

SPATIOTEMPORAL DATA MINING: ISSUES, TASKS AND APPLICATIONS

K.Venkateswara Rao¹, A.Govardhan² and K.V.Chalapathi Rao¹

¹Department of Computer Science and Engineering, CVR College of Engineering, Ibrahimpatnam RR District, Andhra Pradesh, India

kvenkat.cse@gmail.com

chalapatiraokv@gmail.com

²JNTUH, Hyderabad, Andhra Pradesh, India

govardhan_cse@yahoo.co.in

ABSTRACT

Spatiotemporal data usually contain the states of an object, an event or a position in space over a period of time. Vast amount of spatiotemporal data can be found in several application fields such as traffic management, environment monitoring, and weather forecast. These datasets might be collected at different locations at various points of time in different formats. It poses many challenges in representing, processing, analysis and mining of such datasets due to complex structure of spatiotemporal objects and the relationships among them in both spatial and temporal dimensions. In this paper, the issues and challenges related to spatiotemporal data representation, analysis, mining and visualization of knowledge are presented. Various kinds of data mining tasks such as association rules, classification clustering for discovering knowledge from spatiotemporal datasets are examined and reviewed. System functional requirements for such kind of knowledge discovery and database structure are discussed. Finally applications of spatiotemporal data mining are presented.

KEYWORDS

Spatiotemporal data mining, spatiotemporal data mining issues, spatiotemporal data mining tasks, spatiotemporal data mining applications

1. INTRODUCTION

A spatiotemporal object can be defined as an object that has at least one spatial and one temporal property. The spatial properties are location and geometry of the object. The temporal property is timestamp or time interval for which the object is valid. The spatiotemporal object usually contains spatial, temporal and thematic or non-spatial attributes. Examples of such objects are moving car, forest fire, and earth quake. Spatiotemporal data sets essentially capture changing values of spatial and thematic attributes over a period of time. An event in a spatiotemporal dataset describes a spatial and temporal phenomenon that may happens at a certain time t and location x . Examples of event types are earth quake, hurricanes, road traffic jam and road accidents. In real world many of these events interact with each other and exhibit spatial and temporal patterns which may help to understand the physical phenomenon behind them. Therefore, it is very important to identify efficiently the spatial and temporal features of these events and their relationships from large spatiotemporal datasets of a given application domain.

The significance of spatiotemporal data analysis and mining is growing with the increasing availability and awareness of huge amount of geographic and spatiotemporal datasets in many important application domains like

- Meteorology: all kinds of weather data, moving storms, tornados, developments of high pressure areas, movement of precipitation areas, changes in freezing level, droughts.
- Biology: animal movements, mating behavior, species relocation and extinction.
- Crop sciences: harvesting, soil quality changes, land usage management, seasonal grasshopper infestation.
- Forestry: forest growth, forest fires, hydrology patterns, canopy development, planning tree cutting, planning tree planting.
- Medicine: patients' cancer developments, supervising developments in embryology.
- Geophysics: earthquake histories, volcanic activities and prediction.
- Ecology: causal relationships in environmental changes, tracking down pollution incidents.
- Transportation: traffic monitoring, control, tracking vehicle movement, traffic planning, vehicle navigation, fuel efficient routes.

In addition to these individual areas, combinations of phenomena are also of interest. For example what changes in forests can be linked to which kind of animal behavior, which weather developments are responsible for grasshopper infestation. Moreover, some combinations pose particular planning challenges. For example, extreme weather events require rerouting of cars, planes, and ships.

Modeling and representation of spatiotemporal phenomena is complex due to two reasons. First reason is continuous and discrete changes of spatial and non spatial properties of the spatiotemporal objects. The second one is the influence of collocated neighboring spatiotemporal objects on one another. For example spread of fire is influenced by rain and changing wind speed and direction. Understanding spatiotemporal phenomena calls for processing, analysis and mining of vast amounts of spatiotemporal data along spatial, temporal and thematic attribute dimensions at multiple levels of granularity.

Spatiotemporal analysis can be categorized as temporal data analysis, spatial data analysis, dynamic spatiotemporal data analysis and static spatiotemporal data analysis. The temporal data analysis fixes the spatial dimension and analyzes how thematic attributes data change with time. Analysis of rainfall, temperature and humidity of a given region over a period of time is an example of this kind. The spatial data analysis analyzes how thematic attributes data changing with respect to a distance from a spatial reference at a specified time. Study of change in temperature and humidity values when moving away from sea coast at a given time is an example of this type. The dynamic spatiotemporal data analysis fixes thematic attributes dimension and analyzes how spatial properties change with time. Analysis of moving car data, spread of fire are examples of this category. The static spatiotemporal data analysis fixes the temporal and thematic attribute dimensions and studies the spatial dimension. An example of this is finding locations having same rainfall at same time. Analysis of large volume of spatiotemporal data without fixing any dimension is very difficult and complex. However the data mining can be used to uncover unknown patterns and trends within the data.

Spatiotemporal data mining is an emerging research area dedicated to the development and application of novel computational techniques for the analysis of large spatiotemporal databases.

It encompasses techniques for discovering useful spatial and temporal relationships or patterns that are not explicitly stored in spatiotemporal datasets. Usually these techniques have to deal with complex objects with spatial, temporal and other attributes. Both spatial and temporal dimensions add substantial complexity to the data mining process. Classical data mining techniques often perform poorly when applied to spatiotemporal data sets for many reasons. First, spatial data is embedded in a continuous space, whereas classical datasets are in discrete notions like transactions [1]. Second, a common assumption about independence of data

samples in classical statistical analysis is generally false because spatial data tends to be highly auto-correlated. Others include categorization of spatiotemporal patterns, interest measures to quantify them and design of computationally efficient and scalable algorithms to mine their instances.

This paper is organized as follows. Section 2 describes issues and challenges in general for spatiotemporal data mining. Different kinds of spatiotemporal data mining tasks and issues related to those tasks are discussed in section 3. Approach for modeling spatiotemporal data mining application and example applications are presented in section 4. Section 5 concludes the paper.

2. ISSUES AND CHALLENGES

General issues and challenges in representation, processing, analysis and mining of spatiotemporal data are described below.

1. Design and development of robust spatiotemporal representation and data structures is the fundamental issue for spatiotemporal data handling, analysis and mining.
2. The unique characteristics of spatiotemporal datasets are that they carry distance and topological information which require geometric and temporal computation.
3. Spatial and temporal relationships like distance, topology, direction, before and after are information bearing. They need to be considered in spatiotemporal data analysis and mining.
4. Spatial and temporal relationships are implicitly defined. They are not explicitly encoded in a database. These relationships must be extracted from data. There is a trade-off between preprocessing them before the actual mining process starts and computing them on-the fly as and when they are actually needed.
5. Scale effect in space and time is a challenging issue in spatiotemporal data analysis and mining. Scale in terms of spatial resolution or temporal granularity can have a direct impact on the kind and strength of spatiotemporal relationships [2] that can be discovered in datasets.
6. The unique characteristic of spatiotemporal datasets requires significant modification of data mining techniques so that they can exploit the rich spatial and temporal relationships and patterns embedded in the datasets.
7. The attributes of neighboring patterns may have significant influence on a pattern and should be considered. For example, spatiotemporal event like hurricane will have influence on traffic jam pattern.
8. Many rules of qualitative reasoning (ex: transitive property) on spatial and temporal data provide a valuable source of domain independent knowledge that should be taken into account when generating patterns. How to express rules and how to integrate them with spatiotemporal reasoning mechanism is an issue.
9. Visualization of spatiotemporal patterns and phenomena, scalability of data mining methods, data structures to represent and efficiently index spatiotemporal datasets are also challenging issues.
10. Development of efficient techniques for visualization of spatiotemporal knowledge and interaction facilities for gaining an insight of underlying phenomena represented by the knowledge is another challenge. This requires the results of spatiotemporal data mining are to be embedded within a process that interprets the results for further properly structured investigation into reasons behind the results.
11. Development of effective visual interfaces for viewing and manipulating the geometrical and temporal attributes of spatiotemporal data is another challenge.

3. SPATIOTEMPORAL DATA MINING TASKS

Regular structures in space and time, in particular, repeating structures, are often called patterns. Patterns that describe changes in space and time are referred to as spatiotemporal patterns. Spatiotemporal data mining tasks are aimed at discovering various kinds of potentially useful and unknown patterns and trends from spatiotemporal databases. These patterns and trends can be used for understanding spatiotemporal phenomena and decision making or preprocessing step for further analysis and mining. Depending on kind of knowledge to be mined, various spatiotemporal data mining tasks are described in this section.

3.1 Multidimensional analysis of spatiotemporal data

The multidimensional approach for data analysis is based on the concept of facts analyzed with respect to various dimensions. Spatiotemporal data carries multi-dimensional information such as time, location, geometry and non-spatial attributes of spatiotemporal objects. Multidimensional spatiotemporal data model integrates spatial and temporal structures to model the existence of spatial objects over time. It also supports multiple concept hierarchies for the dimensions like time, location and other attributes. This facilitates spatiotemporal data aggregation on the dimensions and dimension hierarchies which results into cuboids of spatiotemporal data cube. This data cube can be used by spatiotemporal on-line analytical processing tools [3] to perform static and dynamic spatiotemporal data analysis as well as temporal and spatial data analysis. Multidimensional model of spatiotemporal data enables to discover evolution rules which describe the manner in which spatial entities change over time. The issue here is development of new methods and techniques for high-dimensional fast analysis and aggregation of spatiotemporal data [4,5,6].

3.2 Spatiotemporal Characterization

Characterization of spatiotemporal data is performed by applying attribute oriented induction based generalization technique. Generalization is performed on spatial, non-spatial and/or temporal attributes. The attribute oriented induction does the aggregation either by attribute removal or attribute generalization. The attribute generalization involves use of concept hierarchies defined on the attribute dimension for data aggregation. Based on the order in which the generalization of attributes is done, there are different types of generalization. Spatial data dominant generalization [7] fixes the temporal dimension and does generalization of spatial attributes first and then proceeds to generalize non-spatial attributes next. Non-spatial data dominant generalization [7] fixes the temporal dimension and performs generalization on non-spatial attributes first, then generalizes spatial attributes next. Similarly spatial dimension can be fixed for characterization of non-spatial attribute data of a particular location over temporal dimension or non-spatial dimension can be fixed to characterize spatial attributes over temporal dimension. Characterization of spatiotemporal data needs the incorporation of statistical techniques used in application domains for computation and presentation. For example, characterization of climatic conditions of a given geographic region over a period of time has to consider correlations, seasonal effects and extreme values over a period of time [8].

3.3 Spatiotemporal Topological Relationship discovery

The topological relationships between two spatial objects at an instance of time can be any one among disjoints, meets, overlaps, contains, covers, intersects and equals. This relationship may change over time. Discovering the time-varying topology among objects involves processing the evolution of the spatial objects and computing topological relationship among them at different points of time [2]. The topological relationships among spatial objects can be represented using a graph in which nodes represent spatial objects and the edges represent topological relationship between the nodes. So discovery of time-varying topology results in producing a series of such graphs representing the topological relationships among the spatial

objects for different time intervals. Experimental program to detect spatiotemporal topological relationships between boundary lines of land parcel is developed in [9].

3.4 Mining Spatiotemporal Topological Relationship Patterns

The topological relationship between two spatial objects may change if geometry or location of any one of the spatial objects changes. The geometry and location changes of spatial objects with time are generally captured and stored in spatiotemporal databases. The changing topological relationship among spatial objects with time is represented using spatiotemporal topological relationship pattern [10]. For example, the topological relationship change between two spatial objects O1 and O2 from time t1 to t4 is shown in Fig., 1. The topological relationship pattern for this example can be represented as D-O-C-T where D, O, C, T corresponds to disjoint, overlaps, contains and touches respectively. Support for such patterns can be computed so that it can be used in decision making. If these patterns appear more than specified number of times, then they are called periodic patterns.

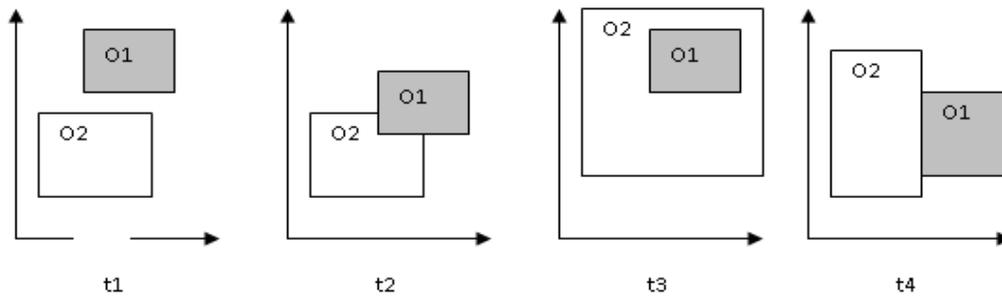


Fig 1:Topological Relationship in an interval [t1,t4]

3.5 Spatiotemporal Neighborhood

Every spatiotemporal object associated with some position(x, y) in space and a valid timestamp (ts). Two spatiotemporal objects o1, o2 are spatial neighbors if the spatial distance between them is less than specified neighborhood threshold value. The spatial distance between o1, o2 can be computed as $\text{SQRT}((o1.x - o2.x)^2 + (o1.y - o2.y)^2)$. Similarly o1 and o2 are temporal neighbors if temporal distance between them is less than specified time window. The temporal distance can be computed as modulus of $(o1.ts - o2.ts)$. The o1 and o2 are spatiotemporal neighbors if they are both spatial neighbors and temporal neighbors. The purpose of the spatiotemporal neighborhoods is to provide regions in data where knowledge discovery tasks such as clustering and outlier detection can be focused. Methods to generate spatial neighborhoods and to discretize temporal intervals are developed in[11] and tested on real life datasets related to sea surface temperature. To capture the concept of “nearby”, a neighborhood set N is defined as a set of objects such that every pair of objects in the set are spatiotemporal neighbors. Neighborhood set computation can be used as a preprocessing step to clustering, outlier detection, and collocation pattern discovery and also in online analytical processing. An algorithm for generation of spatiotemporal dynamic neighborhood is proposed and evaluated in [12] for discovering tele-connected flow anomalies.

3.6 Spatiotemporal Association Rules

Spatiotemporal Association rules (STARs) can be categorized into three types. 1. Spatiotemporal association rules involving moving or migrating objects from one region to another region. 2. Spatiotemporal association rules involving topological relationships. 3. Spatiotemporal association rules which are having thematic information of spatial objects.

3.6.1.1 STARs involving moving objects

This category of STARs [13] describes how objects move between regions over time. A STAR that represents spatial objects satisfying conditions q and migrated from one region, say r_i to another region, say r_j , during time period $[t_1, t_2]$ can be specified as

$(r_i, t_1, q) \Rightarrow (r_j, t_2) [s\%, c\%]$ where s is support of the rule and c is the confidence of the rule.

The support is the number of spatial objects that migrated from region r_i to the region r_j in time period $[t_1, t_2]$ divided by the total number of distinct objects that satisfy q during time period $[t_1, t_2]$ and are contained in any of the region in antecedent or consequent of the rule. The confidence is defined as the ratio of the number of objects migrated to the total number of objects in region r_i at time t_1 . For example, let $R1$ and $R2$ be two spatial regions. $R1$ contains spatial objects a, b, c, d, e, f and $R2$ contains spatial objects g, h, i at time t_1 . Due to migration of the objects, $R1$ contains the objects a, c, d and $R2$ contains b, e, f, g, h, i at time t_2 . The following association rule with support and confidence can be specified.

$(R1, t_1, q) \Rightarrow (R2, t_2) [33\%, 5\%]$ where $s[(R1, t_1, q) \Rightarrow (R2, t_2)] = 3/9 = 0.33$ and

$c[(R1, t_1, q) \Rightarrow (R2, t_2)] = 3/6 = 0.5$

Based on the analysis of migration of objects among regions over time, spatiotemporal regions can be characterized as stationary regions, high traffic regions. The later can further be characterized as sources, sinks and thoroughfares. A region r is a stationary region over time interval TI , if the ratio of number of objects remain in r and total number of objects in r during TI is more than user specified minimum support (say min_sup). A region r is a source if the ratio of number of objects left r to the total number of objects in r during TI is more than min_sup . A region r is a sink if the ratio of number of objects entered r to the total number of objects in r during TI is more than min_sup . If a region is both sink and source, then it is identified as thoroughfare.

3.6.1.2 STARs involving topological relationships

These rules involve spatial topology predicates like $S_overlaps$, $S_Intersects$ and Temporal predicates like T_covers , $T_Overlaps$ or spatiotemporal topological predicates such as $ST_Disjoints$, $ST_Touches$, $ST_Overlaps$. Mining this kind of rules need preprocessing of spatiotemporal data to find topological relationships and organizing those results to apply association rule mining technique or modifying the technique to generate association rules from raw spatiotemporal data. For example, $ST_Overlaps (LandParcel1, Flood, Duration1)$ and $T_Covers(Season1, Duration1) \Rightarrow Yield (LandParcel1, Low, Season1)$ is an association rule of this type.

3.6.1.3 STARs involving Thematic attributes

Association rules involving spatial, temporal features and thematic attributes or non-spatial attributes fall in this category. Some kind of preprocessing may be required while generating this type of rules. For example $rain(R_i, t_1)$ and $neighborhood(R_j, R_i) \Rightarrow flood(R_j, t_2)[s\%, c\%]$ needs neighborhood computation.

3.7 Spatiotemporal data classification

The spatiotemporal data classification is a supervised learning technique. It is a two step process, model building and model Usage. The model building stage takes classified spatiotemporal dataset as input and construct the model using classification techniques like Decision Trees, Neural Networks, Genetic algorithms or Rough sets[14]. Then the model is tested for its accuracy using spatiotemporal test dataset. The model if it is acceptable will be used to classify the new spatiotemporal objects whose class label is unknown. The techniques used for non-spatial data classification need modification to accept spatial objects and their

changes with time. For example, input layer of a neural network based classifier takes attribute values of an object as one record to compute weights of connections and error in back propagation learning technique. But in case of spatiotemporal data, the attribute values of the spatial objects including its location, shape at different timestamps are to be considered as one record for the input layer. Grouping of regions into known categories based on known climate conditions [8] using Bayes' theorem is an example for spatiotemporal classification.

3.8 Trend Prediction or Detection

Trend prediction is an important task in spatiotemporal data mining. The prediction of events occurring at particular geographic locations is very important in several application domains. Examples of problems which require location prediction include crime analysis, cellular networking, and natural disasters such as fires, floods, droughts, diseases, and earthquakes. The location and/or geometry of a moving spatial object are dynamic attributes which are function of time and other non-spatial attributes whose values change continuously. For example, location and geometry of moving cyclone depends on time, wind speed, direction and pressure. Input output pairs denoted by (x_i, y_i) are approximated by a function of the form $y = f(x)$ and used for prediction. For example, it is possible to predict the spread of a disease to different regions based on the geographic locations, highway networks, temperature, wind velocity, time and many other factors using regression and other predictive modeling methods [4]. Spatial Autoregressive Model (SAR) for linear regression and neural network based approach or support vector machines for nonlinear regression are used in prediction of climate conditions [8]. Bayesian statistical approach is used in trend prediction of total mercury in Lake Erie [15].

3.9 Spatiotemporal data clustering

Clustering is one of the major data mining methods for knowledge discovery in large databases. It is the process of grouping large data sets according to their similarity. Spatiotemporal clustering algorithms [16,17] have to consider the spatial and temporal neighbors of objects while extracting the clusters. Spatiotemporal clustering has many variants as described below due to varying requirements of different applications.

1. Clustering of regions or locations based on non-spatial attribute values of spatiotemporal objects over a period of time in a given geographic area. If this is applied to traffic management in a city, the resulting spatiotemporal clusters shows regions of more traffic at different points of time in a day.
2. Clustering of spatiotemporal objects which are moving through the regions over a period of time. If this is applied to moving objects like animals, the resulting clusters shows herd evolution and behavior of animals [18]. If it is applied to user history [19], then the representatives like centroids or medoids of resulting spatiotemporal data clusters give mobility user profile [19,20].
3. Discovering moving clusters [21,22] from spatiotemporal data where the cluster identity remains same but the objects in the cluster may not be same. If this is applied to moving vehicles, the resulting clusters model the behavior of traffic movement in a given region over a period of time.
4. Trajectory clustering [23] is the process of grouping of similar trajectories during a specific time period. One approach for trajectory clustering is partition-and-group framework [24] in which each trajectory is partitioned into a set of line segments and then similar line segments are grouped together to form a cluster. The issues in trajectory clustering are (i) identifying similarity function (ii) how clustering is to be performed. Trajectory clustering can be used in air space monitoring and traffic planning and control applications [25].
5. Shape clustering [26] technique groups the data points based on spatial density. For example, the data points that are packed within a predefined distance can be classified as one group, while the data points that are sparse outside of the neighborhood

distances can be clustered as another group. Then a shape based tracking algorithm [26] can be used to track and monitor those clusters in a sequence of images. An example of this type is monitoring of ocean objects [26].

3.10 Spatiotemporal outlier analysis

Outlier analysis discloses strange objects which appear to be inconsistent with the other objects in the dataset. The outlier objects deviate too much from other observations. The spatiotemporal outlier can be defined as a spatial referenced object whose non-spatial attribute values are significantly different from those of other objects in its spatial and temporal neighborhoods [27,28,29]. For example, a fast moving vehicle over taking many other vehicles over long period of time may not fit into any moving cluster and it can be detected as an outlier. A three step approach proposed in [27] for discovering spatiotemporal outliers is different from general approaches such as distribution based, depth based and distance based described in [28] for outlier detection. An algorithm is proposed in [29] for discovering spatiotemporal outliers and causal relationships between them. An algorithm for spatiotemporal outlier detection is proposed in [30] and used for detecting outlier sequences in precipitation data. A roughest approach is described in [31] for spatiotemporal outlier detection.

3.11 Spatiotemporal Collocation pattern or episode discovery

A spatiotemporal collocation pattern represents two or more object types whose instances are often located in spatial and temporal proximity. A collocation episode is a sequence of spatiotemporal collocation patterns with some common object types across consecutive time slots. Spatiotemporal collocation discovery uncover the existence of two or more types of spatial features that frequently locate together. For example, sets of different types of objects that change directions, speed, and geographic locations in a similar way and move close to each other for some period. An instance of this example is patterns of movement of rabbits and foxes tend to be collocated. Discovering spatiotemporal collocation episodes catch the inter-movement regularities among different types of objects [32]. For example, if a puma is moving near a deer, then a vulture is also going to move close to the same deer with high probability. In a collocation episode, there is a particular object (e.g., deer) called centre feature which participates in a sequence of collocations (e.g., deer-puma, deer-vulture). A two phase mining methodology is proposed in [32] to discover frequent collocation episodes. An algorithm is proposed in [33] to discover zonal co-location patterns for dynamic parameter. Another algorithm is discussed in [34] to discover mixed-drove spatiotemporal co-occurrence patterns.

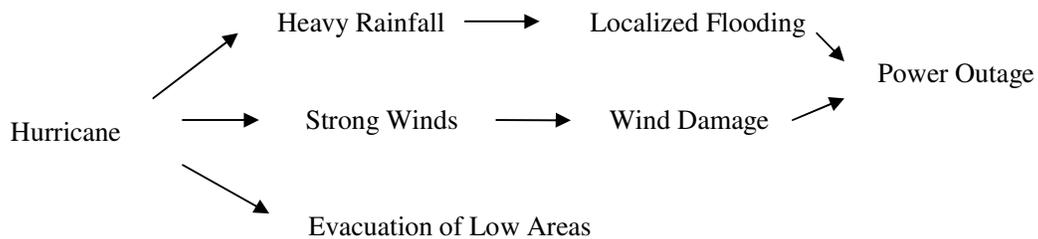
3.12 Discovering Movement Patterns

Movement patterns specify any recognizable spatial and temporal regularity or any interesting relationship in movement dataset. These patterns are classified [35] as generic patterns and behavioral patterns. Detection and description of movement patterns from spatiotemporal data are essential for better understanding of the behavior of moving objects. A sequence of time stamped point locations describing the path of a moving object is called its trajectory. Given a set of trajectories, the grouping dynamics of the moving entities described by their trajectories can be discovered. Interesting grouping dynamics are flock, leadership pattern, convergence or meeting place, periodic pattern and frequent location. The flock pattern describes group of entities moving close to each other for an extended period of time. "close to each other" means inside a circle of some specified radius r . A set of entities can have many flock patterns and even one single entity can be involved in several flock patterns. The leadership pattern is similar to the flock pattern, except that one of the entities was already moving in the specified trajectory for some time before the flock pattern occurs. Convergence or meeting place refers to the specified number of moving objects converge to the same location for specified time steps. "same location" is formalized as a circle of specified radius. A periodic pattern describes

behavior of an entity that shows the same spatiotemporal pattern with some periodicity. A frequent location refers to a frequently visited location which is a region where a single entity spends a lot of time. Approximation algorithms are described to compute flock, leadership and convergence in [36].

3.13 Cascading Spatiotemporal Pattern discovery

Cascading spatiotemporal patterns discovery [37] from a Boolean spatiotemporal event types data set uncovers partially ordered subsets of event types whose instances are located together and occur serially. Spatiotemporal event types and their instances are different. For example, a cyclone is an event-type. It may occur at many different locations at various times. Each event instance is associated with a particular occurrence time and location. The ordering may be total if event instances have disjoint occurrence times. Otherwise, ordering is partial. An examples of CSTP are (i) Analysis of climate science datasets to reveal frequent occurrence of glacier melting, intense flooding with rainfall in some areas and drought in other areas after global warming. (ii) Discovering occurrence of events cyclone, heavy rainfall, strong winds, localized flooding, wind damage and power outage after hurricane warning as shown in figure.



CSTPs occurring after a hurricane warning [37]

CSTPs are useful in applications such as natural disaster planning and climate change. CSTP discovery poses many challenges like Neighborhood enumeration is computationally challenging because of many overlapping spatiotemporal neighborhoods, Exponential Candidate Space and lack of statistically meaningful metrics to quantify interestingness of the patterns.

4. SPATIOTEMPORAL DATA MINING SYSTEM REQUIREMENTS AND APPLICATIONS

4.1 Spatiotemporal Database Structure

Let the S be the space or geographic area for which spatiotemporal data is collected. Assume that S contains regions r_1, r_2, \dots, r_n and each region intern have spatial objects (point, line, polygon) o_1, o_2, \dots, o_k at time t_1 . As time passes, different possible changes are

1. The location of the regions may change.
2. One region may split into two or more regions.
3. Two or more regions may merge in to one region.
4. The region may shrink or expand.
5. The objects in one region may move to some other region.
6. The shape of the objects may change.
7. The location of the object within the region may change.
8. Combinations of above.

The spatiotemporal database stores the spatial objects and the changes happening to them over a period of time at regular and/or irregular intervals.

4.2 System Requirements

1. Spatiotemporal data mining system should provide GUI based environment for user to specify various inputs related to task relevant data, kind of spatiotemporal task or knowledge to be discovered, interesting measures and threshold values applicable to the task and specifying method of visualization of discovered knowledge.
2. The system should push down the user inputs into data mining process as deep as possible to generate the knowledge efficiently.
3. The system should facilitate interactive analysis of data mining results.
4. Major challenge is research and development of scalable, computationally efficient data mining techniques.

Generic Framework for Spatiotemporal Data Mining System proposed in [38] and System architecture, object-oriented modeling, database design developed in [39] can be used as baseline and extended for different tasks of spatiotemporal knowledge discovery.

4.3 Applications

Spatiotemporal database applications related to animal behavior, Traffic management and Agriculture Land Management are described briefly in this section. Some possible spatiotemporal data mining tasks for each application are also identified.

4.3.1 Animal Behavior

The S is the given forest area. This is divided in to different regions. Here the regions may not change their shape over a period of time. The animals are represented as point objects which move from one region to another region over a period of time. Different kinds of knowledge that can be discovered from this kind of spatiotemporal dataset are

1. Spatiotemporal Collocation patterns or episodes
2. Moving clusters
3. Spatiotemporal Outliers
4. Trajectory clusters
5. Prediction of forest fire [40]

4.3.2 Traffic Management

The S is the given city. Different regions are the different areas within S which are connected by routes. Each route can be represented as a poly-line object. The vehicles are represented as can be discovered from this kind of database is

1. Sources, Sinks, Stationary regions and thoroughfares.
2. Spatiotemporal Association rules.
3. Spatiotemporal Clusters.
4. Spatiotemporal outliers.

4.3.3 Agriculture and Land Management

Here S is the given agriculture area. This area may be divided into different land parcels owned by different people. Each region may have one or more sub-regions that representing different crop. These sub-regions can be represented as polygon objects whose characteristics may change over a period of time. The sub-regions with in a region maybe merged and split seasonally. The regions may got split or merge over a period of time due change of ownership or streams. Sometimes a flood which can be represented as polygon objects may overlap with the regions for different durations over a period of time. Different kinds of useful knowledge that can be discovered from this spatiotemporal database are

1. Topological relationships among flood and land parcels.
2. Topological relationship patterns.
3. Hypothesis evaluation for crop rotation.
4. Spatiotemporal classification of land parcels.
5. Spatiotemporal Prediction models for crop yield.

Some other applications of spatiotemporal data mining are Sports Scene analysis, Environmental Monitoring and climate change [8], Monitoring Ocean objects. [26] and Rainfall prediction [41].

5 CONCLUSION

The rapid growth of spatiotemporal datasets due to widespread use of sensor networks and location aware devices as well as domain specific features associated with such dynamic datasets demand research in spatiotemporal data mining tasks. Spatiotemporal data mining poses many challenges and also promising applications in various domains. It is still largely unexplored area of research. In this paper, significance of spatiotemporal data analysis and mining in different domains, issues and challenges related to representation, processing, analysis, mining and visualization are discussed. Nature of spatiotemporal data, how complex it is and need for scalable and efficient algorithms is also presented. Other issues described include the reason for poor performance of classical or traditional data mining algorithms, the need for extensions, and requirements for their change. Spatiotemporal data mining tasks such as multidimensional analysis, characterization, classification, clustering, association analysis and outlier analysis of spatiotemporal data are defined, reviewed and issues in addressing those tasks are discussed. Also concepts and issues in discovering collocation patterns, episodes, cascading spatiotemporal patterns, movement patterns, trends and topological relationships from spatiotemporal data sets are reviewed. Recent research in different spatiotemporal data mining tasks is reported. Spatiotemporal association rules have received some attention. More focus is on spatiotemporal clustering. Classification is still in its infancy. Co-location mining and outlier detection have been addressed. Applications of spatiotemporal data mining tasks in different domains are reported throughout the paper as examples. Spatiotemporal database structure and its application to different domains like animal behavior, traffic management and agriculture land management along with different kinds of knowledge discovery tasks applicable in each domain are discussed.

Future work involves

- Detailed requirements analysis and development of techniques for each of the spatiotemporal data mining tasks.
- Evaluation of the techniques with large datasets in different domains at multiple spatial and temporal granularities.
- Identifying quality measures specific to each of the spatiotemporal data mining tasks.

Future work requires interdisciplinary collaboration of data miners with researchers in different domains to evaluate data mining methods and the discovered results.

It is expected that integration of spatiotemporal data mining to digital earth [42] which can integrate information relating to the atmosphere, hydrosphere, lithosphere, and biosphere of the earth [43] is likely to gain importance in future. Multidimensional modeling, analysis and mining of spatiotemporal data plays a major role in realization of highlighted applications of digital earth related to global climate change, natural disaster prevention, new energy-source development, agriculture and food security, and urban planning and management [43, 44].

REFERENCES

- [1]. Bogorny V & Shekhar S, (2010) "Spatial and Spatio-Temporal Data Mining", *IEEE 10th International Conference on Data Mining (ICDM)*, Sydney, NSW.
- [2]. K.Venkateswara Rao, Dr. A.Govardhan & Dr.K.V.Chalapati Rao, (2011) "Discovering Spatiotemporal Topological Relationships", *The second international workshop on Database Management Systems, DMS-2011*, July, Chennai, India, Springer Proceedings LNCS-CCIS 198.
- [3]. Gabriel Pestana et al, (2005) "Multidimensional Modeling based on spatial, Temporal and Spatio-Temporal Stereotypes", *ESRI International User Conference*, July, Sandiego, California.
- [4]. Jiawei Han, (2003) "Mining Spatiotemporal Knowledge: Methodologies and Research Issues", A position paper, *KDV workshop*.
- [5]. Taher Omran & Maryvonne, (2005) "Multidimensional Structures Dedicated to Continuous Spatiotemporal Phenomena", *BNCOD 2005, LNCS 3567*, pp 29-40.
- [6]. Bruno De C. Leal et.al, (2011) "From Conceptual Modeling to Logical Representation of Trajectories in SGBDOR and DW Systems", *Journal of Information and Data Management*, Vol 2, No 3.
- [7]. Jia-Dong Ren, Jie Bao & Hui-Yu Huang, (2003) "Research on Spatio-Temporal Data Model and Related Mining", *Proceeding of the Second International Conference on Machine Learning and Cybernetics*, Xi'an, 2-5 November.
- [8]. Auroop R Ganguly & Karsten Steinhäuser, (2008) "Data Mining for Climate Change and Impacts", *IEEE International conference on Data mining workshop, ICDMW*, 385-394, 15-19 Dec, Italy.
- [9]. Changbin Wu, (2011) "Detecting Spatio-Temporal Topological relationships between boundary lines of parcel", *International Conference on Remote sensing, Environment and Transportation Engineering*, Nanjing.
- [10]. K.Venkateswara Rao, Dr. A.Govardhan & Dr.K.V.Chalapati Rao, "Mining Topological Relationship Patterns from spatiotemporal Databases", *International Journal of Data Mining and Knowledge Management Process, IJDKP*, Accepted.
- [11]. Michael Mcguire, Vandana J & Aryya Gangopadhyay (2010) "Spatiotemporal Neighborhood discovery for sensor data", *In knowledge discovery from sensor data*, Vol 5840.
- [12]. J M Kang, S Shekar, M Henjum, P Novak & W Arnold, (2009) "Discovering Teleconnected Flow Anomalies : A Relationship Analysis of Spatiotemporal Dynamic Neighborhoods", *In Symposium of Spatial and Temporal Databases SSTD'09*, July 8-10, Aalborg, Denmark.
- [13]. Florian Verhein & Sanjay Chawla, (2005) "Mining Spatio-Temporal Association Ruls, Sources, Sinks, Stationary Regions and Thoroughfares in Object Mobility Databases", *Technical Report Number 574*, The University of Sydney.
- [14]. Bitner T, (2000) "Rough sets in Spatiotemporal data mining", *Proceedings of International workshop on Temporal, Spatial and Spatiotemporal Data Mining*, Lyon, France.
- [15]. M Ekram Azim et.al, (2011) "Detection of the Spatiotemporal Trends of Mercury in Lake Erie Fish Communities: A Bayesian Approach", *ACS Environmental Science & Technology*, 45(6).
- [16]. Derya Birant & Alp Kut, (2007) "ST-DBSCAN: An algorithm for clustering spatio-temporal data", *Data & Knowledge Engineering*, Volume 60, Issue 1, January, Pages 208-221.
- [17]. Manso J A, Times V C, Oliveira G, Alvares L & Bogorny V, (2010) "DB-SMoT: A Direction based Spatio-Temporal Clustering Method", *Fifth IEEE International Conference on intelligent Systems IEEE IS 2010*.
- [18]. Yan Huang, Cai Chen & Pinliang Dong, (2008) "Modeling Herds and Their Evolvments from Trajectory Data", *International Conference on Geographic Information Science, GISCIENCE*.

- [19]. Roberto Trasarti, Fabio Pinelli & Mirco Nanni, (2011) "Mining Mobility User Profiles for Car Pooling", *KDD' 11*, August 21-24, San Diego, California, USA.
- [20]. Vieira M R, Frias Martinez V, Oliver N & Frias Martinez E, (2010) "Characterizing Dense Urban Areas from Mobile Phone-call Data: Discovery and Social Dynamics", *IEEE Second International Conference on Social Computing*, Minneapolis, MN.
- [21]. Panos Kalnis, Nikos Mamoulis & Spiridon Bakirans, (2005) "On Discovering Moving Clusters in Spatio-temporal Data", *Advances in Spatial and Temporal Databases*, Springer Link, Volume 36/33/2005, p364-381.
- [22]. Zhenhui Li, Bolin Ding, Jiawei Han & Roland Kays, (2010) "Swarm: Mining Relaxed Temporal Moving Object Clusters", *36th international Conference on Very Large Data Bases*, September 13-17, Singapore.
- [23]. Susanta Satpathy, Lokesh Sharma, Ajaya K Akasapu & Netreshwari Sharma, (2011) "Towards Mining Approaches for Trajectory Data", *International Journal of Advances in Science and Technology*, Vol. 2, No.3.
- [24]. J.G Lee, J. Han & K. Y Whang, (2007) "Trajectory Clustering: A partition and group framework", *In SIGMOD*.
- [25]. Kharrat A, Zeitouni k, Sandupopa I & Faiz S, (2009) "Characterizing Traffic Density and its evolution through moving object trajectories", *Fifth international conference on signal-Image Technology and internet based systems(SITIS)*, Marrakesh.
- [26]. Yang Cai et al, (2006) "Spatiotemporal Data Mining for Monitoring Ocean Objects", *Proceedings of NASA Data Mining WJPL*.
- [27]. Derya Birant & Alp Kut, (2006) "Spatio-Temporal Outlier Detection in Large Databases", *Journal of Computing and Information Technology*, Vol. 14, No. 4, 291-297.
- [28]. Tao Cheng & Zhlin Li, (2006) "A Multiscale Approach to Detect Spatio-Temporal Outliers", *Transactions in GIS*, 10(2),253-263.
- [29]. Wei Liu, Yu Zheng, Sanjay Chawla, Jing Yuan & Xing Xie, (2011) "Discovering Spatio-Temporal Causal Interactions in Traffic Data Streams", *KDD'11*, August 21-24, San Diego, California, USA.
- [30]. Elizabeth Wu, Wei Liu & Sanjay Chawla, (2010) "Spatio-temporal outlier Detection in Precipitation Data", *Knowledge discovery from sensor data*, Volume 5840, pp 115-133.
- [31]. Alesia Albanese, Sankar K Pal & Alfredo Petrosino, (2011) "A rough set approach to spatio-temporal outlier detection", *Proceedings of 9th international conference on Fuzzy logic and applications*, Springer-verilog, LNCS Volume 6857, pp 67-74.
- [32]. Huiping Cao, Nikos Mamoulis & David W. Cheung, (2006) "Discovery of Collocation Episodes in Spatiotemporal Data", *6th International Conference on Data Mining*, ICDM, p823-827, December, Hongkong.
- [33]. M Celik, J M Kang & S Shekar, (2007) "Zonal Co-location pattern discovery with dynamic parameters", *In proceedings of 7th IEEE international conference on data mining (ICDM)*, Omaha, Nebraska.
- [34]. M Celik, S Shekar, J Rogers & J Shine, (2008) "Mining Mixed-drove Spatio-Temporal Co-Occurrence Patterns", *IEEE Transactions on Knowledge and Data Engineering*, 20(10):1322-1335.
- [35]. Somayeh Dodge, Robert Weibel et al, (2008) "Towards a Taxonomy of Movement Patterns", *Information Visualization* (2008) 7, 240-252.
- [36]. Joachim Gudmundsson et al, (2004) "Efficient Detection of Motion Patterns in Spatio-Temporal Data Sets", *GIS'04*, November 12-13, Washington, DC, USA.

- [37]. Pradeep Mohan, Shashi Shekhar, James A. Shine & James P. Rogers, (2011) “Cascading spatio-temporal pattern discovery”, *IEEE Transactions on Knowledge and Data Engineering*, accepted, Draft digital Id 10.1109/TKDE.2011.146, May 4.
- [38]. K.Venkateswara Rao, Dr. A.Govardhan & Dr.K.V.Chalapati Rao, (2008) “A Generic Framework for Spatio-Temporal Data Mining System”, *Journal of Andhra Pradesh Society for Mathematical Sciences*, Vol 1,No 2,July.
- [39]. K.Venkateswara Rao, Dr. A.Govardhan & Dr.K.V.Chalapati Rao, (2011) “An Object-Oriented Modeling and Implementation of Spatio-Temporal Knowledge Discovery System”, *International Journal of Computer Science & Information Technology(IJCSIT)*, Vol 3, No 2, April.
- [40]. T. Cheng & J.Wang, (2006) “Application of Spatio-Temporal Data Mining and Knowledge Discovery (STDMD) for Forest Fire Prevention”, *ISPRS Commission VII Mid-term Symposium*, Enschede, the Netherlands, 8-11 May.
- [41]. P.SLucio etal, (2007) “Spatiotemporal monthly rainfall reconstruction via artificial neural network – case study: South of Brazil”, *Advances in Geosciences*, 10, 67-76.
- [42]. Karl E Grossner, Michael F Goodchild & Keith C Clarke, (2008) “Defining a Digital earth System “, *Transaction in GIS*, 12(1):145-160.
- [43]. H D Guo, Z Liu & L W Zhu, (2010) “Digital Earth: decadal experiences and some thoughts”, *International Journal of Digital Earth*, Vol 3, No 1, March, 31-46.
- [44]. Huadong Guo, (2010) “Understanding global natural disasters and the role of earth observation”, *International Journal of Digital earth*, Vol 3, No 3, Sept.

Authors

Mr. K. Venkateswara Rao is working as Associate professor in Computer Science and Engineering Department at CVR College of Engineering. He is pursuing his part time PhD from Jawaharlal Nehru Technological University Hyderabad. He received his M.Tech in Computer Science and Engineering from Osmania University. Previously he worked as Software Specialist at Wipro Technologies and as Scientist at ISRO Satellite Centre. His areas of interest includes Databases, Data Warehousing & Mining, Operating systems and Realtime Systems.



Dr.A.Govardhan: received Ph.D. degree in Computer Science and Engineering from Jawaharlal Nehru Technological University in 2003, M.Tech from Jawaharlal Nehru University in 1994 and B.E. from Osmania University in 1992. He is a Professor in CSE and Director of Evaluation of Jawaharlal Nehru Technological University Hyderabad, He has published more than 152 research papers in various national and international Journals/conferences. His research of interest includes Databases, Data Warehousing & Mining, Information Retrieval, Computer

Dr K.V. Chalapati Rao is a Professor of Computer Science & Engg., and Dean, Academics at CVR College of Engineering. Prior to joining the CVR, he served Osmania University as a Professor & Head, Department of CSE and Dean of Engineering. After obtaining his PhD, Dr. Rao joined Electronics Corporation of India Limited and worked in various capacities for 16 years, before joining the Osmania University. He guided number of PhD scholars in areas of Real time systems, Operating Systems, Software Engineering, Distributed Systems, Knowledge and Data Engineering.

