

Study of Temporal Data Mining Applications and Techniques

^[1] G.Sophana, ^[2] R. Sheeba Mary Ananthi, ^[3] M.Dhanalakshmi

^{[1][2]} Associate Professor, ^[3] Assistant Professor,

^{[1][2][3]} Department of Computer Science, Kamaraj College, Thoothukudi

Abstract: - Data Mining is the process of mining patterns in large databases. Temporal Data Mining (TDM) is the key area to mine the sequence patterns of data. TDM is a rapidly evolving area of research and deals with extremity information from temporal data with the time factor. The application of TDM ranges from the prediction of customer behaviour, marketing, medical, communication, agriculture, weather forecast, finance., The techniques involved in TDM are Apriori, Classification, Clustering, GSP, SPADE and PrefixSPAN. This paper focuses on the area of applications and a variety of techniques involved and guide for the selection of applications with appropriate techniques.

Keywords: TDM, GSP, SPADE, Prefix SPAN.

I. INTRODUCTION

The goal of this paper is to give an overview of the techniques and applications proposed within the last decade, to deal with temporal data mining. In addition to that, we aim to classify and organize them in such a way, that it helps to solve real life problems. Temporal data mining is an important part of data mining. It is extraction of implicit, potentially useful and previously unspecified information, from large amounts of data. Temporal data mining deals with data mining of large sequential data sets. Sequential data is the data that is ordered with respect to an index.

II. TEMPORAL DATA MINING

Temporal data mining deals with the extracting useful information from temporal data. The ultimate goal of temporal data mining is to discover hidden relations between sequences and sub-sequences of events that occur within a defined time span [1]. Temporal data mining is concerned with analysing of ordered data streams with temporal interdependencies. Temporal data mining is duration modelling for events in the sequence. Temporal data mining deals with the problem of mining patterns from temporal data which can either be symbolic sequences or numerical time series. The sequence is said to be frequent if its support is higher than the threshold. Sequences are

ordered events, for example Customer shopping sequences. Time series are uniformly separated in time dimension. Time series may be viewed as finding patterns in the data and predicting future values [9] for example Finance. Generally, transaction based databases collect data over some period of time and timestamp for transaction is explicitly available.

Sequential pattern mining

Sequential pattern mining is the mining of frequently occurring ordered events or subsequence as patterns [2]. It is classified into two categories. The first one is Apriori / generate and test approach and another approach is Pattern growth / divide and conquer. Fig.1. Represents the Classification of Sequential pattern mining.

1. Apriori based approach:

Frequent subsets are enhanced one at a time called candidate generation. It follows breadth first search.

2. Pattern growth based approach:

Only projecting frequent prefixes rather than projecting all the possible occurrences of frequent sub sequences from sequence database. It follows depth first traversal

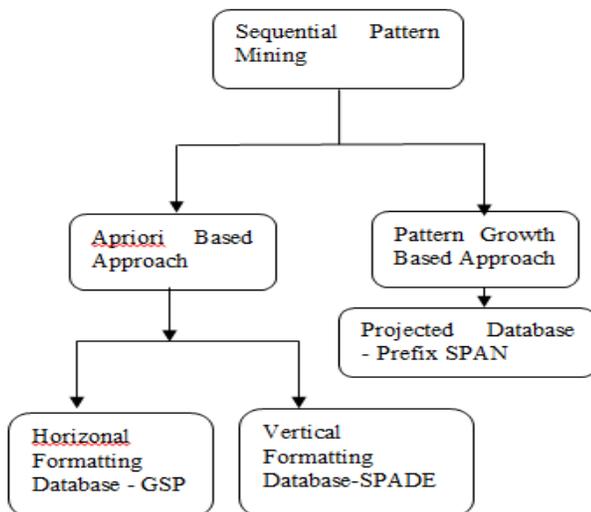


Fig.1. Classification of Sequential pattern mining

III. TECHNIQUES

The various approaches used for temporal data mining are Apriori, Classification, Clustering, GSP (Generalized Sequential Pattern), SPADE (Sequential Pattern Discovery Using Equivalence Classes), PREFIX SPAN (Prefix Projected Sequential Pattern Mining)

A. Apriori

Apriori is an algorithm for association rule learning over transactional databases for mining frequent item set. Multiple scan of database is required. Difficulties at mining long sequential patterns. Apriori generate millions of candidate sets and scan the group action information repeatedly. In association rule mining of transaction database the mining results are about which items are brought together frequently those items must come from same transaction. Sequential pattern mining are to mine the purchase of items in certain order by the same customer from different transaction.

B. Classification

Classification is used to classify objects belonging to a particular group according to their common attributes. Classification in data mining differentiates the parameters to view the clear information [5]. Classification is a supervised learning. In classification there are predefined classes into which the unknown sets of attributes are classified. Classification is used to classify new sample into known classes.

C. Clustering

Clustering groups the data on the basis of a similarity measure. Clustering provides a mechanism to automatically

find patterns in large datasets. Clustering is helpful to divide the information attribute into similar variety of teams that take the high intra similarity and lay to rest similarity. Clustering is a unsupervised learning. In clustering there are no predefined classes into which the unknown set of attributes are classified clustering suggest group, based on patterns from data.

D. GSP (Generalized Sequential Pattern)

GSP follows apriori based approach of sequential pattern mining. It uses downward closure property of sequential patterns. Follows multiple pass candidates generate and test approach. Not suitable for mining long sequential patterns. It is used for extracting frequently occurring sequences. GSP is an apriori based horizontal formatting method [3].

E. SPADE (Sequential Pattern Discovery Using Equivalence Classes)

SPADE was introduced first for discovering set of all frequent sequences. It is used for fast discovery of sequential patterns. SPADE process fixed number of database scans. Each sequence can be associated with a list of objects in which it occurs in conjunction with a related timestamp. Decomposition of smaller pieces independently. SPADE is an apriori based vertical formatting method [7].

Equivalence class: A set of frequent item sets that always occur together in the same set of transactions.

F. PrefixSPAN (Prefix Projected Sequential Pattern Mining) PrefixSPAN, a pattern growth approach for sequential pattern mining as a list of patterns. Prefix span projects database by growing frequent prefixes. Pattern growth algorithm is to avoid the candidate generation step and concentrate the search on a smaller and limited portion of database [8]. No candidate sequence needs to be generated by prefix span. A projected database is smaller than the original database. PREFIX SPAN is a projection based pattern growth method. Sequential Patterns: Patterns between data where the values are delivered in a sequence.

IV. APPLICATIONS

The various applications of temporal mining are Prediction of customer behaviour, Marketing, Medical, Communication, Agriculture, Weather forecast, Finance.

A. Prediction of customer behavior

Customer behavior means the study of individuals, groups or organizations about their process of selecting, securing, using and disposing the products. Customer behavior modelling identifies behavior among groups of customers to predict how similar customers will behave under similar circumstances. Customer behaviour involves services and ideas as well as tangible products. The behaviour of

consumer while shopping or making other marketing decisions. Predictions is possible using decision tree in classification. GSP is used in super market to discover the system of correlation between the customers how they are buying items. The possible percentage of the correlation gives the new knowledge. It is very useful to take decisions.

B. Marketing

Clustering and sequential pattern are used to classify the customers. Clustering techniques which enable customers according to their lifestyle. Client details are classified and sequential pattern mining techniques are applied to spot the client for providing the banking schemes.. Place the product on the shelves based on the sequence order of mined patterns to improve the sales using prefix span algorithm. One may analyze the price fluctuations of stocks to uncover periodic changes in the market. For example, the stock price of a company may follow some patterns every month before it pays its dividend to its shareholders, or before the end of the year.

C. Medical

Sequential pattern mining algorithm can be used to find patterns in medical records. Each sequence is the medical record of the person in a hospital. Patterns could be found such that people who took the medicine X and Y then medicine Z will have a heart attack. Using SPADE algorithm accurate predictions can be made without using the patients entire history. PrefixSPAN is employed for identifying disease trajectory patterns based on frequent subsequences especially for efficiently and effectively searching sequential patterns within huge databases.

D. Communication

Telecommunication companies use data mining algorithms to analyze the profile of their customer based on their communication behavior from the time stamped transactional data and some static data [4]. In the telecommunication industry the data mining algorithms used for network management system as well. Data mining methods have been successfully used to discover optimized solution to learn the network behavior under different network conditions.

E. Agriculture and Weather Forecast

Sequential mining can be applied for weather prediction, Classification of soil, Prediction of yield and Weeds precision detection. Prediction problems may be formulated as classification, association rule mining or clustering problem. Decision tree can be used in predicting the dependent variable like fog and rain [6]. For weather prediction classification and clustering proves to be good

with higher prediction accuracy than the other techniques of data mining. To analyze the change in the natural environment for predicting the occurrences of earthquake one after the other.

F. Finance

Banking system needs to process large amount of data on daily basis related to customer information of their credit card details limit and security details, transaction details and risk profiles. Detecting and preventing activities of involving deception is a consideration of data mining techniques which will assist the organization to focus on the ways and means of analyzing the customer data to identify the patterns that can lead to scams. Data mining models are for pattern detection for the business intelligence and decision based effectiveness for the banking system.

V. CONCLUSION

PrefixSPAN algorithm mines the complete set of patterns. PrefixSPAN is better than GSP and SPADE in performance as it takes lesser time for processing. The processing time maintains the order as PrefixSPAN < SPADE < GSP. GSP is 20 times faster than apriori algorithm. SPADE is faster than GSP. Prefix span is fastest among all algorithms. Memory required by PrefixSPAN is much less than GSP and SPADE. PrefixSPAN results in higher efficiency than all the other algorithms.

REFERENCES

- [1] Mohd. Shahnawaz, Ashish Ranjan, Mohd Danish(2011): Temporal Data Mining: An Overview International Journal of Engineering and Advanced Technology , Vol-1, Issue 1
- [2] Philippe Fournier-Viger, Jerry Chun-Wei Lin, Rage Uday Kiran, Yun Sing Koh, Rincy Thomas(2017) : A Survey of Sequential Pattern Mining, Data Science and Pattern Recognition, ISSN 2520-4165 Vol 1, Number 1 pp 54-70
- [3] Jeff Heaton (2016) : Comparing dataset characteristics that favor the Apriori, Eclat or FP-Growth frequent item set mining algorithms, ISSN: 1558-058X
- [4] Milan Patel, Srushti Karvekar, Zeal Mehta (2014) : CUSTOMER BEHAVIOR MODEL USING DATA MINING International Journal of Advanced Technology in Engineering and Science Vol-2, Issue 1

**International Journal of Engineering Research in Computer Science and Engineering
(IJERCSE)**

Vol 4, Issue 12, December 2017

[5] M. Viswambari, Dr. R. Anbu Selvi (2014) : Data Mining Techniques to Predict Weather: A Survey International Journal of Innovative Science, Engineering & Technology, Vol-1 Issue 4

[6] Rajesh Kumar (2013) : Decision Tree for the Weather Forecasting , International Journal of Computer Applications, Vol-76, Number 2

[7] Manika Verma, Devarshi Mehta (2014) : Sequential Pattern Mining: A Comparison between GSP, SPADE and Prefix SPAN International Journal of Engineering Development and Research, Vol- 2, Issue 3

[8] Roshani Patel, Tarunika Chaudhari (2016): A review on sequential pattern mining using pattern growth approach algorithm based on parameter. International Conference on Wireless Communications, Signal Processing and Networking

[9] Margaret H.Dunham,S.Sridhar :DATA MINING Introductory and Advanced Topics

