Discovering Temporal Features and Relations of Activity Patterns

Ehsan Nazerfard, Parisa Rashidi, and Diane J. Cook
Department of Electrical Engineering and Computer Science
Washington State University
Pullman, WA 99164, USA
{nazerfard, prashidi, cook}@eecs.wsu.edu

***Abstract*—An important problem that arises during the data mining process in many new emerging application domains is mining data with temporal dependencies. One such application domain is activity discovery and recognition. Activity discovery and recognition is used in many real world systems, such as assisted living and security systems, and it has been vastly studied in recent years. However, the temporal features and relations which provide useful insights for activity models have not been exploited to their full potential by mining algorithms. In this paper, we propose a temporal model for discovering temporal features and relations of activity patterns from sensor data. Our algorithm is able to discover features and relations, such as the order of the activities, their usual start times and durations by using rule mining and clustering techniques. The algorithm has been validated using 4 months of data collected in a smart home.**

Keyword—Temporal Association Rules, Temporal Features, Temporal Relations, Rule Mining, Clustering, Smart Homes.

I. INTRODUCTION

Activity discovery and recognition has been recently utilized in many different application domains, such as security and health monitoring of elder adults. Health monitoring of elder adults based on using assistive technologies such as smart homes, has attracted many researchers in recent years due to the increasing aging population. An estimated 9% of adults over age 65 and 50% of adults over age 85 need assistance with their everyday activities. The resulting cost is significant both for the government and for families. To function independently at home, elderly people need to be able to complete the Activities of Daily Living (ADL) [1] such as taking medication, cooking, eating and sleeping.

Assistive technologies such as smart homes can greatly help in monitoring the activities of elder adults at home. By using a large number of different types of sensors deployed in an everyday setting, and using techniques from activity discovery and recognition, smart homes are able to discover and recognize patterns of resident activities.

Activity discovery and recognition has been vastly studied in the context of smart environments [2], [3], [4] and [6]. Despite the increasing progress in this area, a key component in these activities, namely temporal features and relations, has been largely ignored. One of the few works in this area by Rashidi and Cook [5] provides a method for discovering and representing basic temporal features of the activities, such as the usual start times at multiple granularities. However, they do not address discovering other temporal features and relations, such as the relative order of the activities. Discovering the order of the activities can be of great use in many different scenarios, such as for predicting the next activity in an activity reminder application or a home automation system.

In this paper, we provide a framework for discovering and representing the temporal features of activity patterns. Our framework called "DTFRA" (short for **D**iscovering of **T**emporal **F**eatures and **R**elation of **A**ctivities), discovers the usual start times of the activity patterns in form of a normal mixture distribution [7], using a *k*-means clustering technique [8]. DTFRA discovers a similar representation for duration of activities using a normal mixture distribution. We also discover the order of activities using temporal association rule mining techniques.

The temporal information that is discovered by our algorithms can be beneficial in many different applications. Here we mention just a few such applications:

- *Reminder systems:* The discovered temporal information can be used to construct a schedule of activities for an upcoming period. Such a schedule is constructed based on the predicted start time intervals, as well as the relative order of the activities. For example, it can be used as part of an activity reminder system for an Alzheimer patient, in order to automatically generate prompts to remind the patient to initiate activities, if s/he does not follow the usual schedule within the expected start times.
- *Anomaly detection:* Discovering a confidence level for the order relations, as well as finding the typical start times and durations as mixture models allow us to assign a probability to a particular activity occurring in a given interval. Using the assigned probabilities and using the anomaly detection methods [9], one can exploit such information to detect the any anomalous or suspicious activities.
- *Context aware system networks:* Last but not least, such information can be used in designing an activity-aware wireless sensor network. Activity aware sensor networks aim to augment wireless sensor network with

context awareness by taking advantage of the behavioral pattern information such as temporal information. Temporal information allow the sensor networks to act more intelligently by determining which sensors need to be awake at any given time, in order to route data more efficiently with respected to the predicted activities within a given time period.

The rest of this paper is organized as follows. First we will describe the related works. Next we will explain our model in more detail, followed by the results of our experiments on the data obtained from an elderly smart apartment. Finally we end the paper with conclusions.

II. Related Work

In recent years, the data mining community has observed remarkable progress in discovering association rules from transactional data. The notion of association rules was first proposed by Agrawal et al. [10] to capture the co-occurrence of items in transaction events. Because a transaction typically includes a timestamp as well as the transaction items, an interesting extension to the concept of association rule is to augment it with the time dimension. This extension suggests that we may discover different rules when different time intervals are considered. As a result, some rules may hold during some time intervals, while not during the others. When we consider the association rules augmented by their temporal intervals, we refer to them as temporal association rules [11].

Activity patterns in smart environments also include a timestamp. The timestamp indicates when a particular activity has occurred, or more specifically when a specific sensor was triggered. Just like the association rule mining, adding the concept of temporal features to the activity patterns can be quite useful, and in some cases necessary. For example, in a home automation system, the concept of time needs to be taken into account to anticipate the time when activities should or would occur. Despite the potential use of temporal features in activity patterns, this key aspect has usually been ignored and it has not been exploited to its full potential. Gottfried et al. [12] mention that reasoning about time and space in smart homes is necessary, because certain things have to be done at certain times and places, and in association with other contexts. Our study goes beyond just highlighting the importance of using temporal features. Here we present a framework to discover and present the temporal features of activities by inferring temporal association rules among different activities. As another related work, Rashidi and Cook [5] have suggested a method for discovering and representing temporal features of the activities. They propose a method for presenting the usual start times of the activities. However they do not consider other temporal features and relations, such as the order of the activities. Galushka et al. [13] also discuss the importance of temporal features for learning activity patterns; however, they do not exploit such features for learning activity patterns in practice. Jakkula and Cook [14] show the benefit of considering temporal associations for activity prediction. Their main focus is on investigating methods based on using Allen's temporal logic [15, 16] to analyze smart home data, and to use such analysis for event prediction.

In the framework presented in this paper, we will provide a method for discovering various temporal features, and we will present and discuss the results of applying our methods to a real smart home dataset.

III. Model Description

The architecture of our model called DTFRA is illustrated in Fig. 1. It consists of two main components: temporal feature discovery, and temporal relation discovery. Each component will be described in more detail in the following subsections.

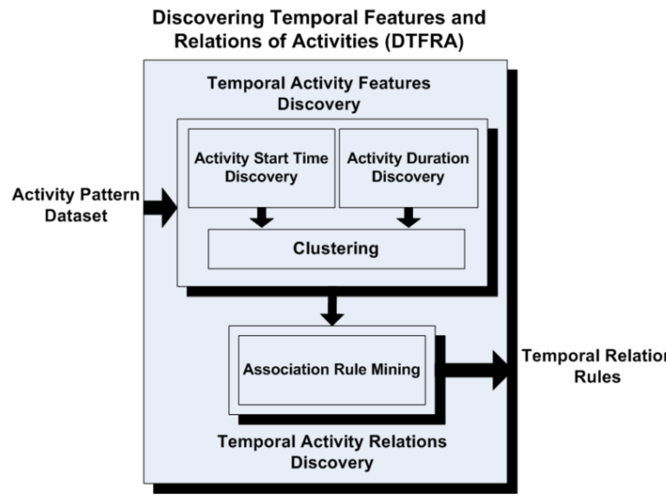

Figure 1. DTFRA architecture.

The input to DTFRA is an activity pattern dataset, and its output is a set of temporal relation rules. To make it easier to follow the description of our model, we will consider the "taking medication" activity as an example throughout our discussions. More activities will be discussed in the Experimental Results section.

A. Activity Pattern Dataset

The input dataset consists of a set of individual sensor events collected from various sensors deployed in the space. Each sensor event consists of a sensor ID and a timestamp Table I provides a sample of collected data.

TABLE I. Sample of collected data. Here, M11 is a motion sensor and D34 is a door senosr.

Sensor ID	State	Timestamp
M11	ON	05/13/2009, 08:33
D34	OPEN	05/13/2009, 08:35

B. Temporal Activity Features Discovery

In this study, we consider two temporal features for every activity: the *start time* of the activity and the *duration* of the activity. For this purpose, we extract the start times of every activity instance, and then we cluster the start times to obtain a canonical representation for the start time of a specific activity. Here we use the *k*-means clustering algorithm to construct a mixture model for each activity a_i. If we denote the start time of an activity instance a_i as t_i, then the probability that t_i belongs to a certain cluster k with parameters $\Theta_k = (\mu, \sigma)$, can be expressed as a normal probability density function shown in Equation (1).

$$prob(t_i|\Theta_k) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t_i-\mu)^2}{2\sigma^2}} \quad (1)$$

The parameters are computed automatically from the available data. As we already mentioned, such a probability measure for example can help us to detect anomalous cases. The results of finding the canonical start times of "taking medication" can be seen in Table II.

TABLE II. THE CANNONICAL START TIMES OF "TAKING MEDIACTION" ACTIVITY.

Cluster #	Mean (hh:mm)	Number of Members	Standard Deviation (hh:mm)
1	9:33	62	0:56
2	7:22	12	1:50
3	20:37	7	1:41

As shown in Table II by calculating the mean and standard deviation of each cluster, the start time of each activity is expressed as a mixture normal distribution (Fig. 2).

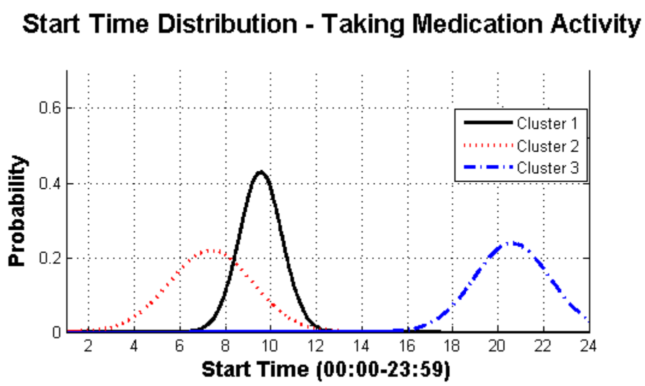

Figure 2. A mixture model for the start time of "taking medication" activity.

Besides the canonical start time of an activity, we will consider the duration of an activity. Similar to our method for finding the canonical start times, by using a *k*-Means clustering technique, we will discover the canonical representation of durations. Table III shows the duration of "taking medication" activities corresponding to the clusters specified in Table II. Furthermore, the mean and standard deviation of duration values of each activity is calculated. Again, the canonical representation is in form of a mixture normal distribution.

TABLE III. DISCOVING USUAL DUARION OF "TAKING MEDICATION".

Cluster #	Mean (hh:mm)	Number of Members	Standard Deviation (hh:mm)
1	0:07	62	0:01
2	0:03	12	0:01
3	0:04	7	0:02

Based on a normal distribution characteristic, we know that a standard deviation from the mean in a normal distribution accounts for about 68% of the values (see Fig. 3). So if we only consider observations falling within this area, some observations that significantly deviate from the other values are automatically left out. Such an observation that is numerically distant from the rest of the data is called "outlying observation" or "outlier".

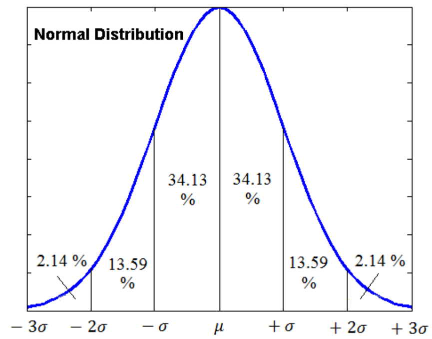

Figure 3. Outlier detection based on normal distribution characteristics.

Taking into account the interval $[\mu_i - \sigma_i, \mu_i + \sigma_i]$ and blending Table II and Table III together, we are left with Table IV. Here μ_i and σ_i are the mean and standard deviation values for the start times of the i^{th} cluster respectively.

TABLE IV. DISCOVING START TIMES AND DUARIONS OF THE "TAKING MEDICATION" ACTIVITY.

Cluster #	Start Time (hh:mm)	Duration (hh:mm)
1	[8:37 – 10:29]	[0:02 – 0:04]
2	[6:32 – 9:12]	[0:06 – 0:08]
3	[18:56 – 22:18]	[0:02 – 0:06]

C. *Temporal Activity Relations Discovery*

After the canonical forms of start time and duration have been discovered, next we will discover the order of activities. The input to this stage is the features discovered in the previous stage, i.e. the canonical start time and durations. The output of this stage is a set of temporal relations between activities. The temporal relations will determine the order of activities with respect to their start time, i.e. for a specific time point, what activity would be the most probable. Such results can be useful in a variety of activity prediction scenarios, such as home automation.

To discover the temporal relations of activities, we use the *Apriori* algorithm. Let's denote an instance i of an activity a by a_i and its successor activity instance in the dataset by b_j. As mentioned in the previous section, each activity instance belongs to specific cluster Θ_k with respect to its start time. To denote an activity a_i belonging to a specific cluster Θ_k, we will show it as a_i^k. We will show the temporal relation "*b* follows *a*" as $a \rightarrow b$. The dataset length will also be shown by $|D|$. Then we can define the *support* of the "follows" relation as in Equation (2) and its *confidence* as in Equation (3).

$$\mathrm{supp}(a^k \rightarrow b) = \frac{\sum_{i,j}(a_k^i \rightarrow b_j)}{|D|} \quad (2)$$

$$\mathrm{conf}(a^k \rightarrow b) = \frac{\sum_{i,j}(a_k^i \rightarrow b_j)}{|a|} \quad (3)$$

Table V shows the discovered temporal relation rules for the "taking medication" activity, with the corresponding confidence. For this example, minimum support is set to 0.1.

TABLE V. DISCOVERING TEMPORAL RELATIONS OF THE "TAKING MEDICATION" ACTIVITY (SUPP > 0.1)

Cluster #	Start Time	Duration	Next Activity	Conf
1	[8:37 – 10:29]	[0:02 – 0:04]	Eating	0.90
2	[6:32 – 9:12]	[0:06 – 0:08]	Eating	0.81
			Meal Preparation	0.11
3	[18:56 – 22:18]	[0:02 – 0:06]	Personal Hygiene	0.50
			Sleeping Not in Bed	0.25
			Eating	0.25

According to the results provided in Table V for "taking medication" activities which occur between 8:37 and 10:29, the next activity is typically "eating" with a confidence of 0.90. This activity usually takes 2 to 4 minutes when performed during this period. For those "taking medication" activities which occur between 6:32 and 9:12, the next activities are "eating" and "meal preparation" with confidences of 0.81 and 0.11 respectively. This activity when performed at this period, usually takes 6 to 8 minutes. Finally those "taking medication" activities occurring between 18:56 and 22:18, usually take 2 to 6 minutes and the next possible activities are "personal hygiene", "sleeping not in bed" and "eating" with the confidences of 0.50, 0.25 and 0.25 respectively.

IV. EXPERIMENTAL RESULTS

In this section, the experimental results of DTFRA are presented. Before getting into the details of our results, we explain the settings of our experiment.

A. *Experimental Setup*

The smart home testbed used in our experiments is a 1-bedroom single resident apartment. The sensor events are generated by motion sensors (these sensor IDs begin with "M") and door/cabinet sensors (these sensor IDs begin with "D"). Fig. 4 shows the motion sensor layout of our smart home testbed. To track people's mobility, we use motion sensors placed on the ceilings and walls, as well as on doors and cabinets. A sensor network captures all the sensor events, and stores them in a database.

For our experiments, we selected ten ADLs which are listed in Table VI.

TABLE VI. LIST OF ADLS WE CONSIDER FOR THE EXPERIMENTS.

1. Bathing	6. Meal Preparation
2. Bed to Toilet Transition	7. Personal Hygiene
3. Eating	8. Sleeping in Bed
4. Enter/Leave Home	9. Sleeping Not in Bed
5. Housekeeping	10. Taking Medication

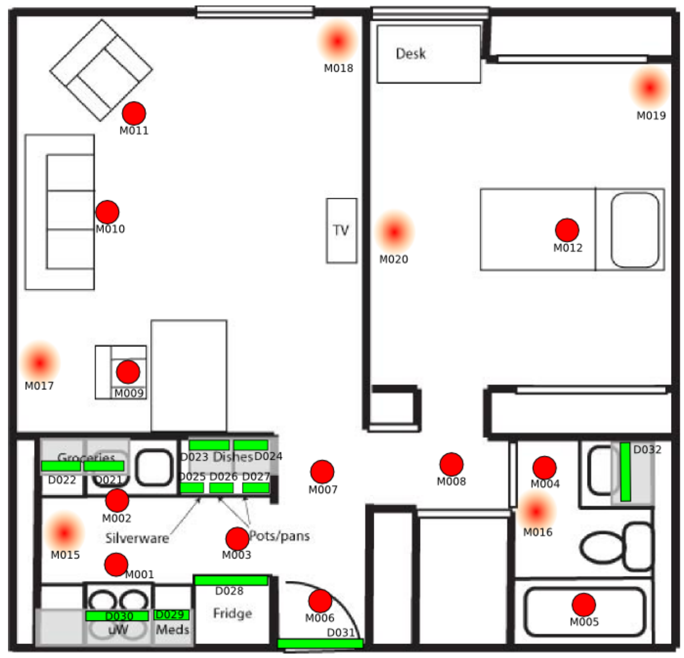

Figure 4. The layout of sensors in the 1-bedroom smart home testbed.

To provide real activity pattern data for our experiments, we have collected data while the resident was living in the smart home. Our training data was gathered over a period of 4 months and more than 250,000 sensor events were generated for our dataset.

B. *Evaluation of DTFRA*

In this section, we provide the results of running DTFRA to discover activity patterns from the smart home testbed. Table VII shows the Discovered Temporal Features and Relations for five ADLs. In the clustering step, we ran k-means three times and picked the best results. Furthermore, for Apriori algorithm, we set the minimum support to 0.1 and only present the next activities with the confidence greater than 0.1 in the results. Moreover, $\boldsymbol{\mu_s}$ and $\boldsymbol{\sigma_s}$ are mean and standard deviation for start time of each cluster, while $\boldsymbol{\mu_d}$ and $\boldsymbol{\sigma_d}$ denote the same parameters for duration of each cluster.

TABLE VII. TEMPORAL RELATION RULES FOR THE FIRST FIVE ADLS (SUPP > 0.1)

	Activity	Cluster	Start Time (hh:mm) $[\mu_s - \sigma_s$ to $\mu_s + \sigma_s]$	Duration (hh:mm) $[\mu_d - \sigma_d$ to $\mu_d + \sigma_d]$	Next Activity	Conf
1	Bathing	1	[0:46 – 01:40]	[0:25 – 0:31]	Sleeping in Bed	0.42
					Bed-Toilet Transition	0.38
					Personal Hygiene	0.20
		2	[08:22 – 11:02]	[0:15 – 0:19]	Personal Hygiene	0.71
					Eating	0.14
		3	[21:55 – 23:03]	[0:04 – 0:10]	Sleeping Not in Bed	0.50
					Personal Hygiene	0.25
					Housekeeping	0.25
2	Bed-Toilet Transition	1	[20:00 – 23:46]	[0:14 – 0:16]	Bed Toilet Transition	0.33
					Sleeping Not in Bed	0.22
					Housekeeping	0.22
		2	[05:09 – 07:49]	[0:08 – 0:10]	Sleeping in Bed	0.95
		3	[00:54 – 03:10]	[0:02 – 0:04]	Sleeping in Bed	0.58
					Bed-Toilet Transition	0.36
3	Eating	1	[06:25 – 11:01]	[0:08 – 0:24]	Housekeeping	0.61
					Personal Hygiene	0.39
		2	[12:19 – 14:47]	[2:18 – 3:20]	Housekeeping	0.60
					Personal Hygiene	0.26
		3	[18:38 – 22:28]	[0:34 – 1:02]	Housekeeping	0.59
					Personal Hygiene	0.27
4	Enter Home	1	[18: 03 – 20:57]	N/A	Personal Hygiene	0.57
					Meal Preparation	0.29
		2	[09:35 – 14:07]	N/A	Personal Hygiene	0.46
					Meal Preparation	0.42
		3	[15:10 – 17:14]	N/A	Personal Hygiene	0.67
					Leave Home	0.14
5	Housekeeping	1	[12:55 – 16:17]	[1:29 – 1:33]	Personal Hygiene	0.59
					Leave Home	0.17
					Sleeping Not in Bed	0.17
		2	[19:40 – 23:08]	[0:01 – 0:03]	Sleeping Not in Bed	0.48
					Personal Hygiene	0.39
		3	[04:33 – 11:53]	[0:23 – 0:39]	Personal Hygiene	0.91

As we mentioned in the introduction section, this information can be used in developing various applications, such as an activity reminder system or an activity-aware wireless sensor network.

In Table VIII the results for the other five activities are presented. Considering the “meal preparation” activity, the results show that for those “meal preparation” activities which occur between 11:59 and 14:11, the next activities are “eating” and “meal preparation” with the confidences of 0.69 and 0.14 respectively. Also, “meal preparation” when performed at this time, usually takes 2:51 to 3:11. Also, “meal preparation” activities occurring between 7:19 and 10:23, usually take 24 to 56 minutes and their next activity is “eating” with confidence of 0.55. Finally, those “meal preparation” activities which occur between 17:38 and 21:44, usually take 4 to 14 minutes and their next activity is “eating” and “personal hygiene” with the confidence of 0.64 and 0.25 respectively.

TABLE VIII. TEMPORAL RELATION RULES FOR THE SECOND FIVE ADLS (SUPP > 0.1)

	Activity	Cluster	Start Time (hh:mm) [$\mu_s - \sigma_s$ to $\mu_s + \sigma_s$]	Duration (hh:mm) [$\mu_d - \sigma_d$ to $\mu_d + \sigma_d$]	Next Activity	Conf
6	Meal Preparation	1	[11:59 – 14:11]	[2:51 – 3:11]	Eating	0.69
					Meal Preparation	0.14
		2	[07:19 – 10:23]	[0:24 – 0:56]	Eating	0.55
		3	[17:38 – 21:44]	[0:04 – 0:14]	Eating	0.64
					Personal Hygiene	0.25
7	Personal Hygiene	1	[06:06 – 11:46]	[0:01 – 0:02]	Personal Hygiene	0.50
					Meal Preparation	0.24
					Leave Home	0.13
		2	[13:38 – 17:04]	[0:05 – 0:13]	Leave Home	0.53
					Personal Hygiene	0.21
		3	[19:23 – 22:37]	[0:04 – 0:16]	Personal Hygiene	0.24
					Sleeping Not in Bed	0.24
					Meal Preparation	0.22
8	Sleeping in Bed	1	[13:59 – 22:11]	[0:37 – 1:45]	Personal Hygiene	0.60
					Meal Preparation	0.40
		2	[01:02 – 03:38]	[4:13 – 6:35]	Bed-Toilet Transition	0.94
		3	[05:22 – 08:14]	[2:22 – 3:28]	Personal Hygiene	0.78
					Bed-Toilet Transition	0.22
9	Sleeping Not in Bed	1	[20:29 – 23:15]	[1:46 – 2:34]	Bed-Toilet Transition	0.50
					Personal Hygiene	0.16
					Sleeping Not in bed	0.14
		2	[00:33 – 02:27]	[0:32 – 1:18]	Bed-Toilet Transition	1.00
		3	[13:42 – 18:58]	[3:11 – 4:23]	Personal Hygiene	0.59
					Meal Preparation	0.19
10	Taking Medication	1	[08:37 – 10:29]	[0:02 – 0:04]	Eating	0.90
		2	[06:32 – 09:12]	[0:06 – 0:08]	Eating	0.81
					Meal Preparation	0.11
		3	[18:56 – 22:18]	[0:02 – 0:06]	Personal Hygiene	0.50
					Sleeping Not in Bed	0.25
					Eating	0.25

I. CONCLUSION

In this paper, we presented DTFRA for discovering temporal features and relations of activity patterns. Our algorithm is able to discover features and relations, such as the order of the activities, their usual start times and durations by using rule mining and clustering techniques. These discoveries can be used in many applications, such as developing activity reminder and anomaly detection systems, as well as designing activity-aware wireless sensor networks.

ACKNOWLEDGMENT

The authors would like to thank Robert Bosch LLC for making available the activity datasets and apartment layouts.

REFERENCES

[1] B. Reisberg, S. Finkel, J. Overall, N. Schmidt-Gollas, S. Kanowski, H. Lehfeld, F. Hulla, S. G. Sclan, H.-U. Wilms, K. Heininger, I. Hindmarch, M. Stemmler, L. Poon, A. Kluger, C. Cooler, M. Bergener, L. Hugonot-Diener, P. H. Robert, and H. Erzigkeit, "The Alzheimer's disease activities of daily living international scale", International Psychogeriatrics, vol. 13, no. 2, pp. 163-181, 2001.

[2] M. Philipose, K. Fishkin, M. Perkowitz, D. Patterson, D. Fox, H. Kautz, and D. Hahnel, "Inferring activities from interactions with objects", IEEE Pervasive Computing, vol. 3, no. 4, pp. 50–57, 2004.

[3] L. Liao, D. Fox, and H. Kautz, "Location-based activity recognition using relational markov networks," in Proceedings of the International Joint Conference on Artificial Intelligence, pp. 773–778, 2005.

[4] N. A. Campbell, ''Mixture models and atypical values''. Math. Geol, vol. 16, pp. 465-477, 1984.

[5] P. Rashidi and D. J. Cook, “Keeping the resident in the loop: Adapting the smart home to the user”, IEEE Transactions on Systems, Man. and Cybernetics journal, Part A, vol. 39, no. 5, pp. 949–959, Sep. 2009.

[6] T. Gu, Z. Wu, X. Tao, H. Pung, and J. Lu, “epsicar: An emerging patterns based approach to sequential, interleaved and concurrent activity recognition.” in Proceedings of the IEEE International Conference on Pervasive Computing and Communication, 2009.

[7] N. A. Campbell, ‘’Mixture models and atypical values’’. Math. Geol, vol. 16, pp. 465-477, 1984.

[8] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-Means Clustering Algorithm", Journal of the Royal Statistical Society, Series C (Applied Statistics), vol. 28, no 1, pp. 100–108, 1979.

[9] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly detection: A survey”, ACM Comput. Surv, vol. 41, no. 3, pp. 1-58, Jul. 2009.

[10] R. Agrawal, T. Imielinski, and A. Swami, "Mining Associations between Sets of Items in Large Databases", ACM SIGMOD Int'l Conf. on Management of Data, Washington D.C., May 1993.

[11] Y. Li, P. Ning, X. S. Wang, and S. Jajodia, "Discovering Calendar-Based Temporal Association Rules," Data & Knowledge Engineering, vol 44, no. 2, pp. 193-218, 2003.

[12] B. Gottfried, H. W. Guesgen, and S. Hubner, “Spatiotemporal reasoning for smart homes”, in Designing Smart Homes, The Role of Artificial Intelligence, pp. 16–34, Springer, 2006.

[13] M. Galushka, D. Patterson, and N. Rooney, “Temporal data mining for smart homes”, Designing Smart Homes. The Role of Artificial Intelligence, Springer, pp. 85–108, 2006.

[14] V. R. Jakkula and D. J. Cook, ‘Using temporal relations in smart environment data for activity prediction”, in Proceedings of the 24th International Conference on Machine Learning, 2007.

[15] J. F. Allen, ‘’Time and time again: The many ways to represent time", International Journal. of Intelligent Systems, vol. 6, no. 4, pp. 341-356, July 1991.

[16] J. F. Allen and G. Ferguson, “Actions and events in interval temporal logic”, Journal of Logic and Computation, vol. 4, pp. 531–579, 1994.