Extracting Structured Knowledge From Sensor Data for Hybrid Simulation

Marco Ortolani

Abstract Obtaining continuous and detailed monitoring of indoor environments has today become viable, also thanks to the widespread availability of effective and flexible sensing technology; this paves the way for the design of practical Ambient Intelligence systems, and for their actual deployment in real-life contexts, which require advanced functionalities, such as for instance the automatic discovery of the activities carried on by users. Novel issues arise in this context; on one hand, it is important to reliably model the phenomena under observation even though, to this end, it is often necessary to craft a carefully designed prototype in order to test and fine-tune the theoretical models.

The work described here proposes to use sensor nodes to capture environmental data related to users' activities; the representation of the environment will rely on an ontology expressed in a well-established ad-hoc formalism for sensor devices. An activity model will be produced by analyzing the effect of users' actions on the collected measurements, in order to infer the underlying structure of sensor data via a *linguistic* approach based on formal grammars. It will finally be shown how such model may be profitably used in the context of a hybrid simulator for wireless sensor networks in order to obtain a scalable and reliable tool.

1 Introduction

Recent developments in sensing technology make it possible to provide non-intrusive and easily deployable solutions for detecting the most diverse observations, spanning from simple phenomena to complex events; this has given rise to the new paradigm of the Internet of Things (IoT), whose fundamental idea is that the objects pervading our everyday environment may make themselves identifiable and may be turned into

M. Ortolani (✉)
DICGIM, University of Palermo, Viale delle Scienze, Edificio 6, 90128 Palermo, Italy
e-mail: marco.ortolani@unipa.it

S. Gaglio and G. Lo Re (eds.), *Advances onto the Internet of Things*,
Advances in Intelligent Systems and Computing 260, DOI: 10.1007/978-3-319-03992-3_11,

smart objects thanks to their ability to communicate with each other and to access other objects' aggregate information [4]. Among available enabling technologies, commonly available wireless sensor networks [2] are noteworthy.

Such developments broaden the horizon of research to include novel fields, such as that of Ambient Intelligence (AmI), a branch of Ambient Intelligence that focuses on adapting the environmental conditions to maximize the user's comfort, and aims to do so transparently by applying methods and ideas borrowed from such fields as pervasive and ubiquitous computing [11, 21]. A key functionality for many AmI applications in a home setting is the automatic recognition of human activities (e.g. cooking, sleeping or working at one's desk), as it represents the premise for the design of complex systems; for instance, this kind of contextual information is fundamental in systems designed for energy saving in buildings, where the activation/deactivation of energy-hungry devices heavily depends on whether the occupants are present or not.

A typical approach to creating predefined models for the most common activities is to use supervised classification methods, but the present proposal substantially differs in that it specifically aims to relieve the designer from the task of creating a detailed model for each activity to track. Even though a network of wireless sensor nodes is apt to represent the connection between the system and the real world, an extensive deployment would not be practically viable, especially during the initial design phase. A common approach to assessing the validity of AmI systems is thus to develop a minimal, albeit fully functional, prototype of the intelligent application to be actually deployed into the environment. This is useful in the light of obtaining more detailed understanding of the semantics of sensor data; in particular, the aim is to infer the *underlying structure* of the environmental sensor readings produced as a consequence of users' interactions with their surroundings, i.e. of users' actions. In other words, the problem will be addressed from an algorithmic perspective, rather than a machine learning one, so that no detailed prior knowledge about the specific application scenario is necessary, other than the readings collected from an environmental sensor network, and the description of the nature of such data in the form of an ontology.

Data acquisition in the system described here will be guided by a tool for *hybrid simulation* of wireless sensor networks, able to exploit models constructed from data sensed by a minimal deployment of actual nodes, and augmented with additional virtual nodes to simulate the behavior of a more complex network. Such a tool may be very effective for identifying users activities starting from sensor data, in order to build reliable models; however, several specific issues arise, with different degrees of complexity, and spanning various architectural levels:

physical-level issues: observed sensor data is likely noisy, either deriving by human inaccuracies (e.g. mistakenly operating one of the actuators), or by the sensor system itself, failing to accurately perform data transmission;

data preprocessing issues: the start and end time of a performed activity is often not easily identifiable, due to the intrinsic ambiguity in the observed sensor data with respect to which activity is taking place; users may perform activities in many different ways, thus making it difficult to infer a general description of an activity;

intelligent analysis: the design of a system able to exploit knowledge about user activities is particularly challenging, also because it requires extensive coverage of the sensor infrastructure; this may be overcome by resorting to simulation, but in this case a reliable predictive model for simulating user activities is also required.

The remainder of the chapter is organized as follows: Sect. 2 provides some technical background about documented techniques for creating usable ontologies for sensors, and for detecting users' activities; Sect. 3 describes the current proposal about a system for structural knowledge extraction from raw sensor data, and Sect. 4 describes an application scenario to an AmI context where users' activities are to be identified and modeled; finally, Sect. 5 presents some conclusions and describes further directions for research.

2 Technical Background

Even though sensor networks provide a practical solution for gathering and analyzing measurements due to complex phenomena [7], the lack of integration and communication between different networks, often prevents the extraction of global knowledge about the monitored phenomena. As pointed out in [26], the main difficulty lies in the fact that sensors encoding of observed phenomena is intrinsically opaque, so metadata play an essential role in managing such kind of data: only the addition of *semantic value* to bare data may provide spatial, temporal, and contextual information essential for analyzing sensor readings.

A powerful tool to address this issue is represented by *ontologies*, which, according to [15], may be defined as a set of formally defined concepts and relations that are relevant to a knowledge domain; roughly speaking, the three three types of semantics mentioned above, and normally associated with sensor data are also useful to define ontologies; moreover, ontological models are suitable to perform advanced tasks, such as representing the sensor domain [22, 25]. SensorML [6], for instance, has been designed as a generic data model expressed in UML for capturing classes and associations that are common to all sensors. Its definition falls within a broader attempt by the Open Geospatial Consortium (OGC)[1] to provide access to sensors via distributed applications and services in order to create some kind of "sensor web" [23]. Instances of classes and associations may be then used to define specific sensor profiles in the light of further processing and integration of the collected readings, even from large number of sensors.

As pointed out in [9], however, SensorML does not include formal definitions of the classes or relations it uses, so it provides no logical or axiomatic-grounded

[1] http://www.opengeospatial.org

theory to account for its conceptualizations; strictly speaking, it therefore cannot be considered to be an ontology. On the other hand, SensorML is able to provide a generic data model for expressing knowledge and data about sensors, as well as sensor metadata and sensed attribute values; moreover, it may be considered as the basic tool to define an actual sensor ontology. Instantiations of classes and associations provided by SensorML can be used to create specific sensor profiles in order to implement higher-level services or to interact with more advanced ontologies (e.g. OWL) [23].

Following the same line of thought, the authors of [25] discuss a Semantic Sensor Web (SSW) in which sensor data is annotated with semantic metadata to increase interoperability as well as provide contextual information essential for situational knowledge, coherently with the standardization efforts of the Semantic Web Activity of the World Wide Web Consortium.[2]

Such frameworks lead the way to new interesting developments; as proposed in [28], given an ontology for an environmental domain, it seems reasonable to suggest that sensor data acquisition can be translated into ontological knowledge acquisition, for instance by means of supervised machine learning to map sensor data into ontological concepts. However, extracting meaningful knowledge from sensor data is still an open challenge, despite the availability of low-power, small-size, wireless technology, and of the related communication and programming models [28].

For the purpose of the present discussion, the focus will be on the extraction of user activity models; in fact, automatic discovery of user activities starting from raw data is a well-studied problem in different contexts, ranging from home automation, to elderly care [29]. As pointed out in [10], from a data mining perspective, activity discovery is often seen as the problem of detecting recurring patterns within a sequence of events; however, there are substantial differences between frequent itemsets detection, and discovery of patterns corresponding to activities, especially since in the latter case, ordering and repetitions of elements must typically be considered. In order to take those factors into account, the authors of [19] use a variant of the *apriori* algorithm [1] to discover sequences of events repeating with regular periodicity, besides patterns related to frequent activities. Another approach, proposed in [8], relies on standard apriori and considers the event sequence as a stream of incoming data; after identifying all sequences of predefined size and support, a transform function maps them into models for activities. A hierarchical description is proposed for such models, and activities are divided into tasks and sub-tasks; the bottom of the hierarchy is represented by activities that cannot be further decomposed; activity recognition, as well as description, is carried on in a bottom-up fashion. A similar approach is described in [18], where the authors address the issue of broken or concurrent activities by considering *emerging patterns*, i.e. those patterns able to capture meaningful differences between two classes of data.

The present proposal, however, argues for the the use of a historically alternative approach, namely the *symbolic* one, for identifying sequences of perceptual inputs and recognizing their underlying *syntactic* structure, which may be used to extract

[2] http://www.w3.org/2001/sw/

a more general pattern representing entire sequences. In this approach, high-level concepts may be identified with computational entities, such as for instance Finite State Automata (FSA), in the simplest case. Inferring models for symbolic learning, however, typically requires the availability of large amounts of source data. When wireless sensor networks are used as the data gathering infrastructure, the issue of expensive deployment cannot be disregarded. In such scenario, simulation is a cost-effective choice for prototyping and testing such applications, as the cost, time, and complexity involved in deploying and constantly changing actual large-scale WSNs for experimental purposes may be prohibitively high. Our proposal for an optimized simulator, consisting of a hybrid approach exploiting data provided by a minimal deployment of actual nodes in order to integrate and validate the data computed in simulation, has been described in [13, 16], and its versatility has been proved by using it in different contexts [17]. Building up on those works, the following discussion will show how to deal with the intrinsic heterogeneity in sensors and data to obtain reliable models for high-level concepts, such as users' activities.

3 Simulating User Activities via a Structural Approach

Considering an indoor environment, fully equipped with sensor devices for capturing all relevant measurements, the actions of the occupants will perturb the environment state; moreover they will likely follow some pattern, not known a priori, which will be reflected into the sensor readings. If we are to infer a model for such actions, we need a detailed knowledge about the nature of the sensing devices, their features and how those are related to the yielded measurements. As previously mentioned, all those characteristics may be effectively represented by means of an appropriate ontology, as long as *contextual information* is accounted for. Generally speaking, ontologies, make it possible to assign data a description of its acquisition and communication policies in order to enhance the integration of the sensor data [5]; moreover, ontologies may be easily fit within a more complex reasoning system. The remainder of this section will show how such an ontology may be used in combination with a hybrid simulator for wireless sensor networks able to generate a sufficiently large number of source data, partly sensed from actual devices and partly generated through reliable models.

3.1 Using a Hybrid Simulator to Capture Activity Data

The design of a WSN-based application is a challenging task due to its dependence on the specific scenario. Although a typical choice for testing such kind of networks requires devising ad-hoc testbeds, this is often impractical as it calls for expensive, and hard to maintain deployment of nodes. On the other hand, simulation is a valuable option, as long as the actual functioning conditions are reliably modeled, and carefully replicated. This is especially true in the context of Ambient Intelligence applications,

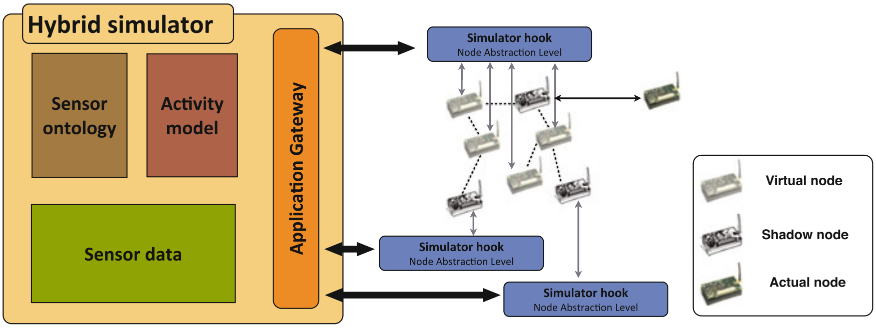

Fig. 1 High-level representation of the structure of the simulation framework

which typically need *scalable* simulations that provide *reliable* sensory data to be used in the preliminary design and test phases.

Although a few general-purpose testbeds have been proposed in the past, WSN applications must be tested on a large scale, and under complex and varying conditions in order to capture a sufficiently wide range of interactions, both among nodes, and with the environment. However, the deployment of a large number of nodes in hostile environments could become prohibitively expensive and practically unfeasible because of maintenance costs. Simulations can address those issues by providing controlled, reproducible environments, and specialized software tools for monitoring and debugging, so that the actual deployment of nodes may be delayed till after algorithms have been thoroughly tested; however they may not deliver fully reliable results, especially due to over-simplistic assumptions about the physical channels, and the node radio models. The issue of implementing an effective simulator for WSNs has been discussed in [16], which describes a framework for supporting the user in early design and testing of a wireless sensor network with an augmented version of a standard WSN simulator, that allows merging actual and virtual nodes seamlessly interacting with each other. One of the key challenges when trying to deliver such kind of realistic simulations is how to ensure fidelity, in terms of ability to reproduce the same behavior both in virtual and real nodes, and accuracy of timing.

Figure 1 shows the architecture of the simulation framework, which will be used also for the present discussion. Once the virtual network is set up by the network administrator, the actual simulation process is handed over to an augmented version of the simulator, customized in order to introduce the notion of *shadow* nodes, i.e. wrappers whose main purpose is to act as interfaces toward their real counterparts, while appearing identical to other virtual nodes from the simulator point of view. The main function of shadow nodes is thus to collect sensed data from the real world, and to re-route communication from virtual nodes to actual ones. In order to ensure correct coordination among virtual, shadow, and actual nodes, simulation timing must be constrained to satisfy at least soft real-time guarantees; this functionality is provided by the application gateway. Introducing actual data into a simulation usually requires constructing complex models, after a considerable amount of data

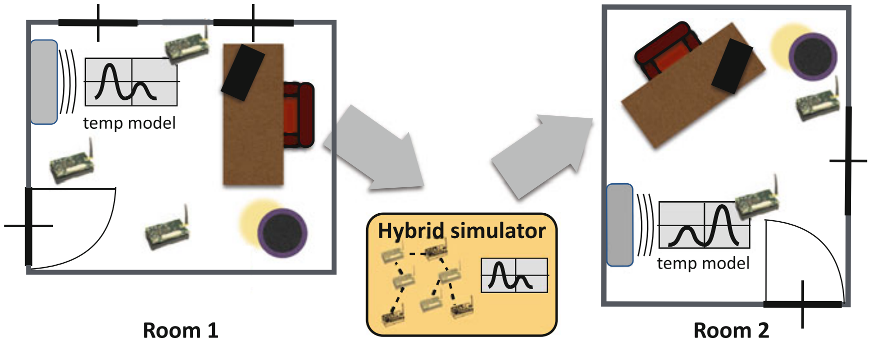

Fig. 2 Scheme of "model transferring"

is preliminarily stored, which is usually very time consuming. The present chapter will instead propose to embed into the hybrid simulator, a description of the nodes characteristics in the form of an ontology, as well as a high-level model for describing information extracted from sensory data, in the form of a linguistic description of user activities.

In order to get basic knowledge about the process or phenomena that generated data, the simulator embeds a "Sensor ontology" submodule, as shown in Fig. 1, with the aim of capturing simple descriptions of the nature of the sensors employed in a specific application scenario, expressed in SensorML formalism. For instance, SensorML can provide a complete and unambiguous description of what is called the *lineage* of an observation, thus delivering a detailed account of the process by which an observation was acquired, possibly also including further postprocessing. This is the first step toward the creation of the basis for intelligent applications grounded into the real world, coherently with the vision of the Internet of Things; in fact, this approach moves the paradigm a step forward in the direction of Semantic Web, i.e. an evolving extension of the World Wide Web in which the semantics of information is formally defined through ontologies, thus allowing for effective automatic interpretation of data content. The most relevant technologies of the Semantic Web include the Resource Description Framework (RDF) data representation model, the ontology representation languages RDF Schema and Web Ontology Language (OWL). A specialization of this paradigm has been proposed for the specific context of the Sensor Web [25]; however, lack of a commonly adopted standard represents an obstacle to a complete realization; to overcome this problem, semantic annotations have further be included so that software applications may be able to "understand" and reason over sensor data consistently, and accurately.

The ultimate purpose for using simulators is the creation of refined models for the sensed data; in particular, this chapter deals with high-level models, which need to be general enough so that they may be exported to different scenarios. The idea is outlined in Fig. 2; initially, reliable models for data sensed in one specific environment taken as reference are extracted, in order to be used as a realistic representation for information coming from virtual nodes. Such models may ideally be “ported” to different sites so that a complex scenario may be realistically simulated. Actual deployment of a limited number of sensor nodes is required only for the reference site, so the approach is both scalable and cost-effective.

3.2 Activity Discovery through Grammar Induction

The focus of the proposal described in this chapter is the recognition of user daily-life activities by simply relying on the analysis of environmental sensory data, with the underlying assumption that a hidden structure may be recognized in such data. A typical approach to creating predefined models for common activities is to use supervised classification methods, i.e. to begin by labelling sample data to train the model, and to go onto look for occurrences of previously found “representatives” in actual data. A few assumptions are implicit with this strategy; namely, all involved users are assumed to carry on the same set of known activities with similar regularity and demeanor; also, a considerable amount of data needs to be collected and consistently labelled, and finally incomplete or discontinuous patterns within original data are not well tolerated. In general, unmediated use of raw sensory data is likely to result in overfitting; in other words, the obtained models are so overly specific for the employed dataset that they are not generalizable to different contexts or users.

The present proposal substantially differs in that it specifically aims to relieve the designer from the task of creating a detailed model for each activity to track; the problem is addressed from an algorithmic perspective, rather than a machine learning one, so that no detailed prior knowledge about the specific application scenario is necessary, other than the readings collected from an environmental sensor network, and the description of the nature of such data in the form of an ontology. The goal is to infer the hidden structure that is inevitably present in data, given that the readings were ultimately caused by the actions of users; this task resembles what human beings do when performing the operations of recognition and classification. In fact, the human brain is very efficient at inferring general patterns from specific examples, a process known as *inductive reasoning*. In the context of algorithmic information theory, an abstract representation for a simplified version of those patterns is provided by formal grammars, and the process of generalization is known as grammatical induction, or grammatical inference; in this case, the specific examples may be regarded as sentences, while the patterns (i.e. sequences with some underlying structure) are captured by the entire grammar.

As pointed out in [27], inferring a language of infinite cardinality from a finite set of examples is not a trivial task, due to the difficulty is finding a good balance between generalization and specialization. A typical technique thus consists in starting with the basic constituting elements of the supposed target language (e.g. symbols directly computed from the raw data) and recursively aggregate them by some kind of merging (e.g. state merging when automata are considered, or non-terminal merging in the case of context-free grammars).

Within computational learning theory, three main formal models have been proposed for inductive inference [24], namely: *identification in the limit* by Gold [14], the *query* learning model by Angluin [3], and finally the *probably approximately correct* (PAC) learning model by Valiant [30].

Identification in the limit regards learning as an infinite process based on the availability of examples from an unknown grammar $\mathscr{G}$ which is to be inferred; positive and negative examples may be provided to the estimation algorithm, which produces a conjecture about the grammar, refining it as long as more examples are provided. If for every set of positive and negative examples from the unknown grammar $\mathscr{G}$, the algorithm guesses a correct grammar which is equivalent to $\mathscr{G}$, and never changes its guess afterwards, then the algorithm is said to identify $\mathscr{G}$ in the limit.

A different approach is the one followed by Angluin, who assumes that an *oracle* is available to the inference algorithm, i.e. some kind of support able to provide the inference algorithm with correct answers to specific queries about the unknown grammar; here, an inference algorithm repeatedly makes queries for the unknown grammar $\mathscr{G}$ to oracles, and eventually halts and outputs a correct grammar in finite time.

Finally, Valiant proposed the probably approximately correct model, which considers a distribution-independent probabilistic model of learning from random examples. In this case, random samples are drawn independently from the input data; the probability distribution is assumed to be arbitrary and unknown. The inference algorithm takes each sample as input and consequently produces a grammar. The quality of identification is measured by an accuracy parameter and a confidence parameter, which may be set by the user.

Not all approaches have the same power, and for instance it has been shown that it is not possible to learn by positive examples alone, with Gold's approach. In the approach proposed here, however, a simplified scenario will be considered and the hybrid simulator will rely on identification in the limit, assuming that the unknown grammar to be inferred may be captured by finite state automata (i.e. it is in fact regular).

4 A Sample Scenario: Predicting User Activities for Energy Saving

In order to highlight the key points of the approach proposed here, a sample scenario will be considered, consisting of an office environment whose occupants' actions are to be identified so that it is possible to reproduce them in analogous contexts; namely, the considered case study will be that of a *smart office* whose overall energy consumption needs to be controlled.

The energy demand for instrumentation in smart offices is nowadays becoming a severe challenge, due to economic and environmental reasons; typical devices with high impact on energy consumption are for instance HVAC systems, computers,

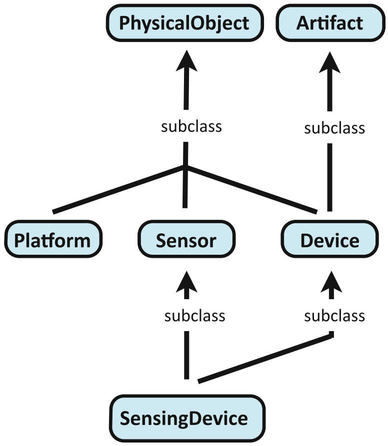

Fig. 3 SensorML representation of a generic sensing device

printers, and so on [12]. Basic usage is heavily related to the occupants' habits, so effective automated approaches must take into account basic information about users, primarily the prediction of their course of actions. In the scenario considered here, it will be assumed that the office premises are made up of smaller areas, which may be considered homogeneous with respect to their intended use and to the environmental conditions therein (e.g. independent rooms, or parts of a larger lab). Users may operate the actuators to modify the environmental conditions of the area they are occupying, and such changes will be reflected into the sensor readings; additional sensors may be deployed to capture the interaction of users with specific devices (e.g. sensor on light switches, weight sensors on chairs, and so on); a complete taxonomy of sensors and their characteristics may be represented by a specialized ontology. The description of available sensors and actuators, and of their main characteristics will be provided to the simulator in the form of a simplified SensorML code, an example of which is reported in Fig. 3.

Analysis of the readings collected by the simulator will be aimed at extracting the underlying pattern identifiable in sensor data and deriving from the action of the users on the environment. The inferred grammar will thus represent a model for the actions themselves. In order to turn raw data into symbols apt to be regarded as the basic elements of a grammar, an approach similar to what described in [10] will be followed.

The main steps of the algorithm for inferring a user activity model are outlined in Fig. 4. Initially, raw sensor triggers are clustered to form basic *events*, which are co-occurrences of triggers, identified via a customized version of an on line algorithm for string matching with wildcards (the alphabet extractor module in the figure). The event sequence needs then to be compressed, and a lossy optimal coding is used for this purpose so that events with low information content will be discarded; in other words, the most relevant patterns will be those that better describe the whole sequence, according to the Minimum Description Length (MDL) principle.

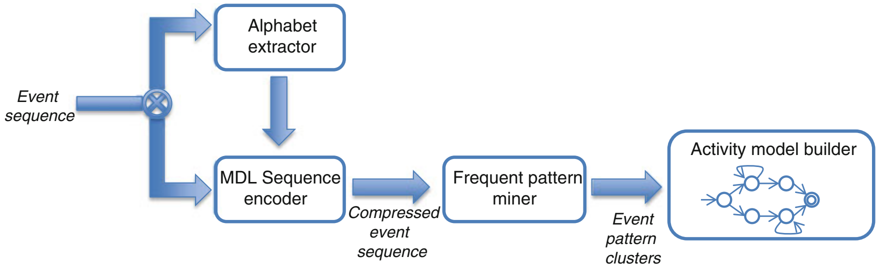

Fig. 4 Outline of the logical blocks for user activity inference

After obtaining a compressed version of the event sequence thanks to the new encoding, the algorithm turns to discovering the most frequent patterns. The proposed approach is similar to Discontinuous Varied-order Sequential Miner (DVSM) [20], which is an apriori-based iterative algorithm. The final frequent pattern set returned by DVSM contains the most relevant patterns, which will be clustered into meaningful classes to obtain the discovered activities, by integrating temporal information with other features of interest, such as composition similarity; this step is accomplished by k-medoids, a variant of the well-known k-means clustering, where representative points are bound to belong to the initial dataset. A dissimilarity measure is computed over all the possible pairs of points, making this algorithm more robust than traditional k-means. The interested reader is referred to [10] for further details.

In the last phase, the features of the obtained clusters are encoded into models representing the discovered activities. The event clusters are regarded as the basic symbols of the unknown target grammar, and positive and negative samples are constructed for the specific domain by using the event patterns, the general SensorML description of the environment, and user knowledge. Following the identification in the limit approach, a grammar representing user actions is finally inferred in the form of a set of FSAs, as depicted in Fig. 4.

5 Conclusion

This chapter described a proposal regarding the use of a tool for hybrid simulation of wireless sensor nodes in the context of Ambient Intelligence scenarios. In particular, considering a *smart office* environment whose occupants' actions are to be identified so that it is possible to model them in order to control the overall energy consumption, the approach proposed here consists in using sensor nodes to capture environmental data related to users' activities. The representation of the environment was assumed to be provided in the form of an ontology expressed in SensorML, a well-established ad-hoc formalism for sensor devices. and it has been shown how an activity model may be produced by analyzing the effect of users' actions on the collected measurements, in order to infer the underlying structure of sensor data via a *linguistic* approach based on formal grammars.

References

1. Agrawal, R., Imieliński, T., Swami, A.: Mining association rules between sets of items in large databases. SIGMOD Rec. **22**, 207–216 (1993)
2. Akyildiz, I., Su, W., Sankarasubramaniam, Y., Cayirci, E.: A survey on sensor networks. IEEE Commun. Mag. **40**(8), 102–114 (2002)
3. Angluin, D.: Queries and concept learning. Mach. Learn. **2**(4), 319–342 (1988)
4. Atzori, L., Iera, A., Morabito, G.: The internet of things: a survey. Comput. Netw. **54**(15), 2787–2805 (2010)
5. Bendadouche, R., Roussey, C., De Sousa, G., Chanet, J.P., Hou, K.M., et al.: Extension of the semantic sensor network ontology for wireless sensor networks: the stimulus-wsnnode-communication pattern. In: Proceedings of the 5th International Workshop on Semantic Sensor Networks (in conj with the 11th ISWC) (2012)
6. Botts, M. et al.: Sensor Model Language (SensorML) for in-situ and remote sensors. Technical report, Open Geospatial Consortium Inc. (2004)
7. Chella, A., Lo Re, G., Macaluso, I., Ortolani, M., Peri, D.: Multi-robot interacting through wireless sensor networks. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) **4733 LNAI**, pp. 789–796 (2007)
8. Chikhaoui, B., Wang, S., Pigot, H.: A frequent pattern mining approach for adls recognition in smart environments. In: Proceedings of International Conference on Advanced Information Networking and Applications, pp. 248–255. IEEE Computer Society (2011)
9. Compton, M., Henson, C.A., Neuhaus, H., Lefort, L., Sheth, A.P.: A survey of the semantic specification of sensors. In: Proceedings of Semantic Sensor, Networks, pp. 17–32 (2009)
10. Cottone, P., Gaglio, S., Lo Re, G., Ortolani, M.: User activity recognition for energy saving in smart homes. In: Sustainable Internet and ICT for Sustainability (SustainIT), 2013, pp. 1–9. IEEE (2013)
11. 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: Proceedings of International Conference on Complex, Intelligent and Software Intensive Systems., pp. 748–753. IEEE (2009)
12. De Paola, A., Lo Re, G., Morana, M., Ortolani, M.: An intelligent system for energy efficiency in a complex of buildings. In: Sustainable Internet and ICT for Sustainability (SustainIT), 2012, pp. 1–5. IEEE (2012)
13. Gatani, L., Lo Re, G., Ortolani, M., Sorbello, F.: A monitoring framework exploiting the synergy between actual and virtual wireless sensors. In: Proceedings of the International Conference on Parallel Processing Workshops, pp. 361–367 (2006)
14. Gold, E.M.: Language identification in the limit. Inf. Control **10**(5), 447–474 (1967)
15. Gruber, T.R., et al.: A translation approach to portable ontology specifications. Knowl. Acquis. **5**(2), 199–220 (1993)
16. Lalomia, A., Lo Re, G., Ortolani, M.: A hybrid framework for soft real-time wsn simulation. In: 13th IEEE/ACM International Symposium on Distributed Simulation and Real Time Applications (2009)
17. Lo Re, G., Milazzo, F., Ortolani, M.: Secure random number generation in wireless sensor networks. In: Proceedings of the 4th international conference on Security of information and, networks, pp. 175–182 (2011)
18. Pung, H.K.: epsicar: An emerging patterns based approach to sequential, interleaved and concurrent activity recognition. 2009 IEEE International Conference on Pervasive Computing and Communications, pp. 1–9 (2009)
19. Rashidi, P., Cook, D.: Keeping the intelligent environment resident in the loop. In: 4th International Conference on Intelligent Environments (2008 IET), pp. 1–9, (2008)
20. Rashidi, P., Cook, D.J., Holder, L.B., Edgecombe, M.S.: Discovering activities to recognize and track in a smart environment. IEEE Trans. Knowl. Data Eng. **23**, 527–539 (2011)
21. Remagnino, P., Foresti, G.L.: Ambient intelligence: a new multidisciplinary paradigm. IEEE Trans. Syst. Man Cybern. B Cybern. - Part A: Systems and Humans **35**(1), 1–6 (2005)

22. Ribino, P., Oliveri, A., Lo Re, G., Gaglio, S.: A knowledge management system based on ontologies. In: Proceedings of the International Conference on New Trends in Information and Service Science, pp. 1025–1033. IEEE (2009)
23. Russomanno, D.J., Kothari, C.R., Thomas, O.A.: Building a sensor ontology: a practical approach leveraging iso and ogc models. In: Proceedings of the International Conference on Artificial Intelligence, pp. 637–643 (2005)
24. Sakakibara, Y.: Recent advances of grammatical inference. Theoret. Comput. Sci. **185**(1), 15–45 (1997)
25. Sheth, A., Henson, C., Sahoo, S.S.: Semantic sensor web. IEEE Internet Comput. **12**(4), 78–83 (2008)
26. Sheth, A., Perry, M.: Traveling the semantic web through space, time, and theme. Internet Comput. **12**(2), 81–86 (2008)
27. Stevenson, A., Cordy, J.R.: Grammatical inference in software engineering: an overview of the state of the art. In: Proceedings of the 5th International Conference on Software Language Engineering, pp. 204–223. Springer, Berlin (2012)
28. Stocker, M., Rönkkö, M., Kolehmainen, M.: Making sense of sensor data using ontology: a discussion for road vehicle classification. In: R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (eds.) Proceedings of the (iEMSs) International Congress on, Environmental Modelling and Software (2012)
29. Tapia, E., Intille, S., Larson, K.: Activity recognition in the home using simple and ubiquitous sensors. In: Ferscha F.M. A. (ed.) Proceedings of Pervasive 2004, vol. vol. LNCS 3001, pp. 158–175. Springer, Berlin (2004)
30. Valiant, L.G.: A theory of the learnable. Commun. ACM **27**(11), 1134–1142 (1984)