Noname manuscript No.
(will be inserted by the editor)

Extracting Spatiotemporal Human Activity Patterns in Assisted Living using a Home Sensor Network

Dimitrios Lymberopoulos · Athanasios Bamis · Andreas Savvides

Received: date / Accepted: date

Abstract This paper presents an automated methodology for extracting the spatiotemporal activity model of a person using a wireless sensor network deployed inside a home. The sensor network is modeled as a source of spatiotemporal symbols whose output is triggered by the monitored person's motion over space and time. Using this stream of symbols, we formulate the problem of human activity modeling as a spatiotemporal pattern-matching problem on top of the sequence of symbolic information the sensor network produces and solve it using an exhaustive search algorithm. The effectiveness of the proposed methodology is demonstrated on a real 30-day dataset extracted from an ongoing deployment of a sensor network inside a home monitoring an elder. Our algorithm examines the person's data over these 30 days and automatically extracts the person's daily pattern.

Keywords Human activity model · spatiotemporal activity patterns · sensor networks

D. Lymberopoulos
Microsoft Research
One Microsoft Way
Redmond, WA
E-mail: dlymper@microsoft.com

A. Bamis
Yale University
Electrical Engineering
New Haven, CT
E-mail: athanasios.bamis@yale.edu

A. Savvides
Yale University
Electrical Engineering
New Haven, CT
E-mail: andreas.savvides@yale.edu

1 Introduction

The growing numbers of aging baby boomers and the increasing healthcare cost [30],[13],[31] obviate the need for automated services that will increase the independence and autonomy of elders living at home. Wireless sensor networks offer a promising technology for realizing such services. On one hand, small wearable devices can collect biometric information, provide feedback and automatically update medical records. On the other hand, other devices deployed inside the living environment, can monitor the actual people over space and time, understand their activities/behaviors and provide responsive services to them.

In this paper we explore the problem of creating a human activity model from data collected by a sensor network deployed inside a home. We do so by deploying sensors in the house infrastructure without requiring the inhabitant to wear any sensors. Our reasoning for constructing the models is based on the fact that human activity is a sequence of actions over space and time. All humans execute a daily cycle in which many activities are periodic and elders living alone in particular, tend to have a highly periodic set of routines that they execute from day to day.

As the monitored person moves from room to room inside the house, a sequence of detected sensing features is produced over time. This sequence represents the monitored person's activity signature and its composed of a set of triplets containing location, time and duration information. To better illustrate this, consider the 7-day data trace, we have acquired through an actual sensor network deployment in an elder's person house, shown in Figure 1. By simply inspecting the sequence of rooms that the person visited over time inside the house, it is clear that patterns, strongly related

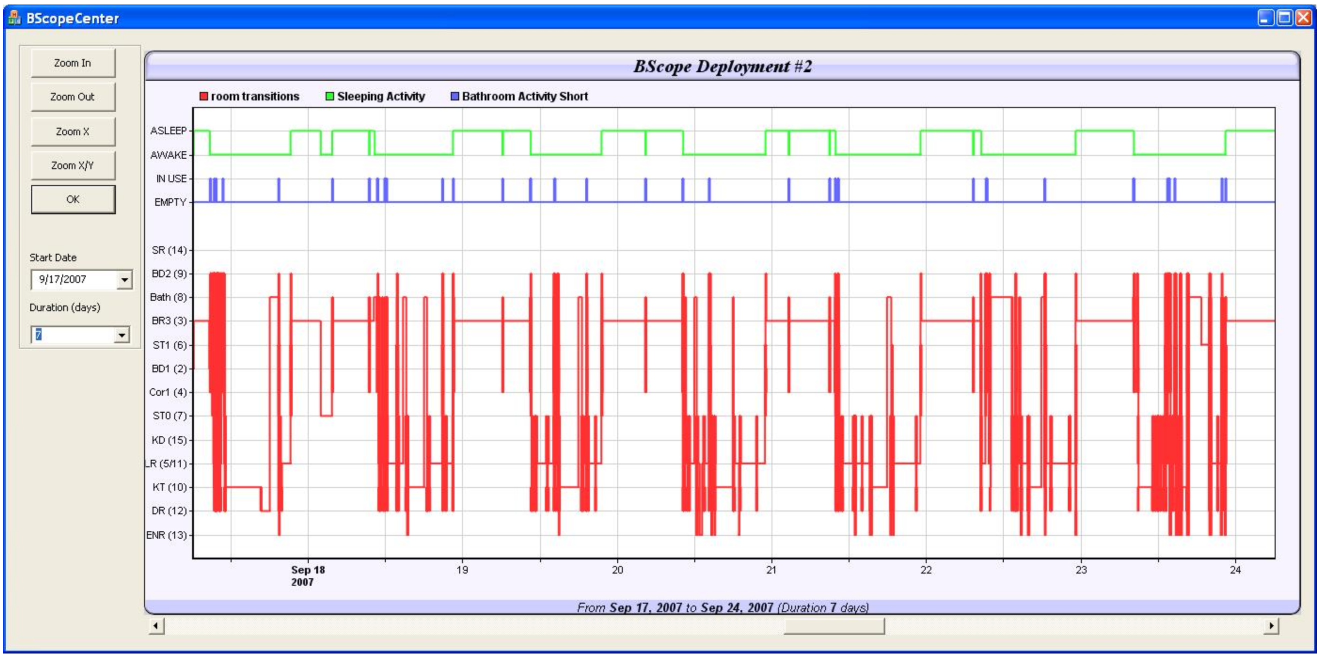

Fig. 1 7-day window of the data trace recorded from our home network deployment.

to the person's activities, start emerging. The sleeping pattern, the bathroom visits, the time that the person is not home and many more patterns provide invaluable information about the person's daily living habits.

The main contribution of this paper is the development of a methodology for automatically identifying the activity model of a person (like the one shown in Figure 9) using a wireless sensor network. First, we model the sensor network as a spatiotemporal symbol generator that is triggered by the monitored person as he moves over space and time. Based on our network model, we formulate the problem of finding the daily activity model of a person as the problem of finding the most probable, network-level, sequences of node-level, sensing features, namely location, time and duration. We propose a methodology for encoding the detected sensing features (location, time and duration) in a way that allows us to apply an exhaustive, yet very efficient, algorithm for automatically discovering sequential patterns based on how frequently they appear in a given data trace. The effectiveness of the proposed methodology is demonstrated using a data trace collected by an actual sensor network deployment of Passive Infrared sensors (PIR) in the house of an elder person living alone for a period of 30 days. Our results, show that: 1) there is a daily activity pattern and 2) we can automatically generate this daily activity model while taking into account both its spatial and temporal characteristics.

The rest of the paper is organized as follows. Section 2 describes the motivation and contributions of this work and Section 3 provides an overview of the related work. In Section 4 we present our testbed along with the type of information it records and we describe in detail the proposed network model. In Section 5 we formulate the problem of human activity model generation as a human spatiotemporal pattern discovery problem. Section 6 describes an exhaustive, yet efficient, pattern finding algorithm and Section 7 provides a novel, simple and scalable way for jointly considering space and time information in the pattern discovery process. In Section 8 we present the results of applying the proposed method on a 30-day dataset, recorded from a deployed sensor network in an actual elder's person. Section 9 concludes the paper.

2 Motivation and Contributions

Our approach to human activity modeling is motivated by, but not limited to, the assisted living application domain. Elder people living alone need continuous medical attention and they currently face a tremendous dilemma. To guarantee they receive the medical attention they need, they have to sacrifice their independence and personal social life by either living in nursing homes or spending a significant fraction of their time in a specialized hospital for continuous monitoring. Not only does this reduce their independence but it also affects, in a negative way, their psychological status. An intelligent sensor network could be used to completely eliminate this dilemma. It could guarantee that elder people living alone receive the medical attention they need, while maintaining their independence and freedom. As

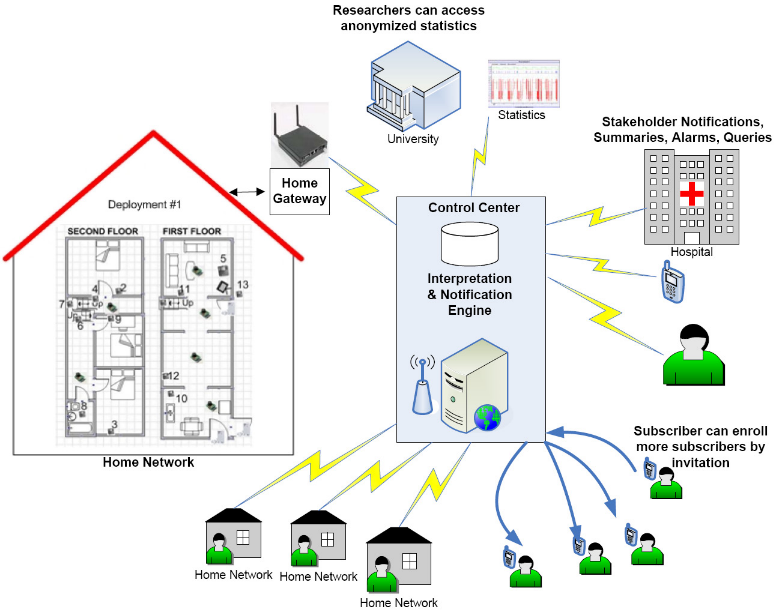

Fig. 2 An intelligent sensor network for continuously monitoring elder people living alone.

it can be seen in Figure 2, a wireless sensor network could be used to continuously monitor the elder person living alone as he moves over space and time inside the house. Based on the interpretation of the sensing data, the network could be used to guard against unsafe activities, post reminders, automate tasks and even initiate conversation with the monitored person. Instead of passively monitoring the home environment, the network could actively use the sensing information to enforce a set of rules such as: make sure that the person is engaging(or not engaging) into the activities that his doctor recommends (or does not recommend), predict the activities that will be performed next and automatically prepare the house for them, detect unusual or abnormal activity and notify the necessary medical personnel and/or family and close friends with detailed information about the status of the monitored person over the last few minutes, hours or days, and more. At the same time, immediate family members and medical personnel would be able to configure and re-task the network, in order to adjust the monitoring process and customize it to the needs of the elder person over time.

The work presented here is one of the steps we have taken towards this direction [16–18]. In particular, the contributions of this paper are the following:

1. We propose a methodology for properly encoding continuous time information in the raw sensing data into discrete spatiotemporal symbols. Our approach allows us to keep the size of the input data set intact while encoding all the necessary information (spatial and temporal) we need to accurately extract activity information.
2. We formulate the problem of extracting human activities from raw sensing data as a sequential pattern search problem and apply an exhaustive, yet very efficient, algorithm to automatically discover sequential patterns based on how frequently they appear over a given time window.
3. We evaluate the proposed methodology using one of our home sensor network deployments. A combination of camera and PIR sensor nodes were used to continuously monitor an elder person living alone for 30 days. After applying our methodology on the 30-day sensor stream we show that the monitored person has a daily activity cycle that we are able to accurately and automatically discover.

3 Related Work

The problem of human activity recognition has been considered in several different domains [11, 22, 20, 23, 14] including wireless sensor networks [10, 25, 29, 24, 1, 5–7]. These approaches usually focus on the robust detection of a single activity either through specification or typical learning techniques on top of labeled data. Our work differs in the sense that *(1)* it provides a general method for discovering multiple activities given a large data trace that contains unlabeled data and *(2)* it takes special care of both spatial and temporal properties of the different activities.

Our previous work in hierarchical human activity recognition based on user-defined probabilistic context-free grammars [16,17] is complimentary to the work presented here. Instead of relying on a user-defined specification of an activity as it was done in previous papers [16], in this chapter we automatically extract the activity as a spatiotemporal pattern given a recorded data set that might contain one or more activities. The two approaches are complimentary in the sense that the automatically discovered patterns could be used to refine the user-defined activity specifications so that more robust activity detection is achieved.

The most closely related work to ours is the work done by Agrawal at. al [2,3]. There, the notion of frequent itemset and frequent sequential pattern discovery in a sequence of events is introduced. Based on the *a-priori* principle, Agrawal was the first to propose efficient algorithms for discovering spatiotemporal sequences of events in large event sequences [3]. Since then, several research efforts led to minor [28,9,27,21] or major [19] modifications of these algorithms. *The major difference in our work lies on the way we extract, encode and use time information in the pattern discovery phase.* In particular, we propose a novel clustering approach for automatically extracting the temporal characteristics of events that can later be used to encode time information on a per-event basis. By doing so, we can combine the efficiency of the a-priori based sequential pattern mining algorithms with the flexibility of exploiting different temporal characteristics on different input events.

Our work also differs from previous research efforts [9,27,4,12,21,26] in the sense that we focus on discovering sequential patterns that are closely related to human activities and not abstract, statistical correlations of events in the input sequence that might not be directly mapped to human activities.

4 Deployment and Network Model

The data considered for this work comes from an ongoing sensor network deployment that monitors an elder living alone. The testbed includes a wide variety of sensors including tracking cameras, door sensors and passive infrared sensors. To derive the activity models presented in this paper we only used the PIR measurements. Every room in the house contains PIR sensors placed in a pattern that can capture the elder's transitions from room to room. Each time a sensor gets triggered it transmits its ID to a home gateway that timestamps and records the sensor ID and uses the data to compute a room-transition function. A 7-day time window of the room transition plot from the actual testbed is shown in Figure 1. From this high level view it is already apparent that the daily activity of the person under observation has regular recurring patterns. The method proposed in this paper will be applied to the complete dataset to extract these recurring patterns and construct a model of the person's daily pattern.

Based on our deployment, the network generates a sequence of triplets of the form: $\{P, T, D\}$ where:

- P: is the phoneme detected by the sensor node (a room identifier for this discussion).
- T: is the actual timestamp at which this phoneme was detected.
- D: is the duration of the phoneme.

For instance, each horizontal line in Figure 1 corresponds to a specific area or room visited by the monitored person inside the house. Vertical lines simply denote the transition from a specific room/area to a new one. Note that each horizontal line encodes exactly three pieces of information: the location of the person at a given time instant denoted by its position on the y-axis (P), the exact time the person visited the location denoted by the start of the line (T), and of course the duration of this visit denoted by the length of the line (D). In essence, every line corresponds to a spatiotemporal triplet $\{P, T, D\}$.

As the person moves over space and time, sequences of such triplets are generated at run-time. Therefore, the output of sensor node i over time will be a time ordered sequence of triplets S^i. Assuming that in a given time frame, sensor node i has generated N_i triplets, its output can be denoted as follows:

$$S^i = <\{P_1^i, T_1^i, D_1^i\}, \{P_2^i, T_2^i, D_2^i\}, \dots, \{P_{N_i}^i, T_{N_i}^i, D_{N_i}^i\}>$$
$$, where : T_1^i < T_2^i < \dots < T_{N_i}^i$$

Consequently, the output O over time of a sensor network with n nodes becomes a collection of such time ordered sequences of triplets:

$$O = \{S^1, S^2, \dots, S^n\},$$
$$|S^1| = N_1, |S^2| = N_2, \dots, |S^n| = N_n$$

where S^i is the time ordered output sequence at node i, containing N_i triplets. For instance, in Figure 1, the waveform at the bottom represents the output O of the sensor network while the two waveforms on the top represent the outputs of the nodes monitoring the bedroom and the bathroom respectively.

We define an *episode* $E(T_{start}, T_{stop})$ as the time-ordered sequence of all triplets in the output sequences of every node in the network that were recorded between T_{start} and T_{stop}. Formally, the episode $E(T_{start}, T_{stop})$ is defined as follows:

$E(T_{start}, T_{stop}) = \{(P_j^i, T_j^i, D_j^i) | (P_j^i, T_j^i, D_j^i) \in S^i,$
$T_{start} \leq T_j^i \leq T_{stop} \forall i, j\}$

Note that each episode is nothing more than a temporal clustering of triplets that might be recorded to one or more sensor nodes. Given the definition of episodes we can express the output O of the sensor network as a collection of episodes:

$$O = \{E(T_1, T_2), E(T_2, T_3), E(T_3, T_4), ...\} \quad (1)$$

To better illustrate our network model, let us consider the simple night/morning activity of the elder person monitored in our deployment. According to the data trace collected, the elder person will go to sleep around 11pm, wake up to go to the bathroom in the middle of the night and then return to sleep, then wake up again in the morning to visit the bathroom and then visit the kitchen to have breakfast. Given that the basic phonemes generated by our sensor network are rooms, a typical sequence of time-ordered phoneme triplets would be the following(duration is in minutes):

$< \{Bed, "11:00pm", 300\}, \{Bath, "4:00am", 5\},$
$\{Bed, "4:05am", 300\}, \{Bath, "9:05am", 10\},$
$\{Kitchen, "9:15am", 30\} >$

The above sequence represents an instance of the night/morning activity pattern of the elder person. If we define an episode as the time-ordered sequences of all phoneme triplets that take place between 10pm and 10am every day, then different episodes would correspond to different instances of the same activity pattern:

$E^1(10pm, 10am) =< \{Bed, "11:00pm", 300\},$
$\{Bath, "4:00am", 5\}, \{Bed, "4:05am", 300\},$
$\{Bath, "9:05am", 10\}, \{Kitchen, "9:15am", 30\} >$

$E^2(10pm, 10am) =< \{Bed, "10:30pm", 270\},$
$\{Bath, "3:00am", 3\}, \{Bed, "3:03am", 330\},$
$\{Kitchen, "8:33am", 20\} >$

$E^3(10pm, 10am) = \ldots$

Note, that in different episodes the absolute time and duration characteristics of the sequences or even the sequences themselves might be different.

5 Extracting Activity Patterns from Data Sequences

The output O of the sensor network is a collection of triplet sequences over time that are temporally organized in episodes. Each episode encodes the spatiotemporal activity of the monitored person in a specific time window. Different episodes provide different instances of the monitored person's activity at different points in time. As a result of this, discovering the similarities across a large set of episodes results into discovering the basic activity model of the monitored person. The type of the model depends on the time duration of an episode. For instance, when an episode is defined as a single day, week or month of activity then the process of discovering activity patterns across a large set of episodes corresponds to the daily, weekly or monthly activity model respectively.

In particular, the more frequently a sequence of phoneme triplets appears on a sequence of episodes the more important it is for the activity model. In general, given a sequence of episodes, we define the frequency f_s of a phoneme triplet sequence as:

$$f_s = \frac{N_s^E}{N^E}$$

where N_s^E is the number of episodes where the sequence s appears at least once and N^E is the total number of episodes. Note that: $0 \leq f_s \leq 1$ and therefore the frequency f_s can be seen as the appearance probability of the sequence s in the input sequence of episodes.

Problem Statement: *Given a sequence of episodes generated by a sensor network, find all the sequences s_i of triplets with frequency $f_{s_i} \geq f_{th}$.*

The goal of the above problem statement is to find the *most frequent sequences of triplets* in a given sensor network output. f_{th} is a user-specified threshold that defines what *most frequent* actually means. All sequences with a frequency higher than f_{th} are considered to be frequent. Frequency can be considered as a measure of how often a sequence of triplets appears in the output of the network. The more often a sequence appears the more probable is that this sequence encodes one of the core(most-performed) activities of the monitored person. Being able to find all these core activities will allow us to build the spatiotemporal model of the person's daily, weekly or monthly activity habits. The ability to construct a model with spatiotemporal characteristics lies on the fact that triplets encode both spatial (phoneme P) and temporal (timestamp T and duration D) information. In other words, triplet sequences are nothing more than a spatiotemporal signature of the monitored person as he moves inside the house.

Note, that the sequences we want to find are not necessarily subsets of the sequences $S^i, i = 1, \ldots, n$ produced by a single sensor node. Instead, the sequences we aim to find are subsequences of episodes which in most of the cases include triplets that belong to different node-level sequences S^i. We call these sequences network-level sequences.

6 Human Activity Discovery

To simplify our discussion, in this section, we will ignore the temporal information included in each triplet generated from a sensor node. In the next section we demonstrate how the proposed approach can be transparently used on top of the spatiotemporal triplets.

A brute-force approach to the problem would be to generate all possible sequences of all possible lengths for all the different episodes, then compute the frequency of each sequence and choose those sequences that have a frequency higher than f_{th}. Even though this consists of an exhaustive search over the input that will find all frequent paterns, it requires to first generate a huge list of candidate frequent sequences and then for each one of these candidates we have to parse all the episodes to compute each candidate's frequency. Even worse, as the number of different phonemes increases and/or the number of observed phonemes in an episode increases, the number of candidate frequent sequences explodes.

Our goal is to reduce the total number of candidate frequent sequences before we even try to compute their frequencies while making sure that *all* frequent patterns will be discovered. To deal with this problem, we take advantage of the *apriori* principle [2,3]: *if a sequence is frequent then all of its subsequences must be frequent.* This argument is very similar to the shortest path argument in a network of nodes. The path between any pair nodes that are located on the shortest path between two nodes is also a shortest path. This observation is very important for two reasons. First, it indicates that all the candidate frequent sequences of size L should be generated by the frequent sequences of size $L-1$. This is due to the fact that the supersequence of any non-frequent sequence will also be a non-frequent sequence. Second, after generating the candidate frequent sequences of size L, every sequence that contains a non-frequent subsequence should be automatically eliminated because it cannot be frequent. Based on these two fundamental observations, Agrawal et. al have designed an efficient algorithm, called the a-priori algorithm, for exhaustively searching the input set of episodes to find all the frequent sequences [2,3].

Figure 3 shows the basic steps of this algorithm. First, the frequency of every sequence of length one is computed. In practice, the frequency of every phoneme is computed and the set of frequent phonemes F_1 is formed by choosing all the phonemes with frequency higher than f_{th}. At the next step, the set of frequent phonemes is used to generate the candidate frequent sequences of length two. In general, the algorithm will iteratively generate all candidate frequent sequences of size L using as input the frequent sequences of size $L-1$.

// Find all frequent sequences of size 1
$L = 1$
$F_L = \{s | f_s \geq f_{th}\}$
while$(F_L != \emptyset)$
{ $L = L + 1$
// Generate candidate frequent sequences of size L
$C_L = candidate\_generation(F_{L-1})$
for *every episode E*
{
// Find all candidate frequent sequences
// that appear in episode E
$C_E = find\_sequences(C_L, E)$
for *every sequence s in* C_E
{
$f_s = f_s + 1$
}
}
$F_L = \{s | s \in C_L and f_s \geq f_{th}\}$
}
$Frequent\_Sequences = \bigcup F_L$

Fig. 3 The a-priori algorithm for discovering the frequent sequences in a set of episodes.

This is done in two discrete steps that have been exhaustively studied in the dta mining domain [2,3,28,32,8]:the candidate generation and pruning steps. First (candidate generation step), every frequent sequence of size $L-1$ is expanded by one frequent phoneme. If there are more than one frequent phonemes then every frequent sequence of size $L-1$ generates a candidate frequent sequence of size L for every frequent phoneme. Second (candidate pruning step), the candidate frequent sequences of size L that contain a non-frequent subsequence of size $L-1$ are immediately eliminated and C_L is formed. As soon as C_L is defined, we pass each episode to find which of the sequences in C_L are contained in that episode. Every time that a sequence in C_L is contained in an episode its frequency is increased by one. After we examine all episodes, the set of frequent sequences of size L (F_L) is formed by keeping only those candidate frequent sequences in C_L with frequency higher than f_{th}. This process continues iteratively until the set of candidate frequent sequences is the empty set. The final output of the algorithm consists of all the frequent sequences of different sizes.

Note, that at each iteration of the algorithm the new candidate frequent sequences are generated by the frequent sequences identified in the previous step. In that way, the overall number of sequences for which we have to compute their frequency is drastically reduced at every step. This reduces the number of passes we have to perform over the set of episodes, which in general might be quite large. Hence, the cost of finding the most frequent patterns while making sure that *all* existing frequent patterns will be discovered is dramatically reduced.

7 Handling Space and Time: A Scalable Approach

The methodology presented in the previous section allows for an exhaustive search over the output of the sensor network for discovering sequences of triplets with a frequency that is higher than a predefined threshold. Even though each triplet contains both location and time information (start time and duration) , so far we have only used the time in a primitive way for sorting out the locations produced by the different nodes in the network. As a result of this, the sequences we discover are nothing more than sequences of spatial features over time. It is apparent that essential temporal information included in the triplets is ignored. For instance, consider the simple example where we want to monitor the bathroom usage from an elder person living inside the house. Knowing that the monitored person visited the bathroom is a useful piece of information. However, knowing when this visit took place is even more important; it is normal for an elder person to visit the bathroom regularly during a day and right after he wakes up, but when a bathroom visit is taking place in the middle of the night its meaning might be totally different. In the same sense, the duration of this activity is also very important. For example, a couple of bathroom visits in the evening or during the day might be considered normal activity, however, lengthier bathroom visits over the night can be used to identify abnormal or possibly emergency situations.

It is clear that even though location, time and duration of an event/activity can independently provide useful information, when these features are combined together we can interpret the same event in a totally different way. Given this, the following question arises: *How can we combine the different spatial and temporal information that a triplet provides with the methodology presented in the previous section in a scalable way?* Answering this question is challenging due to the following reasons:

1. Different activities and even different phonemes require different time and duration resolutions. For instance, monitoring bedroom activity requires a duration resolution that can vary from tens of minutes to several hours. In the same sense, bathroom monitoring requires a duration resolution that varies from a few minutes up to tens of minutes. For the proposed sequence pattern discovery method to be efficient enough in terms of discovering informative patterns, we have to concurrently support all these different resolutions.
2. The support of different time and duration resolutions must have the minimum possible impact on the size of the data set over which the proposed method is executed. Increasing the size of the data set can lead to extremely large execution times of the proposed method and therefore limit its applicability.

Current state-of-the-art methods take advantage of user-defined time windows to guide the pattern discovery process in the sequence of episodes [28,19]. These time windows are used to constrain the scope of the pattern search algorithm in terms of the actual time and duration of a pattern. Even though this approach does not increase the size of the input to the search algorithm, it has a significant drawback; it is not flexible enough to adjust to the different temporal characteristics of different patterns. Using a fixed time window in the search algorithm prevent us from efficiently searching for temporal patterns. On the other hand, using variable window sizes can significantly increase the complexity of the pattern search algorithm.

To deal with this issue, we have designed a flexible, user-configured, hierarchical temporal abstraction layer that encodes both the spatial and temporal information of a triplet into a single spatiotemporal symbol/phoneme. Executing the method described in the previous section on top of these symbols allows us to discover spatiotemporal sequences. Figure 4 shows the main components of the proposed temporal abstraction layer. Note that besides the sensor network input of triplets sequence, the user of the system provides two sets of condition parameters [1]. These are used to hierarchically condition the input triplets based on their timestamps and/or their duration. Both parameter sets include user-specified conditions that could be applied to all, a subset or only a specific triplet of the input sequence. The filtering process for identifying which triplets are subject to a condition parameter is done using their phoneme fields. Initially, every triplet in the input sequence that satisfies the filtering criteria will be conditioned on the actual timestamp it was recorded using the user-defined parameters. This process will convert all or a subset of the input triplets: $\{P, T, D\}$ into a tuple: $\{P^T, D\}$, where P^T is the new phoneme name (provided by the user) that embeds both spatial and absolute time information. Note that if a condition parameter for a specific type of phoneme does not exist, then the input triplet is simply converted to a tuple by ignoring the timestamp field of the triplet. At the immediate next level, the tuples that satisfy the filtering criteria of the user-specified duration condition pa-

[1] The start time and duration condition parameters can be either entered directly by the user based on prior information or they can be automatically extracted from the input phoneme triplets using the methodology we presented in [15].

rameters will be conditioned on their duration values. Again, if a condition parameter for a specific type of phoneme does not exist, then the input tuple is simply converted to a spatiotemporal phoneme by ignoring the duration field of the tuple. This process will convert all or a subset of the input tuples: $\{P^T, D\}$ into a single symbol/phoneme: $\{P^{T,D}\}$.

Note that $P^{T,D}$ now embeds spatial, absolute time and duration information into a single spatiotemporal phoneme. Running the method described in Section 6 on the output sequence of spatiotemporal phonemes allows us to identify spatiotemporal patterns without the need to explicitly process absolute time or duration information. In this way, we manage to identify spatiotemporal patterns by simply conditioning the input triplets based on absolute time, duration or on both absolute time and duration. *This approach can be seen as a phoneme renaming process that allows us to keep the size of the input data set intact while encoding all the necessary information we need. The only incurred overhead has to do with increasing the different number of phonemes used, however, this has no effect on the size of the input to the algorithm and therefore on its complexity. In addition, it provides the necessary flexibility to the user to apply different condition parameters at different phonemes or even different condition parameters at the same phoneme according to the requirements of different activities. Hence, the proposed scheme scales well with both the size of the input data set as well as with the number of activities and phonemes we want to exploit.*

To demonstrate how this temporal abstraction operates, consider the following input sequence of room, time and duration triplets (duration is expressed in minutes):

$$< \{Bed, "11:00pm", 510\}, \{Bath, "7:30am", 25\},$$
$$\{Bed, "7:55am", 15\}, \{Bath, "8:10am", 5\},$$
$$\{Kitchen, "8:15am", 30\} >$$

This sequence shows a typical morning activity. The person went to sleep at $11pm$, slept for $8,5$ hours, then woke up, took a shower, then returned to the bedroom to get dressed, quickly visited the bathroom and finally went to the kitchen to get breakfast. A possible set of absolute time condition parameters in this case could be the following:

Phonemes	Time Range	Spatiotemporal Phoneme
{Bed}	("8:00pm","12:30am")	{Bed_Night}
{Bed}	("5:00am","11:00am")	{Bed_Morning}
{Bath}	("5:00am","11:00am")	{Bath_Morning}

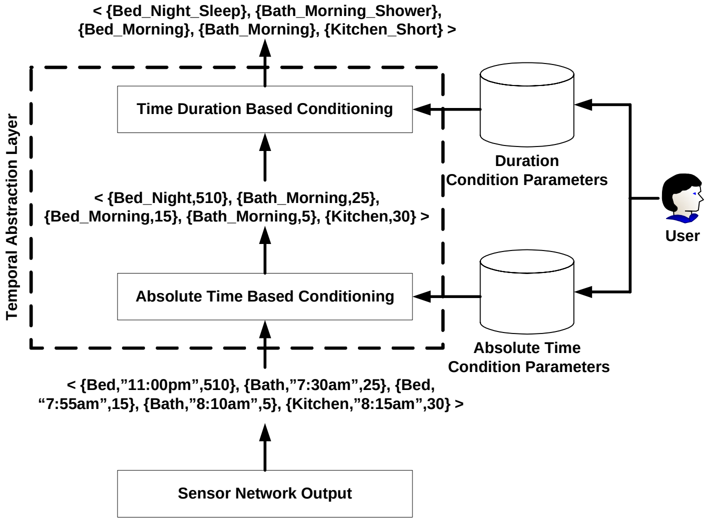

Fig. 4 Outline of the temporal abstraction layer.

After applying this conditioning to the input sequence of triplets we get the following sequence of tuples:

$$< \{Bed\_Night, 510\}, \{Bath\_Morning, 25\},$$
$$\{Bed\_Morning, 15\}, \{Bath\_Morning, 5\}, \{Kitchen, 30\} >$$

Note that all absolute time references have been eliminated and the $Kitchen$ phoneme remained the same since there was no conditioning parameters defined for it. At the immediate next step the set of duration condition parameters is applied. Such a simple set of parameters can be seen next:

Phonemes	Duration Range	Spatiotemporal Phoneme
{Bed}	(120,700)	{Bed_Sleep}
{Bed_Night}	(120,700)	{Bed_Night_Sleep}
{Bath}	(15,30)	{Bath_Shower}
{Bath_Morning}	(15,30)	{Bath_Morning_Shower}
{Kitchen}	(5,40)	{Kitchen_Short}

By applying this set of duration condition parameters on the input sequence of tuples we get the following sequence of spatiotemporal symbols:

$$< \{Bed\_Night\_Sleep\}, \{Bath\_Morning\_Shower\},$$
$$\{Bed\_Morning\}, \{Bath\_Morning\}, \{Kitchen\_Short\} >$$

Note that each symbol in the new output sequence encodes absolute time, location and duration information. For instance, the *Bath_Morning_Shower* indicates a bathroom visit that took place in the morning with a duration large enough to indicate shower activity. In the same way, the *Bed_Night_Sleep* phoneme provides information of a sleeping activity that took place during the night. The time and duration condition parameters for the different phonemes can be either specified by the user or extracted for exploiting the statistics of the raw sensing data as it is demonstrated in Section 8.2.

Due to the semantic information embedded in these symbols, we can use the proposed method for finding way more fine grained, and thus more informative, sequences over space and time. In that way, instead of modifying the algorithm described in Section 6 to explicitly process time information, something that could significantly increase its complexity and would poorly scale with the size of the input data set, we encode the spatiotemporal information into the input phonemes allowing us to implicitly discover spatiotemporal patterns.

8 Experimental Results

We evaluated the proposed spatiotemporal, activity-based frequent pattern mining method on a 30-day data trace collected using the home sensor network deployment described in Section 4. A network of 15 PIR sensors was used to monitor the occupancy of all the different rooms in the house. The set of *phonemes* in this case became the actual rooms that were visited by the monitored elder person over time and every such phoneme was associated with an actual timestamp and a duration interval. To better highlight the value of the proposed approach and to simplify our discussion, we opted to operate on a slightly processed sequence of the visited rooms. In particular, we map the generated sequence of rooms into a sequence of primitive activities by applying a simple set of rules. This process has been successfully demonstrated in our previous work [16–18]. Thus, the actual input phonemes become the different activities: *Sleep*, *Bath*, *Breakfast*, *GetReady*, *Hangout*, and *Out*, that provide information about when the person sleeps, visits the bathroom, has breakfast, gets ready for the day, spends time in the living-room watching TV and when he is out of the house respectively. This mapping is equivalent to the transformation of the raw sensing data (red waveform at the bottom) to the simple activity waveforms (green and blue waveforms on the top) shown in Figure 1.

8.1 Spatial Pattern Discovery

We applied the proposed method on the 30-day sequence of primitive activities, to extract the daily living model of the monitored person. Since, we were interested in the daily living model, we defined the duration of an episode to be the duration of a single 24-hour day. In that way, the input sequence of activities was expressed as a sequence of 30 episodes, where each episode contained an ordered sequence of the five different primitive activities. First, we applied our method while ignoring the temporal characteristics of the input

	Sequence	**Probability**
1	<Sleep,Bath>	68.4%
	<Sleep,Bath,Sleep>	26.3%
	<Sleep,Bath,Sleep,Bath>	21%
	<Sleep,Bath,Breakfast>	47.3%
	<Sleep,Bath,Breakfast,Hangout>	36.8%
2	<Bath,Sleep,Bath>	68.4%
	<Bath,Sleep,Bath,Breakfast>	47.3%
	<Bath,Sleep,Bath,Breakfast,Hangout>	36.8%
	<Bath,Hangout,Out>	42.1%
3	<Breakfast,Hangout>	73.6%
	<Breakfast,Hangout,Bath>	42.1%
	<Breakfast,Hangout,Bath,Hangout,Out>	26.3%
4	<Hangout,Sleep>	26.3%
	<Hangout,Bath,Sleep>	47.3%
	<Hangout,Bath,Hangout>	52.6%
	<Hangout,Out>	84.2%
	<Hangout,Out,Hangout>	57.8%
	<Hangout,Out,Hangout,Bath,Sleep>	26.3%
	<Out,Hangout>	79%

Fig. 5 A subset of the most frequent spatial sequential patterns discovered in the collected trace.

sequence (e.g. no time or duration conditioning was applied). A subset of the extracted frequent sequential patterns can be seen in Figure 5. To facilitate the interpretation of the patterns, we have organized all the similar patterns into chronologically ordered groups. Even though no temporal information was considered, the daily living model of the elder person begins to emerge. From the *Sleep* and *Bath* sequences (pattern groups 1 and 2 in Figure 5), one can see that the sleeping pattern of the monitored person consists of more than one *Sleep* and *Bath* activities. After the person wakes up (alternating sequence of *Sleep* and *Bath* activities), he will have breakfast and then spend most of his time in the living-room watching TV (pattern group 3 in Figure 5). Besides some bathroom visits, the person continues to hangout in the living area until he eventually gets out of the house. After the person returns, he will continue to hangout in the living-room and occasionally visit the bathroom. The day ends by visiting the bathroom for one last time before going to bed (pattern group 4 in Figure 5).

8.2 Spatiotemporal Pattern Discovery

While the patterns in Figure 5 provide basic information about the monitored person's daily living habits, they lack significant temporal information. Without absolute time or duration information the importance of the discovered patterns degrades. For instance, knowing that the person wakes up regularly to visit the bathroom is useful information but it would be even more informative if we knew when this happens (in the middle of the night or in the morning) and if it is periodic or not. In addition, it is important to know that the person leaves the house but it is even more important to know when and for how long. To highlight the im-

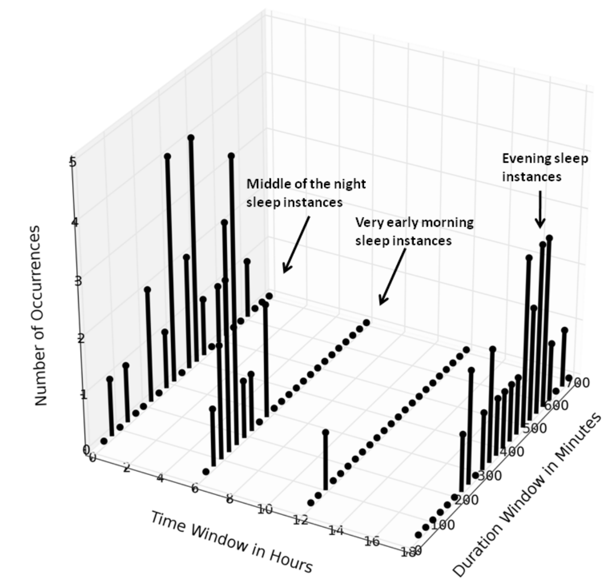

Fig. 6 Time and duration characteristics of the Sleep activity. Time is divided into 4 time windows of 6 hours duration each. Duration is divided into 30-minutes windows. The z-axis represents the number of times the Sleep activity appeared in a specific time and duration window.

	Sequence	Probability
1	<Sleep_N_Long,Bath_M>	21%
	<Sleep_M_Long,Bath_M>	31.6%
	<Sleep_M_Short,Bath_M>	21%
2	<Bath_E,Sleep_E_Long>	52.6%
	<Bath_N,Sleep_N_Long,Bath_M>	21%
	<Bath_M,Sleep_M_Long,Bath_M>	31.6%
	<Bath_M,Sleep_M_Short,Bath_M>	21%
	<Bath_M,Breakfast_M_Long,Hangout_M_Long>	21%
	<Bath_M,Breakfast_M_Short>	26.3%
	<Bath_A,Hangout_A_Long,Out_A_Long>	21%
3	<Breakfast_M_Long,Hangout_M_Long>	52.6%
	<Breakfast_M_Long,Hangout_M_Long,Bath_A>	31.6%
4	<Hangout_M_Long,Bath_A,Hangout_A_Long>	21%
	<Hangout_M_Long,Out_A_Long,Hangout_E_Long>	26.3%
	<Hangout_E_Long,Bath_E,Sleep_E_Long>	42.1%
	<Hangout_E_Long,Sleep_E_Long>	21%
5	<Out_A_Long,Hangout_E_Long>	42.1%
	<Out_A_Long,Hangout_E_Long,Bath_E,Sleep_E_Long>	21%

Fig. 7 A subset of the most frequent sequential patterns discovered in the collected trace when both time and duration conditioning has been applied. The extensions '_M", '_A", '_E", and '_N" stand for morning, afternoon, evening and night.

portance of time and duration information in the pattern discovery process consider Figure 6 where the time and duration characteristics of the *Sleep* activity are shown. Figure 6 shows the number of times the *Sleep* activity appears in a specific time window during a day (4 time-windows of 6 hours duration each) and with a specific duration. Note, that the total number of the *Sleep* instances are more than the total number of days in the recorded data trace. This clearly shows that the sleeping pattern of the elder person consists of multiple *Sleep* instances due to regular bathroom visits. Also, from the time windows where the *Sleep* instances appear we can infer when the person goes to sleep (last time window on the right) and then interrupts his sleep to visit the bathroom (first two time windows on the left correspond to bathroom visits in the night and in the morning respectively). The fact that sleep is interrupted by frequent bathroom visits can also be seen by the duration of the different *Sleep* instances. In general a *Sleep* activity can last from approximately one hour up to approximately 10 hours, however, in most of the cases the duration is limited between 2 and 4 hours. Figure 6 also shows the correlation between the absolute time when the *Sleep* activity happens and its duration. *Sleep* activities at different time-windows have different duration characteristics. As Figure 6 shows, the duration of the *Sleep* activity instances is usually large when the elder person goes for first time to sleep a little bit before midnight and it gradually degrades after each bathroom visit during the night or in the morning.

This combination of spatial and temporal information provides a more detailed insight on the person's daily activity and therefore it should also appear in the discovered patterns. Using the time abstraction layer described in Section 7 we were able to extract more information-rich spatiotemporal patterns by conditioning the sequence of input activities on absolute time and duration. In particular, we partitioned the day in 4 time windows (morning, afternoon, evening and night) and used a rough classification of the input activities into short and long according to their duration. Since different activities have different duration characteristics we used activity-specific duration parameters that we were able to extract using statistical information, like the one shown in Figure 6, for all the different input activities. Figure 7 shows a subset of the discovered spatiotemporal frequent patterns after running the proposed method on the conditioned input sequence of activities. Again, in order to facilitate the interpretation of the patterns, we have organized all the similar patterns into chronologically ordered groups. The information that can be extracted now is more valuable. By looking at the pattern groups 1 and 2 in Figure 7, we can clearly see that the bathroom visits happen once during the night and once during the morning. The same pattern groups show that after the bathroom visit in the morning the person will go to sleep for a small period of one to two hours conversely to the previous night sleeping activity instances. After waking up and having a long, most of the times, breakfast the person will spend most of his time in the living-room watching TV until the afternoon (pattern groups 3 and 4 in Figure 7). It is during the afternoon, usually around 3pm,

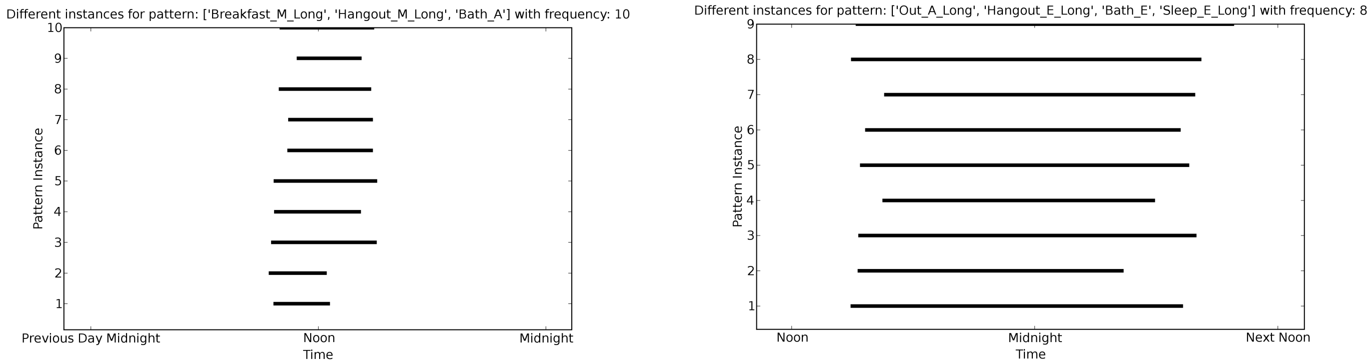

Fig. 8 Different instances of a typical morning and afternoon/evening frequent patterns.

where the person will leave the house for approximately 3 hours to continue watching TV in the living room as soon as he comes back (pattern groups 4 and 5 in Figure 7). During the evening, the person will eventually visit the bathroom for one last time before going to sleep (pattern group 5 in Figure 7).

To show the ability of our approach to find spatiotemporal invariant patterns across different days, we examined the spatiotemporal characteristics of the detected frequent patterns across all the different instances in the 30-day data trace. In particular, for every frequent pattern we detected, we identified all the instances of this pattern in the input data set and recorded the start,stop and duration times for each instance. Figure 8 shows this information for two representative frequent patterns, a morning and an afternoon/evening pattern. Note that for the same pattern, the temporal characteristics across different instances either remain the same or have very small variations. In other words, the start and stop time of the pattern along with pattern's duration do not change significantly across different instances. This verifies that the output of our method captures the most frequent spatiotemporal patterns of the input sequence.

8.3 Model Extraction

The detected frequent sequences represent the spatiotemporal signature of the input data stream and in our case the monitored elder's person core daily activities. Using these frequent sequences one can easily construct the spatiotemporal model of the data stream as follows:

1. **State Extraction**: Every spatiotemporal phoneme contained in *at least one* of the detected frequent sequences becomes a state in the model. Note that phonemes present in the input stream might not be present as states in the generated model if they are not part of at least one frequent sequence. Since each phoneme is spatiotemporal, each state in the model will encode both spatial and temporal information.
2. **Transition Probability Extraction**: For every state in the model and for *all frequent sequences*, count the number of times that each other state appears immediately next in a frequent sequence. By normalizing we get all outgoing transition probabilities for every state in the model. Repeating the same process for every state but considering only the states that appear immediately before in the frequent sequences we get all the incoming transition probabilities for every state in the model.

This two step process leads to the construction of a spatiotemporal context-free model of the input data stream [2] that can be used to either accurately predict future activities over space and time or identify unusual activity sequences at run-time.. Figure 9 shows two such models produced using the extracted frequent sequences of spatiotemporal phonemes like the ones shown in 7. In both cases, each state in the model is named after the spatial activity that this state represents followed by an alphanumeric identifier. This identifier encodes the unique start time and duration characteristics of the spatial activity. As a result, there are multiple "Sleep", "Bath", "Breakfast" etc. states that are differentiated only by the temporal characteristics (start time and duration) of the activity they represent.

By visually comparing the two models in Figure 9, two observations are apparent: (1) the model in Figure 9(a) provides more dense information compared to the model in Figure 9(b) and (2) the number of states in these two models might be different. These observations are justified by the fact that these two models were obtained using a different set of frequent sequences. The

[2] In practice any model building method in the literature can be used

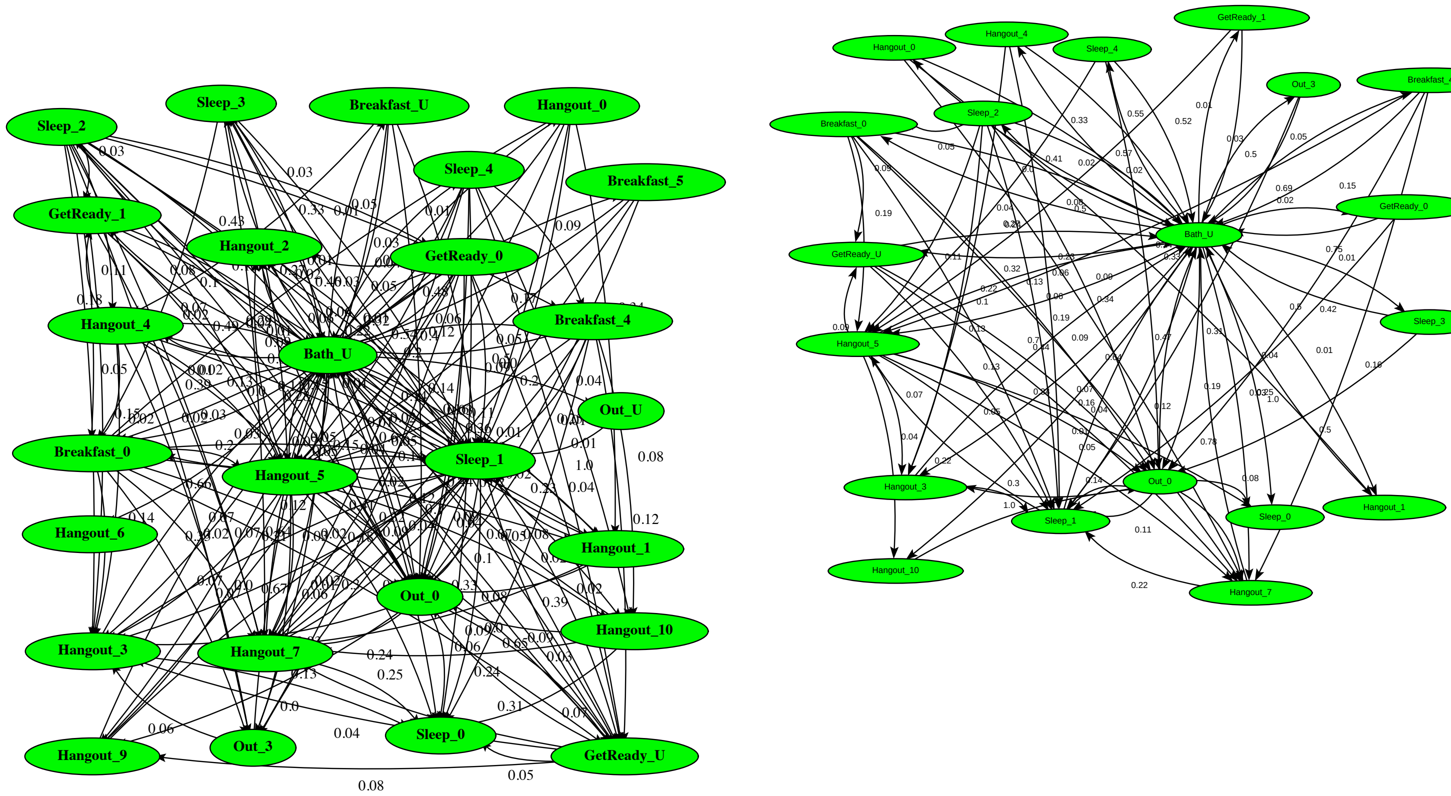

Fig. 9 The generated spatiotemporal daily activity model of the monitored elder person using all frequent patterns with a probability higher than (a) 11% and (b) 18%.

model in Figure 9(a) used all the frequent sequences that appeared in at least 11% of the episodes, while the model in Figure 9(b) used all the frequent sequences that appeared in at least 15% of all the episodes. In general, the lower the frequency threshold, the more frequent sequences are discovered and therefore the more information is captured in the generated model. In addition, the higher the number of frequent sequences used the more states are possible to appear in the model given the two-step model building process described earlier.

9 Conclusion

We have introduced a method for extracting spatiotemporal human activity patterns by properly encoding location, time and duration information into a single phoneme. Our method can be trasnparently used on different people to automatically extract their daily activity model. The 30-day data trace collected from our home sensor network deployment was invaluable in terms of understanding the process, its bottlenecks and requirements, and evaluating the effectiveness of the proposed approach. Our exposure to the real data, revealed that our previous, complimentary work, on grammars [16, 17] is crucial in terms of transforming raw sensing data into a higher level form more appropriate for discovering meaningful patterns. The reason is that due to the noise that low level data always include, you can potentially have infinite permutations making the pattern discovery process extremely difficult. As part of the future work, we will focus more on the automatic discovery of the temporal properties of the model.

References

1. Alarmnet: Assisted-living and residential monitoring network - a wireless sensor network for smart healthcare, http://www.cs.virginia.edu/wsn/medical/.
2. R. Agrawal and R. Srikant. Fast algorithms for mining association rules in large databases. In *VLDB '94: Proceedings*

of the 20th International Conference on Very Large Data Bases, pages 487–499, San Francisco, CA, USA, 1994. Morgan Kaufmann Publishers Inc.

3. R. Agrawal and R. Srikant. Mining sequential patterns. In *ICDE '95: Proceedings of the Eleventh International Conference on Data Engineering*, pages 3–14, Washington, DC, USA, 1995. IEEE Computer Society.
4. Jeffrey Coble, Diane J. Cook, and Lawrence B. Holder. Structure discovery in sequentially-connected data streams. *International Journal on Artificial Intelligence Tools*, 2006.
5. Diane J. Cook. Health monitoring and assistance to support aging in place. *J. UCS*, 2006.
6. Diane J. Cook, G. Michael Youngblood, and Sajal K. Das. A multi-agent approach to controlling a smart environment. In *Designing Smart Homes*, 2006.
7. Sajal K. Das and Diane J. Cook. Designing smart environments: A paradigm based on learning and prediction. In *Proceedings of PReMI*, 2005.
8. M. Hasan, V. Chaoji, S. Salem, N. Parimi, and M. Zaki. Dmtl: A generic data mining template library. In *Workshop on Library-Centric Software Design (LCSD'05)*, 2005.
9. E. Heierman, M. Youngblood, and D. J. Cook. Mining temporal sequences to discover interesting patterns. In *KDD Workshop on Mining Temporal and Sequential Data*, 2004.
10. S. S. Intille, Kent Larson, and E. M. Tapia. Designing and evaluating technology for independent aging in home. In *International Conference oon Aging, Disability and Independence*, 2003.
11. Y. A. Ivanov and A. F. Bobick. Recognition of visual activities and interactions by stochastic parsing. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):852–872, 2000.
12. Istvan Jonyer, Lawrence B. Holder, and Diane J. Cook. Mdl-based context-free graph grammar induction and applications. *International Journal on Artificial Intelligence Tools*, 2004.
13. Ellen Ernst Kossek, Beverly J. DeMarr, Kirsten Backman, and Mark Kollar. Assessing employees' emerging elder care needs and reactions to dependent care benefits. *Personnel Management*, 22, 1993.
14. L. Liao, D. Fox, and H. Kautz. Location-based activity recognition using relational markov models. In *Nineteenth International Joint Conference on Artificial Intelligence*, 2005.
15. D. Lymberopoulos, A. Bamis, and A. Savvides. Extracting temporal characteristics from sensor network data streams. Technical Report TR-05-2008, Yale University, 2008.
16. D. Lymberopoulos, A. Ogale, A. Savvides, and Y. Aloimonos. A sensory grammar for inferring behaviors in sensor networks. In *Proceedings of Information Processing in Sensor Networks (IPSN)*, April 2006.
17. D. Lymberopoulos, T. Teixeira, and A. Savvides. Detecting patterns for assisted living using sensor networks. In *Proceedings of SensorComm*, October 2007.
18. D. Lymberopoulos, T. Teixeira, and A. Savvides. Macroscopic human behavior interpretation using distributed imagers and other sensors. In *to appear in Proceedings of IEEE*, 2008.
19. Heikki Mannila, Hannu Toivonen, and A. Inkeri Verkamo. Discovery of frequent episodes in event sequences. *Data Mining and Knowledge Discovery*, 1(3):259–289, 1997.
20. D. Moore and I. Essa. In *Recognizing multitasked activities from video using stochastic context-free grammar*, pages 770–776, Menlo Park, CA, USA, 2002. AAI.
21. Jyothsna R. Nayak and Diane J. Cook. Approximate association rule mining. In *Proceedings of FLAIRS Conference*, 2001.
22. A. S. Ogale, A. Karapurkar, and Y. Aloimonos. View-invariant modeling and recognition of human actions using grammars. *ICCV'05*, October 2005.
23. S. Park and J. K. Aggarwal. In *Event Semantics in Two-person Interactions*, August 2004.
24. D. J. Patterson, D. Fox, H. Kautz, and M. Philipose. Fine-grained activity recognition by aggregating abstract object usage. In *IEEE International Symposium on Wearable Computers*, October 2005.
25. M. Philipose, K. P. Fishkin, M. Perkowitz, D. J. Patterson, D. Fox, H. Kautz, and D. Hahnel. Inferring activities from interactions with objects. *IEEE Pervasive Computing*, 03(4):50–57, 2004.
26. Jorge C. G. Ramirez, Diane J. Cook, Lynn L. Peterson, and Dolores M. Peterson. An event set approach to sequence discovery in medical data. *Intell. Data Anal.*, 2000.
27. K. Romer. Distributed mining of spatio-temporal event patterns in sensor networks. In *EAWMS / DCOSS*, 2006.
28. P.-N. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining*. Addison-Wesley, 2006.
29. E. Munguia Tapia, S. S. Intille, and K. Larson. Activity recognition in the home setting using simple and ubiquitous sensors. In *PERVASIVE 2004*, 2004.
30. Anneke van den Berg Jeths and Mats Thorslund. Will resources for elder care be scarce? *The Hastings Center Report*, 1994.
31. Kevin D. Willison, G.J.Gavin, and J. Andrews. Complementary medicine and older people: past research and future directions. *Complementary Therapies in Nursing and Midwifery*, 10(2), May 2004.
32. M. J. Zaki, N. Parimi, N. De, F. Gao, B. Phoophakdee, J. Urban, V. Chaoji, M. Al Hasan, and S. Salem. Towards generic pattern mining. In *International Conference on Formal Concept Anaysis*, 2004.