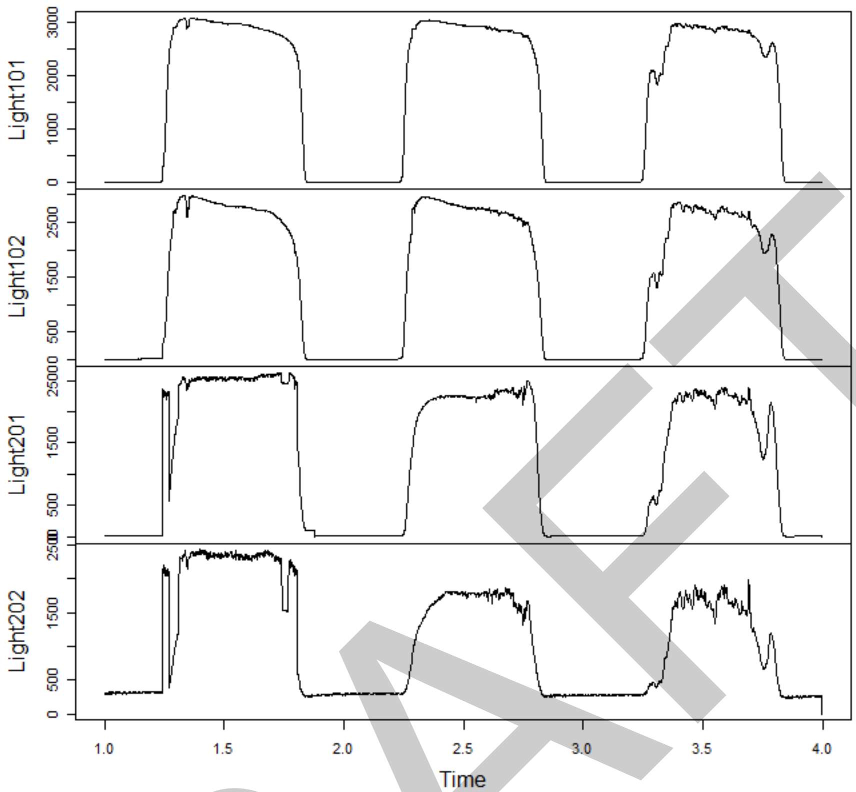

Fig. 4 Time Series for *Light101*, *Light102*, *Light201* and *Light201*

ferences 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 5 and 6 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.

5.2 Event detection and interpretation

Figure 7 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).

Figure 8 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).

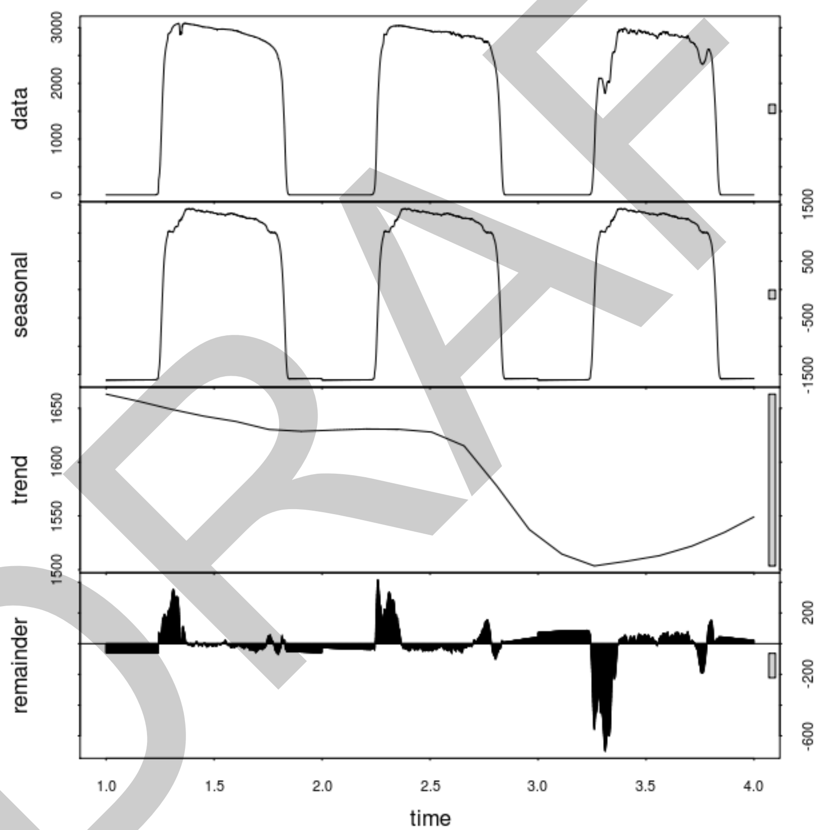

Fig. 5 Decomposition of Light101 Temporal Series

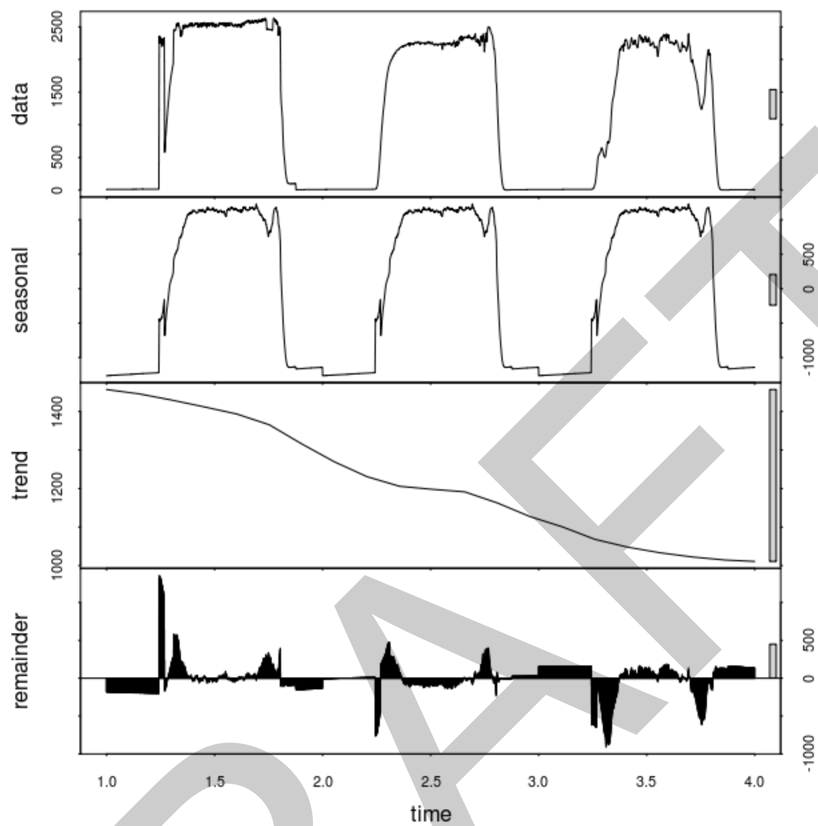

Fig. 6 Decomposition of Light201 Temporal Series

5.3 Clustering of data

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.

In particular, we report here the experiments the experiments related to the clustering module; Figure 9 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, regarding the later time of the day, where the influence of *User2* may be easily

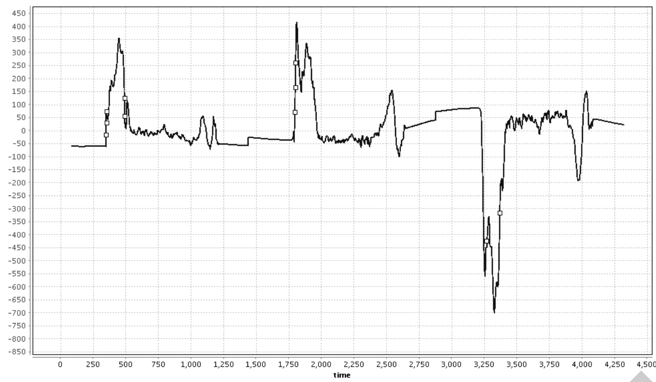

Fig. 7 Light101 Relevant Events: the curve represents the trend of Light, while squares represent the recognized events.

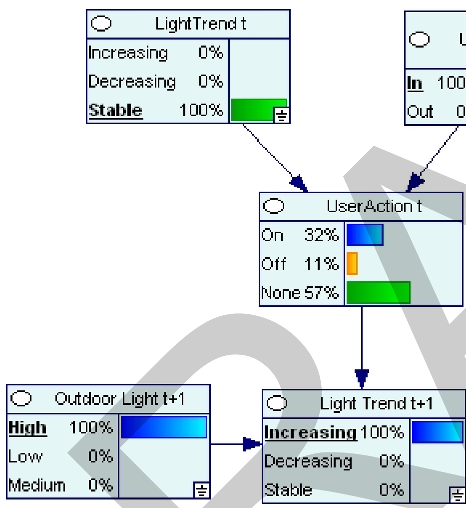

Fig. 8 Probabilist reasoning on a recognized event.

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.

6 On-going work: extension to a distributed approach

This paper described a system aiming to profile users of an intelligent environment; the underlying idea is that the extracted profiles may be used to adapt the environment settings to users preferences, and for this reason further research will be

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. 9 Relation between Users presence and Temperatures and Lights values in a day.

devoted to investigate methods for analyzing and processing real-time information coming from different sources deployed in the intelligent environment. Moreover, an interesting direction for further investigation might regard the implementation of the final system according to a distributed architecture; for instance, it is common to implement the pervasive sensory system necessary to gather environmental measurements by using a separate wireless sensor network for each environment under observation; each network will collect data toward its base station (usually represented by a node, also called *micro-server*, with more computational power than usual sensor nodes), which will be part of a larger backbone network, devoted to ensure connectivity with a remote central storage and processing server. It is thus conceivable to distribute part of the profiling analysis to the local micro-servers, thus relieving the central server from this additional burden; only higher-level information about the profiles will need to be forwarded for further processing. Such architecture is well suited for an agent-based implementation. The analysis of real time data coming from sensors may in fact be performed by a set of autonomous agents, each of which responsible for a particular location. In this way, each agent would perform computations and independently take decisions without overloading the central unit. Moreover agents must be able to exchange information, for example to periodically update users models. In this case, distributed data mining algorithms, would allow agents to individually analyze and process the different informations coming from the multiple sources of information, exchanging and integrating them.

7 Conclusion and future work

In this paper we presented a sensor mining system aimed to profiling occupant behaviors. A specific case of study regards Sensor9k, a project carried out at the Department of Computer Science of Palermo University. Environmental variables are monitored by a sensor network, and a set of modules allow to extract useful information regarding user actions and habits. The system is still under evaluation, but we discusses some preliminary results that appear promising with respect to the detection of basic users' habits based on their use of the actuators, which we implicitly infer through the difference in sensory measurements.

The main problem of evaluation is primarily concerned with the need for more data to analyze in order to better understand users habits. A more detailed and extensive dataset will allow to better validate our system. In future work the inferred information about users behaviors and habits will be formalized in an appropriate user model. The creation of this model will have different applications. In particular the users models can be used to adapt the intelligent environment to the user's preferences, or vice versa in order to identify wrong behavior of employees in order to reduce the energy consumption of the building.

References

1. A. Vasilakos and W. Pedrycz. Ambient Intelligence, Wireless, Networking, Ubiquitous Computing. Artech House Press, MA, USA (2006).
2. Froschl, C. 2008 User Modeling and User Profiling in Adaptive. VDM Verlag.
3. O'Sullivan, D., Smyth, B., and Wilson, D. 2003. Explicit vs implicit profiling: a case-study in electronic programme guides. In Proceedings of the 18th international Joint Conference on Artificial intelligence (Acapulco, Mexico, August 09 - 15, 2003). International Joint Conference On Artificial Intelligence. Morgan Kaufmann Publishers, San Francisco, CA, 1351-1353.
4. Akyildiz, IF and Su, W. and Sankarasubramaniam, Y. and Cayirci, E.: A survey on sensor networks. IEEE communications magazine. Volume 40, Issue 8, 2002, Pages 102–114
5. Shaomin Wu, Derek Clements-Croome, Understanding the indoor environment through mining sensory data–A case study, Energy and Buildings, Volume 39, Issue 11, November 2007, Pages 1183-1191, ISSN 0378-7788, DOI: 10.1016/j.enbuild.2006.07.011.
6. Diane J. Cook, Juan C. Augusto, Vikramaditya R. Jakkula, Ambient intelligence: Technologies, applications, and opportunities, Pervasive and Mobile Computing, Volume 5, Issue 4, August 2009, Pages 277-298, ISSN 1574-1192, DOI: 10.1016/j.pmcj.2009.04.001.
7. M. C. Mozer. Lessons from an adaptive home. In D. J. Cook and S. K. Das, editors,Smart Environments: Technology, Protocols, and Applications, pages 273-298. Wiley, 2004.
8. A. Khalili, C. Wu and H. Aghajan. Autonomous Learning of User's Preference of Music and Light Services in Smart Home Applications. Behavior Monitoring and Interpretation Workshop at German AI Conf, Sept 2009.
9. Barbato, L. Borsani, A. Capone, S. Melzi. Home Energy Saving through a User Profiling System based on Wireless Sensors. ACM Buildsys 2009 (in conjunction with SenSys 2009), Berkeley, CA, Nov. 3, 2009.
10. G. M. Youngblood. Automating inhabitant interactions in home and workplace environments through data-driven generation of hierarchical partially-observable Markov decision processes. PhD thesis, The University of Texas at Arlington, 2005.
11. F. Doctor, H. Hagras, and V. Callaghan. A fuzzy embedded agent-based approach for realizing ambient intelligence in intelligent inhabited environments. IEEE Transactions on Systems, Man, and Cybernetics, Part A, 35(1):55-65, 2005.
12. U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. From data mining to knowledge discovery in databases. AI Magazine(1996),v.17, pp 37-5, AAAI Press.
13. Dr. M.H.Dunham, Data Mining, Introductory and Advanced Topics, Prentice Hall, 2002
14. V. Cantoni, L. Lombardi, P. Lombardi, "Challenges for Data Mining in Distributed Sensor Networks," Pattern Recognition, International Conference on, pp. 1000-1007, 18th International Conference on Pattern Recognition (ICPR'06) Volume 1, 2006

15. Alessandra De Paola, Alfonso Farruggia, Salvatore Gaglio, Giuseppe Lo Re, Marco Ortolani: Exploiting the Human Factor in a WSN-Based System for Ambient Intelligence. CISIS 2009: 748-753
16. Alessandra De Paola, Salvatore Gaglio, Giuseppe Lo Re, Marco Ortolani: Human-ambient interaction through wireless sensor networks. Proceedings of the 2nd IEEE conference on Human System Interactions, Pages 61-64, 2009.
17. M.J. Akhlaghinia, A. Lotfi, C. Langensiepen, and N. Sherkat. Occupant Behaviour Prediction in Ambient Intelligence Computing Environment. Special Issue on Uncertainty-based Technologies for Ambient Intelligence Systems. Volume 2 Number 2, May 2008.
18. T. Fawcett and F. J. Provost. Combining data mining and machine learning for effective user profiling. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD), pages 8–13, 1996.
19. Dong, B. and Andrew, B. Sensor-based Occupancy Behavioral Pattern Recognition for Energy and Comfort Management in Intelligent Buildings. Proceedings of Building Simulation '2009, an IBPSA Conference, Glasgow, U.K.
20. Peter J. Brockwell, Richard A. Davis. Time Series: Theory and Methods. Springer, 1998-09-18. ISBN: 038797429
21. Jolliffe, I. T. Principal Component Analysis. Springer-Verlag. pp. 487. (1986) doi:10.1007/b98835. ISBN 978-0-387-95442-4.
22. Jiawei Han and Micheline Kamber. Data Mining: Concepts and Techniques, 2nd ed. The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor Morgan Kaufmann Publishers, March 2006. ISBN 1-55860-901-6
23. J. B. MacQueen. Some Methods for classification and Analysis of Multivariate Observations", Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, (1967) 1:281-29