A Statistical-Relational Activity Recognition Framework for Ambient Assisted Living Systems

Rim Helaoui, Mathias Niepert and Heiner Stuckenschmidt

Abstract Smart environments with ubiquitous sensing technologies are a promising perspective for reliable and continuous healthcare systems with reduced costs. A primary challenge for such assisted living systems is the automated recognition of everyday activities carried out by humans in their own home. In this work, we investigate the use of Markov Logic Networks as a framework for activity recognition within intelligent home-like environments equipped with pervasive light-weight sensor technologies. In particular, we explore the ability of MLNs to capture temporal relations and background knowledge for improving the recognition performance.

1 Introduction

Simple daily tasks are usually accomplished effortlessly. However, solving them might present a hard struggle for elderly people and individuals with disabilities or cognitive impairments. Due to the rapid population aging, there is an urgent need for more efficient methods towards reliable and continuous healthcare systems with reduced costs. Smart environments with ubiquitous sensing technologies are a promising perspective in this context. A primary challenge for such Ambient Assisted Living (AAL) systems is the automatic recognition of the activities carried out by the human at their domicile. This allows to anticipate and to provide the needed services at home and in time. Current approaches to this problem can be classified according to (1) the learning paradigms used, (2) the relevance of temporal information in the recognition model, and (3) whether the model is linked to real sensor data or simulated data. The current main learning paradigms applied are roughly data-driven or knowledge-based. The work presented in [2] and [13] are two examples that apply knowledge-driven learning algorithms. In their models, a prior rule base is mandatory to react to the incoming information. A formal ontology of the

KR & KM Research Group, Computer Science Institute, University of Mannheim, Germany e-mail: {rim, mathias, heiner}@informatik.uni-mannheim.de

user's intentions is used in [13] to enrich the current domain knowledge und reduce uncertainty. Some recent work emphasizing the importance of the temporal context proposes a novel approach to recognize human activity with constraint-based temporal reasoning framework (OMPS) [8]. Starting from general predefined rules, the system's procedure performs a search in the space of Decision Networks for a set of synchronizations that is applicable given the current state of the Decision network [8]. However, to obtain high level activities, the system resorts to heterogeneous sensor data like a stereo camera on the ceiling with a tracking system. The use of such tracking systems implies some privacy and costs issues which we would like to avoid.
Most of the current approaches to our problem are, nonetheless, data-driven. However, they seldom integrate temporal constraints in their systems. The majority use supervised learning methods like Hidden Markov Models (HMM) to recognize human activities. Obtaining substantial amounts of labelled data is often a bottle-neck for these approaches [3]. On another side, recent efforts to explore unsupervised learning methods have been developed. In [4], web mining is used to automatically extract the probabilities of the co-occurrence of some activities and the used objects. These probabilities assemble a HMM capable of recognizing activities in traces of object data. Increasing interest can be noticed concerning weakly supervised learning like multi-instance learning [3] and semi-supervised learning methods. We conjure that combining both paradigms, i.e. data-driven and knowledge-based, is an appealing direction to solve such non-Markovian problems and capture a rich temporal context. Similar to [5], our work also suggests to use both probabilistic models and relational information to transform the raw sensor data into higher-level descriptions of people's behaviours and activities. In [5] all of the ontology-based methods, web mining and HMM are used to produce a richly structured dataset describing people's daily patterns of activities. Due to the limited HMM temporal dependencies, however we opt for Markov Logic Networks (MLN). Indeed, MLN does not restrict the temporal context to some predefined range and can reason with variably long user's behavioral history. Furthermore, both logical statements as well as probabilistic ones are easily united in one single framework. This allows, for example, to flexibly integrate the existing domain knowledge, which reduces the amount of training data needed(see Section 2). To summarize, many existing approaches rely on manually constructed sets of rules for recognizing activities and/or rarely take temporal knowledge into account. This makes successful recognition of interrupted activities and activities engaging multiple actors rather out of reach. The purpose of this work is to investigate the use of MLN as a framework for activity recognition based on information from pervasive light-weight sensor technology. In particular, we want to test the ability of MLNs to capture temporal relations and background knowledge for improving the recognition performance. To this aim, we address daily human activities within intelligent home-like environments equipped with pervasive light-weight sensor technologies [14].

2 Markov Logic Networks for Activity Recognition

A Markov logic network combine first-order logic and undirected probabilistic models. It is essentially a set of weighted first-order logic formulae. The formulae and their weights can be automatically learnt as proposed in [16] for instance. The real-valued weights represent the confidence one has that each of these formulae hold. Probabilistic inference is performed over the joint probability distribution over all ground atoms. The formulae can be divided into soft and hard constraints. Hard constraints are required to always hold whereas soft ones can be violated in a solution. These weights can be learnt if training data is available. Thus, soft constraints will model the uncertainty of the events and the corresponding activities. Two examples of such soft rules are depicted below, where the predicates *activity* and *sensor* model daily activities and sensor events, respectively, at timestep t.

$$\forall timestep t : activity(settingthetable, t) \rightarrow activity(eating, t+1) \tag{1}$$

$$sensor(Oven, t) \rightarrow activity(cooking, t) \tag{2}$$

Hard constraints allow integrating existing common-sense knowledge, which can reduce the amount of needed training data. An example of such a hard constraint would be that a person cannot eat while sleeping. It can also express relevant physical and temporal constraints like the fact that a person cannot be present at two different places at the same time. Unlike other modelling methods such as HMM, MLN do not require a predefined number of states. They also offer the flexibility to reason with variably long user's behavioral history. Thus, it permits to cover different temporal contexts and explore its effect on the system's recognition and prediction accuracy. MLNs are applied for activity recognition in [6] and [7]. Unlike our framework, however, both approaches rely on video data. Compared to [6], we aim at an expanded set of more complex daily activities, addressing concurrent ones, as well as a richer temporal context. As proposed in [7], missing sensory data is an interesting aspect to be handled. Exploiting common-sense knowledge for a better model is another similarity to our proposed work. The authors, yet, limit their work to simple activities with no explicit consideration of the temporal context.

3 Activity Recognition Framework

The long term goal of this work is a framework towards proactive ambient healthcare systems. The approach applies MLN and relies on lightweight sensor data with real-world deployment. We expect our system to offer a good opportunity to explore the effect of different temporal contexts on the system's recognition and prediction accuracy. We intend to extend the framework to the case of more than one actor and to investigate the viability of the system under limited availability of sensor data. The system can eventually be applied for the case of anticipation and prevention of certain diseases based on behavioral-symptoms such as depression [14]. To achieve these expected contributions, many challenges have to be addressed. Indeed, activities of daily living are usually non sequential. They are often overlapping, alternating or even interrupted. Moreover, most of them share a large subset of sensors and objects. The activation of these sensors do not always follow the same temporal pattern for the same activity. We propose to integrate background knowledge covering many aspects. These can be categorized in three dimensions: Information about the user, about the environment and about time. Table 1 offers a refined overview of the model's background knowledge. In the proposed framework, low level signals

Table 1 Background knowledge for activity recognition using MLN

User	Environment		Time
	Indoors (sensors)	Outdoors	
ID	ID	Weather	Timestamp
Location	Type	Temperature	Part of the day
Mental state (drives[a])	State		Day of the week
			Season of the year
			Holidays and special occasions

[a] The drives correspond to natural needs like sleep, hunger. The sensor's states reflect whether a sensor is on or off for example.

are received as periodic events with a timestamp. Those and the available background knowledge (see Table 1), are described using the corresponding model's predicates. The low level events could be aggregated into actions such as "using the oven" or "entering the bedroom" as examples. The well defined activities of daily living, which we intend to recognize, are then inferred at a higher level. To address the temporal context, we propose to cover both of the qualitative and quantitative aspects. For the latter, we can introduce time intervals of the different events, actions and activities, which also captures their duration. A tolerance range can also be assigned to the start- and end-time. An average window-length can be calculated for each activity and be used as an additional feature during the recognition process [9]. Another option would be to cluster the different durations into classes like "very short, short, long and very long" instead of giving an explicit value. Two components, which can be either an event, an action or an activity, can be distinct, overlapping or alternate. The latter case is a special one of distinct components. To capture those possible relations between two intervals, i.e. the qualitative temporal aspect, the framework can be extended with a subset of the thirteen Allen's base relations. Indeed, for distinct activities we propose to use the relations: "before", and "meets". For the overlapping activities we limit our relations to "overlaps" and "during" (see [10]). Covering the temporal dimension is not only expected to help improve the expressiveness of the model and its recognition accuracy but could also be a good information for some user's states. Indeed, the rapidity of increase in carrying out some activity could be an indicator of its urgency. As already mentioned, there are events, actions and activities that entail the need to assert decisions on other components based on predefined common sense rules as well as physical and temporal constraints (see above). These relationships must be synchronized (propagated) to be kept compatible with the recognized activities. A simple instance would be that the activity "clear the table" entails that "set the table" must have taken place anteriorly. Hard constraints are a good option to express them in our framework. Hence, MLN use those predefined hard constraints and the obtained weighted rules to infer the current activity (see 2). As a beginning to our work, we decided to use a reduced set of sensors that only label some objects in the assisted living home. The activities to be recognized are restricted to one at a time. We do not address the qualitative temporal aspect of the actions and we only consider two temporal qualitative relationships, namely *next* and $after$: For two activities, actions, or events a at a timestamp t and b at timestamp d:

$$t+1=d \Leftrightarrow next(b,a) \; ; \; t+1>d \Leftrightarrow after(b,a) \tag{3}$$

3.1 Experiments

In this section we present some preliminary experiments verifying the viability of the proposed approach for activity recognition. We defined two predicates to describe the events and the activities. These are *sensor(sensor-id, timestamp)* and *act(activity-id, timestamp)*. The possible model uses two soft formulae:

$$\forall s_1, a_1, t : sensor(s_1, t) \rightarrow act(a_1, t) \tag{4}$$

$$\forall a_1, a_2, t \; : act(a_1, t) \rightarrow act(a_2, t+1) \tag{5}$$

The training data consists of the periodic events labelled with their corresponding activities [1], which are considered as the item sets of our rules mining. The sensor-activity and the activity-activity pairs having a given support are extracted and related, as weights to the rules defined above. Formula 4 models the conditional probability of activity a_1 given sensor s_1. Formula 5, on the other hand, captures the probability of activity a_2 being carried out at time step $t+1$ given that activity a_1 occurred at time step t. Since we restrict our activity recognition to one activity at a time in those experiments two more hard constraints have to be added. In the first experiment, the first hard constraint exclude parallel activities. The second rule expresses that at least one activity should take place at each time step. Considering sensor data as observation and the current activity as hidden in our MLN, the system reasons based on the current state as well as the previous ones.

3.2 Dataset and Setting

To train the described system, we considered the publicly available dataset RFID Data[1]. This was collected in a lab outfitted with 60 RFID tag, placed on the different objects involved in performing the set of eleven fine-grained activities: “Using the bathroom”, “Making oatmeal”, “Making soft-boiled eggs”, “Preparing orange juice”, “Making vanilla latte”, “Making tea”, “Making or answering a phone call”, “Using the door”, “Setting the table”, “Eating breakfast” and “Clearing the table”. The user wore two RFID gloves that detected the RFID tag within 2 inches. The collection covered 30 min per day, during ten days. The data is provided in form of timely ordered events relating the sensor ID and the activity being carried out for each timestamp. For the implementation of the proposed activity recognition framework, we used “TheBeast” [15]. The weights of the soft formulae are pre-computed to ensure a faster recognition process. This was achieved by applying an efficient algorithm for learning association rules [11]. Since we tried to extract as much itemsets as possible, we provided very low support and confidence values of 10^{-9} and 10^{-2}, respectively. The MLN inference task has been performed by the ILP solver SCIP [12]. The first day was kept as test data and the remaining ones was used to train the system.

3.3 Results

Our system has been evaluated using precision, recall and the F-measure. First, we only modeled the “next” temporal relationship and used a simple set of two soft rules(see 4 and 5) along with two hard constraints. Those ensure that at each timestamp exactly one activity takes place. An intuitive consequence is that the three mentioned measures will coincide. To avoid redundancy, we only mention the precision value while reporting our results. This first model reached a value of 0.77. The system tried to optimize over the whole test data with a combination of more than 1300 time steps, 10 activities and 59 sensors. Since we are more concerned with the recognition of current activity, we applied time-windows with some length n to our test data. Thus, the model infers using the last n events which, we claim, can alleviates and improve the recognition accuracy, even it reduces the available context. The length of the time window has been chosen experimentally after comparing different values. We fixed n to roughly twice the mean length of an activity, i.e. 160 time steps. The accuracy results with the corresponding activities are depicted in Figure 1. A significant accuracy deterioration is observed starting from the $5th$ window. This corresponds to activities sharing a large set of common sensors. Indeed, “setting the table”, “eating breakfast” and “clearing the table” can hardy be distinguished with provided rules. To address this ambiguity, we insert three hard constraints as a sample common-sense knowledge. Using the “$after$” temporal relationship introduced in Section 3, we state the following: (a) “clear the table” after “eating breakfast”; (b) “eating breakfast” after “setting the table”; and (c) “clear the table” after “setting the table”. The model with these hard constraints raises the accuracy from 0.57 to 0.81 over the second half of the test data (see Table 2) . The overall accuracy was also improved and attained an accuracy of 0.87.

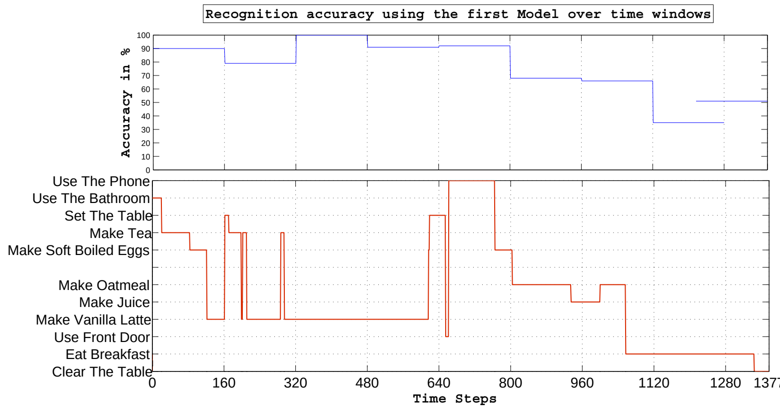

Fig. 1 Inserting time windows generally improves the recognition accuracy. A significant accuracy deterioration is observed starting from the $5th$ window. This corresponds to activities sharing a large set of common sensors: “setting the table”, “eating breakfast” and “clearing the table”.

	Model 1	Model 2
Accuray over the first Time window	0.91	0.90
Accuray over the second Time window	0.57	0.81
Overall accuracy	0.77	0.87

Table 2 Comparing The recognition accuracy of the two proposed models. Model 1 only implements the *next* time-relationship and bases on two soft rules and two hard constraints. Model 2 expands Model 1 with the $after$ temporal-relationship encoded into three new hard constraints.

4 Discussion and Conclusion

MLN offer a high flexibility to expand and to richly describe domain knowledge. They also combine it with learnt probabilistic rules for an adaptive behaviour. Those features make them an appealing approach to activity recognition. In this work, we have presented our preliminary experiments that showed the viability of MLN for this task and their ability to capture and integrate the temporal aspect of the activity for a better inference. Our experiments also showed that the temporal context is a crucial feature that highly influences the recognition accuracy. Indeed, we proposed two simple models. The second covers more temporal relationships than the first and outperformed it over the whole data as one single time window as well as in the case of splitting the test data into two time windows. However, it deteriorated the accuracy from 0.51 to 0.4 when applied to a reduced time window covering the last 160 time steps. The activities coinciding with this time slice are those that share a large set of common sensors.

Future work will focus on a the temporal aspect including the quantitative one. We hypothesize that capturing the order of the activities into the soft rules can improve the recognition model. For example, we merely state that “eating breakfast” can not occur before “setting the table” in the current models and overlook the aspect that the activity following “setting the table” is most probably “eating breakfast”. This is obviously absent in the current models since the majority of activities spread over a large number of time steps so the weights captured by second soft rule (see Rule 5) strictly cover the case that the same activity will most probably take place at the following time step. The first soft rule can also be a subject to improvement, if we cluster the sensors into classes. For example two distinct spoons bearing two different IDs can be grouped under the type spoon. This will lead to a more realistic association between activities and sensors types.

References

1. D. J. Patterson, D. Fox, H. Kautz, and M. Philipose, Fine-grained activity recognition by aggregating abstract object usage, in *ISWC ’05*, 2005.
2. J. C. Augusto,J. Liu,P. McCullagh, H. Wang and J. Yang, Management of uncertainty and spatio-temporal aspects for monitoring and diagnosis in a Smart Home, in *IJCIS*, 2008.
3. M. Stikic and B. Schiele, Activity recognition from sparsely labeled data using multi-instance learning, in *LoCA*. Springer, 2009.
4. D. Wyatt, M. Philipose, and T. Choudhury, Unsupervised activity recognition using automatically mined common sense, in *In AAAI*, 2005.
5. T. Choudhury, M. Philipose, D. Wyatt, and J. Lester, Towards activity databases: Using sensors and statistical models to summarize people’s lives, *Data Eng. Bull.*, vol. 29, pp. 49–58, 2006.
6. R. Biswas, S. Thrun, and K. Fujimura, Recognizing activities with multiple cues, in *HUMO07*, 2007, pp. 255–270.
7. S. D. Tran and L. S. Davis, Event modeling and recognition using markov logic networks, in *ECCV08*, 2008, pp. 610–623.
8. M. Cirillo,F. Lazellotto, F. Pecora, and Saffiotti, Monitoring domestic activities with temporal constraints and components, in *Proc IE09*, 2009, pages 117124.
9. E. M. Tapia, S. S. Intille and K. Larson, Activity Recognition in the Home Using Simple and Ubiquitous Sensors, in *Pervasive*, 2004, pages 158-175.
10. J. F. Allen, Maintaining knowledge about temporal intervals. In: Commun. **26**, 832–843 (1983)
11. F. Bodon, A Survey on Frequent Itemset Mining, Technical Report, Budapest University of Technology and Economic, 2006.
12. T. Achterberg, Constraint Integer Programming, Ph.D. dissertation, Technische Universität Berlin, 2007.
13. J. Hochul, K. Taehwan and C. Joongmin, Ontology-Based User Intention Recognition for Proactive Planning of Intelligent Robot Behavior, In *MUE*, 2008, pp 158-175.
14. R. Helaoui, Towards a Proactive System Based on Activity Recognition, To appear in *PerCom*, 2010.
15. S. Riedel, Improving the accuracy and efficiency of map inference for markov logic, in Proc. of UAI, 2008, 468–475.
16. M. Biba, S. Ferilli and F. Esposito, Structure Learning of Markov Logic Networks through Iterated Local Search., in *ECAI*, 2008, pages 361-365.