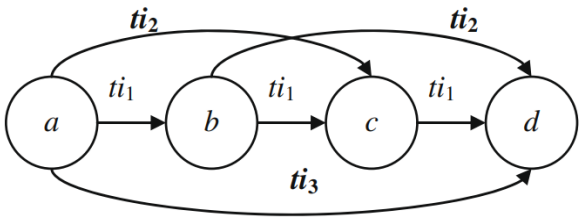
Multi-Time-Interval Sequential Pattern Mining

J Sudarsanan - 181CO222
Sangeeth S V - 181CO246

Overview

Multi-Time-Interval Sequential (MTIS) Patterns are patterns that contain information about the time differences between sequentially occurring events. For example, a person who purchases the first volume of a book series, is likely to purchase the second one a week after, and the third one another 2 weeks later, and so on. Just the time-interval sequential pattern mining algorithms would only be able to get the time interval information between consecutive events. By using multi-time-interval sequential patterns, we seek to establish such time-relations between every pair of events in a sequence.

This example shows four events a, b, c and d, occurring sequentially. The diagram shows how a multi-time-interval pattern needs to establish the time intervals between one event and each successive event so that all the time intervals between any two events are well-defined.



Goals

I. Implementing the algorithms mentioned in the reference paper
   A. MI-Apriori.
   B. MI-PrefixSpan.

II. Testing the implemented algorithms on a real dataset to obtain the patterns it contains.

III. Attempt to improve the MI-Apriori algorithm by extending the FP-Growth algorithm for the multi-time-interval sequential patterns mining task.

Data Mining Task

We would like to take on a sequential pattern mining task, specifically, the problem of multi time interval sequential pattern mining. The dataset is to be in the form of an ordered list of item-sets with time stamps. The same item can occur multiple times with different timestamps, but only once for a particular timestamp. We then assume a set of time intervals and then seek to establish patterns in the dataset using these time intervals.

Data Plan

E-Commerce Purchase History from Jewelry Store:

https://www.kaggle.com/mkechinov/ecommerce-purchase-history-from-jewelry-store.

We'll first be focussing on converting the data to the required ordered list format and then splitting the data into different partitions using appropriate metrics. Then we will attempt to use the algorithms and extract the sequential patterns from the dataset.

Schedule

Task	Schedule
Explore the Data	20 August
Data preparation	27 August
MI-Apriori Algorithm on Synthetic Datasets	10 September
MI-PrefixSpan Algorithm on Synthetic Datasets	24 September
Testing Algorithms on the Dataset	01 October
Prepare some adequate visualisations	15 October
(If Possible) Improve the MI-Apriori Algorithm with the FP-Growth Algorithm	22 October

References

- Ya-Han Hu, Tony Cheng-Kui Huang, Hui-Ru Yang, Yen-Liang Chen, *"On mining multi-time-interval sequential patterns"*, Data & Knowledge Engineering, Volume 68, Issue 10, 2009.