LEARNING TEMPORAL RELATIONS IN SMART HOME DATA

Vikramaditya R. Jakkula and Diane J. Cook.
Washington State University,
EME 206, Spokane Street, Pullman,
Washington – 99164, USA.
{vjakkula, cook}@eecs.wsu.edu.

INTRODUCTION

Temporal rule mining and pattern discovery applied to time series data has attracted considerable interest over the last few years. In this paper we consider the problem of learning temporal relations between time intervals in smart home data, which includes physical activities (such as taking pills while at home) and instrumental activities (such as turning on lamps and electronic devices). The purpose of this work is to identify interesting temporal patterns in order to improve prediction of events based on observed temporal relations in a smart home environment.

We propose one such framework to derive temporal rules from a time series representation of observed inhabitant activities in a smart home, and validate the algorithm using both synthetic datasets and real data collected from the MavHome smart environment. This framework is based on Allen's temporal logic [1]. Allen suggested that it was more common to describe scenarios by time intervals rather than by time points, and listed thirteen relations formulating a temporal logic (before, after, meets, meet-by, overlaps, overlapped-by, starts, started-by, finishes, finished-by, during, contains, equals). These temporal relations play a major role in identifying temporal activities which occur in a smart home. Consider as an example, a case where an elderly person takes pills after eating food. We notice that these two activities, taking pills and eating, share the temporal relation "after" between them. And when the relationship is violated, the relationship is updated to "meets" and an anomaly in activity is noted. The objective of this research is to identify temporal relations among daily activities in a smart home and to enhance prediction and decision making with these discovered relations.

Temporal interval discovery based on Allen's interval relations has several disadvantages when used for knowledge discovery and pattern recognition. One of the major disadvantages is its ambiguous nature. As seen in Figure 1, by applying the notion of temporal relations we can identify relations as A "before" B and B "before" C. Finding the best representation for the identified temporal interval is a current challenge. We can simply visualize this representation as A "before" B "before" C. But the question of whether there are other possible interpretations for this relationship arises and we must choose the best representation.

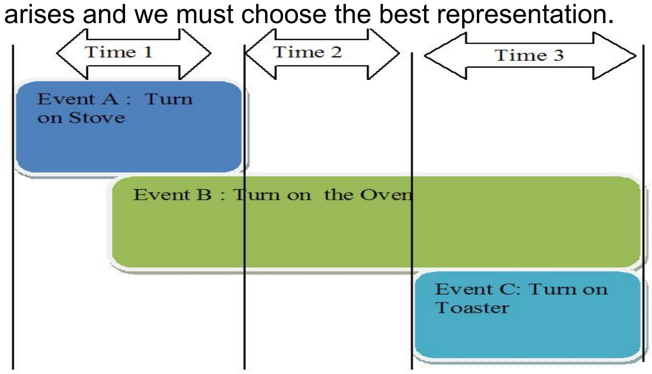

Figure 1:Temporal intervals are labeled as A "before" B "before" C or A "before" B "finishes-by" C.

For example, consider three events A (turn on range top), B (turn on oven), and C (turn on toaster). Figure 1 represents the relationship among the three events A, B, and C. Note that Event A occurs before Event B and Event B occurs before Event C. We can see that A "before" B "before" C is a possible relationship label. However, an alternative representation consistent with the events is A "before" B; B "finishes-by" C. The second interpretation, or relationship, actually changes our perspective of the scenario. In this case when we use the relation B "before" C we know that the event B just occurs before C. In contrast, when we interpret the relationship as B finished-by C, an anomaly can be flagged in cases where B and C do not finish at the same time. If we were to use the earlier relation of B "before" C, such anomalies would not be captured. Thus the relation of B "finished-by" C is a better fit for the relationship illustrated in Figure 1 between events A, B, and C. In this paper, we introduce our temporal relationship discovery algorithm and present experimental results on sample datasets.

Morchen argued that Allen's temporal patterns are not robust and small differences in boundaries lead to different patterns for similar situations [2]. As a possible solution, Morchen presented a Time Series Knowledge Representation, which expresses the temporal concepts of coincidence and partial order.

Although this method appears feasible and computationally sound, it does not suit our smart home application due to the granularity of the time intervals in smart homes datasets. His approach does not involve ways to eliminate noise and the datasets are so huge that computational efficiency would not be the only factor to be considered. Björn, et al. [3] also reason that space and time play essential roles in everyday lives. They discuss several AI techniques for dealing with temporal and spatial knowledge in smart homes, mainly focusing on qualitative approaches to spatiotemporal reasoning.

ENVIRONMENT SENSING & DATA COLLECTION

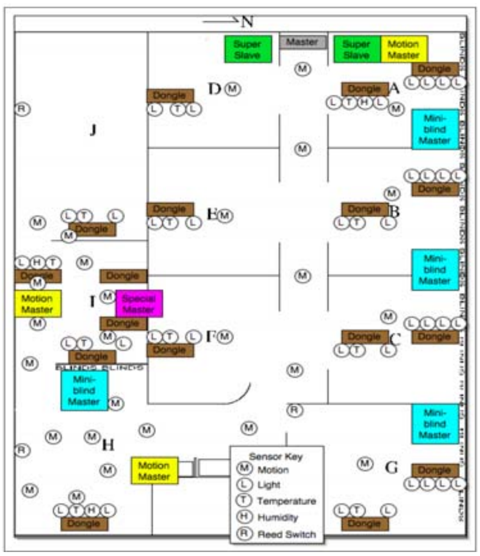

Figure 2: (a) MavPad Argus Sensor Network.

The major goal of the MavHome project [4] is to design a smart environment that acts as an intelligent agent. We define a smart environment as one with the ability to adapt the environment to the inhabitants and meet the goals of comfort and efficiency. In order to achieve these goals the house should be able to predict, reason, and adapt to its inhabitant. In MavLab, sensor network data is the primary source of data collection. The data collection system consists of an array of motion sensors which collect information using X10 devices and our in-house sensor network. Our dataset is based on a single inhabitant working in the MavLab (see Figure 2) and consists of two months of data. The lab consists of a presentation area, kitchen, student desks, and faculty room. There are over 100 sensors deployed in the MavLab that include light, temperature, humidity, and reed switches. In addition, we created a synthetic data generator to validate our approach. We developed a model of a user's pattern which consists of a number of different activities involving several rooms and eight devices. For this paper we generated a data set containing about 4,000 actions representing two months of activity.

EXPERIMENTATION EVALUATION AND RESULTS

Modeling temporal events in smart homes is an important problem and has great advantage to people with disabilities and the elderly. We see that temporal constraints can model normal activities; if a temporal constraint is not satisfied then a potential "abnormal" or "critical" situation may occur. The goal of this experiment is to identify temporal relations in smart home datasets and later use them for decision making (prediction). There are two major problems associated with using Allen's temporal relations. The first problem is the failure of Allen's approach to identify a single most descriptive relation between a pair of events. The second challenge is how to process event relationships in smart home data, which by its nature has a minute time granularity. In our implementation we try to resolve these problems and provide an alternate solution as how the temporal relations can be identified and associated on smart home datasets.

The best way to eliminate ambiguity in identifying the temporal relations is to identify and define the boundary conditions for the thirteen defined intervals before we use it in our algorithm. We illustrate these boundary conditions in Figures 3 through 15, using events X and Y as example events.

X before Y**:** Start Time(X)<Start time(Y); End time(X)<Start time(Y)

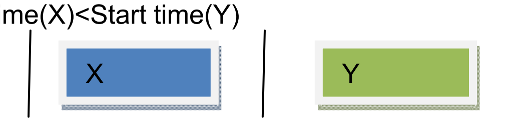

Figure 3: X before Y.

X after Y: Start Time(X)>Start time(Y); End time(Y) < Start time(X)

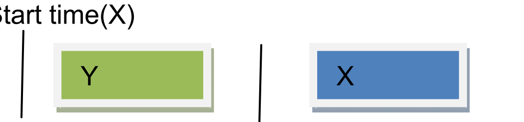

Figure 4: X after Y.

X during Y: Start time(X)>Start time(Y); End time(X) < End time(Y)

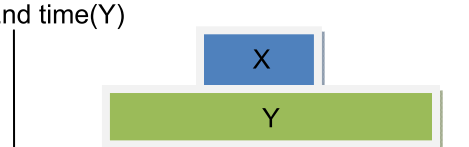

Figure 5: X during Y.

X contains Y: Start time(X)<Start time(Y); End time(X)>End time(Y)

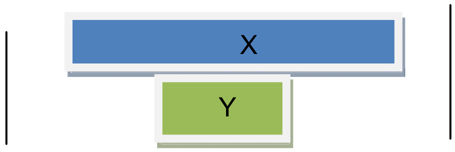

Figure 6: X contains Y.

X overlaps Y: Start time(X)<Start time(Y); Start time(Y)<End time(X); End time(X)<End time(Y)

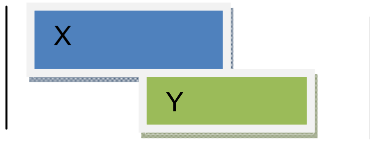

Figure 7: X overlaps Y.

X overlapped-by Y: Start time(Y)<Start time(X); Start time(X)<End time(Y); End time(Y)<End time(X)

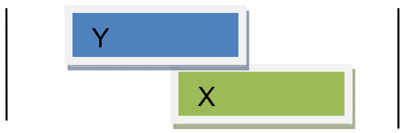

Figure 8: X overlapped-by Y.

X meets Y: Start time(Y) = End time(X)

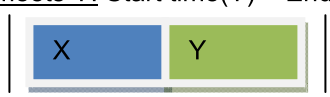

Figure 9: X meets Y.

X met-by Y: Start time(X) = End time(Y)

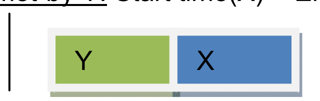

Figure 10: X met-by Y.

X starts Y: Start time(X) = Start time(Y); End time(X) ≠ End time(Y)

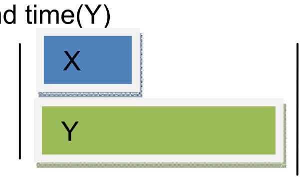

Figure 11: X met-by Y.

X started-by Y: Start time(Y) = Start time(X); End time(X) ≠ End time(Y);

Figure 12: X started-by Y.

X finishes Y: Start time(X)≠start time of Y; End time(X) = End time(Y)

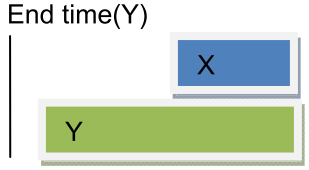

Figure 13: X finishes Y.

X finished-by Y: Start time(X) ≠ Start time(Y); End time(X) = End time(Y)

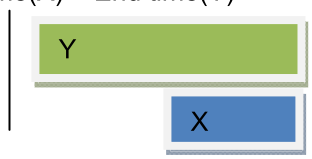

Figure 14: X finished-by Y.

X equals Y: Start time(X) = Start time(Y); End time(X) = End time(Y).

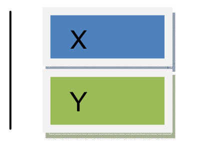

Figure 15: X equals Y.

The thirteen temporal relations are well identified and defined by the boundary conditions as stated above. These conditions are used in the algorithm for identifying temporal intervals. The First step of the algorithm applies the Apriori algorithm [5] as shown in the pseudo code below to identify the most frequent event pairs whose relationships fit one of the thirteen categories. This step addresses the puzzle of handling minute time granularity.

Apriori Algorithm Pseudo-code:

C_k: Candidate itemset of size k
L_k : frequent itemset of size k
L_1 = {frequent items};
for (k = 1; L_k !=∅; k++) **do begin**
  C_{k+1} = candidates generated from L_k;
  for each day t in datasets do
    increment the count of all candidates in C_{k+1}
that are contained in t
  L_{k+1} = candidates in C_{k+1} with min_support
  end
return $\cup_k L_k$;

The next step of the algorithm applies a weighting to the mined temporally-related pairs. This weighting helps us handle ambiguous situations where a temporal relation can be labeled as two or more types.

To resolve the issue, select the temporal label that occurs the most often in the observed data between the targeted pair of events.

$$\text{Weight} = \frac{\text{No of Occurrence of specific Relation}}{\text{Total Identified Relations}} \qquad (1)$$

We apply our data mining algorithm to identify and label frequently-related events in our smart home data. When validating the algorithm, we set the parameters as specified in Table 1. We also note that the algorithm is constrained in this case to only identify event pairs that co-occur in the same day.

Table1: Parameter settings for experimentation.

Datasets	Parameter Setting			
	No of Days	No of Events	No of Intervals Identified	Size of Data
Synthetic	60	8	1729	106KB
Real	60	17	1623	104KB

Table 2 describes the results of varying the minimum support to find the number of itemsets for the real and synthetic datasets. As shown in Table 2, as the minimum support is lowered the runtime correspondingly increases.

Table 2: Minimum Support vs. No of Itemset Identified

Minimum Support	No of Frequent Itemsets	
	Real Dataset	Synthetic Dataset
8	3	2
4	5	3
2	10	4

We then further analyzed the frequent itemsets identified with minimum support of value 2 to identify key temporal relations by using the weight based temporal analyzer algorithm Figure 16 illustrates the identified temporal relations in these datasets. Table 3 displays samples of identified frequent relation pairs.

CONCLUSION

Temporal analysis is smart environments would be a boon for elderly inhabitants, as this would enable them to lead a easier life. The discovery of temporal pattern would help them associate their daily activity and if a violation is found, such as, forgetting to take pills, a quick reminder can be sent and also would help them identify the next set of activities which might occur when we have a prediction component added to this temporal analyzer. Future work would include experimenting with prediction using the discovered temporal patterns. We will also analyze the performance of predictors that use the identified temporal relationship information.

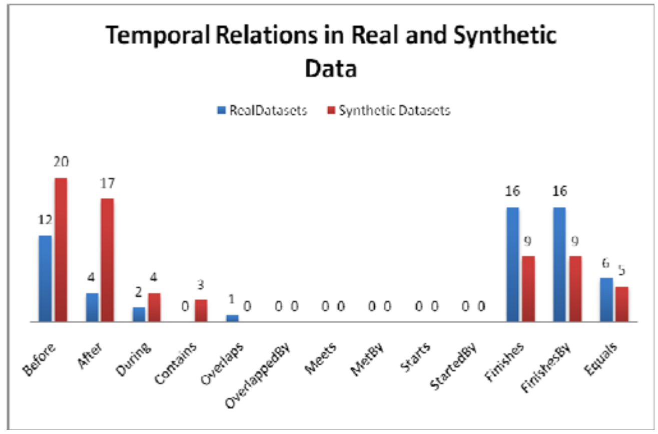

Figure 16: Types of identified temporal relations.

Table 3: Sample of Frequent Relation Pairs.

Frequent Relation Pairs on Real Datasets
Lamp Sensor J16 BEFORE Radio Sensor J11
Lamp Sensor I14 AFTER Lamp Sensor C9
Lamp Sensor I4 EQUALS Lamp Sensor I4
Frequent Relation Pairs on Synthetic Datasets
Cooker Before Oven
Fan After Cooker
Lamp Before Cooker

ACKNOWLEDGEMENTS

This work is supported by NSF grant IIS-0121297.

REFERENCES

[1] James F. Allen, and George Ferguson, *Actions and Events in Interval Temporal Logic*, Technical Report 521, July 1994.

[2] Mörchen, F. *A better tool than Allen's relations for expressing temporal knowledge in interval data*, In Proceedings the Twelveth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Philadelphia, PA, USA, (2006).

[3] Björn Gottfried, Hans W. Guesgen, and Sebastian Hübner, *Spatiotemporal Reasoning for Smart Homes,* Designing Smart Homes, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, Pg 16-34, Volume 4008/2006, July 2006.

[4] G. Michael Youngblood, Lawrence B. Holder, and Diane J. Cook. *Managing Adaptive Versatile Environments*, Proceedings of the IEEE International Conference on Pervasive Computing and Communications. (2005)

[5] Rakesh Agrawal and Ramakrishnan Srikant, *Fast Algorithms for Mining Association Rules*, Proc. 20th Int. Conf. Very Large Data Bases, Morgan Kaufmann,487--499, 1994.