Evident Behavior Analysis on ADLs

Xinran Yu*
The University of Texas at San Antonio
One UTSA Circle
San Antonio, Tx, 78249
xyu@cs.utsa.edu

Turgay Korkmaz†
The University of Texas at San Antonio
One UTSA Circle
San Antonio, Tx, 78249
turgay.korkmaz@utsa.edu

ABSTRACT

Analysis of Activities of Daily Living (ADLs) has been a new research area recently. People perform ADLs in a sequential manner. To analyze different individuals' behavior from ADLs, in this paper, we are using **Super-sequence Frequent Pattern Mining** (SSFPM). Most existing sequence mining research are focusing on looking for frequently occurring sub-sequences, but sometimes, we may also would like to know if there are rare ones occurring before or after the frequent ones. SSFPM can seek not only the frequently happening activity series but also the uncommon ones in-between. So it can extract the evident behavior series as long as the rare events along them. We adopt an already existing algorithm to solve the problem and show interesting results in the experiment section.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database applications—*data mining*

Keywords

ADLs; sequence pattern; graph model

1. INTRODUCTION

Monitoring human Activities of Daily Living (ADLs) has been researched a lot in recent years. This process is based on using sensors to collect different human behaviors [1]. For example, sensors can be deployed in different rooms in a house or be attached to human body [2]. According to the movements and positions of the person, various actives are recorded by the sensor reading. The data collected on ADLs can be used on different applications, e.g., the analysis of the daily living habits can be used in elder people assistance and so on.

Most existing research on ADLs are focusing on methods and algorithms to recognize and activities according the data read from the sensor since sensor readings are noisy [1]. In this paper, we would like to analyze the data that are collected and generated as the ADLs sequence. From the sequence we would like to look into the patterns of different individuals' behaviors and apply the analysis to related applications.

Table 1 shows an example of a person's ADLs in a period of time. We can see that the third column of the different ADLs can be taken as a sequence based on the time order. Through analyzing this ADLs sequence, patterns can be found out and help analysis the individual's behavior.

Frequent pattern mining for sequential dataset has been researched a lot for decades [3]. Traditional sequential pattern mining seeks the frequently occurring ordered (sub)sequences as patterns. One of the first solutions was AprioriAll [4], which extends the similar techniques in Apriori [5]. Candidate sequences are generated by measuring their support when repeating passing over the database. Some of the other well-known sequential mining algorithms are Apriori-based GSP (Generalized Sequential Patterns) [6], Pattern-growth based FreeSpan [7] and PrefixSpan [8], Vertical format-based SPADE [9] and Constraint-based SPIRIT [10]. All of them are mainly about the sub-sequence mining, which is to find the common part of all the sequences in a dataset. Finding out frequent occurring sequences in ADLs can help understand which activities as a series happen a lot for an individual, but sometimes we may want to find out the activities that repeat a lot as well as the activities that do not happen often in between. For example, a person always watch TV after lunch but one day he went to sleep after lunch and then watched TV which is a rare event. Detecting such rarely happening events can help in finding out emergencies as well as other issues. In this paper we are using the definition of **Super-sequence Frequent Patterns** (SSFP) which includes evident events that occurring both frequently and non-frequently as a sequence.

The remaining part of the paper is organized as follows. In Section 2 we introduce super-sequence pattern mining problem and give the graph model. In Section 3 we present SSFPM on two real world ADLs datasets. Finally, we conclude this paper and point out some issues for future research in Section 4.

*Ph.D candidate, Computer Science, The University of Texas at San Antonio
†Associate Professor, Computer Science, The University of Texas at San Antonio

KDD'14, August 24–27, 2014, New York, NY, USA.

Table 1: An example of ADLs dataset

start time	end time	ADL type
02:27:59	10:18:11	Sleeping
10:21:24	10:23:36	Toileting
10:25:44	11:33:00	leaving
11:34:23	11:43:00	Grooming
11:49:48	11:51:13	Spare-time/TV
11:51:41	12:05:07	Toileting
12:06:04	13:06:31	leaving
13:07:04	13:06:31	Grooming

2. PROBLEM FORMULATION AND GRAPH MODEL

Given an ADLs sequence S, we would like to seek an evident set of activities, in which both frequently and non-frequently happening activities are indicated as a series. There is multiple ways to define which factors should be considered for evident series of events. In this paper, to be simple, we only consider the happening times of each transition of activities. For example, if one morning, an individual did toileting after sleeping, there is one transition from sleeping to toileting. If he did these two things in the same order again the second day, there is another same transition and the total transition number from sleeping to toileting now is 2.

SSFPM was introduced by Yu et. al [11] in 2013 and is described as follows:

Definition 1. *Given a sequential dataset S, a sequence length k, and a threshold δ, the **Super-Sequence Frequent Pattern Mining** (SS-FPM) problem is to find the set $P = \{p_1, p_2, ...p_r\}$, where each pattern $p_j = \{w_1, w_2, ..., w_k\}$ satisfies the following condition*

$$support(p_j) = \sum_{i=1}^{k-1} support(w_i w_{i+1}) > \delta.$$

In this paper, ADLs dataset actually is one long sequence. For example, in Table 1, the ADLs sequence has length 8. If we ignore the time period, we can create a graph showing each transition's number, as shown in Figure 1, and we call this graph the ADLs transition graph.

Definition 2. *Given an ADLs sequence S, an* **ADLs Transition Graph** *of S is a directed weighted graph $G = (V, E)$, where $V = \{a_1, a_2, ...a_n\}$ and $E = \{e_1, e_2, ..., e_m\}$. Each edge e_i connects a_x and a_y iff there is a transition from a_x and a_y in S. The weight of edge e_i is the total number of transitions from a_x to a_y in S.*

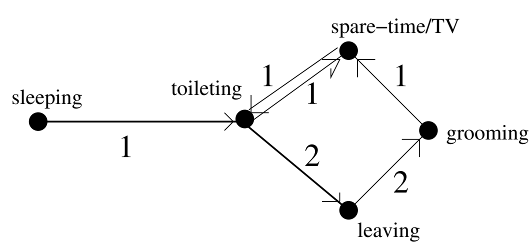

Figure 1: The sequence graph for Table 1

After converting the ADLs sequence to a transition graph, the problem of seeking the most evident series of ADLs becomes the problem of seeking the **heaviest path** of length k. If we would like to seek a length-4 most frequent super-sequence, then our algorithm will found sleeping→ toileting→ leaving→ grooming and toileting→ leaving→ grooming→ spare-time/TV as both paths have the maximum weight of 5. Trying different values of k can help find rare events along the super-sequence patterns. For example, if we set $k = 3$ on Table 1, we will find that toileting→ leaving→ grooming is the heaviest transition path. Both of the transition are 2. When we set $k = 4$ we can find sleeping→ toileting→ leaving→ grooming and toileting→ leaving→ grooming→ spare-time/TV, in which sleeping→ toileting and grooming→ spare-time/TV is relatively rare than toileting→ leaving→ grooming.

Yu et. al [11] proposed a heuristic algorithm on seeking the heaviest paths in a directed weighted graph in polynomial time. In their experiments, they showed efficiency and effectiveness of the algorithm. In this paper, we are going to adopt that algorithm to seek the heaviest paths of length k on ADLs transition graphs.

3. EXPERIMENTS AND ANALYSIS

In this section, we are going to apply the algorithm in [11] in two real world ADLs datasets.

The ADL datasets are called *OrdonezA_ADLs* and *OrdonezB_ADLs*, both of which are obtained from UCI Machine Learning Repository [12].[1] *OrdonezA_ADLs* is at a four-room house and has nine ADLs which are Leaving, Toileting, Showering, Sleeping, Breakfast, Lunch, Snack, Spare_Time/TV and Grooming. The dataset contains the individual's activity list of 248 records in 14 days. *OrdonezB_ADLs* is at a five-room house and has ten ADLs which are Leaving, Toileting, Showering, Sleeping, Breakfast, Lunch, Dinner, Snack, Spare_Time/TV and Grooming. The dataset contains the individual's activity list of 493 records in 21 days.

Since in both of the ADLs datasets have about nine different activities, it would involve loops and repeating sequences if we set the spur-sequence patten length too long. In this experiment, we would like to investigate the super-sequences with length k equals 4 and 5. Table 2 and 3 show the top 5 sequences for length four and five in dataset *OrdonezA_ADLs*. Table 4 and 5 show the top 5 sequences for length four and five in dataset *OrdonezB_ADLs*.

From the tables, we can see that for the *OrdonezA_ADLs* dataset, the routine that happened most often for this individual is using toilet and then spending spare_time or watching TV and then doing grooming and then showering. If we would like to investigate another one activity after grooming, then breakfast happens the most. Also from Table 3 we can see that in the last two rows, the weights on the path are not evenly distributed. For example the last one is saying "grooming→ spare_time/TV → toileting→ showering→ breakfast" and the weight on "toileting→ showering" is only 2. This shows that although there is a lot of showering after toileting and breakfast after showering, it's uncommon that toileting followed by showering. For the second dataset,

[1]The ADL data that we use is from UCI Machine Learning Repository (http://archive.ics.uci.edu/ml/index.html) and was donated by Francisco Javier Ordóñes [12].

in table 4, the most common routine for this individual is showering followed by toileting and then sleeping and at last leaving. We can see that these two individuals have different habits for example the first one rarely use toilet after shower but it is common for the second one. Also for the second person, in table 5, the unevenly distributed sequence happened in row 2, "showering→ toileting→ snack→ leaving→ sleeping". This shows that this person often uses toilet after shower and sleeps when comes back after leaving, but between them he doesn't have snack that often. These behavior patterns can help researchers or health care providers to measure the functional status of a certain group of people or patients, e.g., to provide proper care accordingly.

Table 2: Top 5 frequent super-sequences of length 4 for dataset OrdonezA_ADLs

Super-sequence	weights on the path	total weight
toileting spare_time/TV grooming showering	[20, 26, 12]	58
sleeping toileting spare_time/TV grooming	[12, 20, 26]	58
spare_time/TV grooming showering breakfast	[26, 12, 14]	52
showering breakfast spare_time/TV grooming	[14, 11, 26]	51
breakfast grooming spare_time/TV toileting	[3, 22, 25]	50

Table 3: Top 5 frequent super-sequences of length 5 for dataset OrdonezA_ADLs

Super-sequence	weights on the path	total weight
toileting spare_time/TV grooming showering breakfast	[20, 26, 12, 14]	72
sleeping toileting spare_time/TV grooming showering	[12, 20, 26, 12]	70
spare_time/TV toileting grooming showering breakfast	[25, 13, 12, 14]	64
showering breakfast grooming spare_time/TV toileting	[14, 3, 22, 25]	64
grooming spare_time/TV toileting showering breakfast	[22, 25, 2, 14]	63

4. CONCLUSION AND FUTURE WORK

In this paper, we looked into how to extract the general evident series in an ADLs dataset and used SSFPM to solve this problem by formulating it as a graph model. We then adopted a heuristic algorithm for seeking the heaviest paths in a directed weighted graph to solve it. Accordingly, we did experiments on real world ADLs dataset and found interesting facts.

In the future we would like to look into more ADLs datasets especially with more activities involved ones. Also, we would like to consider more factors when constructing the graph model and seeking the evident super-sequence patterns, e.g., the spending time of the individual at a specific activity, the importance of an activity etc.

5. REFERENCES

[1] T. Gu, Z. Wu, X. Tao, H. K. Pung, and J. Lu, "epsicar: An emerging patterns based approach to sequential, interleaved and concurrent activity recognition," in *Pervasive Computing and Communications, 2009. PerCom 2009. IEEE International Conference on.* IEEE, 2009, pp. 1–9.

[2] M. Perkowitz, M. Philipose, K. Fishkin, and D. J. Patterson, "Mining models of human activities from the web," in *Proceedings of the 13th international conference on World Wide Web.* ACM, 2004, pp. 573–582.

[3] J. Han, M. Kamber, and J. Pei, *Data mining: concepts and techniques.* Morgan kaufmann, 2006.

[4] R. Agrawal and R. Srikant, "Mining sequential patterns," in *Data Engineering, 1995. Proceedings of the Eleventh International Conference on.* IEEE, 1995, pp. 3–14.

[5] R. Agrawal, R. Srikant *et al.*, "Fast algorithms for mining association rules," in *Proc. 20th Int. Conf. Very Large Data Bases, VLDB*, vol. 1215, 1994, pp. 487–499.

[6] R. Srikant and R. Agrawal, *Mining sequential patterns: Generalizations and performance improvements.* Springer, 1996.

[7] J. Han, J. Pei, B. Mortazavi-Asl, Q. Chen, U. Dayal, and M.-C. Hsu, "han2000freespan: frequent pattern-projected sequential pattern mining," in *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM, 2000, pp. 355–359.

[8] J. Han, J. Pei, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal, and M. Hsu, "Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth," in *Proceedings of the 17th International Conference on Data Engineering*, 2001, pp. 215–224.

Table 4: Top 5 frequent super-sequences of length 4 for dataset OrdonezB_ADLs

Super-sequence	weights on the path	total weight
showering toileting sleeping leaving	[34, 25, 21]	80
breakfast showering toileting sleeping	[19, 34, 25]	78
snack showering toileting sleeping	[10, 34, 25]	69
leaving showering toileting sleeping	[9, 34, 25]	68
leaving sleeping toileting breakfast	[24, 27, 14]	65

Table 5: Top 5 frequent super-sequences of length 5 for dataset OrdonezB_ADLs

Super-sequence	weights on the path	total weight
leaving sleeping toileting break-fast shower-ing	[24, 27, 14, 19]	84
showering toileting snack leaving sleeping	[34, 12, 6, 24]	76
snack sleeping leaving show-ering toileting	[12, 21, 9, 34]	76
leaving snack show-ering toileting sleeping	[6, 10, 34, 25]	75
null sleeping leaving show-ering toileting	[7, 21, 9, 34]	71

[9] M. J. Zaki, "Spade: An efficient algorithm for mining frequent sequences," *Machine learning*, vol. 42, no. 1-2, pp. 31–60, 2001.

[10] M. Garofalakis, R. Rastogi, and K. Shim, "Spirit: Sequential pattern mining with regular expression constraints," in *Proceedings of the international conference on very large data bases*, 1999, pp. 223–234.

[11] X. Yu and T. Korkmaz, "Super-sequence frequent pattern mining on sequential dataset," in *Big Data, 2013 IEEE International Conference on*. IEEE, 2013, pp. 52–59.

[12] F. J. Ordóñez, P. de Toledo, and A. Sanchis, "Activity recognition using hybrid generative/discriminative models on home environments using binary sensors," *Sensors*, vol. 13, no. 5, pp. 5460–5477, 2013.