Personal and Ubiquitous Computing
https://doi.org/10.1007/s00779-019-01319-9

ORIGINAL ARTICLE

Hierarchical syntactic models for human activity recognition through mobility traces

Enrico Casella[1] · Marco Ortolani[2] · Simone Silvestri[1] · Sajal K. Das[3]

Received: 30 October 2018 / Accepted: 5 September 2019
© The Author(s) 2019

Abstract
Recognizing users' daily life activities without disrupting their lifestyle is a key functionality to enable a broad variety of advanced services for a Smart City, from energy-efficient management of urban spaces to mobility optimization. In this paper, we propose a novel method for human activity recognition from a collection of outdoor mobility traces acquired through wearable devices. Our method exploits the regularities naturally present in human mobility patterns to construct syntactic models in the form of finite state automata, thanks to an approach known as *grammatical inference*. We also introduce a measure of *similarity* that accounts for the intrinsic hierarchical nature of such models, and allows to identify the common traits in the paths induced by different activities at various granularity levels. Our method has been validated on a dataset of real traces representing movements of users in a large metropolitan area. The experimental results show the effectiveness of our similarity measure to correctly identify a set of common coarse-grained activities, as well as their refinement at a finer level of granularity.

Keywords Grammatical inference · Mobility · Human activity recognition

1 Introduction

Metropolitan areas have witnessed a steady increase in the number of people living therein, which has triggered an unprecedented concentration of resources and services within their boundaries, as well as the deployment of pervasive urban sensing architectures. As a result, Smart Cities [26] have emerged as a paradigm to turn the large amounts of collected data into an asset for city planners and policy makers, transforming the whole city into an intelligent environment with the ultimate goal of improving the citizens' lives. However, smart services such as energy-efficient management of urban spaces, automated surveillance or disaster risk reduction require an understanding of human behavior that goes beyond the mere processing of a collection of measurements from pervasive sensing devices. For this reason, human activity recognition (HAR) [1, 24, 38] has grown into a self-standing branch of machine learning.

Many of the proposals in the literature [1] heavily rely on information gathered from the environment to capture how everyday-life objects are used, or to detect the presence of the users in relevant areas. Typically networked heterogeneous sensors, including cameras, RFid, or contact sensors, needs to be deployed in selected points of interest. However, extensive and pervasive coverage is costly, and often impractical in large outdoor settings, such as Smart Cities. Alternatively, users may be actively involved in the monitoring process, when wearable sensors (e.g., heartbeat and body pressure monitors, or sensors integrated into portable devices, such as smart wristbands) are used to gather precise information about their actions. This, however, might result in intolerable invasiveness for the users.

During the development of this work, M. Ortolani was with the Dept. of Computer Science, Missouri Univ. of Science and Technology, as a Visiting Scholar under a Fulbright Grant. E. Casella was with the Dept. of Computer Science, Missouri Univ. of Science and Technology, as a visiting student.

✉ Marco Ortolani
m.ortolani@keele.ac.uk

1 Computer Science Dept., Univ. of Kentucky, Lexington, KY, USA

2 School of Computing and Mathematics, Keele University, Newcastle, UK

3 Computer Science Dept., Missouri Univ. of Science & Technology, Rolla, MO, USA

Published online: 25 November 2019

In our work, we argue that relevant insight about the users' behavior may be gathered by analyzing their mobility patterns, and we propose a novel approach to complex human activity recognition from the the analysis of outdoor mobility traces. Contextual information is gathered without extensive environmental sensor coverage, and rather just by means of the GPS sensors of commonly available portable devices, namely smart phones, in a completely unobtrusive way. Given the peculiar nature of human mobility, however, grasping the common traits of seemingly unrelated paths may be a daunting task. Nonetheless, advanced services like traffic prediction or mobility optimization require that an activity is captured in its entirety. It is thus essential to know *why* users move across the space (i.e., identify the activity that induced the trajectory), as opposed to just *how* they perform the movement. For instance, a route recommender system might be able to improve the quality of the offered service, and tailor it to the specific user's needs by suggesting shortest routes to people commuting to reach their job, scenic ones to tourists on a sightseeing tour, or less traffic-intensive ones to bikers.

We propose here to represent activities by means of hierarchical models constructed within the framework of Algorithmic learning theory (ALT), i.e., the study of formal languages and their recognizers, automata. Our goal is to capture the natural regularities in mobility traces induced by users' activities by symbolically encoding them, and regarding them as strings generated by an unknown grammar. To this aim, we use grammatical interface (GI) [18], an inductive process capable of selecting the best grammar consistent with the provided samples, which in our case are trajectories encoded in symbolic form, and labeled according to the activities that triggered them. The obtained models are constructed as a composition of simpler models of the same nature, so they are naturally suitable to represent the hierarchical nature of the activities, where each level on the hierarchy corresponds to a level of geographical granularity.

We have validated our approach on a publicly available dataset of real-life trajectories representing movements of users occurring mostly within a large metropolitan area, with occasional long-distance transfers. Experimental results show that our models accurately characterize the users' activities at different geographical granularities, and that the proposed similarity measure can correctly classify them against a taxonomy representing coarse- and fine-grained tasks in everyday life.

The contribution of this paper is threefold: (1) we present a novel tool based on grammatical models to express and extrapolate the underlying semantics of complex activities from mobility traces; (2) we describe a framework showing how the same models can be used to refine the recognition process, both with respect to more specific activities and to finer levels of geographical granularity; and finally, (3) we define a similarity measure that mirrors the intrinsically hierarchical nature of the proposed models, and makes it possible to use them to infer the activity performed by previously unseen users.

The remainder of the paper is organized as follows. Section 2 summarizes the relevant work on activity recognition. We describe our approach to building user activity profiles by grammatical models in Section 3, and the proposed global similarity measure in Section 4. Section 5 presents our experimental results on a dataset of actual mobility traces collected by GPS devices on smart phones. Finally, Section 6 draws the conclusions and discusses on-going research.

2 Related work

Many different methodologies have been proposed to automate the process of human activity recognition. A first, broad distinction can be made according to the taxonomy proposed by [1], between *single-layered*, and *hierarchical* approaches (see Fig. 1). In the former case, human activities are regarded as series of gestures and actions with sequential characteristics; models are usually provided in terms of points in a space-time domain or as sequences of observations. Hierarchical approaches adopt a different viewpoint, in that they attempt to describe high-level human activities in terms of simpler ones, so that the inherent *structure* may emerge. The authors of [2], for instance, show that the overall accuracy of the activity recognition system is improved when a structural

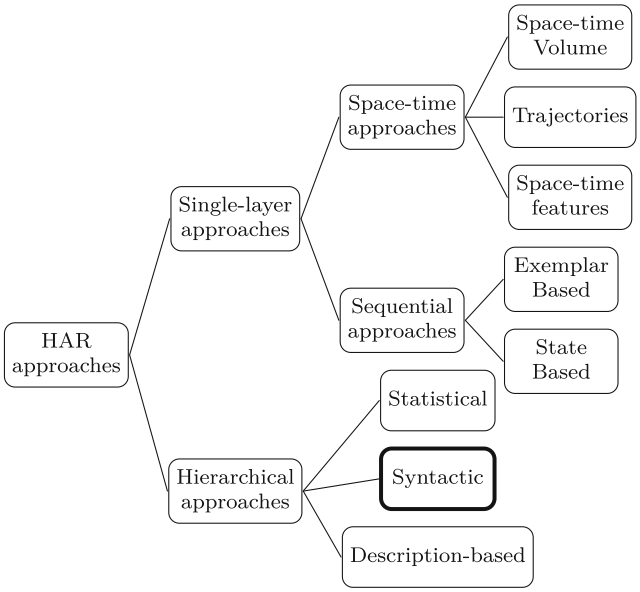


Fig. 1 A taxonomy of HAR methodologies (adapted from [1])

Table 1 Qualitative comparison of HAR approaches. Qualitative categories: (a) Atomic/Complex activities; (b) Personal/Environmental context; c In the Lab/in the Wild data acquisition; (d) Subject- Dependent/Independent profiling

HAR approach	Qualit. categories (a)	(b)	(c)	(d)	Setting	Type of seonsors	Activities recongnized
Bao and Intille [4]	A	P	L	D	Indoor and outdoor	Wearable sensors (biaxial accelerometers)	20 activities: walking, running, washing laundry, reading, vacuuming, ...
Leo et al. [27]	A	E	W	I	Outdoor	Vision-based sensors (static cameras)	4 activities: walking, probing the ground, picking object, damping the ground
Choudhury et al. [10]	A	P	W	I	Indoor and outdoor	Weareable sensors, GPS and wi-fi	10+ activities: walking, cycling, brushing teeth, ...
Chen et al. [8]	C	E	L	D	Indoor	Ambient sensors (contact, motion, tilt and pressure sensors)	8 activities: making tea, making coffee, making pasta, ...
Dernbach et al. [19]	A/C	P	L	I	Indoor and outdoor	Wearable sensors (smart phones)	15 activities: biking, sitting, sweeping, ...
Cook [12]	A/C	E	L	I	Indoor	Ambient sensors (motion, temperature, door and interaction-based sensors)	11 activities: bathing, cook, take medicine, eating, relax, ...
Furletti et al. [20]	A	P	L	I	Outdoor	Vehicle GPS tracking devices	10 activities: training, going home, touring, ...
Saguna et al. [34]	C	P/E	W	I	Indoor	Wearable and interaction-based sensors (smart phones, RFId)	16 activities: getting ready at home, cooking, eating breakfast, ...
De et al. [16]	C	P/E	W	I	Indoor	Wearable sensors, bluetooth beacons	19 activities: walk indoor, run indoor, use refrigerator, clean utensil, cooking, ...
Gaglio et al. [21]	A	E	L	I	Indoor	Kinect camera	8 activities: catch cap, toss paper, take umbrella, walk, phone call, drink, sit down, stand up
Blumrosen et al. [7]	A	E	L	D	Indoor	Kinect camera	2 activities: walking in a complex pattern, repetitive hand tapping
Vaizman and Ellis [39]	A	P/E	W	I	Indoor and outdoor	Smart phones, wearable devices and acoustic sensors	10 activities: bathing, in a meeting, at a restaurant, ...
da Penha Natal et al. [15]	A	P	W	D	Mainly outdoor	Smart phones GPS	13 activities: dining, recreation, shopping, studying, waiting transport, ...
Liono et al. [28]	C	P	W	I	Outdoor	Smartphone sensors	4 activities (riding, walking, drinking, playing) and additional transportation modes
Saini et al. [35]	C	E	L	I	Indoor	Kinect camera	24 basic activities involving 2 person interaction
Younes et al. [43]	C	E	W	I	Indoor	Ambient cameras	8 activities: eating, taking medication, brushing teeth, mopping, using a computer, writing, making phone call, (simulated) driving
Our approach	C	P	W	I	Outdoor	Smart phones	6 activities: travel, work, shopping, sport, social and spare time

representation of data features in the form of a graph is used, and the use of a hierarchy of graphical models, namely layered hidden Markov models, has been proposed by [29].

A work that is closely related to the approach presented here is described in [33]. The system proposed therein maintains the representation of an activity describing how its composing gestures must be concatenated temporally, spatially, and logically. Interestingly, the authors suggest using probabilistic context-free grammars to implement a semantic-level activity recognition algorithm, and they show that this leads to an improvement on the overall system accuracy. However, the syntactical structure itself is not inferred, but rather matched against the occurrences of lower-level components. In our approach, on the other hand, we aim to let syntactical models emerge from data. By building up on previous work [13, 14], where trajectories of individual users were represented in the form of finite state automata, we broaden the scope and focus here on the capture of the intrinsic structure of complex activities starting from mobility traces.

Other systems from the literature on HAR are reported in Table 1 on page 20 along with their discriminating characteristics, such as the targeted setting—either indoor or outdoor—the employed sensors, and the range of recognized activities. We further classify them in terms of some representative *qualitative categories*: (a) their ability to capture the complexity of the activities; (b) the requirements in terms of monitoring equipment to infer the context; (c) the practical applicability of the proposed approach in a real-life setting; and (d) the ability to produce subject-independent models.

Many approaches are characterized by a focus on elementary activities performed in constrained environments or during short periods of time, so they may fail to catch the inherent *complexity* of human behavior. As pointed out by [42], unless only simple, atomic acts are considered, the diversity arising from concurrent or interleaved activities necessarily requires the identification of various levels of abstractions: a high-level activity (e.g., working, shopping or dining out) is to be expressed as a complex sequence of actions, which in turn are series of atomic acts (i.e., primitive patterns, such as gestures) that may be easily singled out. Clearly, this usually implies using some kind of hierarchical model.

A second important parameter to be taken into account regards the use of *context*, which is usually associated with the type of sensors used to detect the behavior of users [32, 40]. The environment may play an active role, as is the case when sensors are deployed in selected points of interest to capture how everyday-life objects are used, or when cameras are used to capture activity-related features such as position, posture, or motion. While pervasive monitoring increases the system precision, it clearly requires possibly costly and hard-to-maintain deployment. On the other hand, when users are actively involved in the monitoring process, a personal context may be obtained by means of wearable sensors, such as heartbeat and body pressure monitors, or sensors integrated into portable devices, such as smart wristbands. In [25], a review of the state of the art in HAR by means of wearable sensors is reported, where the authors assess 28 systems targeting medical, military, or security scenarios. In this case, minimizing the inconvenience for the end-user and preserving their privacy should be the primary concerns. When mobility is the trait of human behavior that is to be modeled, smart phones are a natural candidate as a monitoring device. For example, tracking users exploiting location data gathered through a cellular network has been addressed in [5], with an interesting characterization within an information-theoretic framework, and in [19], where smart phones are used to recognize complex activities in an indoor environment, such as cooking or cleaning, by means of classifiers like multi-layer perceptrons, naïve Bayes, and Bayesian networks.

A further characterizing feature of HAR systems is the manner in which *data collection* is performed. In many cases, only application-specific activities performed in a laboratory setting and in a scripted manner are considered. On one hand, this simplifies the task of obtaining reliable labels for the training data, which is essential when supervised methods are to be used for classification; however, in a real-world scenario, humans perform complex activities in a variety of ways, and such heterogeneity can only be captured by performing experiments "in the wild."

Finally, some of the cited approaches need to be trained and tested for each individual, while others are able to produce a classifier which is valid across different subjects, and can thus deliver subject-independent models for the activities under exam.

As compared with all of the mentioned works, our approach is the only one addressing complex activity recognition in an outdoor setting, with minimal requirement in terms of the used sensors.

3 Building activity profiles hierarchically

In this work, we aim to address the diversity of complex human behavior by building *structural* mobility models, suitable for matching user mobility models against prototypal ones relative to known activities by computing the relative similarity. The underlying assumption is that human trajectories are likely to show a high degree of temporal and spatial regularity which arguably derives from simple, reproducible patterns rather than abstract statistical models, such as basic random walk [23]. The concept has been reaffirmed in [37], where a remarkable lack of variability

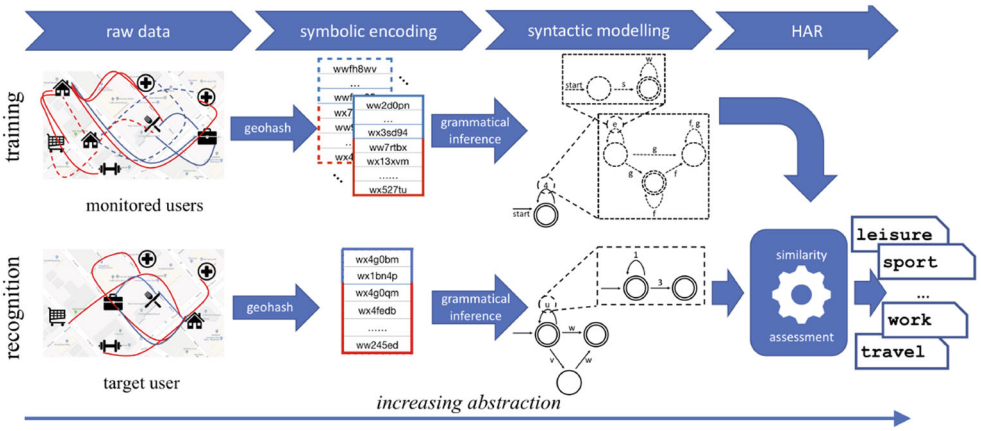


Fig. 2 Outline of the proposed methodology

in predicted travel patterns was observed by measuring the entropy of each individual's trajectory. Hence, it is reasonable to expect that comprehensive models for all trajectories related to the same activity may be inferred by capturing such underlying regularities. In order to do so, we make use of a structural learning approach.

A comprehensive outline of the proposed approach is depicted in Fig. 2. Considering the topmost half of the picture, we assume that a collection of the paths traveled by a group of users is available, either via a public repository or acquired through an ad-hoc application, and labeled according to the activities they were performing. We will show how a proper encoding system allows to turn such raw sequences of geographical locations into strings, i.e., sequences of symbols corresponding to discrete areas. A key feature of human mobility is that it is *intentional*, as it implies an underlying purpose corresponding to the activity that is to be performed, e.g., reaching the workplace, going shopping, and walking in a park. Hence, our syntactic approach allows us to model trajectories sharing an underlying regularity (i.e., reflecting the same underlying activity) by means of finite state automata. As will be discussed in the following, we will use a supervised learning method to this end, and discuss the techniques that can be used in a realistic scenario where data are likely affected by noise.

The lower half of Fig. 2 shows outlines the recognition phase of our method. Once reliable general activity models are produced, the same process may be used on the traces of a previously unseen user. Their model may then be matched against the reference activity models produced earlier, and labelled according to the measure of similarity, which will be presented in Section 4.

3.1 Expressing trajectories symbolically

Mobility traces are generally stored as sequences of locations, possibly coupled with a timestamp depending on the sampling rate of the measurements. A common representation is by pairs of latitude/longitude coordinates.

Since our approach is based on automata designed to recognize a language, the coordinates are to be translated into a symbolic form. To this aim, we selected an encoding system known as *geohash* [3], which assigns a hash string to each latitude/longitude pair in a hierarchical fashion. Considering a specific geographical zone, and the corresponding geohash cell, is equivalent to selecting a specific *granularity* for measurements. Starting with the coarsest granularity (covering the entire globe), any chosen region is divided into 32 subcells identified by a symbol, as shown in Fig. 3. The process may be recursively repeated up to the desired precision.

The collection of all sequences of geohash points, describing the paths travelled by a user while carrying on a specific activity, represents the *raw* description of the activity itself. While coarser granularities may allow us to capture trajectories in their entirety, the corresponding strings would lose detail about the geographical position,

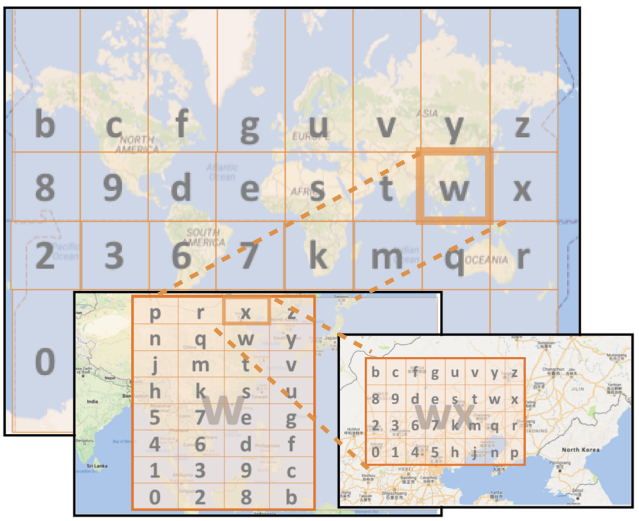
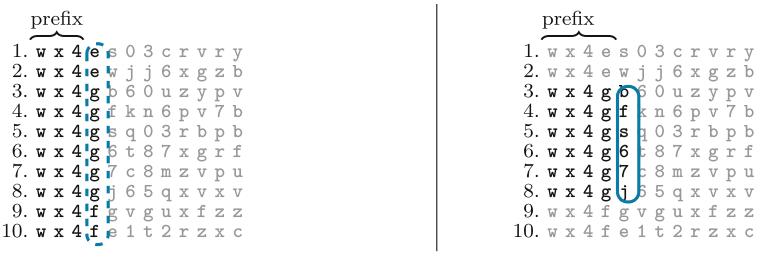


Fig. 3 Hierarchical structure of geohash cells

and the converse holds for finer granularities. For instance, the simple example in Fig. 4 represents a path of 10 locations in a cell covering a metropolitan area defined by the **wx4** geohash cell. In our encoding, this trajectory is defined by the sequence of symbols following the cell geohash prefix, namely **eegggggff**:



If we focused instead on a smaller area covering just a neighborhood within the city, we could consider only the points along the solid line in the figure, characterized by the longer prefix **wx4g**, and resulting in a sub-trajectory described by string **bfs67j**. On the other hand, if we selected a country-wide cell such as **wx** as our base granularity, the sequence of locations would be completely described by sequences of varying length of the symbol **4**, which would be recognized by the automaton equivalent to regular expression 4^*. Although completely accurate, this representation is not detailed, as it only captures the fact that the user stays in the country. A comprehensive model, that is able to capture similarities at the different levels of granularities is hence needed.

3.2 Hierarchical automata models

We now turn our attention to the issue of learning a model for the movements of a group of users reflecting the activity that originated them. Provided a symbolic encoding for the movements, this learning problem may be formulated in terms of GI as the task of inferring the most general recognizer of a given set of strings, i.e., the *minimal* DFA, consistent with the data. In our case, the alphabet for the strings used for training is represented by the geohash symbols. The accepting states of the inferred automaton would identify those strings corresponding to trajectories actually traveled by the user when performing the activity the automaton is intended to recognize. In this work, we make use of *passive learning*, a well-known approach for inferring the automaton recognizing a set of strings provided as training set. It may be regarded as an inductive process of a supervised learning, where inference is formulated as a search in a state space [18].

It is worth pointing out that the chosen symbolic representation for our trajectory strings gives us the freedom to apply the learning algorithm at any desired granularity level. The automaton constructed for a geohash subcell is thus just a specialization for the corresponding transition of the parent model, so the overall model can be conveniently expressed as a composition of automata, as depicted in Fig. 5.

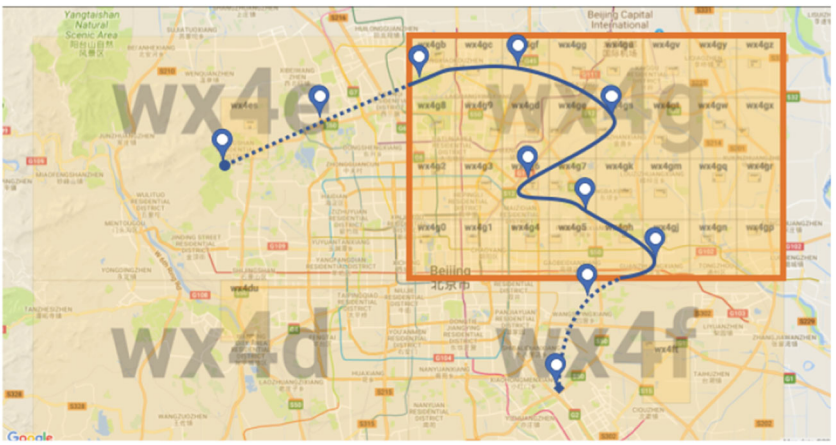


Fig. 4 A geohash trajectory crossing multiple subcells

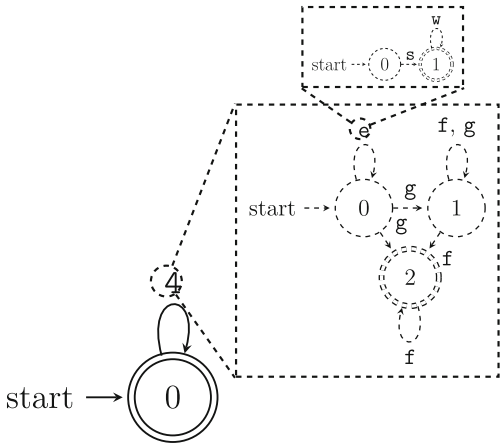


Fig. 5 A hierarchy of automata

3.3 Learning from real-life data

To train a passive learning algorithm, data must be presented to it in the form of *positive* and *negative* examples, i.e., strings that are supposed to be accepted (respectively, rejected) by the classifier. Labels will depend, in our case, by the activity for which the model is being built. Learning begins by formulating an initial hypothesis as the automaton representing the available examples. Note that this is always possible, but one such automaton would trivially recognize just the strings provided as training. The aim is then to have it evolve toward a more general automaton, capable of capturing more strings of the unknown language, which roughly corresponds to the idea of generalization in typical machine learning methods. This may be carried on by means of a structural operator, namely pairwise state merging, that is able to generate automata describing languages larger than the original. From the initial, overfitting, automaton, a lattice of more general recognizers is produced by repeated application of state merging. In order to avoid mistakenly accepting negative samples, overgeneralization needs to be limited, and it may be shown that this amounts to identify a "frontier" in the search space [18]. When dealing with symbolic data generated from real-life measurements, the selection of a meaningful training set of data may be particularly challenging. It is crucial that examples possibly leading to overfitting are discarded, while the most relevant ones are kept. In particular, since the search is constrained only by negative samples, it is essential that they are as representative as possible of the target language, and the order in which they are presented to the algorithm is important.

For coarser granularities (corresponding to geohash prefixes identifying cells larger than a metropolitan area) we chose to use Regular Positive and Negative Inference (RPNI) [31]. This is a passive learning algorithm that performs an exhaustive search of the automata space built via repeated application of the state merging operator, until the frontier of acceptable automata is reached. A remarkable property of RPNI is that it is able to identify in the limit the minimum consistent automaton provided that the learning sample is representative of the unknown model. Its complexity is heavily influenced by the size of the initial automaton, whose width is a linear function of the number of elements in the training set, and whose depth is linear on the size of the longest string in the training set. However, coarser granularities are characterized by few, short strings since many trajectories collapse into the same encoding, so the overall running time of the algorithm is effectively contained.

At finer granularities, on the other hand, we want to account for the fact that our data are not guaranteed to be completely noise-free, which often results into mislabeling. For instance, incomplete or noisy measurements might cause a trajectory string to be assigned to the wrong class, since a small error in a measure corresponds to a potentially very different symbolic encoding. Additionally, structurally similar inputs with contrasting labels typically lead to a needlessly more complex recognizer, which would make RPNI impractical. Hence, at finer granularities, we employed the *Blue** algorithm [36], which specifically addresses potentially mislabeled data by statistically distinguishing between relevant and irrelevant information, which is treated as noise. While the aim is to evolve from the initial, overfitting automaton towards a more compact and general automaton, as in RPNI, here some tolerance to an error in classification is added if it improves generalization. In particular, a generalization by state merging will be deemed as *statistically acceptable*, and consequently the reduction in the size of a DFA is accepted, only if the resulting statistical error does not exceed some chosen threshold. The underlying idea consists in verifying that the proportions of misclassified samples do not increase significantly after a state merging; more specifically, *hypothesis testing* [6] is used to drive statistical inference.

As will be shown in Section 5, the combined use of the two mentioned algorithms allows us to obtain compact models at all granularities, without hindering the overall accuracy.

4 Similarity between activity models

Once reliable models for the activities related to the mobility traces of a group of users are available, we need a method for comparing them. To this aim, we propose a similarity measure between pairs of activity models that takes into account their intrinsic hierarchical nature. Initially, we will focus our attention to the computation of the similarity score between pairs of *local* models, i.e., automata built for the same granularity level. Then, we will show how we can

formulate a comprehensive similarity measure suitable to reliably compare the activity models in their entirety, i.e., considering all granularities of interest.

4.1 Computing local similarity scores

Unlike statistical models, in which similarity is typically assessed by evaluating statistical metrics on a chosen set of features, here activities are described in terms of strings. This implies that we need to assess the similarity between two *languages*. At the same time, we want to account for the nature of the recognizers we are using (i.e., automata), so the chosen measure should also consider similarity in terms of *structure*. This dual aspect is well captured by the similarity measure proposed in [41], which results from the combination of a *linguistic* part and a *structural* part.

The first part of the similarity score is computed using the so-called *w-method* [11]. Considering two automata $\mathcal{A}_1$ and $\mathcal{A}_2$ over the same alphabet, a "representative" set of strings is constructed to be used as probes for the two automata under observation. Roughly speaking, such strings need to ensure that each state and transition of the target automaton is triggered at least once. The score will then depend on how many strings are identically classified by both automata (with respect to our scenario, this part of the measure aims to assess whether geohash trajectory strings are assigned to the same activity according to both automata), and will be computed through classic metrics for classification assessment. More precisely, in order to avoid a bias of the score toward the positive or negative samples, the *linguistic similarity* is expressed as the *F-measure*, i.e., the harmonic mean of precision (prec), and recall (rec):

$$S^{\mathcal{L}} = \frac{2 \cdot \text{prec} \cdot \text{rec}}{(\text{prec} + \text{rec})} \tag{1}$$

The second part of the similarity score aims at comparing the recognizers not in terms of their languages but in terms of their states and transition structures. The measure is based on the "neighbor matching" proposed by [30], whose underlying idea is that the overall structural similarity between two automata $\mathcal{A}_1$ and $\mathcal{A}_2$ can be expressed as the normalized sum of the highest similarity scores between pairs of states. The pairwise state similarity $x_{i,j}$ is computed iteratively in terms of the number of matching incoming and outgoing transitions in the respective neighborhoods of states $i \in \mathcal{A}_1$ and $j \in \mathcal{A}_2$. Finally, the best k states according to the pairwise similarity are selected to compute the overall *structural similarity* measure, as follows:

$$S^{\mathcal{A}} = \frac{1}{n} \sum_{l=1}^{k} x_{f(l)g(l)} \tag{2}$$

where n is the maximum number of states in both automata; $f : \{1, \ldots, k\} \rightarrow \text{states}(\mathcal{A}_1)$ is the enumeration function over the states of $\mathcal{A}_1$, and g the analogous for $\mathcal{A}_2$, that return the ordering for the final best mapping between nodes of the two automata.

By construction, both the linguistic and the structural parts of the similarity score fall within the interval [0, 1], with the upper bound indicating complete similarity. A measure expressing the comprehensive similarity of two automata at the *same granularity* level may thus be obtained as a linear weighted sum of the two parts:

$$S(\mathcal{A}_1, \mathcal{A}_2) = \gamma \, S^{\mathcal{L}} + (1 - \gamma) \, S^{\mathcal{A}} \tag{3}$$

where the parameter γ may be used to fine tune the relative influence of the linguistic and structural parts on the composite measure. A small value of γ would bias the measure towards automata of similar complexity (in terms of number of states and transitions) and thus be useful to disregard small differences in the sequence of symbols in the strings. On the other hand, values of γ close to 1 would produce higher similarities when the same sets of strings are recognized identically by the automata, regardless of their structure.

4.2 Composing local similarity scores into a global measure

In our approach, the activities are modeled by a hierarchical composition of automata, so the formulation for a global similarity score must reflect the same structure.

Referring back to the example in Fig. 5, **wx** was selected as the base granularity so trajectories would be completely described by the automaton equivalent to regular expression `4*`, assuming that all locations fall within the **4** geohash subcell. Even though such automaton would provide a satisfactory model with regard to the description of the user movements, it would not represent their activities as well. In fact, while its statistical precision would be optimal, it would not be able to distinguish between sequences of movements corresponding to the different activities.

With reference to Fig. 6, assume that the solid and dashed lines identify two classes of trajectories, corresponding to two different activities of the user (say, "going to work" and "do shopping") (Table 2). The algorithm of grammatical inference would produce the same recognizer for both activities (i.e., the automaton labeled **wx**, which would simply accept all sequences of **4**'s and reject the others). This implies that they would be completely indistinguishable from each other, at the considered geohash granularity, according to the previously defined similarity measure. On the other hand, we would like to retain the concept that, despite their differences, two activities may show some degree of similarity as they may occur within the same geographical area, or involve movements that are structurally similar. To this end, we can exploit the

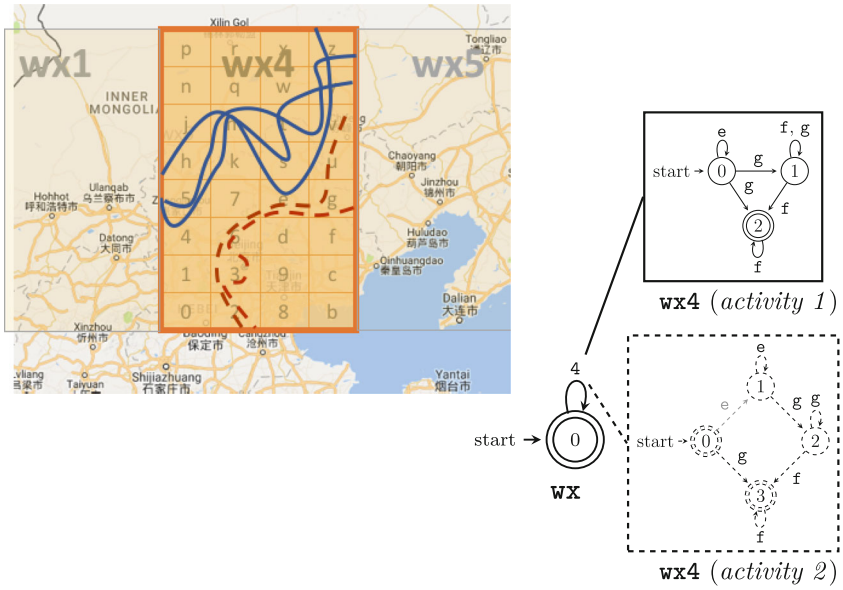


Fig. 6 Capturing dissimilarity at different granularities

hierarchical nature of the automata models, and consider the fact that recognizers for lower granularities are more likely to capture additional details of the user movement. In our example, this means we focus on the **4** subcell, i.e., select points represented by strings sharing the **wx4** prefix. The automata inferred for that granularity would specialize the transition on symbol **4**, and are likely to reflect in their structure the differences in the shape of the trajectories of different activities. The procedure can also be iterated at lower granularities, until we reach such a finer precision that the concept of trajectory would not be representative any longer.

An important question arises in this regard: is it possible to provide a comprehensive formulation of similarity for any pair of hierarchical automata that globally represent the model for the user activity? In real-life settings, it has been shown that people tend to travel frequently on very few paths, and more rarely vary their routes [23]. This means that only a minority of the geohash cells at any given granularity level will contain a non-negligible amount of trajectories, so it is sufficient to refine our models only for those subcells.

In order to obtain a measure of similarity accounting for the hierarchical nature of our models, we compute a *global similarity* S_p^G at prefix p in terms of the "flat" score S_p, as computed by (3). The idea is that we consider S_p as an initial approximation that can be improved by considering all relevant lower-granularity refinements. Formally:

$$S_p^G = S_p + \sum_{c \in \Psi(p)} \phi_c (S_c^G - S_p) \quad \text{with} \quad \sum_{c=1}^{|C(p)|} \phi_c = 1 \tag{4}$$

where $c \in C(p)$ denotes all the possible subcells of prefix p, while $\Psi \subseteq C$ selects the ones that we want to include for refinement, and ϕ_c is the weight modulating the contribution of each subcell.

The intuition behind the above formulation is as follows: human mobility is characterized by the fact that most movements are concentrated in limited areas, not all subcells will provide a relevant contribution to the similarity score. Cells interested by limited, or no movement at all can be excluded, so they do not contribute to the score, which

Table 2 List of the activities used for labeling trajectories

Activity label		Description	Incidence	
Travel		Transit through transportation hubs; movements outside of metropolitan area		1.69%
Work		Movements to and from university buildings, private companies		71.67%
Leisure	Shopping	Malls, food and convenience stores	9.35%	26.64%
	Sport	Use of recreational centers, sport fields, parks, outdoor activities	3.27%	
	Social	Visit to entertainment areas, theaters, museums, friends	4.23%	
	Spare time	Routes within city limits further from other points of interest	9.79%	

would just be determined by the upper-level similarity. On the other hand, if we choose to refine the score at a given granularity p by using all its subcells, then the similarity score would collapse in $S_p^G = \sum_{c \in \mathcal{C}(p)} \phi_c S_c^G$, and hence be a function of the lower refinements only.

It is also worth pointing out that the term S_c^G in the summation of (4), which refers to the similarity of models in a subcell, is recursively computed with the same formula.

5 Experimental study

For validating our proposal, we considered the *Geolife* dataset [44] provided by Microsoft Research Asia. It contains a collection of geographical locations in the form of (*latitude*, *longitude*, *altitude*) triples, representing the movements of 182 users monitored for 5 years in the region of Beijing, China, although routes crossing USA and Europe were occasionally present. More than 17,000 trajectories are stored, acquired trough GPS loggers and smart phones, generally with a high sampling rate of 1 to 5 s in time, and 5 to 10 m in space. However, as shown in Fig. 7, for many of the users, only a small number of trajectories was collected, some of which are too short, or contain clearly erroneous measurements. Hence, for our experiments, we selected 10 representative users, each characterized by more than about 300 non-trivial trajectories. Finally, we disregarded the information about altitude, which was not relevant for our purpose.

5.1 Preliminary data processing

The preliminarily step in our analysis consisted of generating a reliable ground truth for our data. In particular, even though the *Geolife* dataset is anonymized for privacy reasons, the authors of [9] state that a projection of the available data on a map shows that "the volunteers tend to have similar background since they share a common area

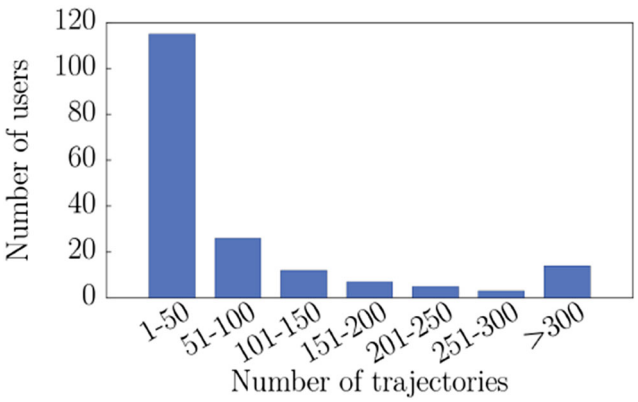


Fig. 7 Distribution of the monitored trajectories per user

with the highest density of visits, which is the assembling place of IT companies. This indicates a high chance that the volunteers may have similar interests to each other."

In order to transform raw locations into trajectories, the same authors suggest that only the so-called *stay points* are considered, which represent groups of nearby positions where a user lingered for a sufficient amount of time. After eliminating outliers, they apply a density-based algorithm to hierarchically cluster the stay points into areas referred to as *regions of interest* (RoI), to which a *location semantics*, that is, the intended functionality of that region (e.g., park, school, workplace, hospital), is associated. Temporal and location semantics together constitute a so-called *T-pattern*, as defined by [22], and an algorithm of frequent sequential pattern mining is used to extract the sequences of places frequently visited by a user and to estimate their similarity with respect to other users. In our experiments, we retained the idea of computing RoIs from users' raw locations, but used them only to semantically label trajectories. The most frequently visited locations were clustered into RoIs. Their proximity to known locations was analyzed and they were tagged as workplaces, transportation hubs, or recreational locations. Trajectories starting, ending, or traversing them have been labelled accordingly, and a set of categories representative of typical users' activities were assigned to them. After this step, however, we retain the native sequence of locations, in geohash encoding, since our aim is to extract the regularities in their original structure.

The actual encoding of a trajectory as a geohash string requires choosing a base granularity or, equivalently, setting a common prefix for the locations. As most of the trajectories in *Geolife* occur in the north-eastern part of China, the shortest possible prefix, `w`, would allow us to capture all of them. Considering that strings falling within geohash cells with a prefix length larger than 5 symbols would span an area roughly as compact as a few city blocks, such prefix was assumed to convey satisfactory precision for our purposes. At the coarser granularity, on the other hand, the high sampling rate provided by *Geolife* is redundant. A series of measurements taken only a few seconds apart in an area of more than 1000 km^2 would be encoded as long repetitions of identical symbols. Therefore, not only would they fail to convey any significant information about the trajectory but also likely hinder the inference process. In our experiments, we chose to adaptively sample the data, with a lower rate (60 s) for prefix length shorter than 4, while keeping the full detail for finer geographical granularities.

Finally, we note that at any chosen granularity, user paths traverse only a potentially small subset of all possible 32 subcells. The degree of coverage, defined as the percentage of trajectories falling into a subcell with respect to the overall number of trajectories at that granularity, thus

Table 3 Weights of subcells for global similarity

Granularity	Subcell (coverage)	
`w`	`wx`	(94.53%)
`wx`	`wx4`	(95.48%)
`wx4`	`wx4g`	(51.36%)
	`wx4e`	(32.71%)
`wx4g`	`wx4g3`	(18.51%)
`wx4e`	`wx4er`	(20.92%)
	`wx4ex`	(15.84%)

provides a good indicator of the relative importance of the subcells. Table 3 shows the subcells used to refine global similarity, and their relative weights, which were used in our experiments as estimates for parameter ϕ_c in (4).

5.2 Accuracy of structural models

The complete list of activities we aim to identify is reported in Table 2 together with a brief description, and the relative percentage of trajectories they refer to. Three of the categories (*travel*, *work*, and *leisure*) are broadly scoped, and capture typical activities performed by users both within, and out of the metropolitan area. As shown, according to our categorization, users are typically involved in work-related activities, consistently with documented in previous works on the same data [9], whereas only a minority of trajectories are relative to travels. In order to assess the potentiality of the proposed similarity measure to discriminate between finer-grained activities, we further specialized the *leisure* activity into 4 sub-categories.

Evidently, the reliability of the similarity measure depends on the accuracy of the inferred models. In a grammatical inference framework, one of the most delicate issues is the generation of negative samples. In general, positive samples will improve the model precision, while negative ones will guide the learning process and limit its overgeneralization. Therefore, mislabeling may have a disruptive effect not only on the accuracy but also on the complexity of the resulting model, which in turn would negatively affect the similarity score. In our case, however, the ground truth assignment provides a reasonable initial choice of positive and negative sample sets. In particular, when inferring the model for one of the activities, trajectories corresponding to the other activities will be used as negative samples. For instance, in the case of the three broadly scoped activities, the negative sample set for *work* would be represented by *leisure*, and *travel*. In general, whatever is the chosen taxonomy, we cannot be expected it to cover the whole range of a user's movements, so we enrich the negative set by excluding cells not covered by *any* trajectory. This has a beneficial side effect for the inference algorithm, as it potentially reduces the size of the alphabet of a specific model. Together with our adaptive subsampling at different prefixes, this allows to keep samples short at lower granularities, thus lowering the overall running time and complexity of the algorithm, while keeping the model simpler.

Models for coarser granularities, where strings show a simpler structure due to the nature of the movements at large scale and to our preprocessing, were inferred using the *RPNI* algorithm [31], whereas for granularities corresponding to prefix lengths 4 and beyond, we used *Blue** [36], whose parameters controlling noise tolerance were selected using a grid search. In all experiments, 75% of the available data was used for training and the rest for the test. In order to obtain unbiased results with regard to the specific subdivision, 10 runs of cross-validation were performed and the average result was reported. The plots in Fig. 8 report the accuracy (expressed as the *F-measure*) of the models obtained for *work*, *leisure*, and *travel* for the 10 selected users with increasing granularity. No trajectories corresponding to a travel-related activity were found for users 41 and 163, so the corresponding bars are missing in Fig. 8c. Performances are usually satisfactory, reaching 0.82 accuracy on an average for *work* and 0.73 for *leisure*, although not homogeneously across the different granularities. Accuracy results are clearly more

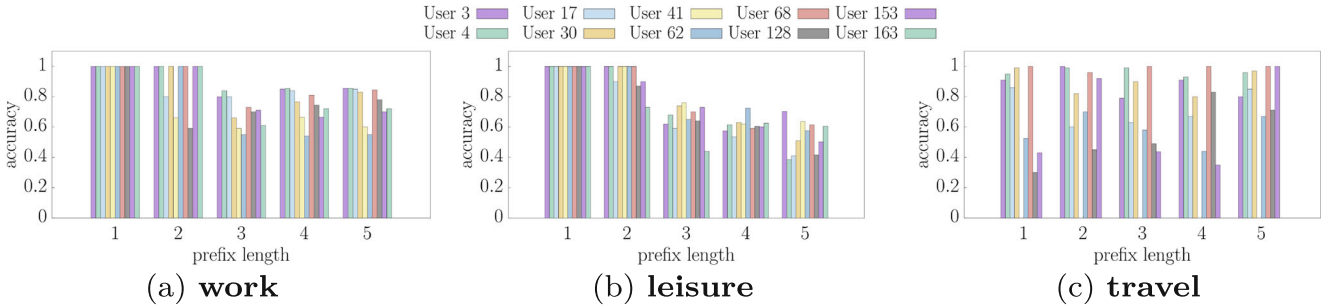


Fig. 8 Accuracy of the individual models for three coarse-grained activities, namely work (**a**), leisure (**b**) and travel (**c**), at varying granularities for all users

stable for the *work* activity, to which the majority of trajectories belong. *travel* models, on the other hand, are less satisfactory. This is likely due to the high imbalance in the training sets, where negative samples greatly outnumbered the positive ones.

5.3 Test scenarios for activity discrimination

Our main interest lies in discerning whether two sets of trajectories were induced by the same activity, which amounts to assessing how close those are to one another in terms of our similarity score. In particular, our goal is to be able to tell activities apart, not only at a coarse-grain level but with a finer precision as well.

We initially consider two test cases, targeting the most representative high-level activities, namely *work* and *leisure*. Models for the two activities are inferred from the trajectories of the 10 selected users, labelled as described earlier. If our similarity measure is sound and captures the underlying semantics of the indicated activities, we would expect any trajectory labelled as *work* to show a higher degree of similarity to any other trajectory with identical label than to any one labelled otherwise. This is indeed confirmed by inspection of the confusion matrix reported in the leftmost part of Fig. 9, where darker colors indicate higher similarity. The matrix clearly shows that, besides reporting full self-similarity, homogeneous activities show a higher similarity score even for different users than heterogeneous activity. This indicates that the proposed measure is reliable, and indeed captures the nature of the activity regardless of who actually performed it.

A second test case was then considered, regarding the ability of the HAR algorithm to refine its outcome. Specifically, we considered the sub-activities of *leisure*: shopping, spare time, social, and sport. In this case, trajectories for all users were considered together when tagged with the same label, and the smaller matrix on the right side of Fig. 9 shows the corresponding results, grouped by activity. Again, the fact that similarities between models for different sub-activities are low indicates that they can be reliably distinguished from each other. However, a closer look at the models producing the extremely low scores for the *shopping* activity revealed that the corresponding model was much more complex than the others (it was in fact a 60-states DFA). This indicates overfitting, and might be a sign of imprecisions during the tagging of the trajectories.

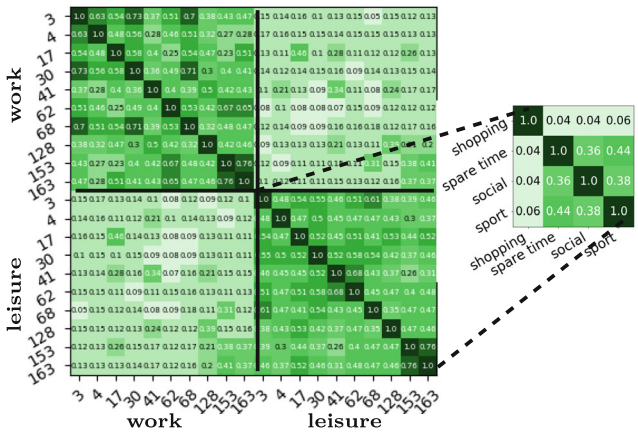


Fig. 9 Similarity for the two most frequent coarse-grained activities of 10 users

Table 4 Similarity of test user against reference models

		Reference		
		Work	Leisure	Travel
Test	W	0.43	0.23	0.08
	Leisure	0.18	0.37	0.17
	Travel	0.05	0.17	0.31

It is worth noting that our results are in accordance to what reported in [9], whose authors computed user profiles based on the same dataset as ours. In their experiments, they calculated the similarity of two sets of activities depending on whether they were performed on weekdays or weekends. Assuming that work activities are mostly performed during weekdays, whereas weekends are typically reserved for leisure, the results reported in Fig. 9 show that the highest pairwise similarities are between users 3-4, as regards *work*, and users 153-163, for *leisure*, which corresponds to the finding in [9].

Finally, our last experiment was conducted by building reference models for all coarse-grained activities, using one of the users as test. The objective is to show how our method might automatically label the activity of a new user, by assessing the similarity score with respect to the reference models. The results reported in Table 4 show that the test models consistently receive a similarity score that associates them with the correct reference model. In other words, the label for the trajectories of the test user could be inferred by assessing the similarity with the reference models, which could be useful for instance in the context of a recommender system application.

6 Conclusions and on-going work

In this work, we presented a method for the recognition of human activities from mobility traces acquired through wearable devices, such as GPS loggers and smart phones. The novelty of our approach lies both in the use of syntactical models to represent user activities, and in the definition of a suitable measure to capture the similarity

among such models by leveraging on their intrinsic hierarchical nature. Our experiments show that the proposed grammatical models are able to accurately discriminate between mobility patterns arising as a consequence of such coarse-grained activities as work, leisure, or travel. Moreover, the same models are suitable to refine the classification and provide a finer distinction into sub-activities. The fact that the very nature of human mobility is hierarchical is mirrored in our formulation for the similarity measure: even though different activities may appear similar in a broader context, we are able to selectively refine the measure and provide a more realistic score by considering lower granularities.

There are, however, open issues worth of further analysis. First of all, more reliable ground truth is likely to produce more accurate classifiers; we plan to refine the RoI-based tagging algorithm and to show the generality of our method in other contexts not necessarily related to outdoor mobility. For instance, besides recognizing common activities, or method could be profitably used to detect anomalous behavior. Moreover, we plan to test the method on additional datasets, possibly expanding the taxonomy of the considered activities, for instance by refining the *work* category further.

Finally, we are investigating alternative methods for the inference of grammatical models, in particular with reference to *active learning* [17]. This paradigm is based on the assumption that an informant, or oracle, may be used to guide inference by a process of queries and assessment. One of the most interesting features is that learning does not need to rely on negative samples, whose selection is usually the weakest part of passive learning methods. In particular, we plan to investigate how an oracle may be constructed by adapting traditional machine learning methods (such as support vector machines, or deep learning algorithms) to our mobility scenario.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Aggarwal JK, Ryoo MS (2011) Human activity analysis: a review. ACM Comput Surv 43(3):1–43
2. Akter SS, Holder LB, Cook DJ (2018) Activity recognition using graphical features from smart phone sensor. In: Proc of the International Conference on Internet of Things, Springer, pp 45–55, ISBN 978-3-319-94370-1
3. Balkic Z, Sostaric D, Horvat G (2012) Geohash and UUID identifier for multi-agent systems. In: Proceedings of the KES International Symposium on Agent and Multi-Agent Systems: Technologies and Applications, Springer, pp 290–298
4. Bao L, Intille SS (2004) Activity recognition from user-annotated acceleration data. In: PERVASIVE 2004, Vol 3001, pp 287–304
5. Bhattacharya A, Das SK (2002) Lezi-update: an information-theoretic framework for personal mobility tracking in pcs networks. Wireless Networks (Special Issue on selected papers from ACM Mobicom '99 papers) 8(2/3):121–135
6. Black K (2011) Business statistics: for contemporary decision making. Wiley, Hoboken
7. Blumrosen G, Miron Y, Intrator N, Plotnik M (2016) A real-time Kinect signature-based patient home monitoring system. In: Sensors
8. Chen L, Nugent CD, Wang H (2012) A knowledge-driven approach to activity recognition in smart homes. IEEE Trans Knowl Data Eng 24(6):961–974
9. Chen X, Pang J, Xue R (2014) Constructing and comparing user mobility profiles. ACM Trans Web 8(4):21
10. Choudhury T, Borriello G, Consolvo S, Haehnel D, Harrison B, Hemingway B, Hightower J, Klasnja P, Koscher K, LaMarca A, Landay JA, LeGrand L, Lester J, Rahimi A, Rea A, Wyatt D (2008) The mobile sensing platform: an embedded activity recognition system. IEEE Pervasive Comput 7(2):32–41
11. Chow TS (1978) Testing software design modeled by finite-state machines. IEEE Trans Softw Eng 4(3):178–187
12. Cook D (2012) Learning setting-generalized activity models for smart spaces. IEEE Intell Syst 27(1):32–38
13. Cottone P, Gaglio S, Lo Re G, Ortolani M (2016) Gaining insight by structural knowledge extraction. In: Proceedings of ECAI European Conference on Artificial Intelligence, vol 285, pp 999–1007
14. Cottone P, Ortolani M, Pergola G (2016) Detecting similarities in mobility patterns. In: Proceedings of the 8th European Starting AI Researcher Symposium (STAIRS 2016), pp 167–178
15. da Penha Natal I, de Avellar Campos Cordeiro R, Garcia ACB (2017) Activity recognition model based on GPS data, points of interest and user profile. In: International Symposium on Methodologies for Intelligent Systems, Springer, pp 358–367
16. De D, Bharti P, Das SK, Chellappan S (2015) Multimodal wearable sensing for fine-grained activity recognition in healthcare. IEEE Internet Comput (Special Issue on Small Wearable Internet) 19(5):26–35
17. de la Higuera C (2005) A bibliographical study of grammatical inference. Pattern Recogn 38(9):1332–1348
18. de la Higuera C (2010) Grammatical inference: learning automata and grammars. Cambridge University Press
19. Dernbach S, Das B, Krishnan NC, Thomas BL, Cook DJ (2012) Simple and complex activity recognition through smart phones. In: Eighth International Conference on Intelligent Environments, pp 214–221
20. Furletti B, Cintia P, Spinsanti L (2013) Inferring human activities from GPS tracks. In: Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing
21. Gaglio S, Re GL, Morana M (2015) Human activity recognition process using 3-d posture data. IEEE Trans on Human-Machine Systems 45:586–597
22. Giannotti F, Nanni M, Pinelli F, Pedreschi D (2007) Trajectory pattern mining. In: Proceedings of the 13th ACM SIGKDD Intl Conf on Knowledge Discovery and Data Mining
23. Gonzalez MC, Hidalgo CA, Barabási A-L (2008) Understanding individual human mobility patterns. Nature 453:779–782
24. Kim E, Helal S, Cook D (2010) Human activity recognition and pattern discovery. Pervasive Computing, IEEE 9(1):48–53
25. Lara OD, Labrador MA (2013) A survey on human activity recognition using wearable sensors. IEEE Commun Surv Tutorials 15(3):1192–1209

26. Leem CS, Kim BG (2013) Taxonomy of ubiquitous computing service for city development. Pers Ubsiquit Comput 17(7):1475–1483
27. Leo M, D'Orazio T, Gnoni I, Spagnolo P, Distante A (2004) Complex human activity recognition for monitoring wide outdoor environments. In: Proceedings of the 17th IEEE International Conference on Pattern Recognition, vol 4, pp 913–916
28. Liono J, Abdallah ZS, Qin AK, Salim FD (2018) Inferring transportation mode and human activity from mobile sensing in daily life. In: Proceedings of the 15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, ACM, pp 342–351
29. Nguyen NT, Phung DQ, Venkatesh S, Bui H (2005) Learning and detecting activities from movement trajectories using the hierarchical hidden markov model. In: IEEE Conference on Computer Vision and Pattern Recognition, vol 2, pp 955–960
30. Nikolić M (2012) Measuring similarity of graph nodes by neighbor matching. Intelligent Data Analysis 16(6):865–878
31. Oncina J, García P (1992) Identifying regular languages in polynomial time. Advances in Structural and Syntactic Pattern Recognition 5(99-108):15–20
32. Rault T, Bouabdallah A, Challal Y, Frédéric M (2017) A survey of energy-efficient context recognition systems using wearable sensors for healthcare applications. Pervasive Mob Comput 37:23–44
33. Ryoo MS, Aggarwal JK (2009) Semantic representation and recognition of continued and recursive human activities. Intl Journal of Computer Vision
34. Saguna S, Zaslavsky A, Chakraborty D (2013) Complex activity recognition using context-driven activity theory and activity signatures. ACM Transactions on Computer-Human Interaction 20(6):1–34
35. Saini R, Kumar P, Roy PP, Dogra DP (2018) A novel framework of continuous human-activity recognition using kinect. Neurocomputing 311:99–111
36. Sebban M, Janodet J-C, Tantini F (2004) Blue*: a blue-fringe procedure for learning dfa with noisy data. In: Proceedings of the Int Conf on Genetic and Evolutionary Computation
37. Song C, Qu Z, Blumm N, Barabási A (2010) Limits of predictability in human mobility. Science 327(5968):1018–1021
38. Turaga P, Chellappa R, Subrahmanian VS, Octavian U (2008) Machine recognition of human activities: a survey. IEEE Trans Circuits Syst Video Technol 18(11):1473–1488
39. Vaizman Y, Ellis K (2017) Recognizing detailed human context in the wild from smartphones and smartwatches. IEEE pervasive computing
40. Varkey JP, Pompili D, Walls TheodoreA (2012) Human motion recognition using a wireless sensor-based wearable system. Pers Ubiquit Comput 16(7):897–910
41. Walkinshaw N, Bogdanov K (2013) Automated comparison of state-based software models in terms of their language and structure. ACM Trans Softw Eng Methodol 22(2):13
42. Yang X, Tian YL (2017) Super normal vector for human activity recognition with depth cameras. IEEE Trans Pattern Anal Mach Intell 39(5):1028–1039
43. Younes R, Jones M, Martin T (2018) Classifier for activities with variations. Sensors 18(10):3529
44. Yu Z, Liu L, Wang L, Xie X (2008) Learning transportation mode from raw gps data for geographic applications on the web. In: Proc of the 17th Int Conf on world wide web, pp 247–256

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.