Journal of Theoretical and Applied Information Technology
20th October 2014. Vol. 68 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MINING FREQUENT PATTERNS FROM SPATIO-TEMPORAL DATA SETS: A SURVEY

[1]GURRAM SUNITHA [2]Dr. A. RAMA MOHAN REDDY
[1]Research Scholar, Department of CSE, S.V.U. College of Engineering, Tirupati, India
[2]Professor, Department of CSE, S.V.U. College of Engineering, Tirupati, India
E-mail: [1]gurramsunitha@gmail.com, [2]ramamohansvu@yahoo.com

ABSTRACT

Space and time are implicit in every activity of life. Every real-world object has its past, present, future and hence is intrinsically tied up with location and time. Storing spatio-temporal attributes in the databases along with the thematic attributes enriches the data and the inherent knowledge stored in the database. Spatio-temporal databases provide description of real-world phenomenon in a spatio-temporal framework and helps to build information systems and services that are more comprehensive and intelligent. This paper highlights the importance and applications of spatio-temporal pattern mining and provides a brief survey of key mining techniques for discovering three types of spatio-temporal patterns – sequential patterns, co-occurrence patterns and cascaded patterns specifically from event data sets and trajectory data sets.

Keywords: *Spatio-Temporal Data Mining, Event, Trajectory, Sequential Patterns, Co-occurrence Patterns, Cascaded Patterns*

1. INTRODUCTION

Almost all of the worldly events are spatio-temporal in nature. Space and time can be treated as independent dimensions. Some data may have only spatial features, some only temporal features, some may have both, and some may have none. Analysis may be done considering only spatial features or only temporal features, but there is an ever-increasing need for the integrated approach. Pushing spatio-temporal features of the data into the mining process provides insight into new dimensions of the discovered patterns. The major goal of data mining is mapping present to future through analysis of past. Different data mining types may develop a different data model, but the ultimate goal of developing a model is either for prediction or for hypothesis testing.

2. SPATIO-TEMPORAL DATABASES

Traditional databases provide only 'what and how' detailing (context-level or semantic information) of the events whereas spatio-temporal databases also provide the 'where and when' detailing (spatio-temporal information). Spatio-temporal databases store data collected from different application areas and from different data sources. The following are some of the types of the spatio-temporal data.

- ✓ *Satellite Image Time-Series Data* - provides information about climate, rainfall, temperature, moving and deforming storms etc.
- ✓ *Geographic Data* - provides information about GIS, LIS, city planning etc.
- ✓ *Earth Science Data* – provides information about rivers, lakes, forests, land coverage, vegetation, ice elevation, fires, droughts, risk evaluation, natural hazards (land erosion, land-slides, avalanche), fauna monitoring, pollution clouds, oil pollution, deforming courses of rivers during floods etc.
- ✓ *Communication Data* – provides information about tele-communications, web logs/ sequences, social networks, social trends etc.
- ✓ *Biological Data* – such as DNA, gene data etc.
- ✓ *Event Data* – provides information about communicable diseases, population, flocks etc.
- ✓ *Transactional data* – provides information about consumer transactions, business data etc.

Journal of Theoretical and Applied Information Technology
20th October 2014. Vol. 68 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

- ✓ *Trajectory data* – provides information about roads, cities, traffic and movement behaviors of humans and vehicles.

3. APPLICATIONS OF SPATIO-TEMPORAL DATA MINING

This section provides an overview of the major application areas of spatio-temporal data mining. Some of the major application areas are –

- ✓ Environmental Monitoring
- ✓ Impact Assessment
- ✓ Resource Management and Decision Support
- ✓ Urban Ecology
- ✓ Administration and Financial Management
- ✓ Real-time Navigational Systems
- ✓ Transportation Scheduling
- ✓ Data Quality and Integrity Enforcement
- ✓ Electronic Navigation Management
- ✓ Trend Analysis and Geo-marketing
- ✓ Traffic Control
- ✓ Demography
- ✓ Epidemic Control etc

[17] has identified all the major application areas that take advantage of spatio-temporal information and has categorized them into 5 application groups– Planning, Engineering, Operations and Maintenance, Construction and Administration.

4. SPATIO-TEMPORAL FREQUENT PATTERN MINING

Spatio-temporal frequent pattern mining can be defined as the process of discovering interesting, useful and non-trivial patterns from large spatio-temporal datasets. The frequent spatio-temporal patterns based on the ordering relationship of the event types in the pattern can be categorized into 3 types - co-occurrence pattern mining, sequential pattern mining and cascaded pattern mining. Co-occurrences are patterns that are unordered, sequential patterns are totally ordered and cascaded patterns are partially-ordered. The ordering of patterns being discussed is related to time. Consider a pattern $<E_1\ E_2,\ ...\ E_n>$ where $E_1,\ E_2,\ ...\ E_n$ are event types. If this is to be considered as a co-occurrence pattern, then the pattern $<E_1\ E_2,\ ...\ E_n>$ does not specify any particular order of occurrence of events. It specifies only that the events of types $E_1,\ E_2,\ ...\ E_n$ occur together in space and time. Sequential patterns strictly follow increasing order of time. The sequential pattern $E_1 \rightarrow E_2 \rightarrow \ ...\ \rightarrow E_n$ specifies that events of type $E_1,\ E_2\ ...\ E_n$ occur in some spatial neighborhood in strictly increasing order of time. i.e., events of type E_1 occur, then later in time, events of type E_2 occur and so on. Sequential patterns can be considered as special case of cascaded patterns where for all pairs of elements in the sequence, one element precedes the other.

As far as our knowledge no one has performed literature review on the topic that has been undertaken in this paper. We have concentrated on briefing the techniques that have played major role in bringing spatio-temporal data mining to prominence. Specifically our review has concentrated on identifying milestone researches in the area of mining ordered patterns from event and trajectory data sets.

The organization of this paper is as follows. Section 5 discusses important techniques to mine sequential patterns. Section 6 discusses major contributions to discover co-occurrence patterns. Section 7 discusses about cascaded pattern mining which is still in its infant stage.

5. SPATIO-TEMPORAL SEQUENTIAL PATTERN MINING

Spatio-temporal sequential pattern mining can be defined as the process of finding order of occurrence of events(event types) from spatio-temporal data sets. [7], [14], [26], [28], [30], [32] have proposed techniques to mine sequential patterns from event data sets.

[28] is the first work done on generating sequential patterns in spatio-temporal data sets. The authors have introduced an algorithm DFS_MINE to mine spatio-temporal patterns called as *location sequences*. A location sequence is a sequence of event types of form $E_1 \rightarrow E_2 \rightarrow ... \rightarrow E_n$ occurred in a location in increasing temporal order. The DFS_MINE algorithm answers queries related to a specific location – influence of one event on the occurrence of subsequent events in the same location. There are many features of the algorithm which make it efficient – lattice-based approach for handling voluminous search space, mining at multi-level granularities of space, depth-first approach for fast-generation of long patterns and maximal frequent sequences etc. But, this algorithm does not aim at reducing the number of database scans.

The outline of the DFS_MINE algorithm is as follows – three lists are maintained (1) MaxFreqList

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

(2) MinNonFreqList and (3) UselessSet. MaxFreqList at any point of time maintains all the maximal frequent patterns discovered till that time. MinNonFreqList contains all the minimal non-frequent patterns discovered till time. UselessSet consists of useless events list for each k-sequence. As and when new sequences are formed, these three lists are updated. A new *(k+1)* sequence is formed, by intersecting an item at all possible positions in a k-sequence as a separate itemset and as part of an existing itemset. Constraints are pushed into intersection process to avoid duplicates. UselessSet of a *k*-sequence is used to take care that useless items are not intersected with the k-sequence while generating *(k+1)*-sequences. This helps in candidate pruning. Once a new sequence is generated it is inserted in to either MaxFreqList or MinNonFreqList accordingly. MaxFreqList gives the final list of frequent location sequences.

[30] introduces a type of sequential patterns called as *flow patterns* which helps in determining the evolutionary phenomena of events in neighborhood space and time. Flow pattern can be defined as a sequence of reflexive event sets where any two consecutive event sets are close neighbors and an event set is the set of events that occurred at the same time. Flow pattern mining tries to define the influence of events occurred at some time t_1 on events occurred at some later time t_2 in some spatio-temporal neighborhood. The disadvantage of the approach proposed in this paper is the assumption that the events will repeat at the same locations over time. Mining flow patters is complex and time-consuming process since the search-space is vast. Hence, three optimization techniques have been proposed to (i) prune infrequent candidates based on the support of individual items (ii) speed up support counting process based on spatial neighborhood of events (iii) delay database scans for all the sequences which has a super-sequence that is frequent.

[32] has introduced a different kind of patterns called as generalized spatio-temporal patterns. A generalized spatio-temporal pattern can be defined as a sequence of relative event sets, such that all relative events sets are close neighbors of each other. The work in [32] has highlighted the importance of addressing relative positions of events rather than absolute positions, in capturing invariant topological relationships within a pattern. The absolute locations of events are converted into relative positions by choosing a reference point. Two measures t-minsup(temporal support) and s-minsup(spatial support) are defined to determine the frequency of the generalized spatio-temporal pattern. Spatial support and temporal support ensure high frequency of repetition of the pattern over space and time respectively. This leads to the extraction of patterns that repeat consistently over space and time. [32] has extended the pattern growth approach and developed an algorithm called GenSTMiner. Two optimization techniques are introduced- (i) conditional projected databases to prune infeasible events and sequences (ii) pseudo projection to reduce the memory requirements. Only one database scan is required to extract 1-generalized spatio-temporal patterns. Patterns of greater length are produced by constructing projected database for each generalized spatio-temporal pattern and mining it recursively. This technique has the disadvantage that it is highly sensitive to the choice of reference point.

[14] has introduced spatio-temporal sequential pattern mining for point based event data set. [28], [30], [32] are frequency-based approaches whereas STS-miner proposed in [14] is a density-based approach to identify statistically significant spatio-temporal sequential patterns. In this approach, a spatio-temporal sequential pattern $A \rightarrow B$ says that events of type *B* are densely clustered around one or more events of type *A*. STS-Miner allows to mine patterns based on one of the temporal predicates - *follow*. The algorithm mines patterns based on the pre-defined thresholds to restrain the spatio-temporal neighborhood of an event. The concept of density ratio calculates the significance value of sequential patterns. Significance measure called as *sequence index* ensures that long sequences are associated with minimal strength. The sequence index of a pattern signifies the cause-effect relationship between different event types. STS-miner approach defines the spatio-temporal neighborhood on events. It handles one event type at a time and does not consider the interaction between event types while generating the patterns. Another algorithm Slicing STS-miner works with the databases that do not fit entirely in the memory.

[26] has concentrated on studying the influence of a spatio-temporal region on its neighboring regions and vice-a-versa. This significantly gives better insight into the evolution of regions considering the events occurred in the region and in its neighborhood regions. The patterns extracted are called *spatio-sequential patterns*. A new interesting measure *spatio-temporal participation index* has been proposed. DFS-S2PMiner algorithm generates spatio-sequential patterns based on both spatial and temporal dimensions. The algorithm goes with

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

depth-first-search and uses pattern-growth approach to construct projections of the database.

Different works have introduced different kinds of sequential patterns. For example, consider the sequential pattern *Rainfall → Mosquito Nests*. In [28], the sequential pattern *Rainfall → Mosquito Nests* is location-specific and says that, in a region *R, Rainfall* is followed by increase in *Mosquito Nests* in temporal proximity. In [30], the same sequential pattern is represented as *<Rainfall(0,1))(Mosquito Nests(1,1))>* says that if *Rainfall* occurs at location *(0,1)*, then *Mosquito Nests* develop at location *(1,1)* in temporal proximity. [14] says that if *Rainfall* occurs at some location *l* and time *t*, then there will be significant increase in the density of *Mosquito Nests* in spatio-temporal proximity of *(l, t)*. [26] calls the patterns as spatio-sequential patterns. For example, *<(Rainfall)(Mosquito Nests.[Mosquito Nests ; Mosquito Nests])>* says that, *Rainfall* occurred in a region *R*, and in temporal proximity, *Mosquito Nests* developed in the same region *R* as well as in two of its spatial neighborhood regions.

Spatio-temporal data can be collected at different scales and granularities. Considering the fact and allowing mining at different levels of granularity may lead to extraction of more complex patterns considering the topological relationships between objects.

The works [14], [26], [28], [30], [32] does not consider spatial scales and relations while mining patterns, whereas [7] introduced another kind of sequential patterns that includes the relationship between spatial objects as well as geographical scales. Spatial relations include the mentioning of spatial directions in the discovered patterns such as *'Rainfall in a region causes humidity in its northern neighborhood region'*. Including spatial scales allows to discover patterns at multiple levels of spatial granularity. Such patterns are called as related sequential patterns and extend the spatio-sequential patterns defined in [26]. Spatio-relational dimension and its hierarchical representation has been defined to include relationships between spatial objects such as Region R_2 is to the north of region R_1. This allows extraction of patterns at different levels of spatio-relational dimensions. Also, spatial granularity hierarchy is defined to include geographical scales. A related sequential pattern of form *R<(.North[(Humidity.Rainfall) (Rainfall)])>* says that in a region *R* and in the region that is to the north of *R*, *Humidity* and *Rainfall* occur at the same time respectively. Later in time, *Rainfall* occurs in the region *R*. Algorithm STR_PrefixGrowth is proposed to extract the sequential patterns.

[1], [9], [22], [16], [19] proposed works to explore trajectory databases enriched with information of moving objects. Prodigious data of movement of pedestrians and vehicles has hidden in itself, the regular movement patterns of humans. Gaining knowledge from trajectories helps in enhancing real-life applications like traffic control, providing information to travelers on shortest routes along with timing information, efficient implementation of location-based services etc.

[1] discovers sequential patterns informative of frequent routes of objects. Techniques are proposed to find frequent spatial regions patterns and long closed patterns. To find frequent singular patterns, the trajectories are segmented using DP line simplification algorithm [5]. Line segments are grouped by considering the spatial similarity of trajectories as well as their temporal closeness. Since the candidate search space of all the combinations of line segments is large, to speed up the process *Growth* heuristic is applied which uses filtering and verification method to assign each line segment to a spatial region. Each line segment may be the seed of the region or the member of the region. Some line segments left out without being assigned to any spatial region are considered to be outliers. This process results in identification of frequent spatial regions. Then, in each original trajectory, the line segments that fall into frequent spatial regions are replaced with region ids and others to outliers. A substring tree is constructed from the converted database and substring tree is scanned to generate frequent closed patterns. The limitation of the approach proposed in this paper is that it cannot perform well on very long trajectories as DP algorithm [5] performs decomposition operation recursively to divide a trajectory into line segments and *Growth* heuristic uses repeated merge operation to assign line segments to spatial regions.

[9] has introduced T-Patterns (Trajectory Patterns) which are sequential patterns of frequently visited spatial objects annotated by temporal information. The temporally annotated sequences provide users information about the traveling time, from one spatial location to another. The process of mining T-Patterns involves the identification of Regions-of-Interest(RoIs) and then extracting T-Patterns from the RoIs. Three scenarios of identifying RoIs are described – (a) static RoIs pre-defined by domain expert (b) static RoIs discovered by performing density-based spatial discretization of trajectory data (c) dynamic RoIs

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

discovered through the ongoing mining process. Algorithm proposed in [8] is extended to extract T-Patterns with static ROIs as well as with dynamic ROIs.

[22] proposes to build a probabilistic model that predicts the probable next location of a moving object. The model is built by mining the trajectory databases to extract frequent trajectory patterns and develop rules to describe the movement phenomenon of the underlying objects. These rules can then be used further for predicting the trajectory of a new object, based on its till time movement. Location prediction framework is extended and rule-matching strategies are proposed for the purpose and the performance of the techniques is proved by stringent experimental evaluation.

[16] proposed discretization as a technique to reduce the size of the spatio-temporal databases and hence increase the efficiency of the sequential pattern mining process. The trajectory data can be defined as a sequence of temporally ordered set of locations through which the object had moved. Trajectory data is continuous projection of movement of objects over space and time, leading to huge accumulation of data which blocks the timely discovery of regular movement patterns of objects. Discretization is used to pre-process such trajectory data. While discretizing the data, care must be ensured to retain spatio-temporal correlations of the data, hence retaining the information of each object's movement. Algorithm STEM(Spatio-TEmporal discretization of Moving object trajectories) discretizes trajectory data by considering their spatio-temporal properties. As the initial step, the trajectories are converted into their approximations, by using DP line simplification algorithm [12] to perform simpler normalization of mapping original trajectory data range to approximation data range. As the next step, such approximations are clustered into logical cells by using pre-clustering phase of BIRCH[37]. Thus the data is discretized to enhance the performance of the sequential pattern mining algorithms.

[19] present Graph-Based Mining (GBM) algorithm to mine frequent trajectory patterns that gives better performance of an order of magnitude than Apriori and PrefixSpan[23] techniques. It uses only one database scan to generate a mapping graph and then creates TI-lists for each node in the mapping graph to store the information about trajectories that are passing through that node. The GBM algorithm supports efficient reduction of search space while extending the trajectory patterns by using adjacency property of geographical space. But, the limitation of using the concept of spatial neighborhood is that the GBM algorithm cannot mine frequent trajectory patterns formed by non-adjacent geographical space.

6. SPATIO-TEMPORAL CO-OCCURRENCE PATTERN MINING

Spatio-temporal co-occurrence pattern mining is the process of finding subsets of event types occurring together in spatio-temporal data sets. It can be defined as finding clique patterns that are prevalent over time. Co-occurrence pattern mining considers space and time in all directions, whereas sequential pattern mining considers space in all directions and time in uni-direction (forward direction). This is because implicitly, a sequential pattern indicates the occurrence of events in increasing temporal order.

Spatio-temporal co-occurrence pattern mining can be categorized depending upon factors such as definition of spatio-temporal neighborhood, mobility of entities and types of entities.

a. *Categorization based on Spatio-temporal neighborhood*

Two variants of co-occurrence pattern mining can be considered based on the definition of spatio-temporal neighborhood of events.

i) *Rigid or local co-occurrence pattern* mining in which co-occurrence is strictly defined with regard to a specific location and time.

ii) *Lenient co-occurrence pattern mining* in which co-occurrence is defined with regard to nearby space and time.

b. *Categorization based on mobility of entities*

Another categorization of co-occurrence pattern mining can be considered based on whether the entities are mobile or immobile. For example, a moving flock of birds can be treated as mobile data and co-occurrence of events *rainfall* and *less temperature* can be considered as immobile data.

c. *Categorization based on entity types*

This categorization is based on whether the moving entities are of same type or of varying type. A herd of sheep is an example for moving entities

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

of same type and prey-predator scenario is an example for moving entities of mixed type [3]. Based upon this categorization, the patterns can be divided into two types –

i) *Flock patterns* which are patterns extracted from the moving entities of same type.

ii) *Collocation episodes* which are patterns extracted from the moving entities of mixed type.

[18], [11], [15], [10] detect flock patterns of moving entities of same type. [31], [2], [34], [3], [29], [25], [4], [24], [33] detect flock patterns of moving entities of different object types.

[18] has defined four variations of spatio-temporal patterns with respect to moving object databases – Flock, Leadership, Convergence and Encounter. REMO stands for RElative MOtion and is the concept developed for analyzing and discovering relative motion patterns in moving object databases. The analysis is based on 3 kinds of motion patterns – Constance, Concurrence, Trend-Setter. Applying a set of spatial constraints on constance, concurrence and trend-setter leads to definition of three spatially constrained REMO patterns – Track, Flock, Leadership. Leadership pattern is an extension of track pattern. The patterns are defined based on the characteristics of the moving objects - direction of movement, change in direction of movement and location of aggregation. Algorithms have been developed for discovering the proposed patterns and implementation issues have been discussed to define the computational complexity of the algorithms.

[11] focused on reducing the computational complexity of the algorithms proposed in [18] for discovering relative motion patterns - flock, leadership, convergence and encounter from moving object databases. The spatio-temporal database is considered to be a snap-shot database and approximation algorithms based on the usage of computational geometry are developed for discovering relative motion patterns but with reduced computational time bounds. The approximation algorithms work either by approximating the size of the region or the size of the subset. Approximation is based on the radius of the region and the number of entities that fall into the region.

[15] tries to find the set of flock patterns by considering flock as a cluster. The time is divided into time-slices and performs spatial clustering of data in a time-slice and accumulates results over different time-slices to identify the flock patterns. All the three algorithms MC1, MC2, MC3 use the enhanced version of density-based clustering algorithm DBSCAN[6] developed to fit the entire time-slice under observation into memory. MC1 is the naïve implementation of the problem and hence has high computational cost. MC2 and MC3 are enhanced versions of MC1. MC2 reduces the computational cost of discovering moving clusters by reducing the union/intersection operations of clusters. MC3 reduces the computational cost by reducing the number of calls to the DBSCAN algorithm, but with reduced accuracy.

[10] proposed detection of longest duration flocks and meetings. A flock may be a fixed-flock where the set of entities remain same during the entire interval whereas a varying-flock is the one in which the entities do not remain same during the interval. Algorithms that mine exact patterns as well as approximation patterns are proposed for fixed flocks and meetings as well as for varying flocks and meetings. It is proved that extracting longest duration flocks is a NP-hard problem.

[31] introduces topological relations into co-occurrence patterns such as cliques, star-like patterns and star-clique patterns. The aim is to consider spatial as well as temporal proximity relationships to discover topological patterns. Algorithm TopologyMiner is proposed to discover topological patterns in depth-first manner. TopologyMiner works based on pattern growth methodology to avoid time-consuming candidate generation process. Also, a concept *summary structure* is defined to handle the memory complexity of storing large spatio-temporal data sets in memory. Instead of storing event instances, this structure emphasizes on storing their count information within a spatio-temporal neighborhood.

[2] examines the trajectory databases of different object types. The idea proposed in this paper varies from [1][9] in its approach of extracting patterns from trajectories of varying objects. They extended the concept of collocation events to define spatio-temporal collocation episodes. A two-phase mining process is defined. In first phase, the trajectories are converted into sequences using a hash-based technique and in second phase, Apriori-based algorithm is used to mine the collocation episodes.

[34] finds temporally co-oriented patterns which give information about the spatial co-orientation of objects over time. The database is considered to be

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

a set of symbolic image sequences and the algorithm TCPMiner mines maximal temporal co-orientation patterns from the given database. TCPMiner is a 3-stage approach that works by converting the problem of mining temporal co-orientation patterns into sequential pattern mining problem. In the first stage, the spatio-temporal database of symbolic image sequences is converted into 2D sequences. In second stage, each 2D sequence is converted into an itemset and the PrefixSpan approach proposed in [23] is used for generating frequent sequences of itemsets. In third stage, sequences generated by second stage are transformed back into symbolic pictures by using methodology similar to Apriori to generate maximal temporal co-orientation patterns.

[3] concentrated on finding co-occurrence patterns in moving object databases. It aims on finding patterns that signify co-occurrence of different object types within a significant space and time. Where traditional co-occurrence pattern mining techniques concentrated on the patterns, the MDCOP-Miner algorithms concentrated on object types and their nature to extract patterns. Spatial and time prevalence measures are defined to signify the spatio-temporal prevalence of the mixed-drove co-occurrence patterns. The composition of spatial and time prevalence measures form the mixed-drove prevalence measure. A pattern is said to be a mixed-drove prevalent if it satisfies the mixed-drove prevalence measure. Two algorithms MDCOP-Miner and FastMDCOP-Miner are proposed that are efficient in terms of computational time required to generate mixed drove co-occurrence patterns (MDCOPs). MDCOP-Miner generates all k-spatial prevalent patterns for all time slots and then prunes the time non-prevalent patterns. FastMDCOP-Miner performs pruning earlier in the process for each time slot, hence performing better than the MDCOP-Miner algorithm.

In contrast to the earlier works, [29] considers behavior of a group of trajectories and discovers group patterns. BFE algorithm proposed in this paper uses grid-based structure and incremental process mining to efficiently compute flock patterns. Four heuristics are deployed to decrease the generation of number of candidate disks.

Earlier works proposed in [13], [38], [36], [35], [2], [3], [29] ignore the fact that the spatio-temporal volume under observation in real-world is generously vast and the number of patterns that satisfy the minimum participation index or prevalence threshold will be least. They do not consider the locality and spread of the co-occurrence patterns, and hence are not applicable for vast spatio-temporal scales. [25] introduces co-occurrence patterns called as SPCOZs, Spread Patterns of spatio-temporal Co-occurrences Over Zones. As the search space of co-occurrence patterns is tremendously vast, an SP-Tree structure is proposed to efficiently index the SPCOZs.

[4] differs from earlier work in considering the lifetime of the spatio-temporal objects. Partial spatio-temporal co-occurrence patterns(PACOPs) are defined as the subsets of object types that exist partially in the database and occur together frequently over space and time. An algorithm PACOP is proposed to mine partial spatio-temporal co-occurrence patterns.

[24] made effort to define spatio-temporal co-occurrence patterns with respect to evolving regions. An evolving region is the one where the shape, size and location of the events continuously evolve over time. Finding frequent co-occurrences of such spatial objects is a complex task. The prevalence of a co-occurrence pattern is determined by measuring its co-occurrence coefficient and prevalence measure. Co-occurrence coefficient is framed from the concepts of co-efficient of areal correspondence and jaccard co-efficient. Participation index defined for co-location patterns from spatial data sets is used as prevalence measure. The algorithm proposed to mine prevalent spatio-temporal co-occurrence patterns is based on Apriori.

[33] is an extension of MDCOP which gives user deliverance from setting threshold values. The disadvantage of the MDCOP algorithm is that the co-occurrence patterns discovered are sensitive to the spatial and time prevalence threshold values which are user-defined. Hence, to compensate the sensitivity of threshold values, an idea is proposed in this paper. The idea is to consider only the top-k% MDCOPs of all the discovered mixed-drove prevalent patterns. The top-k% MDCOPs are the MDCOPs which have highest interestingness within all the discovered MDCOPs and fall into the top-k% of all the MDCOPs when the MDCOPs are sorted in descending order with reference to their mixed-drove prevalence measure. Initially, the naïve approach is discussed, which uses any one of the approaches proposed in [13], [38], [36], [35] to mine co-location patterns for each time slot and then performs post-pruning to discover top-k% MDCOPs. As naïve approach used post-pruning, its computational cost is very high. The algorithm TopMDCOP-Miner proposed in this paper

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

discovers the at most top-k% mixed drove co-occurrence patterns by pruning MDCOPs at each step thus reducing the computational cost of generating MDCOPs.

7. SPATIO-TEMPORAL CASCADED PATTERN MINING

[20] has defined a cascaded spatio-temporal pattern (CSTP) as partially ordered set of event types i.e., any two consecutive event instances in the sequences contributing to a frequent CSTP may have same or disjoint occurrence times. Three challenges in mining CSTPs are discussed – a) Handling exponential effect of event types on candidate search space b) Handling conflicting interests in mining process - statistical correctness and computational complexity c) Handling spatio-temporal neighborhood enumeration scenario where existence of overlapping neighborhoods may lead to high candidate enumeration cost. CSTP Miner provides solution to handle these challenges. Cascade participation ratio and cascade participation index are defined to evaluate the interestingness of the CSTPs. Cascade participation index handles the challenge of mining statistically correct CSTPs. The two filtering strategies - upper-bound filter and multi-resolution spatio-temporal filter handle the challenge of large candidate search space. The anti-monotonic upper bound property of cascade participation index handles the last challenge.

[27] presents which is a simple loop-based technique called as SIAM NL-CSTPM to evaluate the interestingness measure introduced in [20]. On evaluation it was found that the time consumed for calculating the interestingness measure of CSTPs is considerable and is actually higher than the time consumed by candidate generation.

[21] conducted rigorous analysis of CSTPM technique [27] to mine CSTPs and proposed a new algorithm ST partitioning based CSTPM (STP-CSTPM) to reduce the time consumed for calculating the interestingness measure of CSTPs. It also provides cost model for the filtering strategies introduced in [20] and evaluation of the filtering strategies by experimenting with large real-world data sets. Clumpiness degree is introduced as a parameter that helps to set threshold on positive spatio-temporal correlation.

8. OPEN RESEARCH ISSUES

Even after a decade of effort in mining spatio-temporal databases, the area still poses a lot of challenges. Much of the work done in spatio-temporal data mining is based on traditional mining techniques developed for transactional databases. As spatial and temporal relations are not objective in nature, handling spatio-temporal relationships through mining process is a complex and unalike task when compared to transactional databases.

Due to the bigness of spatio-temporal databases, the process of discovering frequent patterns must goes through an exhaustive search space. The search space can be reduced by eliminating useless symbols and sequences. In this regard, heuristic approaches are to be developed for efficient handling of vast search space. Also, Monte Carlo simulation and other novel techniques are needed to be explored to support intelligent selection of threshold values for evaluating the validity, strength and interestingness of the generated sequential patterns.

Mining spatio-temporal databases at varying levels of taxonomy generates frequent patterns that contains elements at different levels which provides an useful and different insight into the patterns. Customizing appropriate functions for defining spatio-temporal neighborhood is yet to be traversed.

9. CONCLUSION

In this paper, the importance of spatio-temporal data mining and its applications is laid sound. A brief survey of the approaches available for mining frequent patterns from spatio-temporal databases specifically from event data sets and trajectory data sets is provided. The research issues yet to be concentrated on are debated.

REFRENCES:

[1] Cao H., Mamoulis N., and Cheung D. W., "Mining frequent spatio-temporal sequential patterns", *Fifth IEEE International Conference on Data Mining*, 2005, pp. 8-pp.

[2] Cao H., Mamoulis N., and Cheung, D. W., "Discovery of Collocation Episodes in Spatiotemporal Data", *Proceedings of the Sixth IEEE International Conference on Data Mining (ICDM '06)*, 2006, pp. 823-827.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[3] Celik M., Shekhar S., Rogers J. P., and Shine J. A., “Mixed-drove spatiotemporal co-occurrence pattern mining”, *IEEE Transactions on Knowledge and Data Engineering,* Vol. 20, No. 10, 2008, pp. 1322-1335.

[4] Celtic M., "Discovering partial spatio-temporal co-occurrence patterns", *IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM)*, 2011, pp. 116-120.

[5] Douglas D. H., and Peucker T. K., “Algorithms for the reduction of the number of points required to represent a digitized line or its caricature”, *Cartographica: The International Journal for Geographic Information and Geovisualization*, Vol. 10, No. 2, 1973, pp. 112-122.

[6] Ester M., Kriegel H. P., Sander J., and Xu X., “A density-based algorithm for discovering clusters in large spatial databases with noise”, *Proceedings of Second International Conference on Knowledge Discovery and Data Mining*, Portland, OR, 1996, pp. 226–231.

[7] Fabrègue M., Braud A., Bringay S., Le Ber F., and Teisseire M., "Including spatial relations and scales within sequential pattern extraction", *Discovery Science*, Springer Berlin Heidelberg, 2012, pp. 209-223.

[8] Giannotti F., Nanni M., and Pedreschi D., “Efficient Mining of Sequences with Temporal Annotations”, *Proceedings of SIAM Conference on Data Mining*, April 2006, pp. 346-357.

[9] Giannotti F., Nanni M., Pinelli F., and Pedreschi D., “Trajectory pattern mining”, *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, August 2007, pp. 330-339.

[10] Gudmundsson J., and van Kreveld M., “Computing Longest Duration Flocks in Trajectory Data”, *Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems (ACM-GIS ’06)*, 2006, pp. 35-42.

[11] Gudmundsson J., Van Kreveld M., and Speckmann B., “Efficient Detection of Motion Patterns in Spatio-Temporal Data Sets”, *Proceedings of the 12th annual ACM International Workshop on Geographic information systems (ACM-GIS ’04)*, 2004, pp. 250-257.

[12] Hershberger J. E., and Snoeyink J., “Speeding up the Douglas-Peucker line-simplification algorithm”, *Proceedings of the 5th International Symposium on Data Handling*, Charleston, SC, August 1992, pp. 134-143.

[13] Huang Y., Shekhar S. and Xiong H., “Discovering Co-location Patterns from Spatial Datasets: A General Approach”, *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, Vol. 16, No. 12, 2004, pp. 1472-1485.

[14] Huang Y., Zhang L., and Zhang P., “A Framework for Mining Sequential Patterns from Spatio-Temporal Event Data Sets”, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 20, No. 4, April 2008, pp. 433-448.

[15] Kalnis P., Mamoulis N., and Bakiras S., “On Discovering Moving Clusters in Spatio-Temporal Data”, *Advances in Spatial and Temporal Databases*, Springer Berlin Heidelberg, 2005, pp. 364-381.

[16] Kang J., and Yong H. S., "Spatio-temporal discretization for sequential pattern mining", *Proceedings of the 2nd International Conference on Ubiquitous Information Management and Communication*, ACM, 2008, pp. 218-224.

[17] Koeppel I. J., and Ahlmer S. D., “Integrating the Dimension of Time into AM/FM Systems”, *Proceedings of AM/FM XVI International Annual Conference*, Aurora, Colorado, 1993.

[18] Laube P., van Kreveld M., and Imfeld S., “Finding REMO - detecting relative motion patterns in geospatial lifelines”, *Developments in spatial data handling*, Springer Berlin Heidelberg, 2004, pp. 201-214.

[19] Lee A. J., Chen Y. A., and Ip W. C., "Mining frequent trajectory patterns in spatial–temporal databases", *Information Sciences*, Vol. 179, No. 13, 2009, pp. 2218-2231.

[20] Mohan P., Shekhar S., Shine J. A., and Rogers J. P., “Cascading spatio-temporal pattern discovery: A summary of results”, *SDM*, 2010, pp. 327–338.

[21] Mohan P., Shekhar S., Shine J. A., and Rogers J. P., "Cascading spatio-temporal pattern discovery", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 24, No. 11, 2012, pp. 1977-1992.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[22] Morzy M., "Mining frequent trajectories of moving objects for location prediction", *Machine Learning and Data Mining in Pattern Recognition*, Springer Berlin Heidelberg, 2007, pp. 667-680.

[23] Pei J., Han J., Mortazavi-Asl B., Wang J., Pinto H., Chen Q., Umeshwar D., and Hsu M. C., "Mining Sequential Patterns by Pattern-Growth: The PrefixSpan Approach", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 16, No. 11, 2004, pp. 1424-1440.

[24] Pillai K. G., Angryk R. A., Banda J. M., Schuh M. A., and Wylie T., "Spatio-temporal co-occurrence pattern mining in data sets with evolving regions", *IEEE 12th International Conference on Data Mining Workshops (ICDMW)*, 2012, pp. 805-812.

[25] Qian F., He Q., and He J., "Mining spread patterns of spatio-temporal co-occurrences over zones", *Computational Science and Its Applications(ICCSA 2009)*, Springer Berlin Heidelberg, 2009, pp. 677-692.

[26] Salas H. A., Bringay S., Flouvat F., Selmaoui-Folcher N., and Teisseire M., "The pattern next door: Towards spatio-sequential pattern discovery", *Advances in Knowledge Discovery and Data Mining*, (7302), 2012, pp. 157-168.

[27] Shine J. A., Rogers J. P., Shekhar S., and Mohan P., "Discovering partially ordered patterns of Terrorism via Spatio-temporal Data Mining", *16th Army conference on Applied Statistics*, Cory, NC, USA, 2010.

[28] Tsoukatos I., and Gunopulos D., "Efficient mining of spatiotemporal patterns", *Springer Berlin Heidelberg*, 2001, pp. 425-442.

[29] Vieira M. R., Bakalov P., and Tsotras V. J., "On-line discovery of flock patterns in spatio-temporal data", *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ACM, 2009, pp. 286-295.

[30] Wang J., Hsu W., and Lee M. L., "Flowminer: Finding flow patterns in spatio-temporal databases", *16th IEEE International Conference on Tools with Artificial Intelligence(ICTAI)*, 2004, pp. 14-21.

[31] Wang J., Hsu W., and Lee M. L., "A Framework for Mining Topological Patterns in Spatio-Temporal Databases", *Proceedings of the 14th ACM International Conference on Information and Knowledge Management (CIKM '05)*, 2005, pp. 429-436.

[32] Wang J., Hsu W., and Lee M. L., "Mining Generalized Spatio-Temporal Patterns", *Database Systems for Advanced Applications*, Springer Berlin Heidelberg, 2005, pp. 649–661.

[33] Wang Z., Peng X., Gu C., and Huang B., "Mining at most top-K% mixed-drove spatiotemporal co-occurrence patterns", *9th Asian Control Conference (ASCC)*, 2013, pp. 1-5.

[34] Wei L. Y., and Shan M. K., "Mining temporal co-orientation pattern from spatio-temporal databases", *Advances in Knowledge Discovery and Data Mining*, Springer Berlin Heidelberg, 2007, pp. 895-903.

[35] Yoo J. S., and Shekhar S., "A Joinless Approach for Mining Spatial Colocation Patterns", *IEEE Transctions on Knowledge and Data Eng. (TKDE)*, Vol. 18, No. 10, 2006, pp. 1323-1337.

[36] Yoo J. S., Shekhar S., Smith J., and Kumquat J. P., "A partial join approach for mining collocation pattern", *Proceedings of the 12th annual ACM International Workshop on Geographic Information Systems*, 2004, pp. 241–249.

[37] Zhang T., Ramakrishnan R., and Livny M., "BIRCH: an efficient data clustering method for very large databases", *Proceedings of the 1996 ACM SIGMOD International Conference on Management of data - SIGMOD '96*, 1996, pp. 103-114.

[38] Zhang X., Mamoulis N., Cheung D. W., and Shou Y., "Fast Mining of Spatial Collocations", *Proceedings of 10th ACM SIGKDD Internationall Conference on Knowledge Discovery and Data Mining*, Seatle, WA, 2004, pp. 384-393.