International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 11, November-2020

A Review on Machine Learning - Spatiotemporal Data Mining: Issues, Tasks and Applications

[1]P. Madhuri, [2]Dr.K. Srinivasa Rao, [3]Dr. S.K. Yadav
Department of Computer Science and Engineering[1, 2, 3]
[1]Research Scholar, Shri Jagdishprasad Jhabarmal Tibrewala University.
[2]Professor and Principal, MLRIT, Hyderabad.
[3]Professor, Shri Jagdishprasad Jhabarmal Tibrewala University.

ABSTRACT: Spatiotemporal data generally involves the state o f an object, an occurrence, or a spatial location over a period. In many application areas, such as traffic control, climate monitoring, and weather prediction, a significa nt amount of spatiotemporal data can be found. These datasets could be gathered in various formats. At different locations at different points of time. Because of the complex nature of spatiotemporal objects and their relationshi ps in both spatial and temporal dimensions, this raises many difficulties in the representation, c ollection, interpretation, and mining of such datasets. Sometimes, spatio-temporal data sets are very board and hard to interpret and view. In this paper we propose several data mining tasks such a s association rules, classification clusters that are analysed and tested to discover information from spatiotemporal datasets. System functional criteria are addressed for certain kinds of information discovery. Finally, applications are presented for spatiotempor al data mining.

Keywords: Data Mining, Temporal Data Mining, Spatial Data Mining, Spatio-Temporal Data Mining

1. INTRODUCTION

Naturally the data mining progress has led to the discovery of application domains within which data m ining can be used. As many of these domains have an inherently temporal or spatial context, to correctly i nterpret the data obtained, the time and/or space dimension must be taken into account in the mining process.

Variable Energy Resources

The ambitious Renewable Portfolio Standards (RPSs) have been mandated by many countries in the world and many sta tes in the United States. Wind energy itself is projected to ri se among various renewable energy resources to provide bet ween 15 to 25 percent of the world's electricity by 2050. According to another report, since 2000, the world's total capacity for wind power has do ubled every three years, reaching an installed capacity of 19 7 GW in 2010 and 369 GW in 2014 (CEC, 2013) (IEA, 201 3). However, the random nature of the wind makes it difficu lt to hit.

The required power balance for its integration into the grid. Accurate forecasts promote the use of ancillary res ources such as frequency control and load following to com pensate for such imbalances (Hao et al., 2013), (Sanandaji et al., 2014). With the growing availability and knowledge of huge numbers of geographical and spatiotemporal datasets i n many significant application domains, such as

- Meteorology: all types of weather info, moving stor ms, tornados, high-pressure area developments, movement of precipitation zone s, freezing level shifts, droughts.
- Biology: movements of animals, mating activity, re location, and extinction of species.
- Crop sciences: harvesting, improvements in soil qu ality, control of land use, seasonal grasshopper infe station.
- Forestry: forest growth, forest fires, patterns of hyd rology, creation of canopies, tree cutting planning, t ree planting planning.
- Medicine: progression of cancer in patients, tracking advances in embryology.
- Geophysics: history of earthquakes, volcanic activit y, and forecasting.
- Ecology: causal correlations in environmental chan ges, monitoring of instances of emissions.
- Transportation: traffic monitoring, regulation, vehi cle movement tracking, traffic planning, navigation of vehicles, fuel efficient routes.

2. ISSUES AND CHALLENGES

List below are general issues and challenges in the representation, collection, analysis, and mining of spatiotem poral data.

1. The fundamental question for spatiotemporal data handlin g, analysis and mining is the design and creation of stable sp atiotemporal representation and data structures.
2. The specific characteristics of spatiotemporal datasets are that they bear geometric and temporal computations that inv olve distance and topological knowledge.
3. Knowledge is carried by spatial and temporal relationship s such as distance, topology, direction, before and after. In s patiotemporal data analysis and mining, they must be considered.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 11, November-2020

4. There are implicitly defined spatial and temporal relations . In a database, they are not encoded directly.
It is important to extract these relationships from dataUntil t he actual mining phase begins, there is a tradeoff between pr eprocessing them and computing them on-the-fly as and when they are required.
5.Scale effects in space and time are a difficult problem in t he study and mining of spatiotemporal data.
Scale may have a direct effect on the type and strength of sp atiotemporal relationships that can be contained in datasets i n terms of spatial resolution or temporal granularity.
6. The unique aspect of spatiotemporal datasets needs major modifications to data mining techniques so that the rich spa tial and temporal relationships and patterns embedded in the datasets can be exploited.
7. Neighboring pattern attributes can have a huge effect on a pattern and should be considered.
For example, spatiotemporal events such as hurricanes can c hange the pattern of traffic jams.
8.Many qualitative reasoning rules on spatial and temporal d ata e.g. transitive property) provide a valuable source of independent domain information that sho uld be considered when producing patterns. How rules are a rticulated and how they can be combined with the process of spatiotemporal reasoning is a challenge.
9.Visualization of spatiotemporal trends and phenomena, sc alability of methods of data mining, representation of data st ructures and effective indexation of spatiotemporal datasets are also difficult problems.
10. Another challenge is the development of appropriate me thods for visualising spatiotemporal information and interact ion facilities to provide an insight into the underlying pheno mena described by knowledge.
This includes incorporating the outcomes of spatiotemporal data mining into a mechanism that interprets the outcomes f or more properly organised analysis into the reasons behind the outcomes.

An Architecture for Data Mining

Just outside of the warehouse, several data mining currently run, requiring extra steps to retrieve, import and an alyse the data. Figure 1 demonstrates the advanced analysis architecture of a large data warehouse.

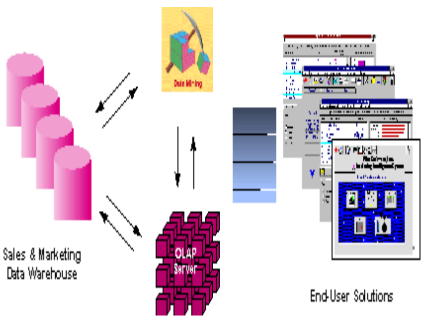


Fig.1 Architecture for Integrated Data Mining

Data warehouse containing a mixture of internal data monit oring all customer interactions, combined with external mar ket data on competitor operation, is the perfect starting point. When accessing the data warehouse, a n OLAP server makes to implement a more sophisticated end-user business model.
To incorporate ROI- focused market analysis directly into this infrastructure, the Data Mining Server must be inte grated with the data warehouse and the OLAP server.

Primary Data Mining Tasks

1. Classificationdiscovery of a function of predictive learnin g that categorizes a data item into one of many predefined gr oups.
2. Regression-discovery of a predictive learning function, which maps a data item to a prediction variable of real value.
3. Clustering is a common descriptive task in which a finite set of categories or clusters is defined in order to classify the data.
4. Summarization an additional descriptive task requiring methods for a set (or subset) of data to find a com pact description.
5. Dependency Modeling-finding a local model that defines significant dependencies in a data set or in a part of a data set between variables or between values of a function.
6. Detection of Shift and Deviation-discovering the most relevant changes in the data collection.

3. MINING SPATIOTEMPORAL TOPOLOGICAL RELATIONSHIP PATTERNS

If the geometry or location of either of the spatial object's changes, the topological relationship between the t wo spatial objects can change. In general, the geometry and position changes of spatial objective are collected and stored in spatiotemporal databases over time.
Using spatiotemporal topological relationship patterns the evolving topological relationship between spatia l objects and time is depicted. The topological relationship s hifts between two spatial objects O1 and O2 from time t1 to t4, for example, is shown in Fig., 1.
For this example, the topological relationship pattern can be expressed as D-O-C-T.
where D O, C, T, respectively, corresponds to disjoints, over laps, contains, and touches. It is possible to compute support for such patterns so that it can be used in decision making. I f these patterns occur more than the number of times stated, then they are known as periodic patterns.

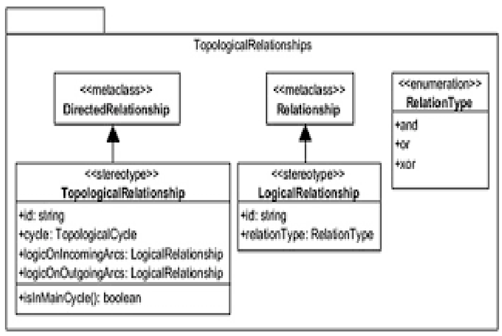


Fig 2: Topological relationship in an interval

4. TEMPORAL DATA MINING

The study of events ordered by one or more-time dimensions is concerned with temporal data mining.
In two broad directions, we distinguish. One concerns the exploration of causal associations between events that are temporally focused. The other concerns the discovery within the same time sequence or among different time sequences with similar patterns. This latter field focuses on the detection of similar pre-specified trends, generally called time series analysis (or trend analysis).

4.1 Mining Temporal Sequences

Temporal data mining's aim is to uncover secret connections between event sequences and subsequences.
Three phases are mainly involved in the exploration of relationships between sequences of events: the representation and modelling of the data sequence in an acceptable form; the description of measures of similarity between sequences; and the application of models and representations to the actual problems of mining.
A sequence consisting of a set of nominal symbols from a given alphabet is commonly referred to as a temporal sequence, and a sequence of continuous elements of real value is known as a time series.

4.1.1 Representation of Temporal Sequence

Time-Domain Continuous Representations

The initial elements, ordered by their moment of occurrence without any preprocessing, are a simple approach to representing a sequence of real-valued elements (time series).
An alternative involves finding a piecewise linear function capable of representing the entire initial series roughly.
The goal is to obtain a representation that can detect important changes in the sequence.

4.1.2 Transformation Based Representations

The key idea of Transformation Based Representations is to transform the initial sequences from time to another domain and then to represent each original sequence by using a point in this new domain.
To transform a series from the time domain to a point in the frequency domain, one proposal uses the Discrete Fourier Transform (DFT).
The Discrete Wavelet Transform (DWT) uses a more recent method to convert each series from the time domain into the time / frequency domain.
The DWT is a linear transformation that decomposes the original sequence into various components of frequency without losing knowledge about the moment of the occurrence of the elements.

4.2 Temporal Data Mining Tasks

In a broad number of applications, data mining has been used. It is possible to group temporary data mining tasks as follows:
Focused market analysis directly focused market analysis directly
(i) estimation, (ii) classification, (iii) clustering, (iv) search & retrieval and (v) discovery of patterns.
The role of predicting time series has to do with estimating (typically) future time series values based on their past focused market analysis directly samples. One needs to construct a predictive model for the data in order to do this.

5.SPATIAL DATA MINING

Objective

The key difference between data mining in relational DBS and spatial DBS is that the characteristics of the neighbours of an object of interest may affect the object and must therefore therefore be considered. Implicit spatial neighbourhood relations that are used by spatial data mining algorithms are defined by the explicit position and extension of spatial objects.
Database Primitives for Mining Spatial Data

A collection of primitive mining database primitives in spatial databases that are sufficient to express most spatial data mining algorithms and can be efficiently assisted by a DBMS.

Efficient Support of DBMS

"Active philtres allow the search to be restricted from a starting object to certain neighborhood paths “leading away.
To facilitate efficient retrieval of database primitives by a DBMS, neighbourhood indices materialise certain neighbourhood graphs.

Software’s for Spatial Data Mining

(1) CrimeSTAT: A Geographic Statistics Method for Crime Event Analysis and its Locations.
Description: Crimestat is commonly used by crime researchers and practitioners as a spatial statistics tool.
The software is Windows-based, and most desktop GIS programmers’ interface with it. The goal is to provide additional statistical resources to assist law enforcement officials and researchers in the field of criminal justice in their attempts to map crime.

(2) Toolbox for Spatial: Statistics for MATLAB/Fortran by K.
Matlab and Fortran toolbox for computing simultaneous and conditional spatial autoregressions and mixed regressive spatially autoregressive models developed by K. Pace: Free software Description: Speed of the Louisiana State University Dept. of Finance.

(3) MATLAB’s Spatial Econometrics Library: Free software
Description: It is a complete collection of expansion functions for spatial analysis, including autoregressive spatial modelling in particular.
(4) TeraSeer's ClusterSeer/BoundarySeeer/SpaceStat: Commercial applications
Description: Spatial clustering, spatial autocorrelation analysis, k-function, and classifications are provided by Tera Seer software.

6.SPATIO -TEMPORAL DATA MINING

The concept of spatial data mining, spatio-temporal data mining, here refers to the extraction of implicit information, spatial and temporal relationships, or other patterns which are not specifically stored in spatio-temporal databases.

Data Mining Strategies "Spatialization" And "Temporalizati on" Spatio-temporal data mining is the confluences of many fields, including spatio-temporal databases, machine learning, statistics, spatial visu alisation, and the theory of information. First, at different levels (scales), spatial and temporal relationships exist among spatial entities. In geographic databases, spatia l relationships, both metric (such as distance) and non-metric (such as topology, direct form etc and temporal relationships (such as before or after), can be explicit or implied.
Second, the intrinsic characteristics of spatio-temporal databases are spatial and temporal dependence and heterogeneity.
Thirdly, the influence of scale in space and time is a compli cated research problem in geographical analysis.

The method of Spatio-Temporal Data mining

The process of data mining typically consists of three stages or steps:

(1) pre-processing or preparation of information.
(2) modelling and validation; and
(3) post-processing or deployment.

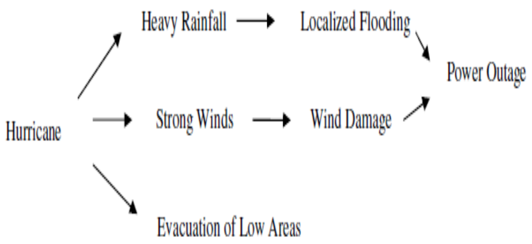
During the first step, according to some constraints imposed by some software, algorithms, or users, the data may need some cleaning and transformation. The secon d stage consists of selecting or creating a model that better r epresents the actions of the application. Finally, the third sta ge consists of using the model to study the application beha viour effectively, which was tested and validated in the seco nd process.

The intrinsic feature of spatio-temporal databases is genetics.

In a study of temporal information exploration, four broad categories of temporality within data are classified. Static (time has to be traced through external knowledge suc h as the creation of databases), Sequences (ordered sequence of events, showing relationships like before and after, or the richer relationships defined as meeting, overlapping, conte mporary), Timestamped (a static data timed sequence taken at more or less frequent intervals), Completely temporal (sp atio-temporal data incorporated, Via, for example, incidents, pro cesses).

Cascading Spatiotemporal Pattern discovery

The discovery of cascading spatiotemporal patterns from a Boolean data set of spatiotemporal event types uncov ers partially ordered subsets of event types whose instances are located and serially occur together. The forms of spatiot emporal events and their cases are distinct.
A cyclone is for example, an event-type. At various time, it can occur at several different places.
Each instance of an event is correlated with a specific place and time of occurrence. If event instances have disjointed oc currence times, the ordering can be total
Ordering is, otherwise, partial. Analysis of climate science d atasets to detect repeated occurrence of ice melting, extreme flooding with rainfall in some areas and drought in other ar eas after global warming are examples of CSTP.
Discovery of cyclone incidents, heavy rainfall, strong winds , localised flooding, wind damage, and post-hurricane warning power outages as shown in figure.



CSTPs occurring after a hurricane warning [37]

SPATIOTEMPORAL DATA MINING SYSTEM REQUIREMENTS AND APPLICATIONS

Spatiotemporal Database Structure

Let the S be the space or geographic region for which s patiotemporal information is gathered.
Suppose S contains regions r1,r2,...,rn and each internal regi on contains spatial objects o1,o2,...,ok at t1. As time passes, various potential adjustments are made,

1. It can change the location of the regions.
2. One area can be divided into two or more regions.
3. In one country two or more regions can merge.
4. It can shrink or expand the area.
5. The objects can travel to some other area in one region.
6. The objects' shape can shift.
7. The object's position within the area can shift.
8. About the combinations above.

At regular and/or irregular intervals, the spatiotemporal data base records the spatial objects and the changes happening t o them over a period.

System Requirements

1. The spatiotemporal data mining framework should provid e the user with a GUI-based environment to specify different inputs related to the specific task data, the type of spatiotemporal task or information to be discovered, interesting measures and threshold values appropriate to the task, and to specify the visualisation method of the knowled ge discovered.
2.To produce the information effectively, the system should force down user inputs as deeply as possible into the data mi ning process.
3. Interactive review of data mining outcomes should be sup ported by the framework.

4.Research and development of scalable, computer-efficient data mining techniques is a major challenge.

7. CONCLUSION

Due to the widespread use of sensor networks and location-aware devices as well as domain-specific features associated with such complex datasets, the rapid growth of spatiotemporal datasets involves research in to spatiotemporal data mining tasks. In different contexts, sp

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 11, November-2020

atiotemporal data mining presents many problems and also exciting applications. The research field remains largely unexplored. This paper explores the importance of the study and mining of spatiotemporal data in various fields, problems and challenges related to representation, processing, analysis, mining and visualisation. It also presents the nature of spatiotemporal data; how complex it is and the need for scalable and effective algorithms. These massive spatio-temporal data sets also conceal attractive information and useful knowledge, perhaps.

Many of the new services strive to provide consumers with more It is evident that it is difficult to manually analyse such data and that data mining could provide useful tools and technology in this context. Spatio-temporal data mining is an increasing field of research dedicated to the development of new algorithms and analytical methods for the efficient analysis of large spatio-temporal databases.

8. REFERENCES

[1] Nectaria Tryfona and Christian S. Jensen, Conceptual Data Modeling for Spatio-temporal Applications, GeoInformatica 3:3, 245-268 (1999)

[2] Tamas Abraham and John F. Roddick, Survey of Spaio-Temporal Databases, GeoInformatica 3:1, 61-99 (1999).

[3] Han, J. and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, San Francisco, 2000.

[4] Cristian S. Jensen, Introduction to Temporal Database Research, Temporal Database Management, 2000.

[5] T. Hadzilacos, Y. Manolopoulos, J.F. Roddick, and Y. Theodoridis, eds. Advances in Spatial and Temporal Databases. Proceedings of the Eighth International Symposium on Spatial and Temporal Databases, SSTD 2003, Santorini Island, Greece, volume 2750 of Lecture Notes in Computer Science. Springer, Berlin, 2003.

[6] Koperski K., " A Progressive Refinement Approach to Spatial Data Mining", PhD Thesis, the School of Computing Science, Simon Fraser University, April 1999.

[7] Koperski, K. and Han, J.: Discovery of Spatial Association Rules in Geographic Information Databases, In Advances in Spatial Databases (SSD'95), pp. 47-66, Portland, ME, August 1995

[8] Lebart L. et al., "Statistique exploratoire multidimensionnelle" , Editions Dunod, Paris, 439 p., 1997.

[9] Lebart, L. (1984) Correspondence analysis of graph structure. Bulletin technique du CESIA, Paris:2, 1-2, pp 5-19.

[10] Longley P. A., Goodchild M. F., Maguire D. J., Rhind D. W., Geographical Information Systems - Principles and Technical Issues, John Wiley & Sons, Inc., Second Edition, 1999.

[11] Cheng, T. and Wang, J., 2008. Integrated Spatio-temporal Data Mining for Forest Fire Prediction. Transactions in GIS, 12(5), pp.591-611.

[12] Cao, L., San, X., Zhao, Y. and Chen, G., 2013. The application of the spatio-temporal data mining algorithm in maize yield prediction. Mathematical and Computer Modelling, 58(3-4), pp.507-513.

[13] Kitamoto, A., 2002. Spatio-temporal data mining for typhoon image collection. Journal of Intelligent Information Systems, 19(1), pp.25-41.

[14] Gokaraju, B., Durbha, S.S., King, R.L. and Younan, N.H., 2011. A machine learning based spatio-temporal data mining approach for detection of harmful algal blooms in the Gulf of Mexico. IEEE Journal of selected topics in applied earth observations and remote sensing, 4(3), pp.710-720.

[15] Benali, H., Escofier, B.: Analyse factorielle lissée et analyse factorielle des différences locales, Revue Statistique Appliquée, 1990, XXXVIII (2), pp 55-76

[16] Cliff A.D., Ord J.K., 1973: "Spatial autocorrelation", Pion, London.

[17] Diggle P.J., 1993, Point process modeling in environmental epidemiology. In Barnett V., Turkman K. (eds) Statistics for the environment, Chichester, John Wiley & Sons, pp 89-110.