TEMPORAL PATTERN DISCOVERY FOR ANOMALY DETECTION IN A SMART HOME

Vikramaditya Jakkula[†], Diane J. Cook [†]

[†]Washington State University
EME 206, Spokane Street, Pullman,
Washington – 99164, USA.
{vjakkula, cook}@eecs.wsu.edu

Keywords: Knowledge discovery, smart homes, anomaly detection, temporal relations, data mining.

Abstract

The temporal nature of data collected in a smart environment provides us with a better understanding of patterns over time. Detecting anomalies in such datasets is a complex and challenging task. To solve this problem, we suggest a solution using temporal relations. Temporal pattern discovery based on modified Allen's temporal relations [5] has helped discover interesting patterns and relations on smart home datasets [10]. This paper describes a method of discovering temporal relations in smart home datasets and applying them to perform anomaly detection process on the frequently-occurring events. We also include experimental results, performed on real and synthetic datasets.

1 Introduction

The need for a robust anomaly detection model is as essential as a prediction model for any intelligent smart home to function in a dynamic world. For an agent to perform anomaly detection, it should be capable of applying the limited experience of environmental event history to a rapidly changing environment, where event occurrences are related by temporal relations. Temporal rule mining has been attracting some considerable attention over the decade [2]. In this paper we consider the problem of anomaly detection based on discovery and application of temporal relations. Anomaly detection can be an important feature of a smart environment. For example, if we are monitoring the well being of an individual in a smart home, that individual has not opened the refrigerator for an entire day, and this is normally a frequent activity, this would be an important anomaly to report to the individual and the caregiver.

Allen suggested that it was more common to describe scenarios by time intervals rather than by time points, and listed thirteen relations comprising a temporal logic: before, after, meets, meet-by, overlaps, overlapped-by, starts, started-by, finishes, finished-by, during, contains, equals [5]. These temporal relations play a major role in identifying temporal activities which occur in a smart home. Consider, for instance, a case where the inhabitant turns the Television (TV) on before sitting on the couch. We notice that these two activities, turning on the TV and sitting on the couch, are frequently related in time according to the "before" temporal relation. Therefore, when the relationship is violated, an anomaly is noted.

A question may arise as to why Allen's temporal relations should be used for generating temporal intervals. The temporal relations defined by Allen form the basic representation of temporal intervals, which when used with constraints become a powerful method of expressing expected temporal orderings between events in a smart environment. In addition, they have an easy naming convention, making it easier to recognize, interpret and use the temporal relations that are identified. There are projects which employ sequential information to predict activities [4], and other methods for identifying suspicious states in a smart environment have been researched [9].

We extend these methods to incorporate valuable information about the interval of time each event spans. While other methods treat each event as a separate entity (including, for example, turning on a lamp and later turning off the same lamp), our interval-based analysis considers these two events as members of one interval. Each interval is expressed in terms of start time and end time values. As a result, temporal relationships between such intervals can be identified and used to perform critical anomaly detection.

We focus only on our objective to develop a model for analysing anomalies in the occurred interval-based events using temporal relations for the most frequently-occurring events, because discovering all possible patterns can be computationally inhibitive. Furthermore, the number of results obtained can be overwhelming, taxing the ability to effectively use the discovered results for practical purposes. In this paper, we introduce a temporal representation to express relationships between interval-based events. We build on this representation to identify frequently-occurring relationships between temporal events, and use the results as the basis for performing anomaly detection. We explain the temporal relations with illustrations and also include a brief description of the temporal intervals formation process. We describe the steps involved in the experimentation for anomaly detection and present the results obtained on real and synthetic datasets.

2 Related work

Morchen argued that Allen's temporal patterns are not robust and small differences in boundaries lead to different patterns for similar situations [8]. 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. [1] also reasons 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.

Other work on anomaly detection was performed on health datasets to check for outliers and drifts in smart homes [3]. This approach is based on regression and correlation on numerical-based health datasets and would not apply to activities which consist of devices or actions, for instance, turning on and off of devices in smart home. Furthermore, this approach considers each event is occurring in a single instant, and therefore overlooks the time interval encompassed by an event.

An extended application of anomaly detection is its use for reminder assistance. Autominder [6], an intelligent cognitive orthotic system for people with memory impairment, employs techniques such as dynamic programming and Bayesian learning to remind individuals about their planned Activities for Daily Living.. Autominder includes a web-based interface for plan initialization and constructs rich models of a inhabitant's activities—including constraints on the times and ways in which activities should be performed—to monitor the execution of those activities. Autominder looks for differences between expected and observed activities, and reasons about whether to issue reminders.

3 Environment sensing and data collection

The MavHome Project is a multi-disciplinary project, which has been engaged in the creation of adaptive and versatile home and workplace environments in the past few years [7]. Its goal 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 with minimum cost in mind. In order to achieve these goals, the house should be able to predict, reason, and adapt to its inhabitant. In MavLab, the sensor network data is the primary source of data collection. This project consists of two implementations of a smart environment. One is a smart apartment called the MavPad and another is at the research lab called the MavLab. The real datasets used for the experimentation process were collected from this project.

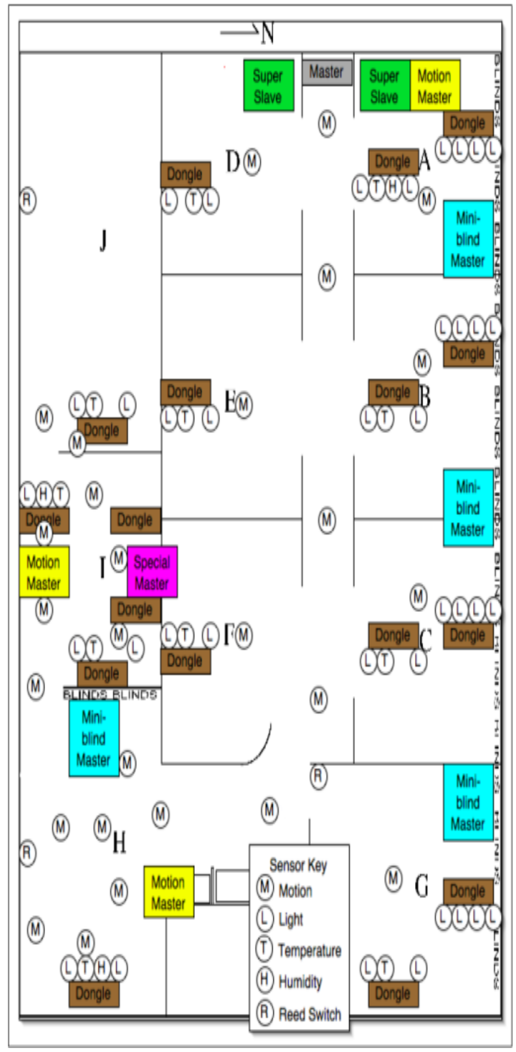

Figure 1: MavLab Argus Sensor Network (M-Motion sensor, L-Light sensor, T-Temperature sensor, H-Humidity sensor, R- Reed switch sensor, S-Smoke sensor, C- Gas sensor).

The data collection system consists of an array of motion sensors, which collect information using X10 devices and the in-house sensor network. Our dataset is collected for an inhabitant working in the MavLab (see Figure 1) 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 activities.

4 Temporal relations

Activities in a smart home include physical activities as well as instrumental activities. These may include walking, sitting on a couch, turning on a lamp, using the coffeemaker, and so forth. We see that these activities are not instantaneous, but have distinct start and end times. We also see that there are well-defined relationships between time intervals for different activities. These temporal relations can be represented using Allen's temporal relations and can be used for knowledge and pattern discovery in day-to-day activities. These discoveries can be used for developing systems which can act as reminder assistants and also help detect anomalies and aid us in taking preventive measures.

Let us consider a scenario which involves a television, fan and a lamp being used in a smart home. We see that the inhabitant turns on the television and after some period of time turns on the fan. As time progresses, feeling cold, the fan is turned off and the individual continues watching the television. Later on, the television is turned off and the individual turns on the lamp to illuminate the room. We see that this scenario involved three activities each defined by interaction with a single device, namely a television, a fan and a lamp. Now we apply Allen's logic to establish the temporal relations among the activities which occurred. The scenario is illustrated in figure 2. These activities can be represented as television "contains" fan and "meets" lamp. We can also represent these relationships as television "meets" lamp and fan "before" lamp.

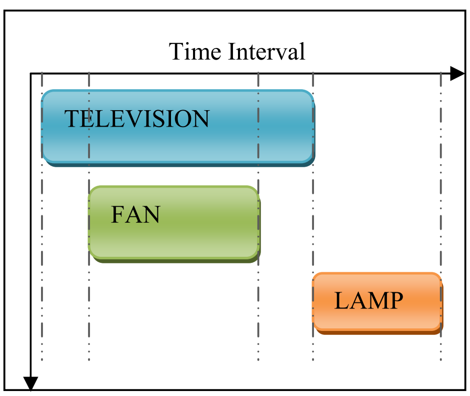

Figure 2: Illustration of Temporal Intervals

Modelling temporal events in smart homes is an important problem and offers advantages to people with disabilities for assisted living. We see that the temporal constraints can model casual activities; if a temporal constraint is not satisfied then a potential "abnormal" or "critical" situation may have occurred. The goal of this experiment is to identify anomalies in activities using temporal relations in smart home datasets. We use constraints based on Allen's temporal relations to model temporal relations in smart home datasets [10]. The thirteen temporal relations are illustrated in Table 1. Consider two general events X and Y; we use this to represent the relations in the table. In Table 1, the interval constraints compare the start time (Start) and end time (End) of the activities, X and Y.

Temporal Relations	Pictorial Representation	Interval constraints
X Before Y	X Y	Start(X)<Start(Y); End(X)<Start(Y)
X After Y	Y X	Start(X)>Start(Y); End(Y)<Start(X)
X During Y	X Y	Start(X)>Start(Y); End(X)<End(Y)
X Contains Y	Y X	Start(X)<Start(Y); End(X)>End(Y)
X Overlaps Y	X Y	Start(X)<Start(Y); Start(Y)<End(X); End(X)<End(Y)
X Overlapped-By Y	Y X	Start(Y)<Start(X); Start(X)<End(Y); End(Y)<End(X)
X Meets Y	X Y	Start(Y) = End(X)
X Met-by Y	Y X	Start(X)= End (Y)
X Starts Y	X Y	Start(X)=Start(Y); End(X)≠End(Y)
X started-by Y	Y X	Start(Y)=Start(X); End(X)≠End(Y)
X Finishes Y	X Y	Start(X)≠start(Y); End(X) = End (Y)
X Finished-by Y	Y X	Start(X)≠Start(Y); End(X)=End(Y)
X Equals Y	X Y	Start(X)=Start(Y); End(X)=End(Y)

Table 1: Temporal Relations representation [10].

5 Experimentation and results

The first step of the process involves identification of the frequent activities, or events, which occur during the day and establishing temporal relations among them. To accomplish this task, we mine the data for frequent itemsets using the Apriori algorithm [10]. Next, we identify observed temporal relations between events in these frequent itemsets. We limit our approach to frequent activities, because the smart home data is so huge that there are many potential anomalies which are just factors of noise and because the datasets are prohibitively large [10]. The final step involves calculating the evidence of the event occurrence, which can be used for calculating the anomaly.

There are two ways to look at anomaly detection using temporal relations. The first step is to detect whether the particular event satisfies the temporal relations which can be used for anomaly detection. These temporal relations which can be used for anomaly detection are listed in Table 2. Let us look at an example where we have three frequent activities and are in the order of toaster , table lamp and radio, turning on and off in the morning. We see that the relation exhibited by them can be toaster "before" table lamp "finishes" radio. Now when the toaster and the radio occur without the table lamp, we can note that this is an anomaly in activity. This method of anomaly detection is based entirely on normative behaviour as observed in the past. As a result, the likelihood of detecting false anomalies increases when there is even the slightest change in inhabitant patterns which actually may not be anomalies.

The second method is a probability-based model which involves calculating the evidence supporting the currently-occurring activity with respect to the previously-occurred activates and determining whether the current activity is an anomaly or not. This evidence calculation draws from the temporal relations which can be used for anomaly detection and which are listed in Table 2. Because of its robustness in the presence of noise, we decide to use this approach for our anomaly detection.

Temporal Relations	Usable for Anomaly Detection
Before	Yes
After	No
During	No
Contains	Yes
Overlaps	Yes
Overlapped-By	No
Meets	Yes
Met-By	No
Starts	Yes
Started-By	Yes
Finishes	Yes
Finished-By	Yes
Equals	Yes

Table 2: Temporal relations which are used for anomaly detection.

Let us focus our attention on events X and Y. We describe the anomaly detection process for event X given information about an event Y that exhibits a temporal relationship with X in the following steps.

Step 1: Learn temporal relations from the observed event history by analyzing the events and deriving the frequent itemsets. We also identify the most current activity (in our example, this is event Y). (Note that this tool can run simultaneously with a prediction component, if one exists).

Step 2: Now we calculate the evidence supporting the occurrence of activity X. The formula to calculate the evidence using temporal relations is given by Equation (1). Note that equation is based on the observed frequency of the temporal relations, specifically those that influence the occurrence of event X.

$$P(X) = P(X|Y) = |\, \text{Before}(X,Y) + \text{Contains}(X,Y) + \text{Overlaps}(X,Y) + \text{Meets}(X,Y) + \text{Starts}(X,Y) + \text{StartedBy}(X,Y) + \text{Finishes}(X,Y) + \text{FinishedBy}(X,Y) + \text{Equals}(X,Y)\, | \;/\; |Y| \quad (1)$$

$$\text{Evidence}_X = P(X) \quad (2)$$

The previous discussion showed how to calculate the likelihood of event X given the occurrence of one other event Y. Now consider the case where we want to combine evidence from multiple events that have a temporal relationship with X. In our example we have observed the start of event A and the start of event B, and want to establish the likelihood of event X occurring. From Equation 1 we can calculate Evidence
Now we have the evidence of A as:

$$P(A|X) = |\, \text{Before}(X,A) + \text{Contains}(X,A) + \text{Overlaps}(X,A) + \text{Meets}(X,A) + \text{Starts}(X,A) + \text{StartedBy}(X,A) + \text{Finishes}(X,A) + \text{FinishedBy}(X,A) + \text{Equals}(X,A)\, | \;/\; |X| \quad (3)$$

Similarly when we have the events occurred as follows: **X A B**
Now the evidence of B is calculated as follows:

$$\begin{aligned} P(B|AUX) &= P(B \cap (AUX)) / P(AUX) \\ &= P(B \cap A) \text{ U } P(B \cap X) / P(A) + P(X) - P(A \cap X) \quad \text{[Association Rule]} \\ &= P(B|A).P(A) + P(B|X).P(B) / P(A) + P(X) - P(A \cap X) \quad \text{[Multiplication Rule]} \quad (4) \end{aligned}$$

And we see that, we can use the previous calculated evidence for calculating newer evidence, based on the equation (4). We see that in this equation (4), uses Association rule and Multiplicative rules to arrive at the final formula which includes previous computed evidences of occurred events. And we use them to calculate the evidence of the most recent occurred event. In this way we compute the evidence of the occurred events and applies to the entire series of events that occurred.

Step 3: Now we finally calculate the anomaly of X by the equation (5) given below.

$$\text{Anomaly}_X = 1 - P(X) \quad (5)$$

Notice that if the event has a probability approaching 1 and has occurred, this is not considered an anomaly. On the other hand, if the probability of the event we just observed is close to 0, then this is an unusual event and should be considered an anomaly. The point at which these anomalies are considered surprising enough to be reported is based somewhat on the data itself. If the probability of an event is based on the occurrence of other events which themselves rarely occur, then the evidence supporting the occurrence of the event is not as strong. In this case, if the event has a low probability yet does occur, it should be considered less anomalous than if the supporting evidence itself appears with great frequency. Consistent with this theory, we calculate the mean and standard deviation of event frequencies over the set of events in the inhabitant's action history. Events are reported as anomalies (or, conversely, the absence of an event) if it does occur and its anomaly value is greater than the mean + 2 standard deviations.

We validate our algorithm by applying it to our real and synthetic datasets. We train the model based on 59 days of data and test the model on one day of activities. We use the training set to form the frequent item sets and identify temporal relations shared between them. The temporal relations formed in these data sets show some interesting patterns and indicate relations that are of interest. The parameter settings pertaining to the training set data are given in Table 3. The parameter settings pertaining to the test set data are given in Table 4.

	Parameter Setting			
Datasets	Number of Days	Number of Possible Events	Number of Intervals Identified	Size of Data
Synthetic	59	8	1703	105KB
Real	59	17	1523	103KB

Table 3: Parameters setting for training set.

	Parameter Setting			
Datasets	Number of Days	Number of Possible Events	Number of Intervals Identified	Size of Data
Synthetic	1	8	17	2KB
Real	1	17	9	1KB

Table 4: Parameters setting for test set.

Next we perform frequent itemset mining and identify the most frequent activities in the training dataset. Then we read these temporal relations into our anomaly detection tool which dynamically calculates evidence for each possible event and outputs anomalies that are detected in the test set data. We manually look at the data and reported results to determine the number of true and false anomalies that are detected.

The results form the real dataset are displayed in Table 5 and the results form the synthetic dataset are displayed in Table 6. These results list anomalies in chronological order based as actual events are observed and evidence for expected events is calculated. The graph in Figure 3 visualizes the anomalies that are detected in each database.

Frequent Event	Evidence	Anomaly	Detected
J10	0.45	0.55	No
J11	0.32	0.68	No
A11	0.33	0.67	No
A15	0.24	0.76	No
A11	0.23	0.77	No
A15	0.22	0.78	No
I11	0.27	0.73	No
I14	0.34	0.66	No
Anomaly Mean	0.7		
Anomaly St. Dev.	0.071764		
Anomaly Cut-off Threshold	0.8435		

Table 5: Anomaly detection in the test set for the real dataset.

We are identifying anomaly in the most frequent activity and we see that the real dataset has a lesser number of frequent activities in the test set compared to synthetic dataset. This is one of the reasons as we had no anomaly detected in the test set of real data. We also note that the total number of distinct events in the synthetic dataset is less compared to real dataset, thus we have used smaller test set for this experiment and subsequently have observed that the anomaly computed in real dataset was lower. But on the synthetic test set we observe, that the anomaly detection performed well, and we could identify anomaly.

Frequent Event	Evidence	Anomaly	Detected
Lamp	0.3	0.7	NO
Lamp	0.23	0.77	NO
Lamp	0.01	0.99	YES
Fan	0.32	0.68	NO
Cooker	0.29	0.71	NO
Lamp	0.45	0.55	NO
Lamp	0.23	0.77	NO
Lamp	0.01	0.99	YES
Lamp	0.23	0.77	NO
Fan	0.3	0.7	NO
Cooker	0.34	0.66	NO
Lamp	0.33	0.67	NO
Lamp	0.2	0.8	NO
Lamp	0.02	0.98	NO
Lamp	0.002	0.998	YES
Fan	0.34	0.66	NO
Cooker	0.42	0.58	NO
Anomaly Mean		0.763412	
Anomaly St. Dev.		0.135626	
Anomaly Cut-off Threshold		1	

Table 6. Anomaly detection in the test set for the synthetic dataset

We have plotted the computed anomaly values of real and synthetic data test sets in the figure 3 below. We see that the spikes visible in the synthetic datasets are clear indication of anomaly.

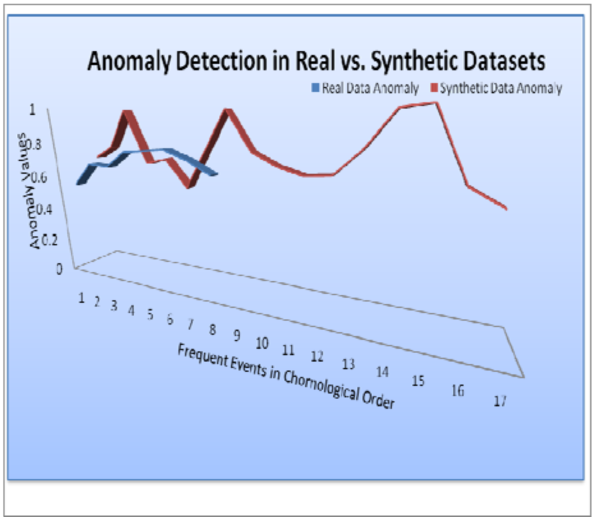

Figure 3: Anomaly detection on test sets of real and synthetic data. The anomaly value is plotted for each possible activity as actual events are observed.

6 Conclusions and future work

Temporal rule based pattern analysis is a niche area in temporal mining world. We notice that the use of temporal relations provides us a unique new approach for anomaly detection. The current approach is currently experimented on small datasets, but we will next validate the performance of our algorithm on larger datasets. Some future directions of this work also include improving activity prediction using temporal relations in smart home data. We will also expand the temporal relations by including more temporal relations, such as until, since, next, and so forth, to create a richer collection of useful temporal relations.

Acknowledgements

This work is supported by NSF grant IIS-0121297.

References

[1] 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.

[2] W. Frawley, G. Paitesky-Shapiro, C. Matheus, *Knowledge Discovery in Databases: An overview,* AI Magazine, Vol 13 n. 3 (57-70). 1992.

[3] Gaurav Jain, Diane J. Cook, and Vikramaditya Jakkula, *Monitoring Health by Detecting Drifts and Outliners for a Smart Environment Inhabitant*, 4th International Conference on Smart Homes and Health Telematics, June 2006.

[4] K. Gopalratnam and D. J. Cook, *Active LeZi: An Incremental Parsing Algorithm for Sequential Prediction*, International Journal of Artificial Intelligence Tools, 14(1-2):917-930, 2004.

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

[6] Martha E. Pollack. *Intelligent Technology for an Aging Population: The use of AI to assist Elders with Cognitive Impairment*. 2005. AI Magazine summer issue. 26(2): 9-24. AAAI Press.

[7] 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.

[8] 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.

[9] Sebastian Lühr, Svetha Venkatesh, Geoff A. W. West, Hung Hai Bu, *Explicit State Duration HMM for Abnormality Detection in Sequences of Human Activity*, PRICAI 2004: 983-984.

[10] Vikramaditya Jakkula, and Diane J. Cook, *Learning temporal relations in smart home data*, Proceedings of the second International Conference on Technology and Aging, Canada, June 2007.