© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 07 Issue: 03 | 23 | April-June 2016
Paper is available at www.jrps.in | Email : info@jrps.in

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

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

ABSTRACT: Temporal Data Mining is a rapidly evolving area of research that is at intersection of several disciplines, including statistics, temporal pattern recognition, temporal databases, optimization, visualization, high-performance computing, & parallel computing. Spatiotemporal data usually contain states of an object, an event or a position within space over a period of time. Vast amount of spatiotemporal data can be found within several application fields such as traffic management, environment monitoring, & weather forecast. These datasets might be collected at different locations at various points of time within different formats. It poses many challenges within representing, processing, analysis & mining of such datasets because of complex structure of spatiotemporal objects & relationships among them in both spatial & temporal dimensions. In this research problems & challenges related to spatiotemporal data representation, analysis, mining & visualization of knowledge are presented. Several kinds of data mining tasks such as 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. Finally applications of spatiotemporal data mining are presented. These applications are related to huge data of processed within IT industries.

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

1. INTRODUCTION

A spatiotemporal object can be defined as an object that has at least one spatial & one temporal property. Spatial properties are location & geometry of object. 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 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. Significance of spatiotemporal data analysis & mining is growing with increasing availability & awareness of huge amount of geographic & spatiotemporal datasets in many important application domains such as

1. Meteorology: all kinds of weather data, moving storms, tornados, developments of high pressure areas, movement of precipitation areas, changes within freezing level, droughts.
2. Biology: animal movements, mating behavior, species relocation & extinction.

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 07 Issue: 03 | 23 | April-June 2016
Paper is available at www.jrps.in | Email : info@jrps.in

3. Crop sciences: harvesting, soil quality changes, land usage management, seasonal grasshopper infestation.
4. Forestry: forest growth, forest fires, hydrology patterns, canopy development, planning tree cutting, planning tree planting.
5. Medicine: patients' cancer developments, supervising developments within embryology.
6. Geophysics: earthquake histories, volcanic activities & prediction.
7. Ecology: causal relationships within environmental changes, tracking down pollution incidents.
8. Transportation: traffic monitoring, control, tracking vehicle movement, traffic Planning, vehicle navigation, fuel efficient routes.

2. LITERATURE REVIEW

There was continued work within early 1980's on AI approaches to managing uncertainty & temporal reasoning for planning & problem solving (Allen&Hayes&Kautz&Koomen, Dean&Mc-Dermott) & analysis of information within clinical databases, generally to detect & eventually to confirm causality (Blum&Downs&Walker&Wiederhold). During early part of this decade, support for time within E-R. model (DeAntonellis&Degli-Antoni&Mauri &Zonta, Klopprogge&Lockemann) & within semantic data models (Hammer&Mc-Leod) surfaced. Work continued within conceptual modeling (Anderson, Ariav, Bubenko&Gustafssongz Karlsson, DeAntonellis&Zonta).

Various aspects of transaction time were explored: hypothetical relations, which are rollback relations with branching transaction time (Ekland&Price, Stonebraker&Woodfill), snapshot relations (Adiba&Lindsay), & optical disk structures (Maier, Kathmann), which was followed by later work (Bulgren&Canas, Easton, Vitter). Overmyer&Stonebraker explored implementing valid time as an abstract data type.

[Clifford & Warren 1983] This paper was a watershed article. paper provided a formal semantics for both an historical data model & a calculus-based query language, first of many to be proposed. (Some felt model was *too* formal because it involved a variant of Montague's complex intentional logic originally formulated within context of computational linguistics.) It was first article on time-oriented databases to appear within a major database journal, thereby introducing this community to topic. Shortly after this article appeared, activity increased significantly.

[Dadam et al. 1984, Lure et al. 1984, Lure et al. 1985] These papers, following Clifford & Warren paper by one to two years, offered a nice counterbalance. They described various aspects of Advanced Information Management project within IBM Heidelberg Scientific Center. prototype temporal DBMS described within these papers was first to support both valid & transaction time, & first to support temporal indexing.

[Stonebraker 1987] Postgres embodies first concrete implementation proposal for optical disks for rollback relations (Stonebraker imaginatively calls rollback operation "time travel"). proposed transaction management & concurrency control algorithms were designed with permanent archiving within mind.

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 07 Issue: 03 | 23 | April-June 2016
Paper is available at www.jrps.in | Email : info@jrps.in

In mid 1980's, Ahn&Snodgrass showed that transaction time & valid time are truly orthogonal, allowing each to be pursued independently. Versioning, which is concerned with transaction time (again, generally branching time), garnered attention (Blanker&Ijbema, Chang&Katz, Chou&Kim, Dittrich, Lu&Verma, Weikum). Ginsburg defined *object histories* & investigated them within depth with his associates, Dong, Gyssens, Kurtman, Tanaka, Tang, & Tian. Work within areas of temporal inferencing (Coelha&Cotta&Lee, Karlsson, Sheng) & integrity constraints across transactions (Abiteboul&Vianu, Casanova, Ceri, Ehrich&Gogolla&Lipeck&Saake, Kung, Mark&ttoussopoulos, Ngu, Tanabe) continued. Algebraic & calculus-based query languages (and their associated data models) incorporating time also started to appear (Arkun&Tansel, Ariav&Beller&Morgan, BenZvi, Clifford&Croker, GadiagzYeung, Snodgrass). [TAIS 1987] This conference was first devoted to temporal databases. papers within this conference emphasized conceptual modeling, with individual papers considering most other topics active at time.

By this point within time (mid-1987), many of important aspects of time-oriented databases had been addressed, if only within an initial fashion. Both algebraic & calculus-based query languages had been defined, prototype implementations existed, & there were solid results within data modeling & conceptual design.

K.Venkateswara Rao1, A.Govardhan2 & K.V.Chalapati Rao (2012) wrote on SPATIOTEMPORAL DATA MINING: ISSUES, TASKS & APPLICATIONS

In this research problems & challenges related to spatiotemporal data representation, analysis, mining & visualization of knowledge are presented. Several kinds of data mining tasks such as association rules, classification clustering for discovering knowledge from spatiotemporal datasets are examined & reviewed. System functional requirements for such kind of knowledge discovery & database structure were discussed.

Weiqiang Lin wrote "An overview of Temporal Data Mining"

Temporal Data Mining is a rapidly evolving area of research that is at intersection of several disciplines, including statistics, temporal pattern recognition, temporal databases, optimization, visualization, high-performance computing, & parallel computing. This paper is first intended to serve as an overview of temporal data mining within research & applications. In this research they have provided an overview of temporal data mining process & some background to Temporal Data Mining. Also they discussed a difficult & fundamental problem, a general analysis theory of temporal data mining & provided some answers to problem. This leads into a discussion on why there was a need for Temporal Data Mining in industry, which has been a major factor in efforts that have gone into building present generation of Temporal Data Mining Systems. They have presented a number of areas which are related to Temporal Data Mining in their objectives & compared & contrasted these technologies with Temporal Data Mining.

Rakesh Kumar, P.K. Suri, & R.K. Chauhan(2005) wrote on Search Engines Evaluation

The volume of world wide web (WWW) is increasing enormously due to a world wide move to

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 07 Issue: 03 | 23 | April-June 2016
Paper is available at www.jrps.in | Email : info@jrps.in

migrate information to online sources. To search some information on WWW, search engines are used, which when presented with queries, return a list of web pages ranked on basis of estimation of relevance. Generally search engines due to abundance of information available on web return millions of pages. But user studies indicate that a common user browses through top 10 or 20 documents only. So it's all-important to get into those top 10 documents. To achieve this web authors are increasingly beginning to rely on underhand techniques to ensure their sites get seen, in turn affecting performance of search engines. existing measures to evaluate these systems' performance are not adequate in current world of highly interactive end-user systems. In this study a metric 'Ranked Precision' is proposed to evaluate performance of search engines.

Richard Snodgrass wrote on "Temporal Databases Status & Research Directions"

It seems somehow fitting to begin this paper on databases that store historical information with a chronology, touching briefly on all work that he was aware of in this area. He discuss in some detail what he consider to be ten most important papers & events in terms of their impact on discipline of temporal databases. These are emphatically not meant to detract from other excellent papers in temporal databases. His goal was to characterize evolution of this field, as an introduction to approximately 350 papers specifically relating time to databases that have appeared thus fax. He then identified & discussed areas where more work is needed.

3. PROBLEM FORMULATION

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

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 07 Issue: 03 | 23 | April-June 2016
Paper is available at www.jrps.in | Email : info@jrps.in

considered. For example, spatiotemporal event like hurricane will have influence on traffic jam pattern.

4. OBJECTIVE OF RESEARCH

a. Collecting datasets from different locations at various points of time in different formats.
b. Presentation of problems & challenges related to spatiotemporal data representation, analysis, mining & visualization of knowledge.
c. Spatiotemporal data mining tasks on data set of IT industry
d. Examining & reviewing several kinds of data mining tasks such as association rules, classification clustering for discovering knowledge from spatiotemporal datasets.
e. Spatiotemporal Topological Relationship discovery.
f. Analysing reason for poor performance of classical or traditional data mining algorithms, need for extensions, & requirements for their change.

5. RESEARCH METHODOLOGY

SPATIOTEMPORAL DATA MINING TASKS

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

Multidimensional analysis of spatiotemporal data

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

6. PROPOSED WORK

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 07 Issue: 03 | 23 | April-June 2016
Paper is available at www.jrps.in | Email : info@jrps.in

Basic steps involved in our proposed work are as follow:

(i) Spatiotemporal Characterization

(ii)Spatiotemporal Topological Relationship discovery

(iii)Mining Spatiotemporal Topological Relationship Patterns

(iv)Spatiotemporal Neighborhood

(v) Spatiotemporal data clustering

9.SCOPE & CONCLUSION

The rapid growth of spatiotemporal datasets due to widespread use of sensor networks & 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 & 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.

[15] M.J.Zaki. Fast mining of sequential patterns in very large databases. Uni. of Rochester Technical report, 1997.

[16] S. a. O.Etzion, editor. Temporal databases: Research & Practice. Springer-Verlag,LNCS1399, 1998.

[17] B. Padmanabhan & A. Tuzhilin. Pattern discovery in temporal databases: A temporal logic approach. In Simoudis et al. [21], page 351.

[18] P.sprites, C.Glymour, & R.Scheines. Causation, Prediction & Search. Springer-Verlag, 1993.

[19] J. Roddick & M. Spiliopoulou. A survey of temporal knowledge discovery paradigms & methods. IEEE Transactions on Knowledge & Data Engineering, 2002.

[20] R.O.Duda & P. Hart. Pattern classification & scene analysis. John Wiley & Sons, 1973.

[21] E. Simoudis, J. W. Han, & U. Fayyad, editors. Proceedings of Second International Conference on Knowledge Discovery & Data Mining (KDD-96).AAAI Press, 1996.