INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 4 (2017) PAGES 2254 - 2259
RECEIVED : 05.07.2017 PUBLISHED : 06.08.2017

Aug 06, 2017

COMPUTATIONAL TECHNIQUES AND CONVENTIONAL DATA SCIENCE TECHNIQUES FOR EXTRACTING SPATIOTEMPORAL PATTERNS

B.Geetha[1] , Dr.J.Dhillipan[2],D.B.Shanmugam[3]

[1]M.Phil, Research Scholar, Dr.M.G.R.Chockalingam Arts College, Arni.

[2] Asst.Prof.,(S.G) & Head, MCA Department, SRM University, Ramapuram Campus, Chennai

[3]Associate Professor, Department of MCA, Sri Balaji Chockalingam Engineering College , Arni

ABSTRACT

Explosive growth in geospatial and temporal data as well as the emergence of new technologies emphasize the need for automated discovery of spatiotemporal knowledge. Spatiotemporal data mining studies the process of discovering interesting and previously unknown, but potentially useful patterns from large spatiotemporal databases. It has broad application domains including ecology and environmental management, public safety, transportation, earth science, epidemiology, and climatology. The complexity of spatiotemporal data and intrinsic relationships limits the usefulness of conventional data science techniques for extracting spatiotemporal patterns. We review recent computational techniques and tools in spatiotemporal data mining, focusing on several major pattern families: spatiotemporal outlier, spatiotemporal coupling and tele-coupling, spatiotemporal prediction, spatiotemporal partitioning and summarization, spatiotemporal hotspots, and change detection.

Keywords: Spatiotemporal data, Earth science, Climatology.

1. INTRODUCTION

The Data mining methods that are combined with Geographic Information Systems (GIS) for carrying out spatial analysis of geographic data. Spatial data mining research by developing new techniques for point pattern analysis, prediction in space–time data, and analysis of moving object data, as well as by demonstrating applications of genetic algorithms for optimization in the context of image classification and spatial interpolation. To address these challenges, spatial data mining and geographic knowledge discovery has emerged as an active research field, focusing on the development of theory, methodology, and practice for the extraction of useful information and knowledge from massive and complex spatial databases There is an urgent need for effective and efficient methods to extract unknown and unexpected information from spatial data sets of unprecedentedly large size, high dimensionality, and complexity. The availability of vast and high-resolution spatial and spatiotemporal data provides opportunities for gaining new knowledge and better understanding of complex geographic phenomena, such as human–environment interaction and social– economic dynamics, and address urgent real-world problems, such as global climate change and pandemic flu spread.

However, traditional spatial analysis methods were developed in an era when data were relatively scarce and computational power was not as powerful as it is today. Facing the massive data that are increasingly available and the complex analysis questions that they may potentially answer, traditional analysis methods often have one or more of the following three limitations. First, most existing methods focus on a limited perspective (such as univariate spatial autocorrelation) or a specific type of relation model (e.g., linear regression). If the chosen perspective or assumed model is inappropriate for the phenomenon being analyzed, the analysis can at best indicate that the data do not show interesting relationships, but cannot suggest any better alternatives. Second, many traditional methods cannot process very large data volume. Third, newly emerged data types (such as trajectories of moving objects, geographic information embedded in web pages, and surveillance videos) and new application needs demand new approaches to analyze such data and discover embedded patterns and information.

2. RELATED WORK

The spatial data mining can be used to understand spatial data, discover the relation between space and the non space data, set up the spatial knowledge base, excel the query, reorganize spatial database and obtain concise total characteristic etc.. The system structure of the spatial data mining can be divided into three layer structures mostly such as the Figure 1 show The customer interface layer is mainly used for input and output, the miner layer is mainly used to manage data, select algorithm and storage the mined knowledge, the data source layer, which mainly includes the spatial database and other related data and knowledge bases, is original data of the spatial data mining. Global autocorrelation: The most common way of summarizing a dataset is to apply elementary statistics, such as the calculation of average, variance, etc., and graphic tools like histograms and pie charts.

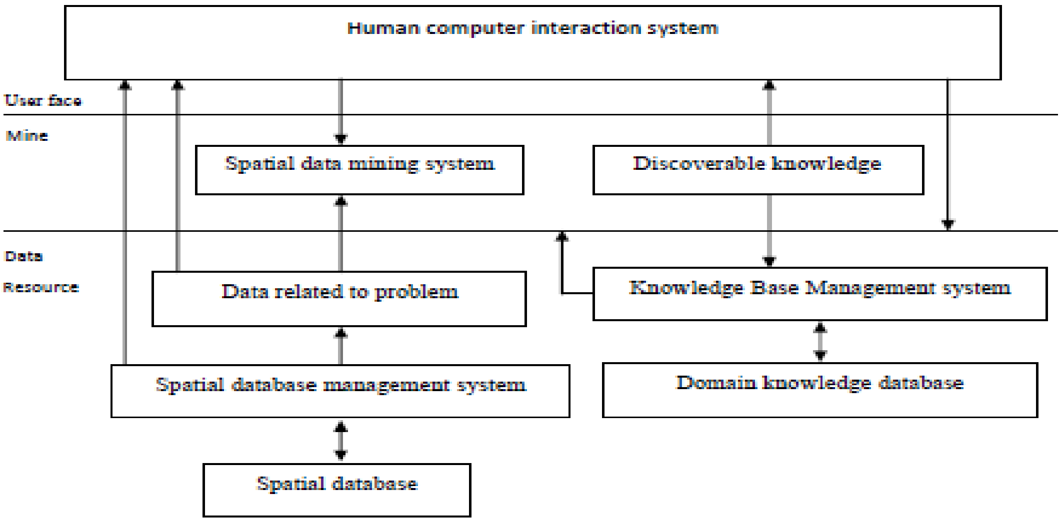

Fig.1. The systematic structure of spatial data mining

New methods have been developed for measuring neighborhood dependency at a global level, such as local variance and local covariance, spatial auto-correlation by Geary, and Moran indices. These methods are based on the notion of a contiguity matrix that represents the spatial relationships between objects. It should be noted that this contiguity can correspond to different spatial relationships, such as adjacency, a distance gap, and so on. The first technique performs a smoothing by replacing each attribute value by the average value of its neighbors. This highlights the general characteristics of the data. The other contrasts data by subtracting this average from each value. Each attribute (called variable) in statistics can then be analyzed using conventional methods. However, when multiple attributes (above tree) have to be analyzed together, multidimensional data analysis methods (i.e. factorial analysis) become necessary. Their principle is to reduce the number of variables by looking for the factorial axes where there is maximum spreading of data values. By projecting and visualizing the initial dataset on those axes, the correlation or dependencies between properties can be deduced. In statistics and especially in the above methods, the analyzed objects were originally considered to be independent. The extension of factorial analysis methods to contiguous objects entails applying common Principal Component Analysis or Correspondence Analysis methods once the original table is transformed using smoothing or contrasting techniques.

3. EVALUATION

A noteworthy trend is the increasing size of data sets in common use, such as records of business transactions, environmental data and census demographics. These data sets often contain millions of records, or even far more. This situation creates new challenges in coping with scale. The challenges and impacts can be classified into three main areas, namely, geographic information in knowledge discovery, geographic knowledge discovery in geographic information science and geographic knowledge discovery in geographic research. The huge volume of spatial data . So retrieval and storage is difficult. The spatial data types/structures are complex .Expensive spatial processing operations. The mass data storaged in spatial database includes spatial topological, nospatial properties and objects appearing variety on the time .

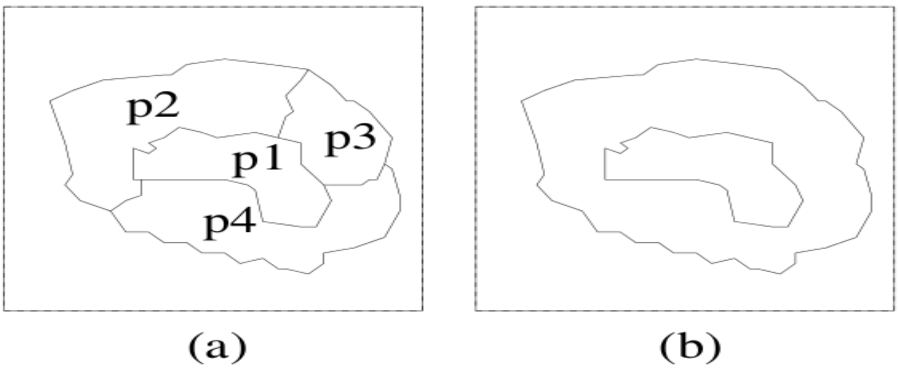

Fig.2. Shadow Ring

The main knowledge types that can be discovered in the spatial database are: general geometric knowledge, spatial distribution rules, spatial association rules, spatial clustering rules, spatial characteristic rules, spatial discriminate rules, spatial evolution rules etc.. For land use dynamic monitoring, according to the knowledge mined in the spatial database, there are following several applications. Spatial data mining technique makes use of general geometric knowledge, spatial distribution rules, spatial association rules, spatial evolution rules to get many factors about terrain, prevent or control flood, preventive pollution during the city planning for providing good data environment in city construction. For data mining of large data sets to be effective, it is also important to include humans in the data exploration process and combine their flexibility, creativity, and general knowledge with the enormous storage capacity and computational power of today's computers. Visual data mining applies human visual perception to the exploration of large data sets. Presenting data in an interactive, graphical form often fosters new insights, encouraging the formation and validation of new hypotheses to the end of better problem-solving and gaining deeper domain knowledge. Visual data mining often follows a three step process: Overview first, zoom and filter, and then details-on-demand.

4. ANALYSIS

Visual analytics is the science of analytical reasoning supported by interactive visual interfaces. A more specific definition would be: 'Visual analytics combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex datasets' (Thomas, J., Cook, K., 2005). It is an integral approach combining visualization, human factors, and data analysis, which allow users to combine their knowledge with the automation data processing and analysis which done by computer to explore and gain more information from the data.

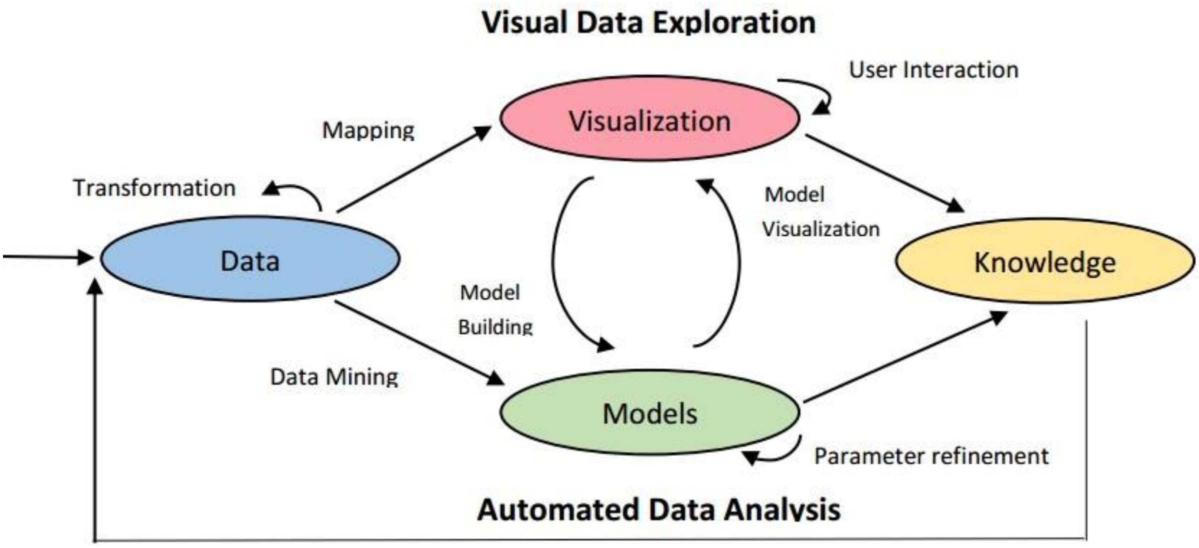

Fig.3. Analysis Process

The visual analytics process combines automatic and visual analysis methods with a tight coupling through human interaction in order to gain knowledge from data. Different data sources need to be integrated first in the preprocessing steps (e.g. data cleaning, normalization, etc) before the analytics step could be executed. Then the analyst could apply automatic analysis methods using data mining techniques to generate models of the original data which then could be combined with visualization methods where the analyst interacts with visualization of the data model to define which data model could generate a better result based on certain parameters. Maps for presenting an overview of the spatial distribution of moving objects within a selected time interval are commonly used. The reason is that the moving objects, which are often shown in their raw form like points, are very numerous and consequently their positions have to be presented in an aggregated way.

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 4 (2017) PAGES 2254 - 2259
RECEIVED : 05.07.2017 PUBLISHED : 06.08.2017 Aug 06, 2017

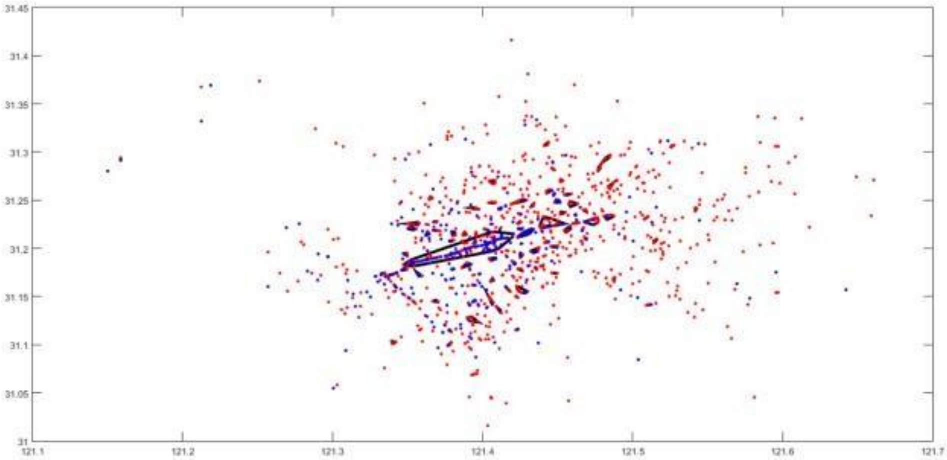

Fig.4. Output Formation

The pie chart on the bottom of the graph is located near Shanghai South Railway Station. And the pie chart close to Hongqiao airport is located on one big overpass.

CONCLUSION

This thesis provides an over view of current research in the field of spatiotemporal data mining from a computational perspective. Spatiotemporal data mining has broad application domains including ecology and environmental management, public safety, transportation, earth science, epidemiology, and climatology. However, the complexity of spatiotemporal data and intrinsic relationships limits the usefulness of conventional data science techniques for extracting spatiotemporal patterns. We provide taxonomy of different spatiotemporal data types and underlying spatiotemporal statistics. We also review common spatiotemporal data mining techniques organized by major output pattern families: spatiotemporal outlier, spatiotemporal coupling and tele-coupling, spatiotemporal prediction, spatiotemporal partitioning and summarization, spatiotemporal hotspots, and change detection.

REFERENCES

[1] Agarwal, P., & Skupin, A. (2008). Self-organising maps: Applications in geographic information science. Chichester: Wiley.

[2] Anselin, L., Syabri, I., & Kho, Y. (2006). GeoDa: An introduction to spatial data analysis. Geographical Analysis, 38(1), 5–22.

[3] Appice, A., Ceci, M., Lanza, A., Lisi, F. A., & Malerba, D. (2003). Discovery of spatial association rules in geo-referenced census data: A relational mining approach. Intelligent Data Analysis, 7(6), 541–566.

[4] Brimicombe, A. J. (2007). A dual approach to cluster discovery in point event data sets. Computers Environment and Urban Systems, 31(1), 4–18.

[5] Cheng, T., & Wang, J. (2009). Accommodating spatial associations in DRNN for space–time analysis. Computers, Environment and Urban Systems, 33(6), 409–418.

[6] Chernoff, H., & Rizvi, M. H. (1975). Effect on classification error of random permutations of features in representing multivariate data by faces. Journal of American Statistical Association, 70, 548–554.