Activity Recognition Approaches for Smart Cities

The City4Age use case

Aitor Almeida & Gorka Azkune
MORElab
DeustoTech-Deusto Foundation, University of Deusto
Bilbao, Spain
{aitor.almeida, gorka.azkune}@deusto.es

***Abstract*— Activity Recognition is an important ingredient that allows the interpretation of elementary data. Understanding which activity is going on allows framing an elementary action (e.g. "a movement") in a proper context. This paper presents an activity recognition system designed to work in urban scenarios, which impose several restrictions: the unfeasibility of having enough annotated datasets, the heterogeneous sensor infrastructures and the presence of very different individuals. The main idea of our system is to combine knowledge- and data-driven techniques, to build a hybrid and scalable activity recognition system for smart cities.**

Keywords—Smart Cities; activity recognition; Ambient Assisted Living

I. Introduction

As a result of the growth of urban population worldwide [1], cities are consolidating their position as one of the central structures in human organization. This concentration of resources and services around cities offers new opportunities to be exploited. Smart Cities [2][3] are emerging as a paradigm to take advantage of these opportunities to improve their citizens lives. Smart Cities use the sensing architecture deployed in the city to provide new and disruptive city-wide services both to the citizens and the policy-makers. The large quantity of data available allows improving the decision making process, transforming the whole city in an intelligent environment at the service of its inhabitants. One of the target groups for these improved services is, so-called, the "young-old" category, whose age varies from 60 to 69 [4], who are starting to develop the ailments of old age. The early detection of frailty and Mild Cognitive Impairments (MCI) is an important step to treat these problems. In order to do so, the Ambient Assisted Living (AAL) [5] solutions must transition from the homes to the cities.

City4Age[1] is a H2020 research and innovation project with the aim of enabling age-friendly cities. The project aims to create an innovative framework on ICT tools and services that can be deployed by European cities in order to enhance the early detection of risk related to frailty and MCI, and provide personalized intervention that can help the elderly population to improve their daily life promoting positive behaviour changes. As part of the tools created for the framework we have developed a series of algorithm for activity recognition and behaviour modelling. The recognized activities and the behaviour variations are then used to ascertain the frailty and MCI risks levels of the users and to recommend relevant interventions that will help the users to palliate those risks. In the past we have worked on creating single-home activity recognition algorithms [6][7], but in the case of City4Age we have created algorithms that take into account the large scale scenarios of the project. The City4Age project is being deployed in six different cities (Montpellier, Singapore, Lecce, Madrid, Athens and Birmingham), each one with different requirements and scenarios, which involve the need to identify heterogeneous activities and behaviours.

The objectives of the City4Age activity recognition algorithms is to automatically evaluate and recognize the user activities, based on the combination of the deployed sensor data and spatial and temporal information. As explained in Section 4, the City4Age algorithms do not work directly with low level sensor data, but with their projection to the action-space. In this paper we present an analysis of the challenges and problems of for the activity recognition in Smart Cities and we describe the City4Age approach to deal with them, as a guideline for future projects.

II. Related Work

There are two main monitoring approaches for human activity recognition; namely, vision-based and sensor-based monitoring. For a review of vision-based approaches, [11] can be consulted. When approaching human activity recognition in intelligent environments, sensor-based activity recognition is the most used solution [12], since vision-based approaches tend to generate privacy concerns among the users [13]. Sensor-based approaches are based on the use of emerging sensor network technologies for activity monitoring. The generated sensor data from sensor-based monitoring are mainly time series of state changes and/or various parameter values that are usually processed through data fusion, probabilistic or statistical analysis methods and formal knowledge technologies for activity recognition. There are two main approaches for sensor-based activity recognition in the literature: data-driven and knowledge-driven approaches.

The idea behind data-driven approaches is to use data mining and machine learning techniques to learn activity models. It is usually presented as a supervised learning

[1] http://www.city4ageproject.eu/

approach, where different techniques have been used to learn activities from collected sensor data. Data-driven approaches need big datasets of labelled activities to train different kinds of classifiers. The learning techniques used in the literature are broad, going from simple Naive Bayes classifiers [14] to Hidden Markov Models [15], Dynamic Bayesian Networks [16], Support Vector Machines [17] and online (or incremental) classifiers [18].

Although supervised learning reports excellent performance, the need of large-scale labelled datasets results on scalability problems for practical deployments. It seems unfeasible to obtain enough labelled data for real world scenarios, since the involved users and activities may be too numerous.

However, there are some efforts in the community directed to solve this problem. For instance, Rashidi and Cook tried to overcome the problem of depending on manually labelled activity datasets in [19]. They use an unlabelled dataset, where they extract activity clusters using unsupervised learning techniques. Those clusters are used to train a boosted Hidden Markov Model, which is shown to be able to recognise several activities. However, there are three fundamental problems in this approach: (i) the modelled and recognised activities do not have any semantic meaning which makes harder for humans to understand what a user is actually doing, (ii) activity granularity, since the clusters found may refer to chaining activities such as washing dishes after having lunch, as only one activity, and (iii) the performance of current systems are still far from supervised learning approaches.

In order to maintain the scalability of unsupervised learning approaches, but overcome the posted problems, we take ideas from knowledge-driven activity recognition approaches. Knowledge-driven activity recognition is based on real world observations that the list of objects and functionalities to perform an activity are always very similar. For example, to prepare coffee, a liquid container is needed alongside with some coffee and sugar. Although different people may use different coffee brands, some may add milk and some may prefer white sugar to brown sugar, there are some essential concepts that are always present for every activity. The idea is to use this prior knowledge to create rough activity models. The implicit relationships between activities, related temporal and spatial context and the entities involved (objects and people) provide a diversity of hints and heuristics for inferring activities.

The first step for knowledge-driven systems is to acquire the needed contextual knowledge. This is usually achieved using standard knowledge engineering approaches. Depending on the nature of the acquired knowledge, different approaches can be distinguished. Some researchers use logic-based approaches for activity recognition, as [20]. Others adopt ontology-based approaches which allow a commonly agreed explicit representation of activity definitions independent of algorithmic choices, thus facilitating portability, interoperability and reusability. Good examples can be found in [21] and in [22]. A very recent work can be found by [23], where authors use Dempster-Shafer theory to combine uncertainty reasoning and ontologic reasoning. However, their system has only been evaluated in controlled laboratory experiments.

Building on the ideas introduced by [19] about unsupervised learning, City4Age presents a scalable activity recognition system for real-world deployments. However, using knowledge-driven ideas, the approach does not suffer from activity granularity and lack of semantic meaning problems. And most notably, the obtained performance is comparable to supervised learning approaches. A similar philosophy is followed by [24] and [25]. Both systems combine data- and knowledge-driven approaches in a different way. For instance, [24] produce very detailed ontologic models based on OWL2 to later map the knowledge to Markov Logic Networks (MLN), which allows them to use probabilistic reasoning. Their knowledge engineering effort is too high for City4Age scenarios, where 6 cities should be modeled with many users, requiring very detailed activity models which question the generality of the approach. Furthermore, their evaluation is performed on datasets with low number of activities (8 at most) and they do not address the *idle* activity, i.e., time segments where no activities occur even though sensor activations appear.

On the other hand, [25] presents a very interesting system called USMART. They segment sensor activations using the semantic similarity between the fired sensors. Based on previously modelled ontologic activity models they define the sufficient conditions for a sensor sequence to be mapped to an activity. Those activity models are very similar to the EAMs we use in City4Age, but the overall knowledge engineering effort to model environments and sensors is much higher than ours. Once sensor segments are extracted, they use semantic reasoning to recognise activities. This imposes some burdens in their recognition capacities. As authors admit, they need at least a sensor event that uniquely describes an activity, in order to distinguish it from others. The evaluation they present does not consider the *idle* activity and they do not cover all the activities monitored in the used datasets.

The activity recognition system of City4Age can address the enumerated problems for [24] and [25]. Due to the nature of the presented scenarios, it is very important to keep modelling efforts low. That implies that the activity recognition algorithm has to work with roughly modelled contexts. Furthermore, City4Age also demands a system with high level of generalization, since the number of different users, sensor infrastructures and activities is very high. As a real-world deployment, we also have to consider the *idle* activity. Humans spend a lot of time of the day without performing any meaningful activities. Those situations cannot be ignored in City4Age.

III. THE CITY4AGE APPROACH TO ACTIVITY RECOGNITION

A. *Actions, Activities and Behaviors*

User's behaviour is composed by a large collection of defining elements, making it a complex structure. In order to

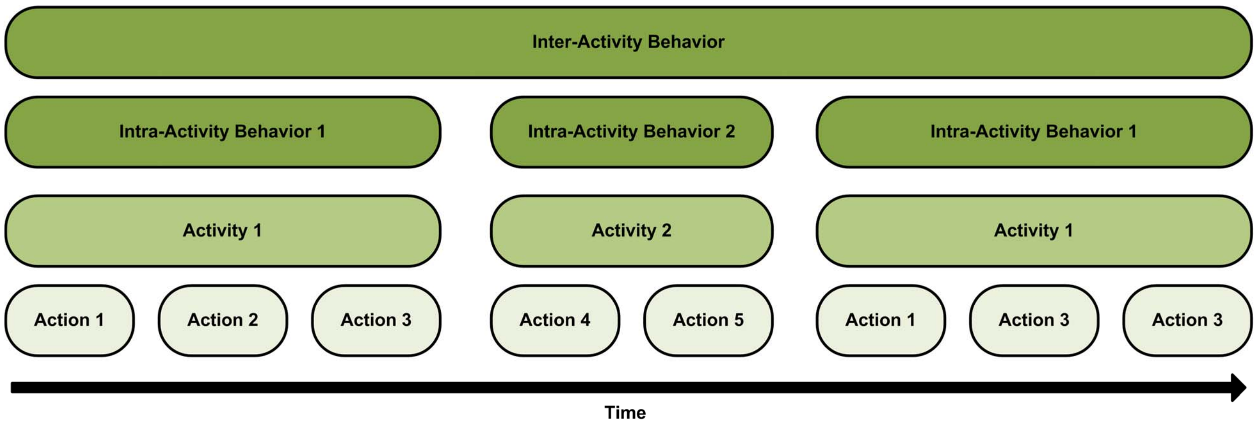

Figure 1 Elements of the user behaviour

properly describe it, we have defined a series of concepts based on the ones proposed in [8] to describe it: *Actions*, *Activities* and *Behaviours*. *Actions* describe the simplest behaviours; while *behaviours* describe the most complex conduct (see Figure 1).

The different elements of the user behaviour are:

- *Actions* are temporally short and conscious muscular movements done by the users (e.g. taking a cup, opening the fridge. . .).
- *Activities* are temporally longer, but finite, and are composed by several *actions* (e.g. preparing dinner, taking a shower, watching a movie…).
- *Behaviours* describe how the user performs those *activities* in different moments. We have identified two types of *behaviours*. The *Intra-Activity Behaviours* describe how a single *activity* is performed by a user in different moments (e.g. while the user is preparing dinner, sometimes he takes all the ingredients before starting, while in other occasions he takes the ingredients when he needs them). *The Inter-Activity Behaviours* describe how the user chains different *activities* (e.g. on Mondays after having breakfast she leaves the house to go to work, but on the weekends she goes to the main room).

B. Projection of the sensor data to the Action-Space

One of the defining characteristics of our City4Age algorithms is that they work on the Action-Space instead of the Sensor-Space. To be able to work with a more flexible representation of the information in the intelligent environments, we map the raw sensor data to *actions* (see Figure 2). The advantage of working with the Action-Space is that different sensor types may detect the same *action* type, simplifying and reducing the hypothesis space. This is even more important when using semantic embeddings to represent those actions in the model, as the reduced amount of *actions* produce more significant embedding representations.

Traditionally in activity recognition or behaviour modelling tasks, the inputs (when using *actions*) have being represented as IDs, strings or one-hot-vectors. The problem with this type of representations is that they do not contain any information about the *action* meaning. Using only a one-hot-vector is not possible to compute how similar two actions are and that information is not available for the model that will use the actions. A similar problem occurs in the area of Natural Language Processing (NLP) with the representation of words. The solution to this is to use embeddings[9] to represent the words, and we have used the same approach for the actions.

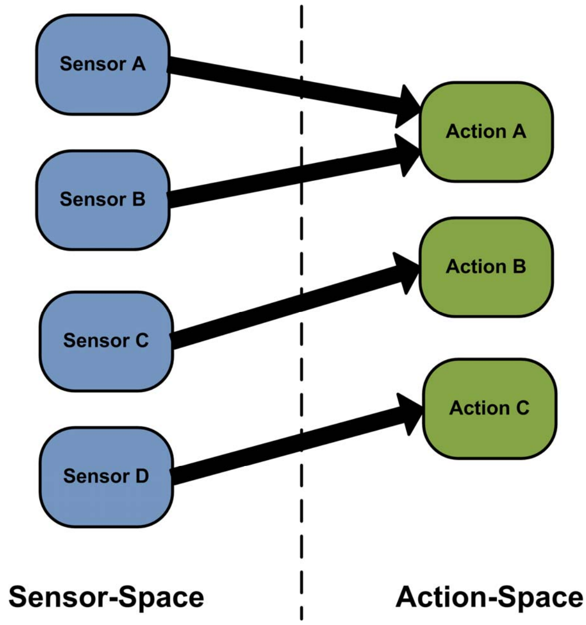

Figure 2 Projection from the Sensor-Space to the Action-Space

While one-hot-vectors are sparse and the features of the model increase with the *action* dictionary size, embeddings are dense and more computationally efficient, with the number of features staying the same, no matter the number of *action* types. Most significantly for our models, embeddings provide semantic meaning to the representation of the actions. Each action is represented as a point in a multidimensional plane, which place them at a distance of the other actions, thus providing relations of similitude and significance between them.

In our models we use the word2vec implementation in Gensim[2] to calculate the embedding values for each *action* in the dataset. We represent each *action* with a vector of 50 float values, due to the small number of action instances compared with the number of words that are usually used in NLP tasks. Instead of providing the values directly to our models we have included an Embedding layer in it. In this layer we store the procedural information on how to transform an *action* ID to its embedding. Adding this layer allows us to train it with the rest of the model and in that way to fine tune the embedding values to the current task, improving the general accuracy of the model.

C. Hybrid approach to Activity Recognition

The scientific community has developed two main approaches to solve activity and behaviour recognition, namely the data-driven and knowledge-driven approaches. Data-driven approaches make use of large-scale datasets of sensors to learn activity models using data mining and machine learning techniques. On the other hand, knowledge-driven approaches exploit rich prior knowledge in the domain of interest to build activity models using knowledge engineering and management technologies.

For knowledge-driven activity recognition systems, a widely recognized drawback is that activity models are usually static, i.e. once they have been defined, they cannot be automatically adapted to users' specificities [10]. This is a very restrictive limitation, because it is not generally possible to define complete activity models for every user. Domain experts have the necessary knowledge about activities, but this knowledge may not be enough to generate complete models in all the cases. To make knowledge-driven activity recognition systems work in real world applications, activity models have to evolve automatically to adapt to users' varying behaviours'. It turns out that model adaptability and evolution are aspects that can be properly addressed by data-driven approaches.

In City4Age we have created a scalable and hybrid activity recognition system to work in scenarios similar to the ones targeted in City4Age. We call the activity recognition system hybrid because it combines data- and knowledge-driven approaches. The key idea is to use data-mining techniques to find the most frequent action patterns in the unlabelled dataset produced after monitoring a person's activity in an intelligent environment (smart city and smart home). Those patterns reflect specific executions of activities.

In order to know what activities are being performed in a given action pattern, we use Expert Activity Models (EAM). EAMs are knowledge-based computational models where the previous knowledge about target activities is represented. The spirit of EAMs is not to have a detailed activity model for a given person, but rather to represent a generic activity with minimum knowledge.

With the purpose of discovering the activities for a given action pattern and a set of EAMs, we have developed a pattern-model matching algorithm. This algorithm is posed as a maximisation problem, where the objective is to find the set of EAMs that better explains the given action pattern. We use *actions*, locations, duration and starting time to address the maximisation problem.

```
"MakeCoffee" : {
        " actions ": ["hasContainer",
        "hasCoffee"],
        "duration": 300,
        "start":[[7:00-10:00], [13:00-15:00]],
        "locations":["Kitchen"]
}
```

Figure 3 Example of an EAM for hypothetical activity MakeCoffee.

IV. CONCLUSIONS AND FUTURE WORK

In this paper we have presented the problems related with activity recognition in smart cities, analysing how the large amount of unlabelled data presents new challenges. We have presented a multilevel model to describe the user conduct in smart environments, composed by *actions*, *activities*, *Intra-Activity Behaviours* and *Inter-Activity Behaviours*. We have also discussed the approach of the City4Age algorithms, by working in the Action-Space instead of the Sensor-Space to reduce the hypothesis space. We have also proposed the usage of semantic embeddings the represent actions instead of IDs or one-hot-vectors, allowing in this way for a more semantically meaningful representation of the input data. Finally, we have discussed the City4Age hybrid approach to activity recognition, by combining both the data-driven and knowledge-driven approaches generally used in the literature.

As future work we plan to validate these ideas in the 6 pilots that are being developed within the project. This will allow us to test them in a heterogeneous environment, providing an excellent validation bed case for the developed algorithms.

ACKNOWLEDGMENT

This work has been supported by the European Commission under the City4Age project grant agreement (number 689731). We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research.

REFERENCES

[1] U. Nations, "World urbanization prospects: The 2014 revision, highlights. department of economic and social affairs," Population Division, United Nations, 2014.

[2] A. Caragliu, C. Del Bo, and P. Nijkamp, "Smart cities in europe," Journal of urban technology, vol. 18, no. 2, pp. 65–82, 2011.

[3] J. M. Shapiro, "Smart cities: quality of life, productivity, and the growth effects of human capital," The review of economics and statistics, vol. 88, no. 2, pp. 324–335, 2006.

[4] Forman, D. E.; Berman, A. D.; McCabe, C. H.; Baim, D. S.; Wei, J. Y. (1992). "PTCA in the elderly: The "young-old" versus the "old-old"". Journal of the American Geriatrics Society. 40 (1): 19–22. doi:10.1111/j.1532-5415.1992.tb01823.

[2] https://github.com/RaRe-Technologies/gensim

[5] Rashidi, Parisa, and Alex Mihailidis. "A survey on ambient-assisted living tools for older adults." IEEE journal of biomedical and health informatics 17.3 (2013): 579-590.

[6] Gorka Azkune, Aitor Almeida, Diego López-de-Ipiña, Liming Luke Chen. (2015) "Extending knowledge-driven activity models through data-driven learning techniques" In Expert Systems with Applications. vol. 42. p. 3115-3128. DOI: 10.1016/j.eswa.2014.11.063.

[7] Gorka Azkune, Aitor Almeida, Diego López-de-Ipiña, Liming Luke Chen. (2015) "Combining Users' Activity Survey and Simulators to Evaluate Human Activity Recognition Systems" In Sensors. vol. 15. p. 8192-8213. DOI: 10.3390/s150408192.

[8] Chaaraoui, A. A., Climent-Prez, P., & Flrez-Revuelta, F. (2012). A review on vision techniques applied to human behaviour analysis for ambient-assisted living. Expert Systems with Applications, 39(12), 10873-10888.

[9] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems (pp. 3111-3119).

[10] Chen, L., Nugent, C., & Okeyo, G. (2014). An ontology-based hybrid approach to activity modeling for smart homes. IEEE Transactions on human-machine systems, 44(1), 92-105.

[11] Weinland, D., Ronfard, R., Boyer, E.: A survey of vision-based methods for action representation, segmentation and recognition. Computer Vision and Image Understanding 115(2), 224-241 (2011). DOI 10.1016/j.cviu.2010.10.002

[12] Chen, L., Hoey, J., Nugent, C., Cook, D., Yu, Z.: Sensor-based activity recognition. IEEE Transactions on Systems, Man, and Cybernetics-Part C 42(6), 790-808 (2012). DOI 10.1109/TSMCC.2012.2198883

[13] Yilmaz, A., Javed, O., Shah, M.: Object tracking: A survey. Acm computing surveys (CSUR) 38(4), 13 (2006)

[14] Bao, L., Intille, S.: Activity recognition from user-annotated acceleration data. In: L.N.i.C. Science (ed.) Pervasive Computing, pp. 117. Springer Berlin Heidelberg, Vienna, Austria (2004). DOI 10.1007/978-3-540-24646-6 1

[15] Kasteren, T.V., Noulas, A.: Accurate activity recognition in a home setting. In: Proceedings of the 10th International Conference on Ubiquitous computing, pp. 19 (2008)

[16] Oliver, N., Garg, A., Horvitz, E.: Layered representations for learning and inferring office activity from multiple sensory channels. Computer Vision and Image Understanding 96(2), 163-180 (2004)

[17] Brdiczka, O.: Learning situation models in a smart home. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 39(1), 56-63 (2009)

[18] Ordoñez, J., Iglesias, J., Toledo, P.D.: Online activity recognition using evolving classifiers. Expert Systems with Applications 40(4), 1248-1255 (2013)

[19] Rashidi, P., Cook, D.: COM: A method for mining and monitoring human activity patterns in home-based health monitoring systems. ACM Transactions on Intelligent Systems and Technology (TIST) 4(4), 64 (2013). DOI 10.1145/2508037.2508045

[20] Chen, L., Nugent, C.: A logical framework for behaviour reasoning and assistance in a smart home. International Journal of Assistive Robotics and Mechatronics 9(4), 20-34 (2008)

[21] Riboni, D., Bettini, C.: COSAR: hybrid reasoning for context-aware activity recognition. Personal and Ubiquitous Computing 15(3), 271-289 (2011). DOI 10.1007/s00779-010-0331-7

[22] Chen, L., Nugent, C., Wang, H.: A knowledge-driven approach to activity recognition in smart homes. Knowledge and Data Engineering, IEEE Transactions on 24(6), 961-974 (2012). DOI 10.1109/TKDE.2011.51

[23] Noor, M.H.M., Salcic, Z., Wang, K.I.K.: Enhancing ontological reasoning with uncertainty handling for activity recognition. Knowledge-Based Systems 114, 47-60 (2016). DOI 10.1016/j.knosys.2016.09.028

[24] Riboni, D., Sztyler, T., Civitarese, G., Stuckenschmidt, H.: Unsupervised recognition of interleaved activities of daily living through ontological and probabilistic reasoning. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp â€™16, pp. 1-12. ACM Press, New York, New York, USA (2016). DOI 10.1145/2971648.2971691

[25] Ye, J., Stevenson, G., Dobson, S.: USMART: An Unsupervised Semantic Mining Activity Recognition Technique. ACM Transactions on Interactive Intelligent Systems 4(4), 1-27 (2014). DOI 10.1145/2662870

[26] Cook, D.J., Krishnan, N.C., Rashidi, P.: Activity discovery and activity recognition: a new partnership. IEEE transactions on cybernetics 43(3), 820 (2013). DOI 10.1109/TSMCB.2012.2216873