Using Association Rule Mining to Discover Temporal Relations of Daily Activities

Ehsan Nazerfard, Parisa Rashidi, and Diane J. Cook

School of Electrical Engineering and Computer Science
Washington State University
Pullman, WA 99164, USA
{nazerfard,prashidi,cook}@eecs.wsu.com

Abstract. The increasing aging population has inspired many machine learning researchers to find innovative solutions for assisted living. A problem often encountered in assisted living settings is activity recognition. Although activity recognition has been vastly studied by many researchers, the temporal features that constitute an activity usually have been ignored by researchers. Temporal features can provide useful insights for building predictive activity models and for recognizing activities. In this paper, we explore the use of temporal features for activity recognition in assisted living settings. We discover temporal relations such as order of activities, as well as their corresponding start time and duration features. To validate our method, we used four months of real data collected from a smart home.

Keywords: Association Rule Mining, Temporal Relations, Clustering, Smart Homes.

1 Introduction

The projection of age demographics shows an increasing aging population in the near future. Today, approximately 10% of the world's population is 60 years old or older, and by 2050 this proportion will be more than doubled. Moreover, the greatest rate of increase is among the "oldest old", i.e. people aged 85 and over [1]. This results in a society where medical and assistive care demands simply cannot be met by the available care facilities. The resulting cost is significant for both government and families. At the same time, many researchers have noted that most elderly people prefer to stay at home rather than at care facilities.

To provide in-home assisted living, smart homes can play a great role. A smart home is a collection of various sensors and actuators embedded into everyday objects. Data is collected from various sensors and is analyzed using machine learning techniques to recognize a resident's activities and environmental situations. Based on such information, a smart home can provide context-aware services. For example, to function independently at home, elderly people need to be able to complete Activities of Daily Living (ADL) [2], such as taking medication, cooking, eating and sleeping. A smart home can monitor how completely

B. Abdulrazak et al. (Eds.): ICOST 2011, LNCS 6719, pp. 49–56, 2011.

and consistently such activities are performed, to raise an alarm if needed, and to provide useful hints and prompts.

An important component of a smart home is an activity discovery and recognition module. Activity recognition has been studied by many researchers [3][4]. Unfortunately, most activity recognition methods ignore temporal aspects of activities, and solely focus on recognizing activities as a sequence of events. Exploiting temporal aspects of activities can have tremendous potential applications, especially in an assisted living setting. For example, consider the following classic scenario which shows how temporal features can be useful in an assistive living setting:

Marilla Cuthbert is an older adult who is living alone and is in the early stages of dementia. Marilla usually wakes up between 7:00 AM and 9:00 AM every day. After she wakes up (detected by the activity recognition module), her coffee maker is turned on through power-line controllers. If she doesn't wake up by the expected time, her daughter is informed to make sure that she is alright. Marilla should take her medicine within at most one hour of having breakfast. If she doesn't take her medicine within the prescribed time or takes it before eating, she is prompted by the system. The rest of her day is carried out similarly.

The above scenario shows how temporal features can be useful in an assisted living setting. 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, as well as the relative order of the activities.

In this paper, we propose a framework for discovering and representing temporal aspects of activity patterns, including temporal ordering of activities and their usual start time and duration. We refer to the proposed framework as "TEREDA", short for "TEmporal RElation Discovery of Daily Activities". TEREDA discovers the order relation between different activities using temporal association rule mining techniques [5]. It represents temporal features such as the usual start time and duration of activities as a normal mixture model [6], using the Expectation Maximization (EM) clustering method [7]. The discovered temporal information can be beneficial in many applications, such as for home automation, for constructing the schedule of activities for a context-aware activity reminder system, and for abnormal behavior detection in smart homes.

2 Related Work

The concept of association rules was first proposed by Agrawal et al. [8] to discover what items are bought together within a transactional dataset. Since each transaction includes a timestamp, it is possible to extend the concept of association rules to include a time dimension. This results in a new type of association rules are called temporal association rules [5]. This extension suggests that we might discover different rules for different timeframes. As a result, a rule might be valid during certain timeframe, but not during some other timeframes.

Activity pattern dataset in smart homes also include a timestamp. The timestamp implies when a particular activity has performed, or more specifically when a specific sensor was triggered. Similar to association rule mining, considering the concept of temporal features to the activity patterns can be quite useful. For instance, in a home automation setting, we can determine when a certain activity is expected to occur and which activities are most likely to occur next. Despite the potential use of temporal features in activity patterns, this key aspect is usually neglected and has not been exploited to its full potential.

One of the few works in this area is provided by Rashidi et al. [4], in the form of an integrated system for discovering and representing temporal features. They only consider start times of activities at multiple granularities, and do not address discovering other important temporal features and relations such as the relative order of activities. Galushka et al. [9] 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 [10] show the benefit of considering temporal associations for activity prediction. Their main focus is on investigating methods based on using Allen's temporal logic to analyze smart home data, and to use such analysis for event prediction.

3 Model Description

The architecture of TEREDA is illustrated in Fig. 1. TEREDA consists of two main components: the temporal feature discovery component and the temporal relation discovery. Each component will be described in more depth in the following sections. The input dataset consists of a set of sensor events collected from various sensors deployed in the space. Each sensor event consists of an ID and a timestamp. To make it easier to follow the description of our model, we consider an example "wash Dishes" activity throughout our discussions.

3.1 Temporal Activity Features Discovery

For each activity, we consider the *start time* of the activity and the *duration* of the activity. After extracting the start times of all instances of a specific activity, we cluster the start times to obtain a canonical representation. For this purpose we use the Expectation Maximization (EM) clustering algorithm [7] to construct a normal mixture model for each activity.

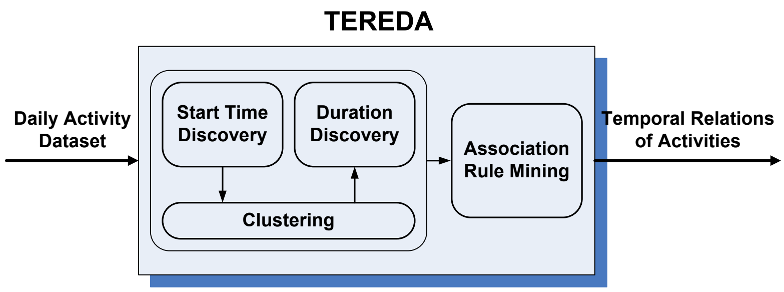
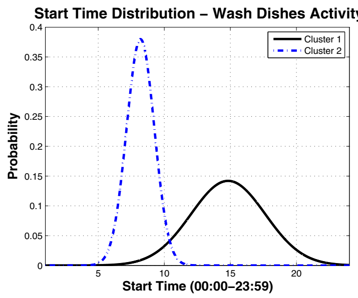


Fig. 1. TEREDA architecture

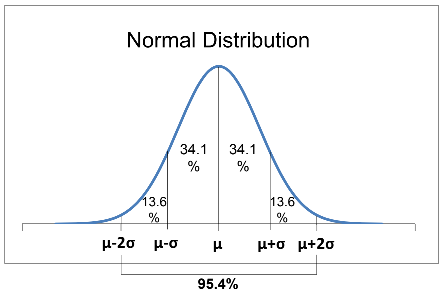
Lets denote the start time of activity a_i by t_i. Then the probability that t_i belongs to a certain cluster k can be expressed as a normal probability density function with parameters $\Theta_k = (\mu, \sigma)$ as in Equation 1. Here μ and σ are the mean and standard deviation values, calculated for each cluster of start times.

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

The parameters of the mixture normal model are computed automatically from the available data. The results of finding the canonical start times of the "Wash Dishes" activity can be seen in Fig. 2a as a mixture of two normal distributions. According to the normal distribution features, the distance of "two standard deviations" from the mean accounts for about 95% of the values (see Fig. 2b). Therefore if we consider only observations falling within two standard deviations, observations that are deviating from the mean will be automatically left out. Such observations that are distant from the rest of the data are called "outlying observations" or "outliers".



(a) A mixture model for the start time of "Wash Dishes" activity



(b) Outlier detection based on normal distribution features

Fig. 2. A mixture normal model and normal distribution characteristics

Besides the start time, we also consider the duration of an activity. For each resulting cluster of the start time discovery step, we calculate the average duration of all instances fallen in that cluster.

3.2 Temporal Activity Relations Discovery

Discovering the *temporal relations* of activities is the main component of TEREDA. The input to this stage is the features discovered in the previous stage, i.e. the canonical start times 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 times, i.e. for a specific time what are the most likely activities that follow a specific activity. Such results can be useful in a variety of activity prediction scenarios. To discover the temporal relations of activities, we use the Apriori algorithm [8].

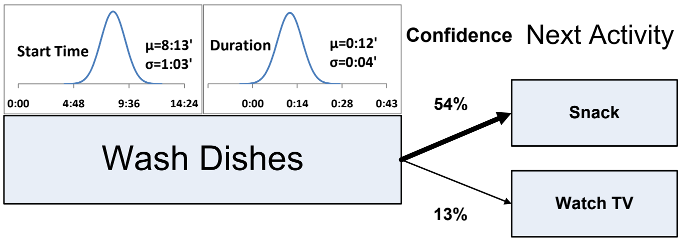


Fig. 3. Temporal relations of the 1st cluster of "Wash Dishes"

To describe the temporal relation discovery component more precisely, lets denote an instance i of an activity a by a_i. The successor activity of a_i in the dataset is denoted by b_j, where j refers to the instance index of activity b. Also as mentioned in the previous section, each activity instance belongs to a specific cluster Θ_k that is defined by the start time of the activity instance. Furthermore, to show that an activity a_i belongs to a specific cluster Θ_k, we denote it by a_i^k. Finally, we will show the temporal relation "b follows a" as $a \rightarrow b$.

Denoting the mean and standard deviation of cluster k as μ_k and σ_k, we refer to the number of instances of all activities and activity a falling within $[\mu_k - 2\sigma_k, \mu_k + 2\sigma_k]$ interval as $|D^k|$ and $|a^k|$ respectively. Then we can define the support of the "follows" relation as in Equation 2 and its confidence as in Equation 3.

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

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

The result of this stage is a set of temporal relation rules corresponding to each cluster. Fig. 3 illustrates the discovered temporal relation rules, whose confidence values are greater than 0.1, for the first cluster of the "Wash Dishes" activity. According to Fig. 3, if "Wash Dishes" activity occurs in the time interval $[7\colon 35, 8\colon 51]$, it usually takes between 4 to 20 minutes and the next activities are typically "Snack" with a confidence of 0.54 and "Watch TV" with a confidence of 0.13.

4 Experimental Results

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

4.1 Experimental Setup

The smart home testbed used in our experiments is a 1-bedroom apartment hosted two married residents who perform their normal daily activities. The

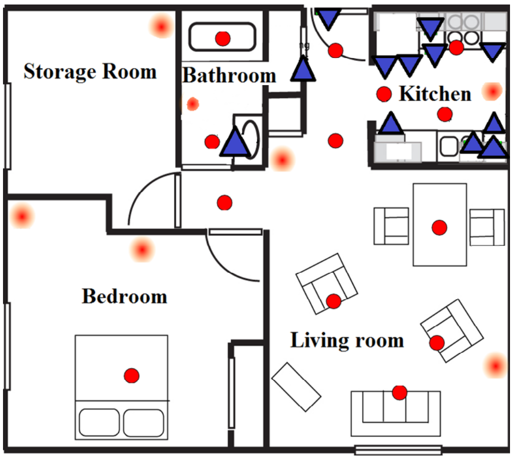


Fig. 4. The sensor layout in the 1-bedroom smart home testbed, where circles represent motion sensors and triangles show door/cabinet sensors

Table 1. EM clustering and Apriori parameters

EM clustering	Max Iterations = 100
	Min Standard Deviation = $1.0e-6$
	Seed = 100
Apriori	Min Support = 0.01
	Min Confidence = 0.1

sensor events are generated by motion and door/cabinet sensors. Fig. 4 shows the sensor layout of our smart home testbed. To track the residents' 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. Our training data was gathered over a period of 4 months and more than 480,000 sensor events were collected for this dataset[1]. For our experiments, we selected 10 ADLs including: Cook Breakfast, R1 Eat Breakfast, R2 Eat Breakfast, Cook Lunch, Leave Home, Watch TV, R1 Snack, Enter Home, Wash Dishes, and Group Meeting; where R1 and R2 represent the residents of the smart home.

Moreover, Table 1 depicts the parameter values for the Expectation Maximization (EM) clustering and Apriori association rule mining in our experiments.

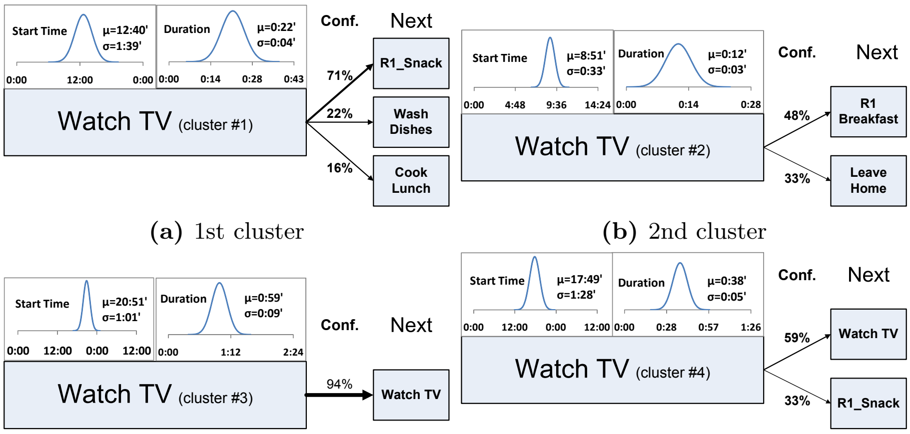
4.2 Validation of TEREDA

In this section, we provide the results of running TEREDA on the smart home dataset. Table 2 shows the number of clusters corresponding to start times of each activity after running TEREDA on the dataset.

[1] Availavle online at http://ailab.eecs.wsu.edu/casas/datasets.html

Table 2. Number of start time clusters for each activity

Activity	# of clusters	Activity	# of clusters
Cook Breakfast	2	R1 Eat Breakfast	2
Cook Lunch	1	R1 Snack	5
Enter Home	3	R2 Eat Breakfast	2
Group Meeting	1	Wash Dishes	2
Leave Home	2	Watch TV	4



(a) 1st cluster **(b)** 2nd cluster **(c)** 3rd cluster **(d)** 4th cluster

Fig. 5. Temporal relations of the "Watch TV" activity

The discovered temporal relations for "Watch TV" activity is illustrated in Fig. 5. As mentioned in subsection 3.1, in order to handle outliers, we only retain values within the [μ-2σ, μ+2σ] interval for start time of each activity. According to Fig. 5a, if activity "Watch TV" occurs in the [9: 22, 15: 58] timeframe, the activity takes approximately 14 to 30 minutes. This activity is typically followed by the "R1 Snack", "Wash Dishes" or "Cook Lunch" activities. The likelihood of occurrence of "R1 Snack" is 0.71, while "Wash Dishes" and "Cook Lunch" occur with likelihoods of 0.22 and 0.16, respectively. Furthermore, Fig. 5b indicates that when "Watch TV" occurs between 7: 45 and 9: 57, it takes approximately 6 to 18 minutes and the next likely activities are "R1 Breakfast" with a confidence of 0.48 and "Leave Home" with a likelihood of 0.33. Moreover, Fig. 5c shows that when the "Watch TV" activity is performed within the timeslice [18: 49, 22: 53], it usually takes around 41 to 1: 17 minutes. In this case, "Watch TV" is most likely followed again by "Watch TV" with a confidence of 0.94. Finally, Fig. 5d shows temporal features and relations of the "Watch TV" activity when it occurs within the [14: 53, 20: 45] interval. In this time interval, the "Watch TV" activity takes approximately 28 to 48 minutes and the next activities are "Watch TV' with a confidence of 0.59 and "R1 Snack" with a confidence of 0.33.

5 Conclusions and Future Work

In this paper, we introduced TEREDA to discover the temporal relations of the activities of daily living. The proposed approach is based on association rule mining and clustering techniques. TEREDA also discovers the usual start time and duration of the activities as mixture normal model. As a future direction, we are planning to use these discoveries in developing activity reminder and abnormal behavior detection systems.

References

1. Pollack, M.: Intelligent Technology for the Aging Population. AI Magazine 26(2), 9–24 (2005)
2. Kautz, H., Etzioni, O., Weld, D., Shastri, L.: Foundations of assisted cognition systems. Technical report, University of Washington, Computer Science Department (2004)
3. Philipose, M., Fishkin, K., Perkowitz, M., Patterson, D., Fox, D., Kautz, H., Hahnel, D.: Inferring activities from interactions with objects. IEEE Pervasive Computing 3(4), 50–57 (2004)
4. Rashidi, P., Cook, D.J.: Keeping the resident in the loop: Adapting the smart home to the user. IEEE Transactions on Systems, Man and Cybernetics, Part A 39(5), 949–959 (2009)
5. Li, Y., Ning, P., Wang, X.S., Jajodia, S.: Discovering Calendar-Based Temporal Association Rules. Data and Knowledge Engineering 44(2), 193–218 (2003)
6. Campbell, N.A.: Mixture models and atypical values. Math. Geol. 16, 465–477 (1984)
7. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society, Series B (Methodological) 39(1), 1–38 (1977)
8. Agrawal, R., Imielinski, T., Swami, A.: Mining Associations between Sets of Items in Large Databases. In: ACM SIGMOD Int'l Conf. on Management of Data, Washington D.C (1993)
9. Galushka, M., Patterson, D., Rooney, N.: Temporal Data Mining for Smart Homes. In: Augusto, J.C., Nugent, C.D. (eds.) Designing Smart Homes. LNCS (LNAI), vol. 4008, pp. 85–108. Springer, Heidelberg (2006) ISSN 0302-9743
10. Jakkula, V.R., Cook, D.J.: Using temporal relations in smart environment data for activity prediction. In: Proceedings of the ICML Workshop on the Induction of Process Models (2007)