Context-aware Activity Recognition through a Combination of Ontological and Statistical Reasoning

Daniele Riboni Claudio Bettini

EveryWare Lab, D.I.Co., Università di Milano
{riboni,bettini}@dico.unimi.it

Abstract. In the last years, techniques for activity recognition have attracted increasing attention. Among many applications, a special interest is in the pervasive e-Health domain where automatic activity recognition is used in rehabilitation systems, chronic disease management, monitoring of the elderly, as well as in personal well being applications. Research in this field has mainly adopted techniques based on supervised learning algorithms to recognize activities based on contextual conditions (e.g., location, surrounding environment, used objects) and data retrieved from body-worn sensors. Since these systems rely on a sufficiently large amount of training data which is hard to collect, scalability with respect to the number of considered activities and contextual data is a major issue. In this paper, we propose the use of ontologies and ontological reasoning combined with statistical inferencing to address this problem. Our technique relies on the use of semantic relationships that express the feasibility of performing a given activity in a given context. The proposed technique neither increases the obtrusiveness of the statistical activity recognition system, nor introduces significant computational overhead to real-time activity recognition. The results of extensive experiments with data collected from sensors worn by a group of volunteers performing activities both indoor and outdoor show the superiority of the combined technique with respect to a solely statistical approach. To the best of our knowledge, this is the first work that systematically investigates the integration of statistical and ontological reasoning for activity recognition.

1 Introduction

There is a general consensus on the fact that effective automatic recognition of user activities would greatly enhance the ability of a pervasive system to properly react and adapt to the circumstances. Among many applications of activity recognition, a special interest is in the pervasive e-Health domain where automatic activity recognition is used in rehabilitation systems, chronic disease management, monitoring of the elderly, as well as in personal well being applications (see, e.g., [1–3]).

Example 1. Consider the case of Alice, an elderly person undergoing rehabilitation after having been hospitalized for a minor heart attack. In order to help Alice in correctly following the practitioners' prescriptions about the physical activities to perform during rehabilitation, the hospital center provides her with a monitoring system that continuously keeps track of her physiological data as well as of her activities. In particular, physiological data (e.g., heart rate and blood pressure) are acquired by wearable sensors that transmit them through a wireless link to the monitoring application hosted on her mobile phone. Similarly, accelerometer data provided by a smartwatch are transmitted to the monitoring application and merged with those provided by the accelerometer integrated in her mobile phone to automatically infer her current physical activity. On the basis of physiological data and performed activities, the monitoring application provides Alice with alerts and suggestions to better follow her rehabilitation plan (e.g., "please consider to take a walk this morning", or "take some rest now"). Moreover, those data are reported to the medical center on a daily basis for further processing.

Of course, for such a system to be effective, the activity recognition module must provide very accurate results. In fact, if activities are wrongly recognized, the monitoring system may draw erroneous conclusions about the actual adherence of the patient to the practitioners' prescriptions, as well as provide error-prone statistics about the health status of the patient.

A research direction consists in devising techniques to recognize activities using cameras with the help of sound, image and scene recognition software (see, e.g., [4, 5]), but this is limited to very confined environments and often subject to serious privacy concerns, clearly perceived by the monitored users.

Alternative activity recognition techniques are based on data acquired from body-worn sensors (e.g., motion tracking and inertial sensors, cardio-frequency meters, etc) and on the application of statistical learning methods. Early attempts in this sense were mainly based on the use of data acquired from multiple body-worn accelerometers (e.g., [6, 7]). One of the main limitations of these early systems relied on the fact that they did not consider contextual information (such as current location, environmental conditions, surrounding objects) that could be usefully exploited to derive the user's activity (for simplicity, in the rest of this paper we refer to this kind of data as *context*). As a consequence, later approaches were aimed at devising activity recognition systems taking into account the user's context. For instance, in [8] a method is proposed to classify physical activities by considering not only data retrieved from a body-worn accelerometer, but also environmental data acquired from several other sensors (sound, humidity, acceleration, orientation, barometric pressure, . . .). Spatio-temporal traces are used in [9] to derive high-level activities such as *shopping* or *dining out*. Observations regarding the user's surrounding environment (in particular, objects' use), possibly coupled with body-worn sensor data, are the basis of many other activity recognition systems (e.g., [10, 11]).

Most of these systems rely on the application of supervised learning algorithms that, in order to perform well, need to be trained with a sufficiently large amount of labeled data. Indeed, with an insufficient set of training data, to consider a wide set of context data would be ineffective, if not counterproductive, since the classifier could draw erroneous predictions due to the problem of overfitting. For instance, in [8] some available context data are discarded in order to avoid this problem, that is one of the main reasons why activity recognition systems do not perform well out of the laboratory. Since training data are very hard to acquire, systems relying on supervised learning are prone to serious scalability issues the more activities and the more context data are considered. For instance, suppose to consider as the only context data the user's current symbolic location (e.g., *kitchen*, *dining room*, *mall*, *park*, etc). Even in this simple case, in order to gain good recognition results a sufficiently large set of training data should be acquired for each activity in any considered location. Of course, such a large set of training data is very hard to obtain. Moreover, when we consider as context not only location but also environmental conditions and surrounding objects, the task of collecting a sufficient amount of training data is very likely to become unmanageable, since training data should be acquired in any possible contextual condition. This challenging issue has been addressed (e.g., in [12]) by means of a combination of supervised and unsupervised learning techniques. Even if similar techniques can be adopted to mitigate the problem, it is unlikely that they can provide a definitive solution.

In this paper we investigate the use of ontological reasoning coupled with statistical reasoning in order to address the above-mentioned problem. The intuition behind our solution is that very useful hints about the possible activities performed by a user based on her context can be obtained by exploiting symbolic reasoning without the need of any training data. Besides, statistical inferencing can be performed based on raw data retrieved from body-worn sensors (e.g., accelerometers) without the need to acquire them under different context conditions during the training phase; indeed, given a performed activity, these data are mainly independent from the user's context. Hence, by coupling symbolic reasoning with statistical inferencing it is possible to perform activity recognition using a comprehensive set of information in a sufficiently scalable way.

In particular, our technique consists in the use of semantic relationships and constraints to express the feasibility of performing a given activity in a given context. For this reason we have defined an ontology that models activities, artifacts, persons, communication routes, and symbolic locations, and that expresses relations and constraints between these entities. To the best of our knowledge this is the first work that systematically investigates the integration of statistical and ontological reasoning for activity recognition. Extensive experimental results with data collected by volunteers show the superiority of the proposed technique with respect to a solely statistical approach.

The rest of the paper is organized as follows: Section 2 presents an overview of the proposed activity recognition system; Section 3 illustrates the technique for combining ontological reasoning and statistical activity recognition; Section 4 presents experimental results; Section 5 concludes the paper.

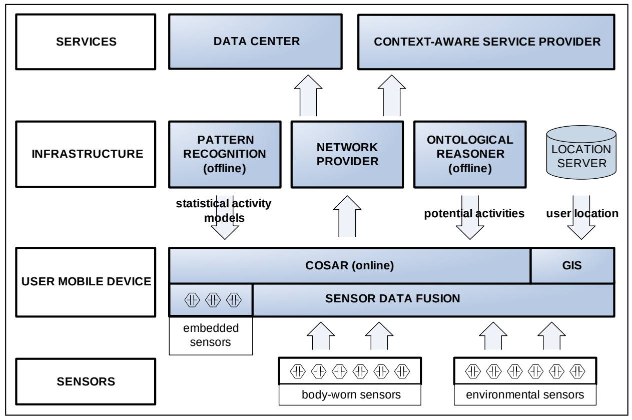


Fig. 1. The *COSAR* system

2 The *COSAR* activity recognition system

The proposed activity recognition system is graphically depicted in Figure 1. The lower layer (SENSORS) includes body-worn sensors (providing data such as accelerometer readings and physiological parameters) and sensors spread in the environment.

Data provided by environmental and body-worn sensors are communicated through a wireless connection to the USER MOBILE DEVICE, and merged with sensor data retrieved by the device itself (e.g., data provided by an embedded accelerometer) to build a *feature vector* that will be used to predict the user's activity. The device also continuously keeps track of the current physical location provided by a GPS receiver. When the GPS reading is not available or not sufficiently accurate (e.g., indoor), localization is performed by an external LOCATION SERVER (e.g., a GSM triangulation system provided by the network operator, or an RFID system). The GIS module is in charge of mapping the physical location reading to the most specific symbolic location that correspond to that physical location. This information will be used by the COMBINED ONTOLOGICAL/STATISTICAL ACTIVITY RECOGNITION module (COSAR) to refine the statistical predictions.

The INFRASTRUCTURE layer includes a PATTERN RECOGNITION module that is in charge of deriving a statistical model of the considered activities, which is communicated offline to the COSAR module. This layer is also in charge of performing ontological reasoning to calculate the set of activities that can be potentially performed in a given context. This set is also communicated offline to the COSAR module. In addition, the infrastructure layer includes a network provider offering the connectivity necessary to exchange data between modules at different layers, and, in particular, to communicate activity information to remote data centers or context-aware service providers. With respect to efficiency issues, we point out that the most computationally expensive tasks (i.e., ontological reasoning and pattern recognition to build a statistical model of activities) are executed offline. Note that privacy issues are of paramount importance in this domain; however, their treatment is outside the scope of this paper. Preliminary work on integrating privacy preservation in a context-aware middleware can be found in [13].

3 Combining ontological reasoning and statistical activity recognition

In this section we illustrate how ontological reasoning is coupled with statistical inferencing in order to recognize the user's activity.

3.1 Statistical activity recognition with a temporal *voted* variant

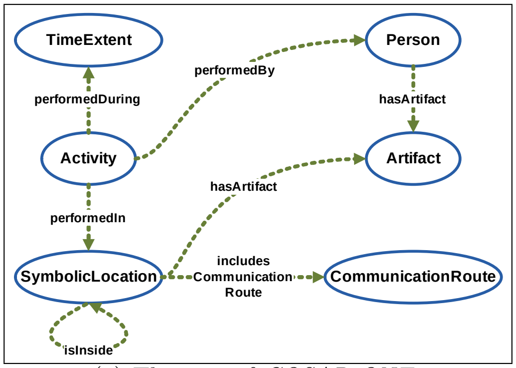
As illustrated in the introduction, the most common approach to activity recognition is to make use of supervised statistical learning methods. Roughly speaking, these methods rely on a set of preclassified activity instances that are used in a training phase to learn a statistical model of a given set of activities. The obtained model is then used to automatically classify new activity instances.

Activity instances are characterized by a *duration*; i.e., the temporal resolution at which activity instances are considered. Each activity instance is represented by means of a *feature vector*, in which each feature corresponds to a given measure (typically, a statistics about some measurements retrieved from a sensor or from a set of sensors during the duration of the activity instance).

Even if significant exceptions exist (e.g., Hidden Markov Models and Linear Dynamical Systems [14]), it is worth to note that most models adopted by statistical learning algorithms implicitly assume independence between each pair of instances to be classified. As a consequence, the prediction of an instance i_2 does not depend on the prediction of another instance i_1. However, when considering activity instances the above-mentioned assumption does not hold. In fact, persons do not continuously switch among different activities; instead, they tend to perform the same activity for a certain lapse of time before changing activity. Similarly to other approaches in the literature, we exploit this characteristic to improve the classification result of statistical activity recognition systems by introducing a *voted* variant. Since a thorough description of this variant is outside the scope of this paper, we only mention that it is based on a time window to classify each activity instance considering some of the previous activity instances; this technique can be applied to a large class of statistical learning algorithms.

3.2 Ontological reasoning to identify potential activities based on context

Even if our technique can be applied to any kind of context data, in the rest of this paper we concentrate on location information. Indeed, location is an important case of context information, and the current symbolic location of a



(a) The core of *COSAR-ONT*

Class	Descendants
Activity	35
Artifact	43
CommunicationRoute	14
Person	4
SymbolicLocation	30
TimeExtent	11

(b) Number of classes

Fig. 2. The *COSAR-ONT* ontology

user can give useful hints about which activities she can or cannot perform. Moreover, from a practical perspective, localization technologies are more and more integrated in mobile devices and buildings; hence, differently from other context data, location information is available in many situations.

Rationale. Even if in theory the set of possible activities that can be performed in a given symbolic location could be manually specified by a domain expert, this method would be clearly impractical. Indeed, even considering a few tens of activities and symbolic locations, the number of their combinations would quickly render this task unmanageable. Moreover, this task should be repeated each time the characteristics of a symbolic location change (e.g., when an artifact is added to or removed from a room).

Our solution is based on the use of an OWL-DL [15] ontology to represent activities, symbolic locations, communication routes, artifacts, persons and time granularities, as well as relations among them. To this aim we have developed a novel ontology, named *COSAR-ONT*, whose main classes and properties are graphically depicted in Figure 2(a). Figure 2(b) shows the number of descendants of the main classes of the ontology. In particular, COSAR-ONT includes 30 symbolic locations and 35 activities. Figure 3 shows part of the activities modeled by our ontology; the rightmost activities in the figure are those used in the experimental evaluation of our system (Section 4). The set of locations and activities in our ontology is obviously non exhaustive; however, we believe that this ontology can be profitably used to model many health care scenarios. Moreover, the ontology is easily extensible to address additional application domains. In order to illustrate our technique we introduce the following example.

Example 2. Consider the activity `BrushingTeeth`, and the task of automatically inferring the set of symbolic locations in which such activity can reasonably be performed. One possible definition of the considered activity is the following:

$$\texttt{BrushingTeeth} \sqsubseteq \texttt{PersonalActivity} \sqcap \forall\, \texttt{performedIn}.\,(\,\exists\, \texttt{hasArtifact}.\texttt{Sink}\,) \sqcap \ldots$$

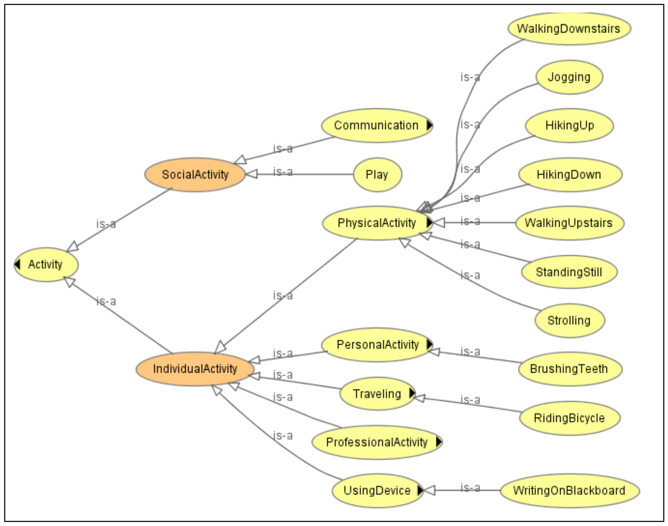


Fig. 3. Part of the *COSAR-ONT* activities

According to the above definition, `BrushingTeeth` is a subclass of `PersonalActivity` that can be performed only in locations that contain a `Sink` (that is defined as a subclass of `WaterFixture`); other restrictions may follow, but they are not considered here for simplicity. Now consider two symbolic locations, namely `RestRoom` and `LivingRoom`, defined as follows:

$$\texttt{RestRoom} \sqsubseteq \texttt{Room} \sqcap \exists\, \texttt{hasArtifact.Sink} \sqcap \ldots$$
$$\texttt{LivingRoom} \sqsubseteq \texttt{Room} \sqcap \neg\exists\, \texttt{hasArtifact.WaterFixture} \sqcap \ldots$$

According to the above definitions, `RestRoom` is a `Room` that contains a sink, while `LivingRoom` is a `Room` that does not contain any `WaterFixture` (once again, other details about the definition of these classes are omitted)[1]. Given those ontological definitions it is possible to automatically derive through ontological reasoning the set of symbolic locations in which the activity `BrushingTeeth` can be performed. To this aim, the following assertions are stated and added to the assertional part of the ontology (called *ABox*):

$$\texttt{BrushingTeeth(CURR\_ACT)}; \texttt{RestRoom(CURR\_LOC\_1)}; \texttt{LivingRoom(CURR\_LOC\_2)}$$

The above assertions create an instance of activity `BrushingTeeth` identified as `CURR_ACT`, an instance of location `RestRoom` identified as `CURR_LOC_1`, and an instance of location `LivingRoom` identified as `CURR_LOC_2`. Then, in order to understand if a given

[1] Note that, due to the open-world assumption of description logic systems [16] and, consequently, of OWL-DL, it is necessary to explicitly state those artifacts that are not present in a given location. This is simplified by considering in the definition of symbolic locations only artifacts that characterize the activities to be discriminated and using the artifact ontology to exclude whole classes of artifacts, as done in the `LivingRoom` example with `WaterFixture`.

activity instance a can be performed in a given location l it is sufficient to add an assertion to the ABox stating that activity a is performed in location l, and then to check if the ABox is consistent with respect to the terminological part of the ontology by performing a *consistency checking* reasoning task:

```
performedIn(CURR_ACT,CURR_LOC_1); isABoxConsistent()
```

The above statements are used to verify if activity `BrushingTeeth` can be performed in location `RestRoom`. In this case the consistency check succeeds, since the declared constraints on the execution of `BrushingTeeth` (i.e., the presence of a sink) are satisfied by the considered location. The same statements, substituting `CURR_LOC_1` with `CURR_LOC_2` verify if activity `BrushingTeeth` can be performed in location `LivingRoom`. In this case the consistency check does not succeed, since the definition of `LivingRoom` states that no `WaterFixture` is present in that location. As a consequence, since `Sink` has been defined as a subclass of `WaterFixture`, the ontological reasoner infers that no sink is present in `LivingRoom`, thus violating the constraints for the execution of activity `BrushingTeeth`.

Algorithm for the Derivation of Possible Activities (*DPA*). The DPA algorithm takes as input an empty ABox and the terminological part of the ontology (called TBox) that describes classes and their properties. The output of the algorithm is a matrix M whose rows correspond to symbolic locations in the TBox, columns correspond to activities in the TBox, and $M_{i,j}$ equals to 1 if activity corresponding to column j is a possible activity in location corresponding to row i according to the TBox; $M_{i,j}$ equals to 0 otherwise.

As a first step, the terminological part of the ontology is classified to compute the hierarchy of the atomic concepts of the TBox. Then for each pair $\langle l_i, a_j \rangle$, where l_i is a symbolic location and a_j is an activity in TBox, the algorithm creates three assertions s_1 = "$a_j(\mathcal{A})$", s_2 = "$l_i(\mathcal{L})$", and s_3 = "performedIn($\mathcal{A}, \mathcal{L}$)" to state that activity a_j is performed in location l_i, and adds them to the ABox. Then, the ABox is checked for consistency, and $M_{i,j}$ is set with the result of the test (1 if the check succeeds, 0 otherwise). Finally, assertions s_1, s_2 and s_3 are retracted from the ABox in order to remove the possible inconsistency that would affect the result of future consistency checks.

An example of the output of the DPA algorithm with a subset of locations and activities modeled by COSAR-ONT is given in Table 1.

3.3 Coupling ontological reasoning with statistical inferencing

Rationale. We illustrate our technique by means of an example.

Example 3. Suppose that user Alice is taking a stroll on a path that goes across the wood near home wearing the sensor equipment of the monitoring system. As explained before, the system (deployed on her mobile phone) continuously keeps track of her current activity, as well as of her current symbolic location (that in this case is `Wood`). The system also knows the matrix M that was calculated offline by the DPA algorithm.

Considering a single activity instance i and the statistical model of m different activities $a_1, \ldots, a_m$, the statistical classifier of the system returns a m-length *confidence*

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Garden	0	0	0	1	1	1	1	0	0	0
HospitalBuilding	1	0	0	0	0	1	0	1	1	1
Kitchen	1	0	0	0	0	1	0	0	0	1
Laboratory	0	0	0	0	0	1	0	0	0	1
LivingRoom	0	0	0	0	0	1	0	0	0	0
Meadow	0	0	0	1	1	1	1	0	0	0
RestRoom	1	0	0	0	0	1	0	0	0	0
UrbanArea	0	0	0	1	1	1	1	1	1	0
Wood	0	1	1	1	1	1	1	0	0	0

Columns: 1=**brushingTeeth**; 2=**hikingUp**; 3=**hikingDown**; 4=**ridingBycicle**; 5=**jogging**; 6=**standingStill**; 7=**strolling**; 8=**walkingDownstairs**; 9=**walkingUpstairs**; 10=**writingOnBlackboard**

Table 1. Part of the M matrix of potential activities

vector $\overrightarrow{s}_i$ in which the j^{th} element $\overrightarrow{s}_i^{(j)}$ corresponds to activity a_j and its value corresponds to the *confidence* of the classifier regarding the association of i to a_j, such that $0 \leq \overrightarrow{s}_i^{(j)} \leq 1$ and $\sum_{j=1}^{m}(\overrightarrow{s}^{(j)}) = 1$. For instance, suppose that the considered activities are those shown in Table 1 (the j^{th} column of the table corresponds to activity a_j), and that $\overrightarrow{s}_i = \langle 0, 0, 0.16, 0, 0, 0, 0.39, 0.45, 0, 0 \rangle$. In this case, the maximum confidence value (0.45) corresponds to activity `WalkingDownstairs`, followed by `Strolling` (0.39) and `hikingDown` (0.16). The confidence value corresponding to the other seven activities is 0. Hence, considering the statistical prediction alone, the classifier would erroneously conclude that user Alice is walking downstairs.

However, looking at matrix M one can note that `WalkingDownstairs` is not a feasible activity in the current location of Alice. The rationale of the COSAR technique is to discard those elements of $\overrightarrow{s}_i$ that correspond to unfeasible activities according to M, and to choose the activity having maximum confidence among the remaining elements (or one such activity at random if the maximum confidence corresponds to more than one activity). In this case, the COSAR technique consists in discarding activities `BrushingTeeth`, `WalkingDownstairs`, `WalkingUpstairs` and `WritingOnBlackboard`, and in choosing activity `Strolling`, since it is the one that corresponds to the maximum confidence among the remaining activities. Hence, in this case the COSAR technique correctly recognizes Alice's activity.

Handling location uncertainty. Every localization technology is characterized by a certain level of inaccuracy. As a consequence, the mapping of a physical location reading to a symbolic location is prone to uncertainty. For instance, if the physical location is retrieved from a GSM cell identification system, the area including the user may correspond to different symbolic locations, such as a `HomeBuilding`, a `HospitalBuilding` and a `Park`.

Uncertainty in location is taken into account by our system. In particular, if the user's physical location corresponds to n possible symbolic locations $l_1, \ldots, l_n$, the possible activities that can be performed by the user are calculated as those that can be performed in at least one location belonging to $\{l_1, \ldots, l_n\}$.

Example 4. Suppose that Alice forgot her GPS receiver at home. Consequently she relies on a GSM cell identification service, which provides coarse-grained location information. In particular, the service localizes Alice within an area that includes both a `Wood` and a `UrbanArea`. Hence, our system calculates the set of Alice's possible activities as the union of the set of activities that can be performed in woods and the set of activities that can be performed in urban areas. Considering matrix M derived by the DPA algorithm and shown in Table 1, possible activities for Alice are those that correspond to columns 2 to 9, included. Therefore, with respect to the scenario depicted in Example 3, in this case `WalkingDownstairs` and `WalkingUpstairs` are possible activities (since urban areas may include steps).

The *COSAR-voted* algorithm. At first, the vector of predictions $\overrightarrow{C}'$ is initialized. Then, the process of actual activity recognition starts. For each activity instance to be recognized, the LOCATION SERVER is queried to obtain the symbolic location corresponding to the current physical location $\overrightarrow{l}_i$ of the user. Note that, if the location server provides location information at a coarse grain, more than one symbolic location can correspond to the user's physical location. Then, sensor data are retrieved from sensors and a feature vector f_i is build by the SENSOR DATA FUSION module. The feature vector is used to classify the corresponding activity instance according to the statistical model provided by the PATTERN RECOGNITION module, obtaining a confidence vector $\overrightarrow{s}_i$. According to $\overrightarrow{s}_i$, to the possible symbolic locations $\overrightarrow{l}_i$, and to the matrix M obtained by the DPA algorithm, the combined ontological-statistical prediction $\overline{c}_i$ is calculated; as explained before, $\overline{c}_i$ is the possible activity according to M having highest confidence in $\overrightarrow{s}_i$. Finally, the voted variant is applied to obtain the voted prediction $\overline{c}'_i$ considering $\overline{c}_i$ and the bag of (non-voted) predictions $\{\overline{c}_{i-1}, \ldots, \overline{c}_{i-k}\}$ of the k most recently classified activity instances. In particular, $\overline{c}'_i$ is set to the prediction having the maximum multiplicity in $\{\overline{c}_i, \overline{c}_{i-1}, \ldots, \overline{c}_{i-k}\}$. If multiple predictions exist which have the maximum multiplicity, one of them is chosen at random. Then, prediction $\overline{c}'_i$ is added to the vector of predictions $\overrightarrow{C}'$.

4 Experimental evaluation

In order to validate our solution we performed an extensive experimental evaluation comparing our technique with a purely statistical one. We point out that the symbolic location is used as a feature only in the experiments performed with the purely statistical technique (named *statistical* and *statistical-voted* in the following). In the experiments with the COSAR technique (named *COSAR* and *COSAR-voted*) location is not used as a feature by the statistical classifier; instead, it is used by the ontological module only.

4.1 Experimental setup

The experiments concerned the recognition of 10 different activities performed both indoor and outdoor by 6 volunteers (3 men and 3 women, ages ranging

Fig. 4. Sensors used for the experiments

from 30 to 60) having different attitude to physical activities. While performing activities, volunteers wore one sensor on their left pocket and one sensor on their right wrist to collect accelerometer data, plus a GPS receiver to track their current physical location. This setup reproduces the situation in which data are acquired from an accelerometer embedded in a fitness watch, and from an accelerometer and a GPS receiver embedded in a mobile phone. Since in the current implementation of our system the GIS module is only simulated, physical locations were manually mapped to symbolic locations.

Each activity was performed by 4 different volunteers for 450 seconds each. Overall, each activity was performed for 30 minutes; hence, the dataset is composed of 5 hours of activity data. The dataset is published on the web site of our project[2] and can be freely used to reproduce the experiments, or as a testbed for evaluating other techniques.

Accelerometer data were retrieved using *Small Programmable Object Technology (SPOT)* by Sun® Microsystems. SPOTs (shown in Figure 4 together with the used GPS receiver) are sensor devices programmable in Java Micro Edition; they are equipped with a 180 MHz 32 bit processor, 512K RAM/4M Flash memory, and IEEE 802.15.4 radio with integrated antenna. They mount a 3-axis accelerometer, and sensors for light intensity and temperature.

Samples from accelerometers were taken at 16Hz, and the time extent of each activity instance was 1 second; hence, the dataset is composed of 18000 activity instances. For each activity instance, accelerometer readings were merged to build a feature vector composed of 148 features, including means, variances, correlations, kurtosis, and other statistical measures. Preliminary experiments (not reported here for lack of space) suggest that in our case feature selection does not improve classification accuracy; however, feature selection can still be useful to reduce CPU usage at run time, hence we will consider this issue in future work.

Statistical classification was performed using *Weka*[3], a Java-based toolkit that provides APIs for implementing several machine learning algorithms. The

[2] `http://everywarelab.dico.unimi.it/palspot`
[3] `http://www.cs.waikato.ac.nz/ml/weka/`

(a) Evaluation of statistical classifiers

Classifier	**Accuracy**
Bayesian Network	72.95%
C4.5 Decision Tree	66.23%
Multiclass Logistic Regression	80.21%
Naive Bayes	68.55%
SVM	71.81%

(b) Overall accuracy

Classifier	**Accuracy**
statistical	80.21%
statistical-voted	84.72%
COSAR	89.20%
COSAR-voted	93.44%

(c) Error reduction

versus →	**statistical**	**statistical-voted**	**COSAR**
statistical-voted	22.79%		
COSAR	45.43%	29.32%	
COSAR-voted	66.85%	57.07%	39.26%

Table 2. Summary of experimental results

COSAR ontology was developed using *Protégé*[4], while *RacerPro*[5] was used to perform ontological reasoning. Since the sensor devices we used lacked a Bluetooth interface we could not establish a direct connection between a mobile device and the sensor devices themselves. For this reason experiments were executed on a desktop workstation. However, at the time of writing we are working on the implementation of the COSAR-voted algorithm for devices supporting Java Micro Edition.

In order to evaluate recognition rates we performed 4-folds cross validation, dividing the dataset in 4 subsamples such that each subsample contains 450 instances for each activity. Ideally, an out-of-the-box activity recognition system should be able to recognize one person's activities without the need of being trained on that person. Hence, in order to avoid the use of activity data of the same user for both training and testing we ensured that activity instances regarding a given volunteer did not appear in more that one subsample.

4.2 Results

Exp. 1) Statistical classification algorithms: The first set of experiments was only aimed at choosing a statistical classification algorithm to be used in the subsequent experiments. In general, since in many applications activity recognition must be performed on-line, the choice of a classification algorithm should privilege not only good recognition performance, but also very efficient classification procedures. Indeed, in many cases, the activity recognition algorithm must be executed on a resource-constrained mobile device.

In this first experiment we compared classification techniques belonging to different classes of pattern recognition algorithms (i.e., Bayesian approaches, decision trees, probabilistic discriminative models and kernel machines). Experimental results on our data (shown in Table 1(a)) show that, among the considered techniques, Multiclass Logistic Regression with a ridge estimator (MLR),

[4] `http://protege.stanford.edu/`
[5] `http://www.racer-systems.com/`

(a) Confusion matrix

classified as →	1	2	3	4	5	6	7	8	9	10
1	**1336**	4	1	11	8	304	0	33	2	101
2	4	**1551**	219	5	14	0	1	1	5	0
3	0	382	**1376**	4	3	1	31	2	1	0
4	1	5	10	**1738**	23	0	0	23	0	0
5	13	3	17	21	**1664**	1	7	73	1	0
6	32	5	3	0	290	**1254**	17	34	126	39
7	0	0	78	0	304	3	**917**	383	115	0
8	0	0	1	0	0	0	2	**1762**	35	0
9	0	5	0	4	0	1	16	144	**1629**	1
10	14	61	1	16	7	485	0	1	5	**1210**

(b) Precision / recall

prec.	recall
95,43%	74,22%
76,93%	86,17%
80,66%	76,44%
96,61%	96,56%
71,94%	92,44%
61,20%	69,67%
92,53%	50,94%
71,74%	97,89%
84,89%	90,50%
89,56%	67,22%

Columns: 1=**brushingTeeth**; 2=**hikingUp**; 3=**hikingDown**; 4=**ridingBycicle**; 5=**jogging**; 6=**standingStill**; 7=**strolling**; 8=**walkingDownstairs**; 9=**walkingUpstairs**; 10=**writingOnBlackboard**

Table 3. Results for the statistical classifier

(a) Confusion matrix

classified as →	1	2	3	4	5	6	7	8	9	10
1	**1622**	0	0	0	0	178	0	0	0	0
2	0	**1443**	171	19	34	14	119	0	0	0
3	0	268	**1284**	22	2	13	211	0	0	0
4	0	4	7	**1787**	1	1	0	0	0	0
5	0	0	0	134	**1640**	9	6	8	3	0
6	0	3	0	26	9	**1738**	21	1	2	0
7	0	0	0	69	9	54	**1597**	67	4	0
8	4	0	0	0	0	1	0	**1753**	42	0
9	24	0	0	0	0	26	0	107	**1643**	0
10	0	0	0	0	0	251	0	0	0	**1549**

(b) Precision / recall

prec.	recall
98,30%	90,11%
83,99%	80,17%
87,82%	71,33%
86,87%	99,28%
96,76%	91,11%
76,06%	96,56%
81,73%	88,72%
90,55%	97,39%
96,99%	91,28%
100,00%	86,06%

Columns: 1=**brushingTeeth**; 2=**hikingUp**; 3=**hikingDown**; 4=**ridingBycicle**; 5=**jogging**; 6=**standingStill**; 7=**strolling**; 8=**walkingDownstairs**; 9=**walkingUpstairs**; 10=**writingOnBlackboard**

Table 4. Results for the COSAR classifier

outperform the other techniques, gaining recognition rates higher than 80%. Hence, our choice for the statistical classification algorithm was to use MLR [17], a classification technique belonging to the class of probabilistic discriminative models [14], having the advantage of being particularly computationally efficient at classification time.

Exp. 2) Statistical technique: Table 3 shows the confusion matrix and precision/recall measures for the statistical technique evaluated in the first set of experiments. As expected, when data from accelerometers are used and the symbolic location is used as a feature, many misclassifications occur between activities that involve similar body movements; e.g., instances of *strolling* are often classified as instances of *walking downstairs*.

Exp. 3) Statistical-voted technique: We evaluated the *voted* variant of the statistical classification algorithm by simulating the case in which a user performs each activity for 7.5 minutes before changing activity. With this technique, the accuracy of activity recognition is 84.72% (see Table 1(b)), which results in an error reduction rate of 22.79% with respect to the statistical technique (see Table 1(c)). Due to lack of space we do not report the confusion matrix and precision/recall measures for this experiment; however, this technique does not significantly reduce the number of misclassifications between activities involving similar movements.

Exp. 4) COSAR technique: The use of the COSAR technique considerably improves the recognition rate with respect to the solely statistical techniques. In particular, the recognition rate of COSAR is 89.2%, which results in an error reduction of 45.43% with respect to the statistical technique, and of 29.32% with respect to the statistical-voted technique. Looking at the confusion matrix (Table 4), we note that COSAR avoids many misclassifications between activities characterized by similar body movements but different locations in which they are typically performed (e.g., *brushing teeth* versus *writing on a blackboard*, and *strolling* versus *walking up/downstairs*).

Exp. 5) COSAR-voted technique: Finally, the voted variant of COSAR (evaluated with the same setup as in *Exp. 3*) further improves classification results, gaining a recognition rate of 93.44%, an error reduction of 39.26% with respect to the COSAR technique, and of 66.85% with respect to the statistical technique.

5 Conclusions and future work

In this paper we proposed the integration of statistical and ontological reasoning for activity recognition. Our technique relies on modeling context data in ontologies and using derived semantic relationships expressing the feasibility of performing a given activity in a given context. Results from extensive experiments with data acquired by volunteers confirm the effectiveness of our technique.

Even if in the current implementation of our system we focused on location data to enact ontological reasoning, our technique can be easily extended to consider a wider class of context data. In particular, future work includes an extension of our technique to consider the temporal characterization of activities (e.g., duration), as well as their temporal relationships (i.e., the probability that a given activity a_i is followed by an activity a_j). To this aim we plan to design a temporal extension of our ontology, and to investigate the use of a probabilistic framework such as Hidden Markov Models. Moreover, since in many situations available context data may be insufficient to unambiguously determine the activity performed by a user, we are investigating the use of fuzzy ontologies to cope with uncertainty and fuzziness.

Acknowledgments

This work has been partially supported by a grant from Sun® Microsystems. The authors would like to thank the volunteers that collaborated to the collection of data used in our experiments.

References

1. Chang, K.H., Liu, S.Y., Chu, H.H., Hsu, J.Y.J., Chen, C., Lin, T.Y., Chen, C.Y., Huang, P.: The Diet-Aware Dining Table: Observing Dietary Behaviors over a Tabletop Surface. In: Proc. of Pervasive 2006. Vol. 3968 of LNCS, Springer (2006) 366–382
2. Tentori, M., Favela, J.: Activity-Aware Computing for Healthcare. IEEE Pervasive Computing **7**(2) (2008) 51–57
3. Amft, O., Tröster, G.: Recognition of dietary activity events using on-body sensors. Artificial Intelligence in Medicine **42**(2) (2008) 121–136
4. Oliver, N., Horvitz, E., Garg, A.: Layered Representations for Human Activity Recognition. In: Proc. of ICMI-02, IEEE Comp. Soc. (2002) 3–8
5. Brdiczka, O., Crowley, J.L., Reignier, P.: Learning Situation Models for Providing Context-Aware Services. In: Proc. of HCI 2007. Vol. 4555 of LNCS, Springer (2007) 23–32
6. Golding, A.R., Lesh, N.: Indoor Navigation Using a Diverse Set of Cheap, Wearable Sensors. In: Proc. of ISWC-99, IEEE Comp. Soc. (1999) 29–36
7. Kern, N., Schiele, B., Schmidt, A.: Multi-sensor Activity Context Detection for Wearable Computing. In: Proc. of EUSAI 2003. Vol. 2875 of LNCS, Springer (2003) 220–232
8. Lester, J., Choudhury, T., Kern, N., Borriello, G., Hannaford, B.: A Hybrid Discriminative/Generative Approach for Modeling Human Activities. In: Proc. of IJCAI-05, P.B.C. (2005) 766–772
9. Liao, L., Fox, D., Kautz, H.A.: Location-Based Activity Recognition using Relational Markov Networks. In: Proc. of IJCAI-05, P.B.C. (2005) 773–778
10. Wang, S., Pentney, W., Popescu, A.M., Choudhury, T., Philipose, M.: Common Sense Based Joint Training of Human Activity Recognizers. In: Proc. of IJCAI-07. (2007) 2237–2242
11. Stikic, M., Huynh, T., Laerhoven, K.V., Schiele, B.: ADL Recognition Based on the Combination of RFID and Accelerometer Sensing. In: Proc. of Pervasive Health 2008, IEEE Comp. Soc. (2008) 2237–2242
12. Huynh, T., Schiele, B.: Towards Less Supervision in Activity Recognition from Wearable Sensors. In: Proc. of ISWC 2006, IEEE Comp. Soc. (2006) 3–10
13. Pareschi, L., Riboni, D., Agostini, A., Bettini, C.: Composition and Generalization of Context Data for Privacy Preservation. In: Proc. of PerCom 2008 Workshops, IEEE Comp. Soc. (2008) 429–433
14. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer (2008)
15. Horrocks, I., Patel-Schneider, P.F., van Harmelen, F.: From SHIQ and RDF to OWL: The making of a Web Ontology Language. Journal of Web Semantics **1**(1) (2003) 7–26
16. Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F., eds.: The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press (2003)
17. Le Cessie, S., van Houwelingen, J.: Ridge Estimators in Logistic Regression. Applied Statistics **41**(1) (1992) 191–201