SPATIO TEMPORAL CASCADE: A SURVEY

P.Ashwin,
Master of Engineering,
Department of Computer science,
K.S.R.College of Engineering,
Tiruchengode.

V.Vennilla.M.E(Ph.D),
Assistant Professor
Department of Computer science,
K.S.R.College of Engineering,
Tiruchengode.

Abstract-The trajectory of any dynamic object is associated with space and time. The idea of motion would turn meaningless if time is separated from space or vice versa. A novel cascade of space and time was made and named as spatiotemporal (ST) in the domain of data mining. Recurrent events identified from any Boolean ST data set required optimizations as the cascade mining is computationally complex. Partially ordered sets from a given Boolean ST set are considered to mine patterns. The user area of interest over the entire data set tends to be very small so it is sufficient if patterns of user interest alone are mined. Consider the data set related to tsunami is mined for patterns related to death then there is no need to retrieve patterns related to asset damage or water level rise as they are irrelevant to user interest. Focus here lies on reducing the cost of optimization along with use of effective filtering techniques to outperform the existing system. A cascade participation index (CPI) is computed through bottleneck analysis over the data set to measure user interest. Based on this value and the directed graph representation the filtering of irrelevant cascade spatiotemporal patterns (CSTP) can be done having the miner nested in the process. Patterns identified from partial sets are not complete and require an enumeration of patterns from spatial neighbors to be done to ensure completeness. The result has to be extended to real time data sets to prove validation of content.

Index terms – spatiotemporal, Boolean spatiotemporal data set, partially ordered sets , cascade participation index , cascade spatiotemporal pattern.

I. INTRODUCTION

Any process carried out today will have a time and space complexity. These two constraints will have to be optimized in order to improve efficiency. For example, if an employee is to be located in a company it is necessary to know the bounds of the company with respect to the location. The space is first known but this alone can't be sufficient enough to identify the employee, which poses a requirement of spatial partitioning. To have spatial partitions we can split the space into working section, store room, rest room etc. This partitioning helps us to identify the employee better.

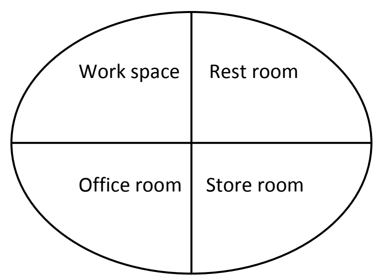

Figure 1 –Spatial mining

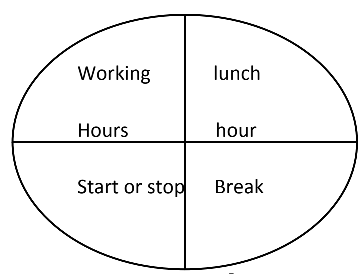

Figure 2 – Temporal mining

The employee can be identified better by considering the time constraint too. During the working hours the employee has got to be in the work space and may be found in the rest room during the break or lunch time. The combination of time and space can be represented as

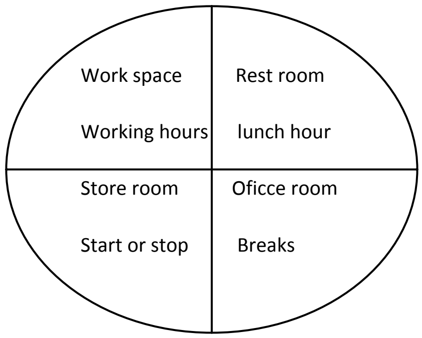

Figure 3 –Spatiotemporal cascade

In the process of identifying an employee the figure 3 seems to efficiently handle the features than the figure 1 AND 2. This certain idea gave data miners a trail to cascade space and time as spatiotemporal (ST) in the process of data mining. Of no doubt this trail was implemented in real time and the surprise was that the experimental results outperformed the available techniques in the same mining. The attention of many people in this domain got fixed towards this.

II. SPATIOTEMPORAL MINING

The spatiotemporal mining[1] helps to identify patterns from a data set with respect to space and time. Patterns[3] are recurrent events in a data set that help us to derive valid conclusions. The client can make use of queries to perform transactions on the database. The point is that, how efficiently can the query be processed? Much research has been done in query processing and its optimization but the ocean of mining is too vast to put an end to it. Many new inventions have been put forward in this area and most of them have gained global acceptance. The main focus lies on reducing the number of retrievals with respect to a query so that the user may be given a better response rate. For example if a student named Jack has to be found from a school then each and every student has to be compared with Jack to identify him. In this case the number of references and retrievals rate is very high. If an additional detail of Jack is available as he is studying in the eight standard then the process of searching gets simplified very much.

All the patterns with respect to standard eight are said to be relevant and all other remaining patterns are said to be irrelevant. The school may have any number of students with the name Jack but our interest is with the student in standard eight which is the relevant pattern and all other patterns can be called as irrelevant. Simply the irrelevant patterns can be filtered and only the relevant patterns can be taken into consideration as it is very simple and quick to identify Jack from standard eight than searching for Jack in the entire school. The entity Jack is not static in order to find him in the particular location. For example Jack may be in the class, playground, laboratory, auditorium or similar locations with respect to his time-table which means that time also can play a vital role in the search. This paved the way for temporal mining to identify the patterns temporary to a location. This concludes that spatial mining and temporal mining have an evident role in data mining. Many algorithms have been framed in these areas to satisfy various types of requirements.

III. CASCADING AND ITS NEED

Technology progress is at its peak, moving ahead at a rapid rate. User requirements keep varying with time which makes the developers to try innovative tasks and give a vital contribution to technology advancement. Ultimately all the developers strive to attain user satisfaction Main challenges for the developer remain in the areas of device portability, battery backup and, uninterrupted access. Relating these challenges to our domain we can say that a dynamic approach is essential in mining patterns. Both spatial and temporal mining deal with dynamic objects. Any dynamic tracking is mapped onto a certain space and a certain time. Some instances can turn meaningless if space is separated from time or vice versa .A best example is the crime pattern mining from crime data sets in which any crime attractors or generators are considered within a particular place and in a particular time and similar is the case with climate pattern mining in which the climate over a particular area during a particular duration is considered. To handle such data sets the new concept of combining spatial and temporal mining as spatiotemporal(ST) mining was introduced.

IV. AVAILABLE TECHNIQUES

The emergence of spatiotemporal mining[7] did not have the expected reach. The main reason for this lies in the fact that spatial mining is a separate process and temporal mining is a separate process and if we are going to combine these two into a single process , it would prove to be complex. The cost and time of spatial and temporal mining will definitely be more than the previously available statistics as the time required to perform both these tasks will be more than any of the individual mining and so is the case with cost, hence optimization in this regard turned as the focus for research.

V.BOOLEAN SPATIOTEMPORAL DATA SET

A Boolean spatiotemporal data set is the platform for this research. Mining is done to find a pattern from a data set. The result will only indicate if the pattern is found in the data set or not. It is true if the pattern is found and false if the pattern is not found and that is why a Boolean spatiotemporal data set is considered for such type of mining. To mine for patterns in a Boolean ST data set initially totally ordered sets or unordered sets are considered. In our example of finding Jack from a school, we say that the entire school is a totally ordered set. Similarly if Jack is searched in only selected classes of the school then it is mining from unordered sets where a class is a subset and any class of choice will make it unordered.

VI. MIXED DROVE

Initially a mixed drove approach[2] was used to identify cascade spatiotemporal patterns(CSTP). The cascade was simply related to patterns that co-occurred. This just combined the ideas of the two types of mining and laid it as a single concept and implemented it in real time operations. As discussed earlier this concept had to perform spatial as well as temporal operations on a single data set which increased the time requirement and the cost of this process too went high. In simple terms it proved to be computationally complex.

VII. CSTAG

To improvise on the computational complexity the next trail was made and the concept[3] of cSTAG mining was put forward. Tracking the dynamic object was done to understand patterns. Clusters were used to depict the location and time of the object movement, by means directed acyclic graphs. Although this simplified the mining process, it had its own drawbacks. The major problem was faced with cluster formation and its size. Since the pattern s dynamic no static means to frame a radius for cluster size could be framed. The next problem was with the representation of data. Generally graphs are used for this purpose and represent the details as static entities whereas our concern lies with dynamic entities which could not be effectively represented and explained.

VIII. CASCADE ST PATTERN MINING (CSTPM)

A new idea of mining cascaded spatiotemporal patterns[6] from data set was proposed and had a great reach. The mined patterns were in a combinatorial format having a confluence of space and time. Such patterns had a significant role in CSTPM [8]as they could satisfy plenty of constraints placed by the user. A nested loop based approach was used to mine such patterns. In the nested approach the miner was also given a role to play in the optimization of retrievals. Any optimization done in this mining would be to reduce the number of retrievals. From a large data set it would be easy for an user to work with a few retrieved patterns than carrying out the same work with more retrieved patterns. With a progress being attained in this mining many people began to agree with this process. The idea that the reduction of cost associated with optimization of mining can also serve as a optimization technique lead to a immense development in CSTPM. Yet another milestone in this mining was reached and the idea that patterns can identified from partially ordered subsets was revealed. The processing of such partially ordered sets was an easy task but forming such sets was a bit tedious.

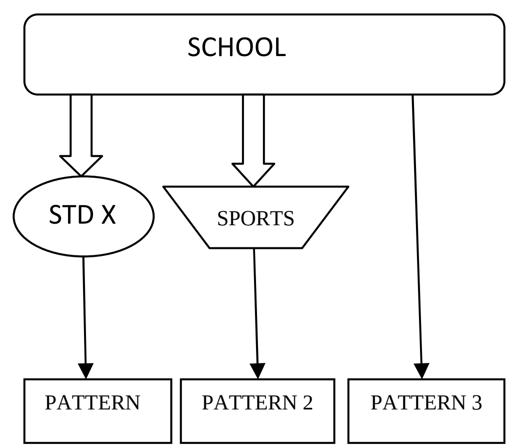

Figure 4 - Mining from sets 1) Unordered set 2) Partially ordered set 3) Totally ordered set.

The user who accesses the data set will not want to retrieve the entire data in the set but instead the user will be interested only in a portion of the entire retrieval so a measure to find the user interest needs to be framed. Cascade participation index (CPI) served this purpose and was formulated as the ratio of number of instances in the retrieved pattern to the number of instances in the data set. The CPI only considered the positive correlation factor in computations. A nested loop based mining was performed having the miner nested in the mining process. This helps us to get a rough idea of the user interest. A study of the bottleneck analysis could support in the calculation of CPI.

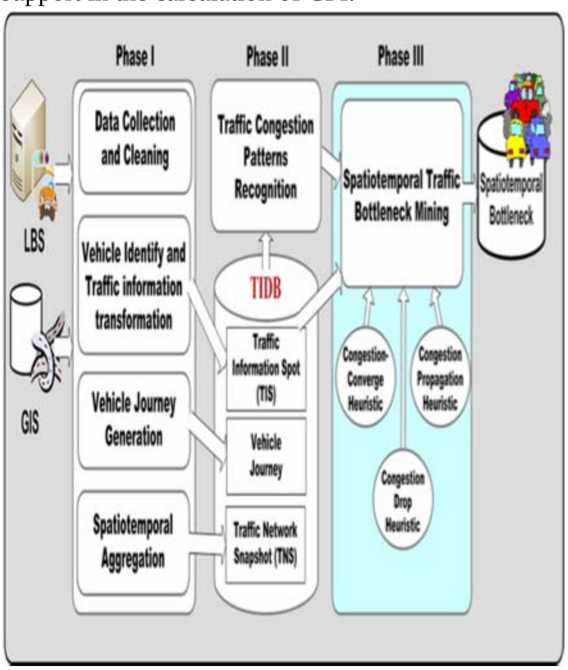

Figure 5 – Spatiotemporal bottleneck

The bottleneck analysis could give us an idea about to which portion is the user interest focused. For example for a traffic mining denoted in figure 5 we can say that the user is only considered with the spatiotemporal traffic pattern and is neither worried about the cause of the traffic nor a reduction in the traffic. Under the bottleneck analysis we can say that a partitioning of the ST space would simplify things as only user interested patterns are required here.

A partitioning names ST CSTP was framed to partition the ST space to identify cascaded patterns. For example if the user is in search of a bike in his office campus it would be wise enough if his search is carried out in bike parking zone alone than to search for it in the whole campus. Next step is to justify our result and denote the extent to how far it is correct and valid in the real time. Partially ordered subsets are not complete by themselves so our attempt to mine patterns from then would always raise the question of completeness. Of no doubt the pattern is going to be incomplete only but then how to complete it? The data is represented as directed acyclic graphs where each node is represented as a spatial neighbor. An enumeration of spatial neighbors in the surroundings of user interest would help us to complete patterns. Filtering has been done using multi resolution filters and upper bound filters and based on user interest to decide upon which spatial neighbors are to be considered for mining. Most of the conclusions are made from statistical interpretations of data. Always a complexity tends to exist between the computation and its interpretation which needs to be compromised. Once all of the above challenges are met a fruitful result can be obtained. The output was checked by using various data sets and of various sizes and the output ensured its efficiency at all cases.

VI. CONCLUSION

The invention of spatiotemporal mining proved to be a vital asset to data mining. Since the patterns are considered on the time as well as on the space domain this concept could be applied to many real time applications and results were highly extensible. The idea to reduce the cost involved in optimization lead to a significant outcome. Partially ordered subsets could also be mined for patterns than the usually used sets. The partitioning of data sets with respect to space and time as a single cascaded factor helped to handle dynamic entities more efficiently. The user participation in a data set supported for filtering irrelevant events with respect to user interest. The results were extended to more than one data set to ensure validation of mined events. On the whole this strategy outperformed the existing ones and proved its dominance over others.

In future the task of establishing statistical significance of patterns and improving the filtering efficiency can be dealt with as many drawbacks tend to exist with the multi resolution filters and the join operations performed.

REFERENCE:

[1] H. Cao, N. Mamoulis, and D. Cheung, “Discovery of Collocation Episodes in Spatiotemporal Data,” Proc. Sixth Int’l Conf. DataMining (ICDM ’06), pp. 823-827, 2006.
[2] M. Celik, S. Shekhar, J.P. Rogers, and J.A. Shine, “Mixed-Drove Spatiotemporal Co-Occurrence Pattern Mining,” IEEE Trans. Knowledge and Data Eng., vol. 20, no. 10, pp. 1322-1335, Oct. 2008.
[3] H. Cao, N. Mamoulis, and D. Cheung, “Discovery of Periodic Patterns in Spatiotemporal Sequences,” IEEE Trans. Knowledge andData Eng., vol. 19, no. 4, pp. 453-467, Apr. 2007.
[4] J. Chan, J. Bailey, and C. Leckie, “Discovering Correlated Spatio-Temporal Changes in Evolving Graphs,” Knowledge and InformationSystems, vol. 16, no. 1, pp. 53-96, 2008.
[5] G. Webb, “Discovering Significant Patterns,” Machine Learning, vol. 68, no. 1, pp. 1-33, 2007.
[6] P. Mohan, S. Shekhar, J.A. Shine, and J.P. Rogers, “Cascading Spatio-Temporal Pattern Discovery,” Technical Report TR 11-010,Dept. of CS and Eng., Univ. of Minnesota, 2011.
[7] D.M. Morens, G.K. Folkers, and A.S. Fauci, “The Challenge of Emerging and Re-Emerging Infectious Diseases,” Nature, vol. 430, pp. 242-249, July 2004.
[8] P. Mohan, S. Shekhar, J.A. Shine, and J.P. Rogers, “Cascading Spatio-Temporal Pattern Discovery: A Summary of Results,” Proc.SIAM Int’l Conf. Data Mining (SDM), pp. 327-338, 2010.
[9] J.A. Shine, J.P. Rogers, S. Shekhar, and P. Mohan, “Discovering Partially Ordered Patterns of Terrorism via Spatio-Temporal Data Mining,” Proc. 16th Army Conf. Applied Statistics, 2010.
[10] J.A. Shine, J.P. Rogers, S. Shekhar, and P. Mohan, “Cascade Models for Spatio-Temporal Pattern Discovery,” Proc. First USACE Research and Development Conf., 2009.
[11] J.F. Allen, “Towards a General Theory of Action and Time, ”Artificial Intelligence, vol. 23, no. 2, pp. 123-154, 1984.
[12]M.F.Worboys,“Event-Oriented Approaches to Geographic Phenomena,” Int’l J. Geographical Information Science, vol. 19, no. 1, pp. 1-28, 2005.
[13] M.S. Scott and K. Dedel, “Assaults in and Around Bars,” Problem Oriented Guides for Police, Problem Specific Guides, second ed., vol. 1,pp. 1-78, US Dept. of Justice, 2006.
[14] Committee on Strategic Advice on the U.S. Climate Change Science Program; Nat’l Research Council: Restructuring Fed. Climate Research to Meet the Challenges of Climate Change. The Nat’l Academies Press, 2009.