© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 07 Issue: 03 | 24 | April-June 2016
Paper is available at www.jrps.in | Email : info@jrps.in

INVESTIGATION OF ISSUES, TASKS & APPLICATIONS OF TEMPORAL DATA MINING IN IT INDUSTRIES

REENA DHULL, Research Scholar, reenadhull15@gmail.com
Er. Urvashi Garg, Assistant Professor

ABSTRACT : In our research problems & challenges related to spatiotemporal data representation, analysis, mining & visualization of knowledge have been presented. Many type of data mining tasks like association rules, classification clustering for discovering knowledge from spatiotemporal datasets are examined & reviewed. System functional requirements for such kind of knowledge discovery & database structure are discussed. So applications of spatiotemporal data mining are presented. Such applications are related to huge data of processed within IT industries. Temporal Data Mining is a rapidly evolving area of research that is at intersection of several disciplines, consisting statistics, temporal pattern recognition, temporal databases, optimization, visualization, high-performance computing & parallel computing. Spatiotemporal data generally consists states of an object and event or a position within space over a period of time.

KEYWORDS: *Spatiotemporal data mining issues, spatiotemporal data mining tasks,KMean, Neural Network spatiotemporal data mining applications, Datamining, Fuzzy logic, Spatiotemporal data mining*

1. INTRODUCTION

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

2. ISSUES & CHALLENGES

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

1. Design & development of robust spatiotemporal representation & data structures is fundamental issue for spatiotemporal data handling, analysis & mining.
2. unique characteristics of spatiotemporal datasets are that they carry distance & topological information which require geometric & temporal computation.
3. Spatial & temporal relationships like distance, topology, direction, before & after are information bearing. They need to be considered in spatiotemporal data analysis & mining.
4. Spatial & 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 actual mining process starts & computing them on-the fly as & when they are actually needed.
5. Scale effect in space & 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 kind & strength of spatiotemporal relationships that can be discovered in datasets.
6. unique characteristic of spatiotemporal datasets requires significant modification of data mining techniques so that they can exploit rich spatial & temporal relationships & patterns embedded in datasets.
7. attributes of neighboring patterns may have significant influence on a pattern & should be considered. For example, spatiotemporal event like hurricane will have influence on traffic jam pattern.
8. Many rules of qualitative reasoning on spatial & temporal data provide a valuable source of

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 07 Issue: 03 | 24 | April-June 2016
Paper is available at www.jrps.in | Email : info@jrps.in

domain independent knowledge that should be taken into account when generating patterns. How to express rules & how to integrate them with spatiotemporal reasoning mechanism is an issue.

9. Visualization of spatiotemporal patterns & phenomena, scalability of data mining methods, data structures to represent & efficiently index spatiotemporal datasets are also challenging issues.

10. Development of efficient techniques for visualization of spatiotemporal knowledge & interaction facilities for gaining an insight of underlying phenomena represented by knowledge is another challenge. This requires results of spatiotemporal data mining are to be embedded within a process that interprets results for further properly structured investigation into reasons behind results.

11. Development of effective visual interfaces for viewing & manipulating geometrical & temporal attributes of spatiotemporal data is another challenge.

3. PROPOSED WORK

Basic steps involved in our proposed work are as follow:

(i) 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. attribute oriented induction does aggregation either by attribute removal or attribute generalization. attribute generalization involves use of concept hierarchies defined on attribute dimension for data aggregation. Based on order in which generalization of attributes is done, there are different types of generalization. Spatial data dominant generalization fixes temporal dimension & does generalization of spatial attributes first & then proceeds to generalize non-spatial attributes next. Non-spatial data dominant generalization fixes temporal dimension & performs generalization on nonspatial 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 incorporation of statistical techniques used in application domains for computation & presentation. For example, characterization of climatic conditions of a given geographic region over a period of time has to consider correlations, seasonal effects & extreme values over a period of time.

(ii)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 & equals. This relationship may change over time. Discovering time-varying topology among objects involves processing evolution of spatial objects & computing topological relationship among them at different points of time. topological relationships among spatial objects can be represented using a graph in which nodes represent spatial objects & edges represent topological relationship between nodes. So discovery of time-varying topology results in producing a series of such graphs representing topological relationships among spatial objects for different time intervals. Experimental program to detect spatiotemporal topological relationships between boundary lines of land parcel is developed in.

(iii)Mining Spatiotemporal Topological Relationship Patterns

The topological relationship between two spatial objects may change if geometry or location of any one of spatial objects changes. geometry & location changes of spatial objects with time are generally captured & stored in spatiotemporal databases. changing topological relationship among spatial objects with time is represented using spatiotemporal topological relationship pattern. For example, topological relationship change between two spatial objects O1 & O2 from time t1 to t4 is shown in following Figure. Topological relationship pattern for this example can be represented as D-O-C-T where D, O, C, T corresponds to disjoints, overlaps, contains & touches respectively. Support for such pattern scan 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.

(iv)Spatiotemporal Neighborhood

Every spatiotemporal object associated with some position(x, y) in space & a valid timestamp (ts). Two spatiotemporal objects o1, o2 are spatial neighbors if spatial distance between them is less than specified

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 07 Issue: 03 | 24 | April-June 2016
Paper is available at www.jrps.in | Email : info@jrps.in

neighborhood threshold value. spatial distance between o1, o2 can be computed as SQRT ((*o*1.*x* - *o*2.*x*) 2 + (*o*1.*y* - *o*2.*y*) 2). Similarly o1 & o2 are temporal neighbors if temporal distance between them is less than specified time window. temporal distance can be computed as modulus of (o1.ts – o2.ts). o1 & o2 are spatiotemporal neighbors if they are both spatial neighbors & temporal neighbors. purpose of spatiotemporal neighborhoods is to provide regions in data where knowledge discovery tasks such as clustering & outlier detection can be focused. Methods to generate spatial neighborhoods & to discrtize temporal intervals are developed in & tested on real life datasets related to sea surface temperature. To capture concept of "nearby", a neighborhood set N is defined as a set of objects such that every pair of objects in set are spatiotemporal neighbors. Neighborhood set computation can be used as a preprocessing step to clustering, outlier detection, & collocation pattern discovery & also in online analytical processing. An algorithm for generation of spatiotemporal dynamic neighborhood is proposed & evaluated in for discovering tele-connected flow anomalies.

(v) Spatiotemporal data clustering

Clustering is one of major data mining methods for knowledge discovery in large databases.
It is process of grouping large data sets according to their similarity. Spatiotemporal clustering algorithms have to consider spatial & temporal neighbors of objects while extracting 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, 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 regions over a period of time. If this is applied to moving objects like animals, resulting clusters shows herd evolvement & behavior of animals. If it is applied to user history, then representatives like centroids or medoids of resulting spatiotemporal data clusters give mobility user profile.

3. Discovering moving clusters from spatiotemporal data where cluster identity remains same but objects in cluster may not be same. If this is applied to moving vehicles, resulting clusters model behavior of traffic movement in a given region over a period of time.

4. Trajectory clustering is process of grouping of similar trajectories during a specific time period. One approach for trajectory clustering is partition-and-group framework in which each trajectory is partitioned into a set of line segments & then similar line segments are grouped together to form a cluster. 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 & traffic planning & control applications.

5.Shape clustering technique groups data points based on spatial density. For example, data points that are packed within a predefined distance can be classified as one group, while data points that are sparse outside of neighborhood.

4. K-MEAN CLUSTERING PROCESS FOR SPATIAL DATABASE

Suppose we have following data set

2	5	6	8	12	15	18	28	30

Suppose K=3

C1=2

C2=12

C3=30

2	5	6	8	12	15	18	28	30
C1				C2				C3

So cluster according to distance are as follow

12-5>5-2

So cluster for data point 5 is C1

6-2>12-6

So cluster for data point 6 is C1

In same way cluster would be assigned

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 07 Issue: 03 | 24 | April-June 2016
Paper is available at www.jrps.in | Email : info@jrps.in

2	5	6	8	12	15	18	28	30
C1	C1	C1	C2	C2	C2	C2	C3	C3

Data member of C1 are 2,5,6

Data Member for C2 are 8,12,15,18

Data Member for C3 are 28,30

So mean of cluster C1 is (2+5+6)/3=4.3

So mean of cluster C2 is (8+12+15+18)/4=13.25

So mean of cluster C3 is (28+30)/2=29

Now distance would be recalculated with new mean and the cluster of data point would be changed according to new distance

2	5	6	**8**	12	15	18	28	30
C1	C1	C1	**C2**	C2	C2	C2	C3	C3

For example take 8 from C2 cluster

5. Comparative analysis of result between OLD AND NEW K-MEAN

Number of record	Old Cluster	New Cluster
1000	2	1
1500	2	2
2000	3	2
2500	3	3
3000	4	3
4000	6	4
5000	8	5
6000	9	6
7000	9	7
8000	12	8
9000	13	9
10000	14	10

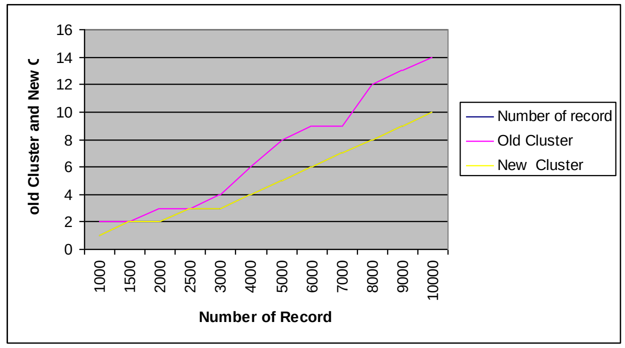

Fig 2 Analysis of old and new cluster

5. SYSTEM REQUIREMENTS

Hardware Requirement

CPU: 1GHZ above
RAM: More than 1 GB.
HARDISK: 5 GB. Free space
Monitor: High Resolution
Keyboard, Mouse

Software Requirement

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 & threshold values applicable to task & specifying method of visualization of discovered knowledge.
2. system should push down user inputs into data mining process as deep as possible to generate knowledge efficiently.
3. system should facilitate interactive analysis of data mining results.
4. Major challenge is research & development of scalable, computationally efficient data mining techniques.

6. AREA OF APPLICATIONS

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

Animal Behavior

The S is given forest area. This is divided into different regions. Here regions may not change their shape over a period of time. animals are represented as point objects which move from one region to another region over a period of time. Different kinds

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 07 Issue: 03 | 24 | April-June 2016
Paper is available at www.jrps.in | Email : info@jrps.in

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

Traffic Management

The S is given city. Different regions are different areas within S which are connected

by routes. Each route can be represented as a poly-line object. vehicles are represented as

can be discovered from this kind of database is

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

Agriculture & Land Management

Here S is 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.

1. Topological relationships among flood & 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.

7.SCOPE & CONCLUSION

Spatiotemporal data mining poses many challenges & also promising applications in various domains. It is still largely unexplored area of research. In this research importance of spatiotemporal data analysis & mining in different domains, issues & challenges related to representation, processing, analysis, mining & visualization are discussed.

Nature of spatiotemporal data, how complex it is & need for scalable & efficient algorithms is also presented. Other issues described include reason for poor performance of classical or traditional data mining algorithms, need for extensions, & requirements for their change. Spatiotemporal data mining tasks such as multidimensional analysis, characterization, classification, clustering, association analysis & outlier analysis of spatiotemporal data are defined, reviewed & issues in addressing those tasks are discussed. Also concepts & issues in discovering collocation patterns, episodes, cascading spatiotemporal patterns, movement patterns, trends & 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 & outlier detection have been addressed. Applications of spatiotemporal data mining tasks in different domains are reported throughout paper as examples.

Spatiotemporal database structure & its application to different domains like animal behavior, traffic management & agriculture land management along with different kinds of knowledge discovery tasks applicable in each domain specially in case of IT industry are discussed.

REFERENCES

[1] D. Brillinger, editor. Time Series: Data Analysis & Theory. Holt, Rinehart & Winston, New York, 1975.

[2] P. Cheeseman & J. Stutz. Bayesian classification (AUTOCLASS): Theory & results. In U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, & R. Uthurusamy, editors, Advances in Knowledge Discovery & Data Mining. AAAI Press / MIT Press, 1995.

[3] T. Fulton, S. Salzberg, S. Kasif, & D. Waltz. Local induction of decision trees: Towards interactive data mining. In Simoudis et al. [21], page 14.

[4] B. R. Gaines & P. Compton. Induction of metaknowledge about knowledge discovery. IEEE Trans. On Knowledge & Data Engineering, 5:990–992, 1993.

[5] C. Glymour, D. Madigan, D. Pregibon, & P. Smyth. Statistical inference & data mining. Communications of ACM, 39(11):35–41, Nov. 1996.

[6] F. H. Grupe & M. M. Owrang. Data-base mining - discovering new knowledge & competitive advantage. Information Systems Management, 12:26–31, 1995.

[7] J. Han, Y. Cai, & N. Cercone. Knowledge discovery in databases: An attribute-oriented approach. In Proceedings of 18th VLDB Conference, pages 547–559, Vancouver, British Columbia, Canada, Aug. 1992.

[8] J. W. Han, Y. D. Cai, & N. Cercone. Data-driven discovery of quantitative rules in relational databases.

Ieee Trans. On Knowledge & Data Engineering, 5:29– 40, Feburary 1993.
[9] J. W. Han, Y. Yin, & G.Dong. Efficient mining of partial periodic patterns in time series database. IEEE Trans. On Knowledge & Data Engineering, 1998.
[10] D. Heckerman, H. Mannila, D. Pregibon, & R. Uthurusamy, editors. Learning bayesian networks: combineation of knowledge & statistical data. AAAI Press, 1994.
[11] D. Heckerman, H. Mannila, D. Pregibon, & R. Uthurusamy, editors. Proceedings of Third International Conference on Knowledge Discovery & Data Mining (KDD-97). AAAI Press, 1997.
[12] E. Keogh & P. Smyth. A probabilistic approach to fast pattern matching in time series databases. Page 126.
[13] A. Ketterlin. Clustering sequences of complex objects. In Heckerman et al. [11], page 215.
[14] C. Li & G. Biswas. Temporal pattern generation using hidden markov model based unsuperised classifcation. In Proc. of IDA-99, pages 245–256, 1999.