Human Activity Recognition Using Place-based Decision Fusion in a Smart Home Context Management System

Fusion de décisions par lieu pour reconnaître l'Activité Humaine et sa prise en compte dans la gestion de contexte

Julien Cumin[1,2], Fano Ramparany[1], Grégoire Lefebvre[1], James L. Crowley[2]

[1] Orange Labs, France, *{julien1.cumin, fano.ramparany, gregoire.lefebvre}@orange.com*
[2] Laboratoire d'Informatique de Grenoble, Université Grenoble Alpes & Inria, France *{firstname.lastname}@inria.fr*

RÉSUMÉ. Notre article décrit les résultats de travaux utilisant l'information de localisation afin de reconnaître les activités domestiques. Nous explorons l'utilisation de l'apprentissage supervisé pour reconnaître les activités situées dans la banque de données du projet Opportunity, couplé à une phase de fusion de décision. Nos résultats montrent que l'utilisation d'information de localisation permet une diminution substantielle du taux d'erreur ainsi que du coût de calcul de la reconnaissance d'activité par rapport aux approches classiques, pour lesquelles tous les capteurs sont utilisés et toutes les activités sont possibles. L'utilisation de l'information de localisation pour contrôler le processus de reconnaissance permet d'obtenir un score F1 de $92.70\% \pm 1.26\%$, et requiert seulement 73 millisecondes of temps de traitement par activité. Ces résultats montrent qu'organiser la reconnaîssance d'activité autour d'un modèle de contexte basé sur la localisation permet le passage à l'échelle de services sensibles au contexte d'activité dans les environnements domestiques.

ABSTRACT. This paper describes the results of experiments where information about places is used in the recognition of activities in the home. We explore the use of place-specific activity recognition trained with supervised learning, coupled with a decision fusion step, for recognition of activities in the *Opportunity* dataset. Our experiments show that using place information to control recognition can substantially improve both the error rates and the computation cost of activity recognition compared to classical approaches where all sensors are used and all activities are possible. The use of place information for controlling recognition gives an F1 classification score of $92.70\% \pm 1.26\%$, requiring on average only 73 milliseconds of computing time per instance of activity. These experiments demonstrate that organizing activity recognition with place-based context models can provide a scalable approach for building context-aware services in smart home environments.

MOTS-CLÉS. reconnaissance d'activité, apprentissage automatique, modèle de contexte, localisation.
KEYWORDS. activity recognition, machine learning, context model, localisation.

1. Introduction

The arrival of low-cost computing and wireless communications has provided the potential for a technological rupture in home technologies. In theory, it has became possible to provide “smart” home services for applications such as environmental control, energy efficiency, security, entertainment, active healthy ageing and assisted living. Activity recognition from environmental sensors is generally recognised as a key enabling technology for such services. However, to date, this vision of the “smart” home remains a technology of the future. The complexity and scalability of activity recognition from environmental sensors has emerged as an important barrier to the emergence of practical systems and services [7].

A scalable approach for smart-home services requires the use of context [3], where context can be defined as any information that can be used to characterise situation [10]. For smart home services, time-of-day, place, identity of inhabitants and activity are key elements of context information for providing appropriate services.

Time-of-day, place, identity and activity are abstract semantic concepts. Each of these provides key information that can condition the suitability or appropriateness of smart home services. Time-of-day refers to periods such as morning, evening or night, as well as day of the week and summer vacation or Christmas holidays. Time-of-day is strongly correlated with local time and date, with only minor variations in sequence and boundary that can be inferred from activities of inhabitants. Places are generally defined as region of space where specific classes of activities occur and can be easily inferred from location information. In a home, identity refers not only to the identity of the inhabitant but also to their position within the family for each other person (Father, mother, child, family-friend, etc). Social role is a static property that is easily determined from the identity of an inhabitant.

Activity in the home is the most difficult to determine. Activity refers to the collections of actions that are performed in order to accomplish a task. Activity recognition is challenging both because the number of activity classes can be very large, and because the manner in which an activity occurs can vary from one individual to the next. Even for a single individual, an activity may be highly variable. In addition, it is not unusual for individuals to perform several activities in parallel, interleaving the actions of the individual activities.

Human activity recognition is currently a hot topic in computer vision. Certainly, image sequences can be a rich source of information about activity. However, the use of cameras for activity recognition is generally not well accepted by inhabitants, because of privacy concerns [19]. An alternative is to recognise activities based on a large number of environmental sensors. In particular, instrumenting an electrical system to monitor electrical use converts every electrical switch into a sensor. This information can be enriched by infrared presence detectors, switches on doors, wearable sensors, or even smartphones carried by the inhabitant as in [13]. The result is a large number of simple data elements that can be used to construct systems to monitor activity.

Many authors have speculated that context information, such as time-of-day, identity and place can be used to organize smart services. In this paper we report on experiments that show the extent to which context can improve error rates and execution time of recognition of activity. In particular, we focus on place as an organizational element for activity recognition. We note that activity is highly dependent on place. For example, the activity "*Cooking*" is very likely to happen in the place "*Kitchen*". Therefore, place information would appear valuable to improve activity recognition.

In this paper we investigate a place-based approach to activity recognition, which relies on multiple supervised classification models, one for each place in the home, as well as a decision fusion step.

We also elaborate on the integration of this approach within a general context management system.

In Sect. 2. we present a summary of the state of the art of activity recognition in the home, and discuss differences that exist between those works and our approach. In Sect. 3. we present details of our approach. The experiments we lead to evaluate this approach are presented in Sect. 4.. In section 5.2. we describe how the approach integrates into a generic context management system, after which we conclude in Sect. 6. on the suitability of our approach for human activity recognition in smart homes.

2. State of the Art

Activity recognition in smart homes is a very active research subject. Here, we are particularly interested in approaches which use low level data, as opposed to image-based techniques. Approaches based on machine learning are naturally very popular in this field, most of which are supervised learning approaches. We are however starting to see some works emerge that are based on unsupervised techniques, as in [6]. Although they are simpler to use than unsupervised approaches, supervised approaches are still not, to this day, sufficiently accurate to provide an information of activity that is reliable enough in order to generate context-based services which are useful to inhabitants [3]. More efforts are thus still needed on this research topic.

Some recent supervised activity recognition approaches are based on *deep learning*. For example, we find the work of Ordóñez and Roggen [12] which seek to exploit, on low level data of smart homes, the capabilities of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) neural networks, CNNs being very effective on signals such as images, and LSTMs being capable of modeling the temporal dimension of data. This work is very convincing, in particular because they are applied on a dataset that is more limited in its number of instances than typical datasets used for deep learning. The number of labeled instances is still substantial, which makes the acquisition, training and tuning processes very hard tasks. It is indeed very difficult to obtain labeled data of inhabitants of homes for a commercial application. Providing sufficient information is very uncompelling for users, and activities that they perform in their home will probably evolve frequently too. The necessity of having a large number of labeled data and long training times are thus definitive drawbacks for such applications.

On the other hand, the literature proposes approaches that are based on *ontologies*, as in [1] or [5]. In the latter, activity recognition is based on expert knowledge of smart home environments : sensors, rooms of the home, inhabitants, activities and sub-activities, etc. Expert knowledge is more reliable and not annoying to obtain for users, compared to labeled data provided by inhabitants. This approach also allows to have a logical and formal view of the home, which can be used in other applications than activity recognition (e.g. energy management), whereas a machine learning approach is limited to the application for which it was trained. However, those ontology-based approaches also have drawbacks : they rely on expert knowledge, which ought to be as exhaustive as possible (which is very expensive) in order for the system to work properly for a random home. Consequently, we can expect those approaches to be somewhat efficient in general but way too rigid to perfectly adapt to the specificities of each possible inhabitant, which greatly impacts the capability of the system to provide services that correspond to inhabitants' needs.

Lastly, there are hybrid approaches which are based on machine learning, but that attempt to exploit expert information in order to improve the performances of the recognition algorithm. This is for example the case in the work of Wu et al. [21], where the localization of the person is estimated so as to reduce the set of possible activities that can correspond to the current instance. Here, the only expert information needed are the positions of sensors throughout the home, as well as the sets of activities possible in each room. Reducing the set of possible activities based on the localization of the person performing the activity is a technique that we also use in the work presented in this paper. However, we believe that reducing the set of sensors used as well, based on their location, can simplify the classification task. Lastly, instead of relying on an estimation of the localization of the person, we believe that it is simpler and more robust to classify the instance with all local models simultaneously, and then let a decision fusion step decide which of the classes corresponds to the instance. This allows to alleviate the need for localization estimation (which can add errors if not accurate enough), while allowing the possibility of recognizing simultaneous activities which would happen in different places of the home (not covered in this paper).

Location and activity are only two of the many dimensions of context. Semantic modeling makes it possible to aggregate the various dimensions of context information into a unified framework. The introduction of semantic modeling techniques and languages such as RDF, RDFS and OWL to express sensors data has improved the interoperability of this data and improved its ability to characterize specific contexts [9], [18] and makes it possible to express in a single model : activities, location, user profiles and temporal aspects. These contexts can thus be shared among use cases and lead to better genericness in the solutions.

3. A Place-based Approach to Activity Recognition

3.1. *Places and Motivations for the Approach*

Inhabitants of a home have *routines*, that is, sequences of activities that they perform in repeated fashion during their time in the home. Those activities are performed in what we can call *places*, such as a bedroom or a bathroom, which reciprocally get associated to a set of activities by the inhabitant : for example, the activity "*Brushing your teeth*" will be unique to the place "*Bathroom*". Although it is possible that occupants perform activities in rooms that are not usually used for such activities (e.g. brushing your teeth in the living room), such events are by nature exceptional, and thus are not easily exploitable to provide context-aware services based on the routines of the occupants. As such, we can ignore them in this work. Moreover, every place very often corresponds to exactly one room of the home ; a finer granularity does not seem very useful for anything but large rooms where activities would be very diverse in different parts of the room.

It is obvious that, by proposing an activity recognition approach based on places, we need to have *a priori* knowledge of the existing places in the home (the correspondence between rooms and places making this step relatively simple), as well as both the distribution of sensors in the places and the activity classes that can happen in each place. If those information seem difficult to obtain in current smart homes, we can conceive that, for all but the activity classes, those information will be readily available : indeed, the constructor of proper smart homes could directly fill in the distribution of sensors that they installed in the rooms, as well as a set of places based on those rooms. As for the distribution of activity classes, this information seems to go in pair with the knowledge itself of the activity classes, which is typically assumed to be given by the user in supervised approaches, such as the one we present in this work.

As presented in Sect. 2., supervised activity recognition approaches are typically "global" approaches : in order to classify a new instance of activity, a classifier trained in advance will try to decide the correct class, among all possible activity classes of the home, based on all sensors available. Here, we propose a "local" approach which exploits the information available on places, by building a different classifier for each place ; this classifier of a place will only use the sensors in that place as inputs, and will only have to model activity classes which can happen in that place. An additional step of decision fusion (presented in Sect. 3.3.) allows to take a final decision for the entire home.

That way, a classifier specific to one place has a simpler model to learn, because of the reduced number of available sensors and decidable classes, as opposed to a global approach where the model can become so complex that learning is too difficult. Consequently, parametrization of classifiers is greatly simplified, and computing times during the learning step ought to be shortened. Besides, every classifier being independent from place to place, it is possible to parallelize the learning step between all places. Thus, one can retrain a subset of classifiers instead of the entire global model, if some changes happened in the home (e.g. a new activity class exists, or a new sensor was installed).

3.2. *Place-based Activity Recognition*

Suppose that there are three places in a home (see Figure 1). We can identify, for each place P_i, the data sources (i.e. sensors) $S_j^{(i)}$ that are in that place. Note that some sensors can appear in more than one place (e.g. bodily-worn sensors) ; it is thus possible that for two places P_i and $P_{i'}$, we have $S_j^{(i)} = S_{j'}^{(i')}$.

We can then associate a classifier C_i to each place, which will classify an activity instance using only the $S_j^{(i)}$ sources as inputs. Moreover, C_i can only decide the classes that can happen in P_i ; thus, if the current activity instance does not happen in P_i, then C_i should ideally decide the dummy class *None*.

The *None* class is a source of difficulty during the training and testing phases. It indeed represents in our approach three different situations : no activity is happening, an unknown activity is happening, or an activity from another place is happening. The instances' data of the *None* class for a place will thus be much more varied than for other activity classes.

We assume here that only one activity can happen at a time in the home. It would be possible, with our place-based approach, to recognize activities that happen simultaneously in two different places by simply removing the decision fusion step ; this would not be feasible as straightforwardly using the classical global approach.

3.3. *Decision Fusion*

In order to combine the decisions taken by the classifiers of each place, we can use multiple approaches of decision fusion that can be found in the literature [8]. We only retain the two best decision fusion methods that we tested, both based on the principles of *Stacking* [20], where the problem of decision fusion is considered to be a classification problem, only where the input data are itself decisions from classifiers. Therefore, it is possible to fuse the decisions taken in all places by using the confidence of each classifier as *inputs to the stacking classifier*. The two stacking classifiers that we have retained, after preliminary experiments, to perform the decision fusion step are the MultiLayer Perceptron (MLP) and the Support Vector Machine (SVM).

Since we propose to use a fusion step to get a global decision, for the entire home, on which activity class the current instance belongs to, it seems natural to exploit even more this fusion step by using multiple classifiers in each place. Therefore, looking back at the example of Figure 1, we can imagine that we have three classifiers per place, instead of just one, and thus fuse nine decisions instead of three (three

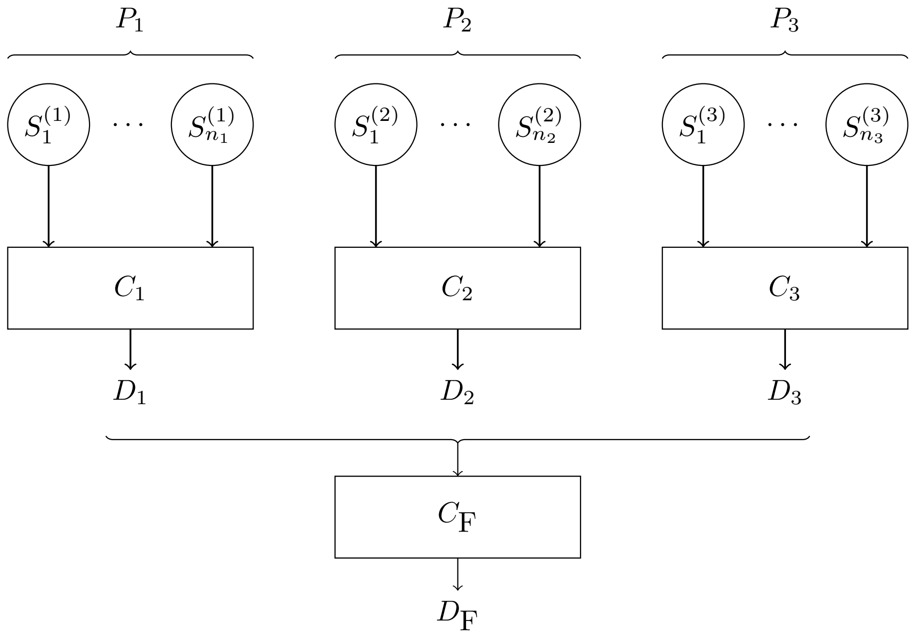

Figure 1. *Data flow of place-based decision fusion.*

decisions would be taken per place). This can lead to better performances of the system, by ensuring that more than one classifier give their opinions on the class of the current instance, and thus combine the strengths of different kinds of classifiers. Decision fusion is also directly usable in the standard global approach, where we would this time for example have three classifiers which would classify the current instance using all available data sources (which is the classical use of decision fusion).

4. Experiments

4.1. *The Opportunity Dataset*

Opportunity is proposed by Roggen et al. [17] to be a reference dataset for the evaluation of algorithms related to human activity recognition in the home, such as classification or automatic segmentation of activities. In this dataset, each of the four inhabitants has performed by themselves five sessions of activities of daily living (see Figure 3), during which they performed the activities by following a brief description of the session, with no specific restrictions. Each inhabitant also performed a “Drill” session, during which they perform 20 times a precise sequence of 17 activities.

The activities of *Opportunity* are performed in a unique room (see Figure 2), in which both its elements (drawers, forks and knives, doors, etc.) and the inhabitant themselves are instrumented, with 39 inertial sensors (19 on the inhabitant, 20 in the environment) and 13 state-change sensors (all in the environment). *Opportunity* offers multiple levels of labeling of activities ; we are only interested here in the 17 mid-level activities, namely *Clean Table*, *Drink from Cup*, *Open Dishwasher*, *Close Dishwasher*, *Open Drawer 1*, *Close Drawer 1*, *Open Drawer 2*, *Close Drawer 2*, *Open Drawer 3*, *Close Drawer 3*, *Open Fridge*, *Close Fridge*, *Toggle Switch*, *Open Door 1*, *Close Door 1*, *Open Door 2* and *Close Door 2*. A dummy activity *None* is used when there is no activity or when no other activity fits. This class also corresponds to the *None* class mentioned in Sect. 3.2., used for activities that do not happen in a certain place.

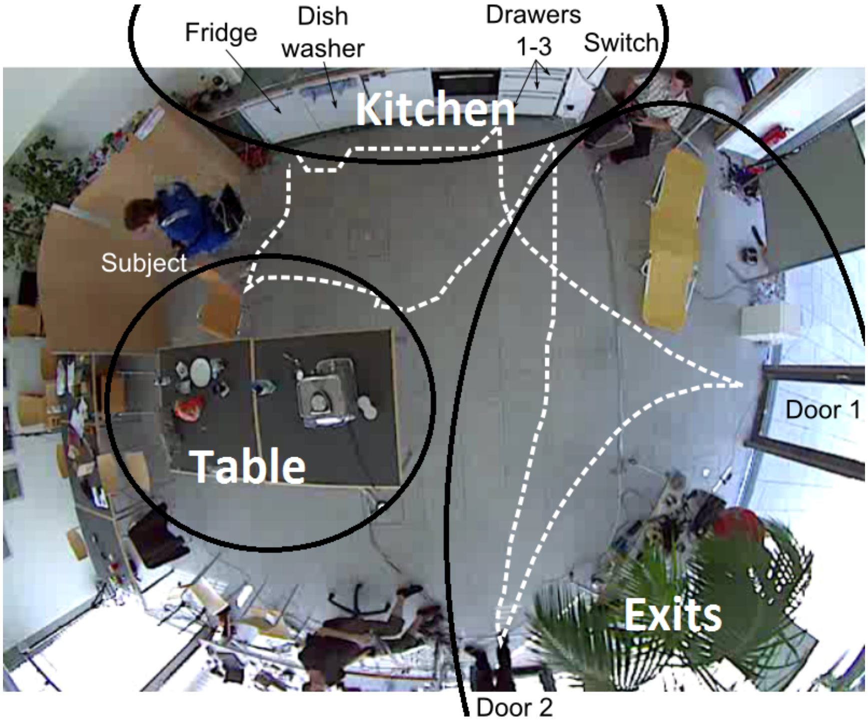

Figure 2. *Opportunity's environment, annotated with instrumented objects and places.*

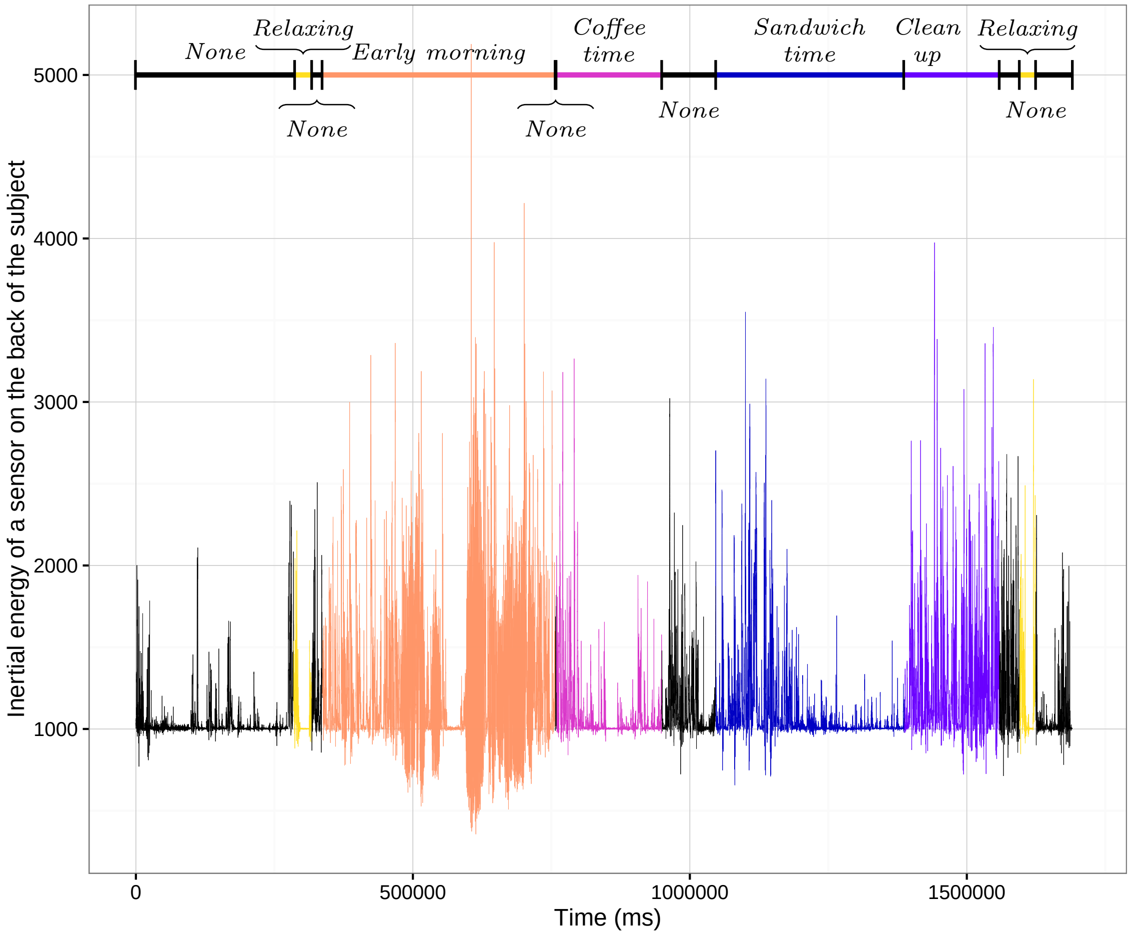

Figure 3. *High level activities during a session of activities of daily living.*

Place	Classifier		
	MLP	SVM	BN
Table	$\mathbf{98.97\% \pm 0.48\%}$	$98.77\% \pm 0.54\%$	$98.70\% \pm 0.48\%$
Kitchen	$\mathbf{94.06\% \pm 1.58\%}$	$93.78\% \pm 1.32\%$	$91.79\% \pm 1.27\%$
Exits	$99.15\% \pm 0.39\%$	$\mathbf{99.24\% \pm 0.34\%}$	$98.34\% \pm 0.62\%$

Parameters :
- **MLP** 80 hidden neurons, 100 epochs, 0.2 learning rate, 0.1 momentum.
- **SVM** $C = 1000$, $\gamma = 0.01$.
- **BN** K2 search, *SimpleEstimator* estimator.

Table 1. *F1 scores of classifiers for each place.*

4.2. *Experimental Protocol and Data Preprocessing Strategy*

To experimentally evaluate our approach on the *Opportunity* dataset, we assume the segmentation of each activity instance to be known. Therefore, the beginning and the end of each instance are marked by the transition between two labels of different activities, at two successive timesteps. We use a 10-fold random cross-validation where each fold contains, for each of the 18 classes (including *None*), 72 training instances, 22 test instances and 18 validation instances (used to optimize the parameters of classifiers). Those instances are randomly selected for each fold among the four inhabitants. Following the recommendations made in [4], we decided not to use the localization data and quaternion data, for all following experiments.

Data are preprocessed such that missing values are interpolated using cubic splines. A low-pass filter is applied on the data and they are normalized so that the average value of each sensors is 0 and its standard deviation is 1. For classifiers that require a feature vector of fixed size as input, we construct that vector by resampling the data into 20 samples, and then concatenating each sample one after the other. The information of duration of the instance as well as its start timestamp are prepended to the vector.

We evaluate our approach using three standard classification models : the MultiLayer Perceptron (MLP), the Support Vector Machine (SVM) and the Bayesian Network (BN). Those classifiers use feature vectors of fixed size as input, and their implementations are taken from the `Weka` library [11]. We had also used Hidden Markov Models (HMM) and Dynamic Time Warping (DTW) during our experiments, but they proved to be respectively not accurate enough and too slow ; we will thus not report the results of those two methods in the rest of the paper.

We define three places in the *Opportunity* experiment (see Figure 2) :

- *Table* : represents the table in the center-left part of the room ; contains the 12 sensors placed on the objects that are on the table, as well as the 19 sensors on the inhabitant. Activities *Clean Table*, *Drink from Cup* and *None* can happen in this place.
- *Kitchen* : represents the kitchen counter ; contains the 18 sensors on the fridge, drawers, dishwasher, light switch, and the 19 sensors on the inhabitant. Activities *Open/Close Dishwasher/Fridge/Drawer1/Drawer2/Drawer3*, *Toggle Switch* and *None* can happen in this place.
- *Exits* : represents the two doors in the room ; contains the two sensors on those doors, the sensor placed on the lazy chair next to one of the doors, and the 19 sensors on the inhabitant. Activities *Open/Close Door1/Door2* and *None* can happen in this place.

Approach	Classifier MLP	SVM	BN
Table, Kitchen, Exits (Fusion)	$\mathbf{92.52\% \pm 1.25\%}^1$	$\mathbf{91.78\% \pm 1.37\%}^2$	$89.14\% \pm 1.27\%^3$
Home	$90.21\% \pm 1.62\%$	$90.05\% \pm 1.64\%$	$\mathbf{90.61\% \pm 1.37\%}$

Parameters on *Home* :
- **MLP** 500 hidden neurons, 200 epochs, 0.2 learning rate, 0.1 momentum.
- **SVM** $C = 100$, $\gamma = 0.0005$.
- **BN** K2 search, *SimpleEstimator* estimator.

Decision fusion in places :
- [1] **SVM stacking** $C = 100$, $\gamma = 0.01$.
- [2] **MLP stacking** 100 hidden neurons, 100 epochs, 0.2 learning rate, 0.1 momentum.
- [3] **MLP stacking** 20 hidden neurons, 100 epochs, 0.2 learning rate, 0.1 momentum.

Table 2. *F1 scores of classifiers on the* Home *configuration or of decision fusion of classifiers of the same type in all places.*

We will also use the *Home* configuration for comparison's sake, which corresponds to the classical approach where all sensors are used and all activities that can happen in the home are decidable.

Our protocol is quite different from the usual protocol that is used on the *Opportunity* dataset [12], which comes from a challenge. The protocol of this challenge only uses the sensors on the inhabitant, which would not allow us to validate the benefits of a significant part of our approach, which is that each place's model only uses the sensors of the place they are in. Moreover, this protocol does not cross-validate its results and requires an additional segmentation step (which might skew the results) ; it is thus not well-adapted to validate an activity recognition approach, which should not be optimized for a specific dataset.

4.3. *Results*

We present in Table 1 the F1 scores of classifiers for each place. We can observe that activity recognition in the places *Table* and *Exits* is relatively "easy", since all classifiers manage to reach scores above 98%. The task seems more difficult in the place *Kitchen*, which can be explained by the fact that 12 classes can happen in this place (including *None*), whereas only 5 and 3 respectively can happen in *Table* and *Exits*. Moreover, some classes of *Kitchen* are very similar (e.g. *Open Drawer 1* and *Open Drawer 2*), which makes it difficult to distinguish them from the available data.

We present in Table 2 the F1 scores of classifiers on the *Home* configuration and the F1 scores of the fusion of decisions taken on the 3 places, when all places uses the same type of classifier. Those results allow us to see that the place-based approach we propose, fusing decisions taken on each place, attains significantly better scores than the classical global approach (e.g. for the MLP, $92.52\% \pm 1.25\%$ versus $90.21\% \pm 1.62\%$), for all tested classifiers but the Bayesian Network (BN), for which the *Home* configuration attains slightly better scores ($89.14\% \pm 1.27\%$ versus $90.61\% \pm 1.37\%$). We can also observe that our approach produces more stable results : the standard deviations recorded are smaller for all tested classifiers.

Finally, we present in Table 3 the F1 scores of decision fusion of multiple classifiers for each place

	Classifier		
	MLP	SVM	BN
Approach	Fusion		
Table + Kitchen + Exits	**$92.70\% \pm 1.26\%^1$**		
Home	$91.62\% \pm 1.59\%^1$		
Classifiers' parameters : see Table 2.			
Fusion : [1] **SVM stacking** $C = 1, \gamma = 0.1$.			

Table 3. *F1 scores of decision fusion for local and global approaches.*

Classifier	Phase	Table	Kitchen	Exits	Home
		Model			
MLP	Training	947.65 ± 160.77	732.83 ± 60.04	561.71 ± 30.78	11250.06 ± 1593.57
	Testing	12.64 ± 1.56	9.84 ± 1.22	8.49 ± 1.99	20.70 ± 1.19
SVM	Training	24.42 ± 0.23	19.11 ± 0.16	12.75 ± 0.23	35.37 ± 0.48
	Testing	6.56 ± 0.06	12.49 ± 0.13	4.21 ± 0.03	29.47 ± 0.96
BN	Training	19.06 ± 0.34	13.87 ± 0.28	11.34 ± 0.25	26.49 ± 0.37
	Testing	8.75 ± 0.06	6.71 ± 0.13	5.49 ± 0.07	11.73 ± 0.10
Classifiers' parameters : see Table 1 and Table 2.					

Table 4. *Average computing times in seconds of classifiers in training and testing phase for each model, for an entire fold of cross-validation.*

and for *Home*. SVM stacking fusion on the set of decisions taken by the three types of classifiers used previously in each place (9 decisions) reaches an F1 score of $92.70\% \pm 1.26\%$, which is slightly better than decision fusion using only the MLP in the three places ($92.52\% \pm 1.25\%$, see Table 2). SVM stacking fusion on the three classifiers in the *Home* configuration only reaches an F1 score of $91.62\% \pm 1.59\%$. We present in Figure 4 the confusion matrix of one fold of test of the place-based three classifiers decision fusion approach. We find as expected that most confusions happen for very similar activities (e.g. *Open Drawer 1* and *Close Drawer 1*), and for the class *None*.

4.4. *Computing Times*

Besides an improvement of performances in classification, the approach we propose also allows, thanks to the reduction of the number of sensors used and activity classes per model, to reduce computing times. We present in Table 4 the training and testing computing times for the three classifiers in the three places and the global model *Home*. Those computing times were evaluated on a 4-cores Intel i7 2.8 GHz processor with 16 GB of RAM. The computing times of the final decision fusion step are very stable and less than 1 second for one fold of training or testing, for both the place-based approach and for *Home* ; we will thus ignore them in this analysis. If we assume to have four computing cores, as is common in

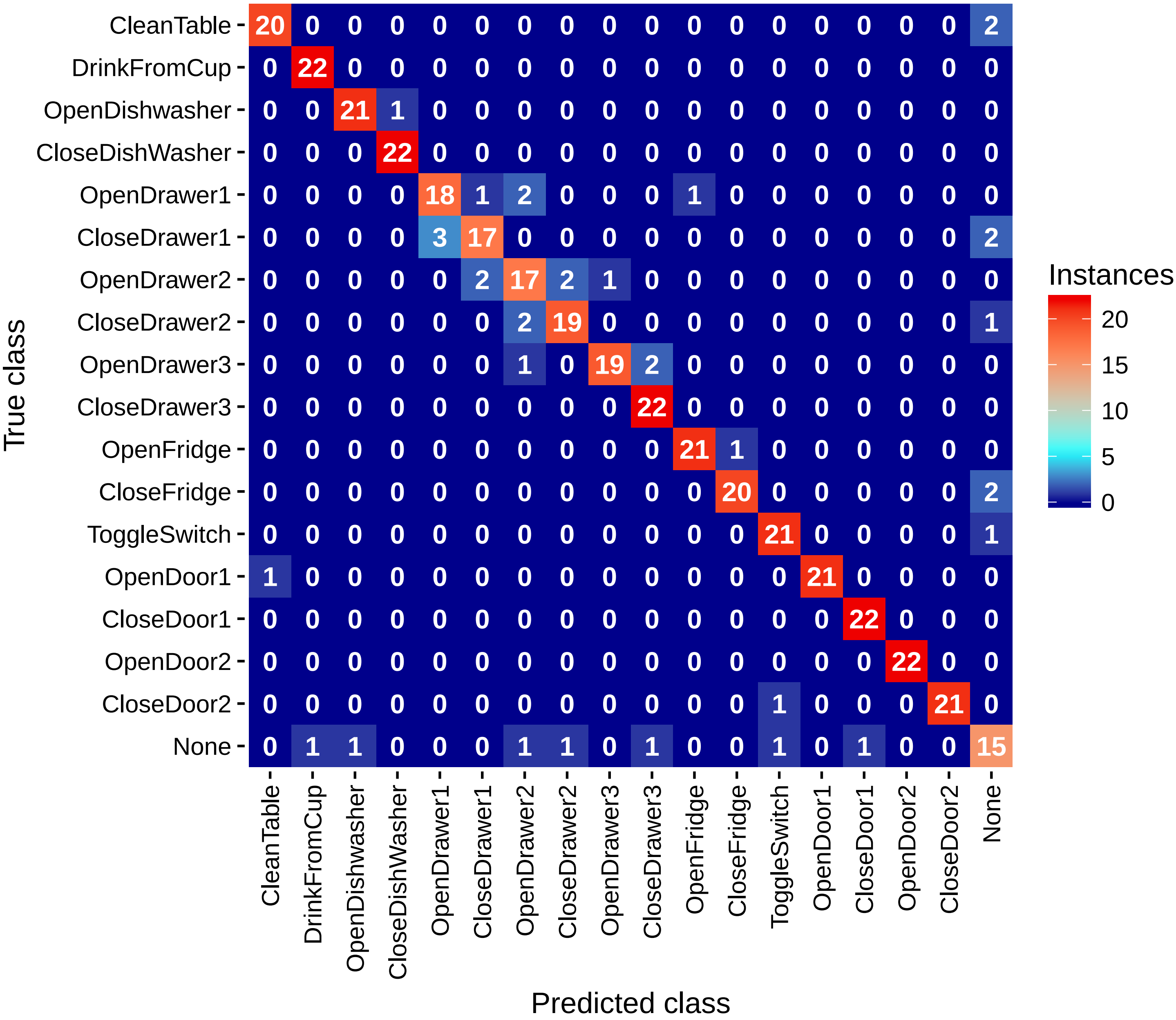

Figure 4. Confusion matrix of one fold of test of our best configuration.

personal computers nowadays, we can parallelize the three classifiers for the *Home* configuration ; thus, to process all instances of one fold, we will need on average as much time as the slowest classifier, e.g. the SVM in the testing phase (29.47 seconds). In our place-based approach, we can parallelize the three places ; thus, to process all instances of one fold, we will need on average as much time as the place for which the sum of the computing times of its three classifiers is the greatest, e.g. *Kitchen* in the testing phase ($9.84 + 12.49 + 6.71 = 29.04$ seconds). Since there are $22 \times 18 = 396$ instances per fold, our approach requires on average 73.33 milliseconds to process one instance, which is much shorter than the duration of the instances themselves.

We clearly see, on this dataset, that our approach is slightly faster for the testing phase. For the training phase, it is much faster because of the MLP ; classifiers for which their training complexity grows quickly compared to the number of instances, inputs and classes will thus greatly benefit from our approach. The more sensors and activities, the more complex a global model needs to be in order to perform well, and thus the more our approach is appropriate.

5. Context Modeling

5.1. *Representing Context using FLOD*

Location and activity are only two of the many dimensions of context. Semantic modeling makes it possible to aggregate the various dimensions of context information into a unified framework. The introduction of semantic modeling techniques and languages such as RDF, RDFS and OWL to express sensors data has improved the interoperability of this data and improved its ability to characterize specific contexts [9], [18] and makes it possible to express in a single model : activities, location, user profiles and temporal aspects. These contexts can thus be shared among use cases and lead to better genericness in the solutions.

We model and represent context using a Semantic Context Management Service (CMS) [15]. This CMS has been adapted to the IoT domain through the FLOD semantic broker enabler [14]. FLOD has kept the same philosophy as the CMS. More specifically as depicted in Figure 5, it serves as a mediation platform between context sources which are entities that provide context information and context aware application. Context aware applications are applications that benefit from context information for adapting their behaviour. This mediation plaform implements and provides functions for *collecting* rough data from context sensors, for formatting these data in RDF and *aggregating* them with previously collected data. This aggregation results into a *semantic context model*. This context model can then be queried by context aware applications through the standard SPARQL query language, or processed by *reasoning* mechanisms. FLOD inherits its reasoning capabilities from its using of semantic modeling techniques. One particular policy that FLOD adopts is to ensure that the RDF documents contained in its context model comply to well defined ontologies.

As prescribed by cognitive engineering methodologies and good practices, we have looked up existing ontologies that cover our universe of discourse, i.e. the scope of concepts and issues that we need to model in our application. We ended up using :
– the DogOnt ontology [2] to model home environments.
– the ActivO ontology [16] to model human activities.

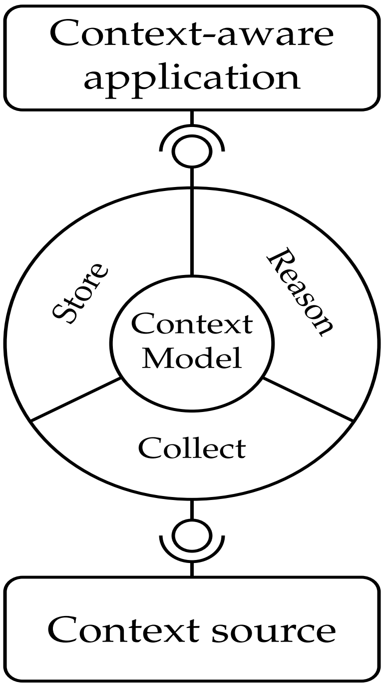

Figure 5. FLOD Semantic Context Broker functional architecture

By using these two ontologies and instantiating their classes and relations we could model the Opportunity environment shown in Figure 6, as a RDF document from which we show a fragment displayed in Figure 7.

In this graph we can see the activities Open/CloseDoor1, Open/CloseDoor2 and None that can take place in the place Exits. These activities are instances of the class PersonalActivity which is a concept of the ActivO ontology. The place Exits is an instance of the class Lobby which is one concept of the DogOnt ontology. In the same place we can see that there are two sensors which are A5_7 and A5_8. These sensors are instances of the class A5 that we have defined in our local ontology. The class A5 is a subclass of the class Sensor which is one concept of the DogOnt ontology.

The FLOD system models and registers this kind of information, but more importantly can answer specific queries about this information. For instance, a SPARQL query that retrieves potential activities from places is formulated as follows :

```
PREFIX do: <http://elite.polito.it/ontologies/dogont.owl#>
PREFIX act: <http://webmind.dico.unimi.it/CARE/locont.owl#>
PREFIX shd: <http://www.orange.com/ontologies/shd#>

SELECT *
WHERE {
  shd:Exits act:canHostActivity ?activity
}
```

In this query we look for all targets ?activity of canHostActivity relations which source is Exits. On the basis of the context model displayed in Figure 7, the result returned by FLOD will be the list : OpenDoor1, CloseDoor1, OpenDoor2, CloseDoor2 and None.

Another SPARQL query that retrieves sensors from places is formulated as follows :

```
PREFIX do: <http://elite.polito.it/ontologies/dogont.owl#>
```


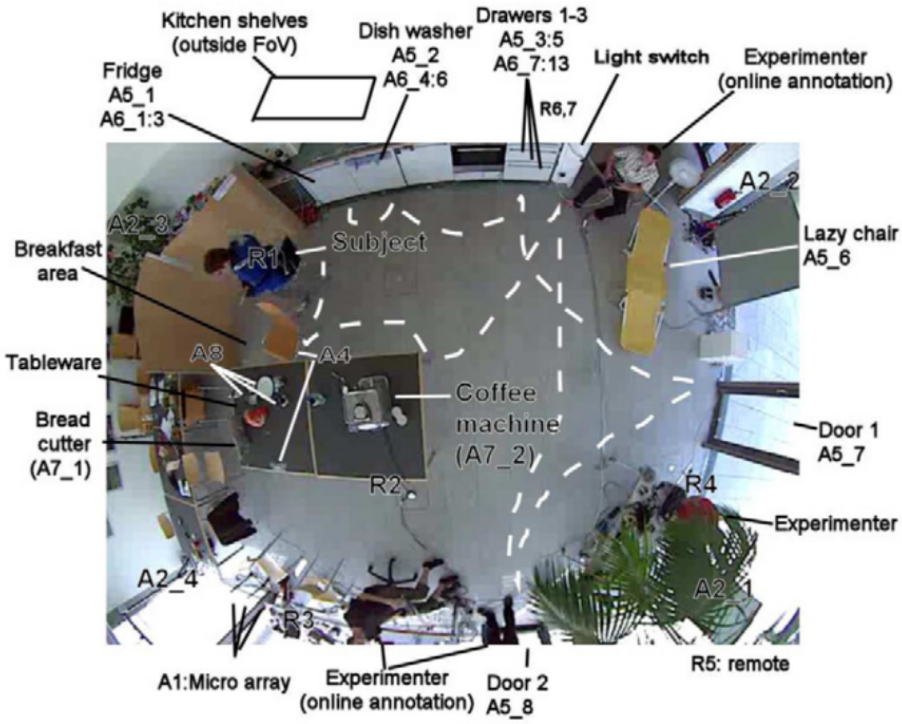

Figure 6. Map of the environment

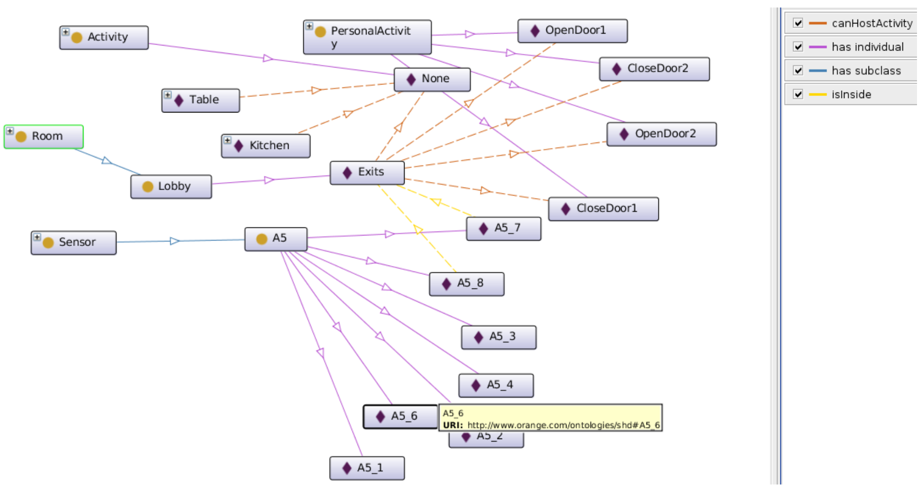

Figure 7. Context RDF model

```
PREFIX act: <http://webmind.dico.unimi.it/CARE/locont.owl#>
PREFIX shd: <http://www.orange.com/ontologies/shd#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT *
WHERE {
  ?s rdf:type act:Sensor .
  ?s act:isInside shd:Exits
}
```

In this query we look for all sources ?s of isInside relations which target is Exits. On the basis of the context model displayed in Figure 7, the result returned by FLOD will be the list : A5_7 and A5_8.

5.2. *Context modeling enrichment*

Our approach like any machine learning process aims at generating a runtime system, in our case one or more classifiers, for instance, a classifier for each place. In this section we elaborate on architectural consideration. More precisely we describe mechanisms for integrating our Activity Recognition Decision System into our FLOD Context Management System. From a high level perspective, FLOD will provide the ML system with the information required to setup the training and testing instances for each ML process. This interaction takes place during the initialization stage. Which means that it occurs only once as long as the home instrumentation (sensors and their places) doesn't change. Once the ML processes have produced the classifiers, each classifier behaves as new sensors and thus provide FLOD with activity information on the basis of lower level sensors. The architecture displayed in Figure 8 illustrates these mechanisms.

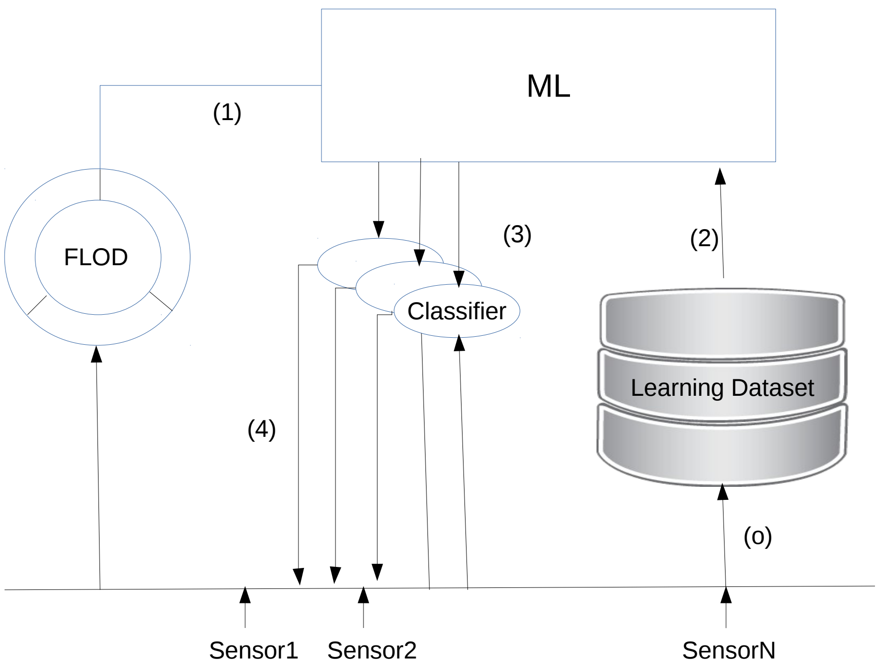

Figure 8. *FLOD ML interaction.*

More specifically, during the acquisition stage (0), sensors data are collected and stored in a global database, while the house inhabitants label their activities. Our Activity recognition system will then query FLOD to get the set of places P_i, and the set of related activities $C_{i,j}$ and sensors $S_{i,k}$ during stage (1). FLOD Northbound interface is used for this retrieval. For our Opportunity use-case there are three places. For instance, for the Exits place, we filter the data produced by the two sensors located in the Exits place : namely the a5_7 and a5_8 sensors. The ML system will use these parameters to split the global data base into as many training and testing sets as there are places. These training and testing sets are used to learn the classifiers (2) which are activity recognition runtimes during stage (3). Each classifier is then used online to interpret dynamically the stream of sensors data into activities. The activities detected are then pushed into FLOD through its southbound interface to enrich the context model with activity context (4). Because these activities are located, the classifier also provides location context to FLOD.

Thus our activity recognition machine learning system nicely interfaces with our semantic context management system and in addition to enriching the context model with activity context, it can also provide location context.

6. Conclusion and Perspectives

We presented in this paper an original approach to improve the performances of supervised algorithms to recognize activities of inhabitants of smart homes, using another piece of context information : place. By defining places in the home, which contain sensors and are the place of realization of certain activities, we can greatly reduce the size of the input data and the number of decidable classes for a classifier, instead of building a classifier which uses all sensors of the home and that ought to recognize all possible activity classes. A decision fusion step allows to combine the decisions taken in each different place in order to attain a global decision. Our approach does not require the knowledge or an estimation of the localization of inhabitants in the home. In fact, our approach can actually help estimate that localization by observing in which places activities are recognized. We have evaluated our approach on the *Opportunity* dataset, by comparing it to the classical global approach where all sensors are used to recognize all possible activities. On this dataset, our approach reaches better classification scores while being faster, whether it be in the training or testing phase.

We have applied our approach on places. but we could also imagine to apply it on qualitative time periods (*Morning*, *Afternoon*, etc.) or on the identity of inhabitants. The usefulness of such granularities for activity recognition remains unknown. The approach we propose requires *a priori* knowledge of places, the sensors and the activities they contain. Even though those information could be obtained through smart home contractors and inhabitants, we can hope to discover those information automatically based on data, in an unsupervised fashion. Advances on that subject seem essential to improve the acceptability of smart home solutions for the average user.

We have explored the process of learning and recognizing human activities from low level sensor data. We believe that the solution that we have developed for this recognition task can be applied to the recognition of other type of situation, beside activity recognition. Among them, people identity recognition, localisation and counting would be worth investigating. On the more general level of context modeling, we have elaborated a process for inserting activity information within a generic context model. In previous work, we have followed a similar approach for inserting presence information and agenda events information. Therefore, the work we describe here adds to the assessment that our modeling and integration approach is generic.

We also plan to assess the overall architecture that integrates our activity recognition learning system with a context management system on real data and setting involving end users.

Using semantic modeling in the context management system makes it easy to add a new sensor in a room. It is thus straightforward to identify which classifiers needs to be relearned, because we can distinguish the activity classifiers that might be impacted by this change from those which are not. Another interesting feature offered by a semantic approach to context management is the handling of exception. For example what if someone is brushing her/his teeth in the kitchen instead of the doing it in the bathroom ? which is the more conventional place for this. Identifying such exception could lead to the updating of the ontology so that this activity (i.e. brushing teeth) could be located in the kitchen (in addition to the bathroom).

This contextual information allows other services to adapt and personalize their responses to the real uses of the home and the real needs of the occupants. We can think of home automation services (heating systems, lighting management, etc.) but also multimedia services (e.g. a music recommendation system following user activities.).

Bibliography

[1] Allègre, W., Burger, T., Berruet, P., Antoine, J.Y., et al. : Modèle de supervision d'interactions non-intrusif basé sur les ontologies. In : EGC. pp. 285–290 (2012)

[2] Bonino, A. : Dogont - ontology modelling for intelligent domotic environments. In : Proceedings of the International Semantic Web Conference (2008)

[3] Brdiczka, O., Crowley, J.L., Reignier, P. : Learning situation models in a smart home. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 39(1), 56–63 (2009)

[4] Chavarriaga, R., Sagha, H., Calatroni, A., Digumarti, S.T., Tröster, G., Millán, J.d.R., Roggen, D. : The opportunity challenge : A benchmark database for on-body sensor-based activity recognition. Pattern Recognition Letters 34(15), 2033–2042 (2013)

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

[6] Cook, D.J., Krishnan, N.C., Rashidi, P. : Activity discovery and activity recognition : A new partnership. IEEE transactions on cybernetics 43(3), 820–828 (2013)

[7] Crowley, J.L., Coutaz, J. : An ecological view of smart home technologies. In : European Conference on Ambient Intelligence. pp. 1–16. Springer (2015)

[8] Cumin, J., Lefebvre, G. : A priori data and a posteriori decision fusions for human action recognition. In : 11th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP) (2016)

[9] De, S., Barnaghi, P., Bauer, M., Meissner, S. : Service modelling for the internet of things. In : Proceedings of the 3rd Workshop on Software Services (2011)

[10] Dey, A.K. : Understanding and using context. Personal and ubiquitous computing 5(1), 4–7 (2001)

[11] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H. : The weka data mining software : an update. ACM SIGKDD explorations newsletter 11(1), 10–18 (2009)

[12] Ordóñez, F.J., Roggen, D. : Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. Sensors 16(1), 115 (2016)

[13] Ouchi, K., Doi, M. : A real-time living activity recognition system using off-the-shelf sensors on a mobile phone. In : International and Interdisciplinary Conference on Modeling and Using Context. pp. 226–232. Springer (2011)

[14] Ramparany, F., Marquez, F., Soriano, J., Elsaleh, T. : handling smart environment devices, data and services at the semantic level with the fi-ware core platform. In : on Big Data, I.I.C. (ed.) Proceeding of the 1st Workshop on Semantics for Big Data on the Internet of Things (SemBIoT 2014). pp. 14–20 (2014)

[15] Ramparany, F., Poortinga, R., Stikic, M., Schmalenströer, J., Prante, T. : An open context information management infrastructure - the ist-amigo project. In : of Engineering, I.I., T.e. (ed.) Proceedings of the 3rd IET International Conference on Intelligent Environments (IE'07). pp. 398–403. University of Ulm, Germany (september 24-25 2007)

[16] Riboni, D., Bettini, C. : Owl 2 modeling and reasoning with complex human activities. In : Journal of Pervasive Mobile Computing, vol. 7, pp. 379–395. Elsevier (2011)

[17] Roggen, D., Calatroni, A., Rossi, M., Holleczek, T., Forster, K., Troster, G., Lukowicz, P., Bannach, D., Pirkl, G., Ferscha, A., Doppler, J., Holzmann, C., Kurz, M., Holl, G., Chavarriaga, R., Sagha, H., Bayati, H., Creatura, M., del R Millan, J. : Collecting complex activity datasets in highly rich networked sensor environments. In : 2010 Seventh International Conference on Networked Sensing Systems (INSS). pp. 233–240 (Jun 2010)

[18] Sorici, A., Picard, G., Boissier, O., Zimmerman, A., Florea, A. : : CONSERT : Applying semantic web technologies to context modeling in ambient intelligence. Computers & Electrical Engineering (2015)

[19] Townsend, D., Knoefel, F., Goubran, R. : Privacy versus autonomy : a tradeoff model for smart home monitoring technologies. In : 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 4749–4752. IEEE (2011)

[20] Wolpert, D.H. : Stacked generalization. Neural networks 5(2), 241–259 (1992)

[21] Wu, C., Khalili, A.H., Aghajan, H. : Multiview activity recognition in smart homes with spatio-temporal features. In : Proceedings of the Fourth ACM/IEEE International Conference on Distributed Smart Cameras. pp. 142–149. ACM (2010)