International Journal of Enhanced Research in Management & Computer Applications
ISSN: 2319-7471, Vol. 6 Issue 12, December-2017, Impact Factor: 3.578

Exploratory Visualization Technique in Spatio –Temporal Data Mining

Srikanth Lakumarapu[1], Dr. Rashmi Agarwal[2]

[1]Research Scholar, Department of Computer Science & Engineering, Madhav University
[2]Associate Professor, Department of Computer Science & Engineering, Madhav University

ABSTRACT

Spatio-temporal data sets are often very large and difficult to analyze and display. Since they are fundamental for decision support in many application contexts, recently a lot of interest has arisen toward data-mining techniques to filter out relevant subsets of very large data repositories as well as visualization tools to effectively display the results. Spatiotemporal data mining studies the process of discovering interesting and previously unknown, but potentially useful patterns from large spatiotemporal databases. In this paper we propose a data mining system to deal with very large spatio -temporal data sets. Within this system ,new techniques have been developed to efficiently support the data-mining process ,address the spatial and temporal dimensions of the data set, and visualize ,interpret results .In particular two complementary 3D visualization environments have been implemented .One exploits "Google Earth" to display the mining outcomes combined with map and other geographical layers, while the other is a Java 3D-based tool for providing advanced interactions with the data set in a non-geo-referenced space, such as displaying association rules and variable distributions.

Keywords: Data mining, Spatio-temporal data, Exploratory visualization, 3D visualization, spatiotemporal patterns.

I INTRODUCTION

During the last decade, our ability to collect and store data has far outpaced our ability to process, analyze and exploit it. Many organizations have begun to routinely capture huge volumes of historical data describing their operations, products and customers. At the same time scientists and engineers in many fields have been capturing increasingly complex experimental data sets, such as terabytes of data received daily from space-borne instruments, high spatial, temporal and spectral-resolution remote sensing systems, and other environmental monitoring device .Some researchers estimate that about 80% of the data stored in corporate databases integrate spatial information , leading to huge amounts of geo-referenced information that need to be analyzed and processed. These data sets are often critical for decision support, but their value depends on the ability to extract useful information for studying and understanding the phenomena governing the data source. Therefore, the need for efficient and effective techniques for mining and analyzing spatio-temporal data sets has recently emerged as a research priority.

Data mining techniques have been proven to be of significant value for spatio-temporal applications. It is a user-centric, interactive process where data mining experts and domain experts work closely together to gain insight on a given problem. In particular spatio-temporal data mining is an emerging research area, encompassing a set of exploratory,

Computational and interactive approaches for analyzing very large spatial and spatio-temporal data sets. Several open issues have been identified ranging from the definition of mining techniques capable of dealing with spatial-temporal information to the development of effective methods for interpreting and presenting the final results.

Visualization techniques are widely recognized to be powerful in this domain, since they take advantage of human abilities to perceive visual patterns and to interpret them. To address these issues we have developed a system for exploratory spatio-temporal data mining. The aim of this system is on one hand to enable data- mining tools to provide some form of localization in the data being analyzed, and, on the other hand to interactively visualize in 3D the outcome of the mining process, thus leading to greater effectiveness and significance of the results. To achieve these goals the

system includes a data mining engine that can integrate different data mining algorithms and two complementary 3D visualization tools.

Societal importance: Spatiotemporal data mining techniques are crucial to organizations which make decisions based on large spatial and spatiotemporal datasets, including NASA, the National Geospatial-Intelligence Agency [6], the National Cancer Institute [7], the US Department of Transportation [8], and the National Institute of Justice [9]. These organizations are spread across many application domains. In ecology and environmental management [10– 13], researchers need tools to classify remote sensing images to map forest coverage. In public safety [14], crime analysts are interested in discovering hotspot patterns from crime event maps so as to effectively allocate police resources. In transportation [15], researchers analyze historical taxi GPS trajectories to recommend fast routes from places to places. Epidemiologists [16] use spatiotemporal data mining techniques to detect disease outbreak. There are also other application domains such as earth science [17], climatology [18], precision agriculture [19], and Internet of Things [20]

II.SYSTEM ARCHITECTURE

This section describes the architecture of the system. we first discuss the main concepts of the data-mining process, and then introduce the main components of the system ,namely the mining engine and the visualization tools.

The spatio-temporal data-mining process

The data mining process usually consists of three phases or steps

1) Pre-processing or data preparation. 2) Modeling and validation 3) Post processing or deployment

During the first phase the data may need some cleaning and transformation according to some constraints imposed by some tools, algorithms, or users .The second phase consists of choosing or building a model that better reflects the application behavior.

The third step of using this model evaluated and validated in the second phase, to effectively study the application behavior. The mining process for spatial data is more complex than for relational data in terms of both the mining efficiently and the complexity of possible patterns that can be extracted from spatial data sets. Therefore, new techniques are required to efficiently and effectively mine spatial data sets. Especially in spatial data mining the third phase is so important that some researchers incorporated most of its processes into phase two such as automatic and interactive visualization of data, and called it interactive data mining (IDM).

Exploratory spatio-temporal data-mining system

The proposed system for mining large spatio-temporal data sets describes the behavior of some natural phenomena, which have been monitored and recorded at several time instants. Our system relies on a standard three-tier architecture, including a data store at the back end, an application server, and two visualization components at the front end.

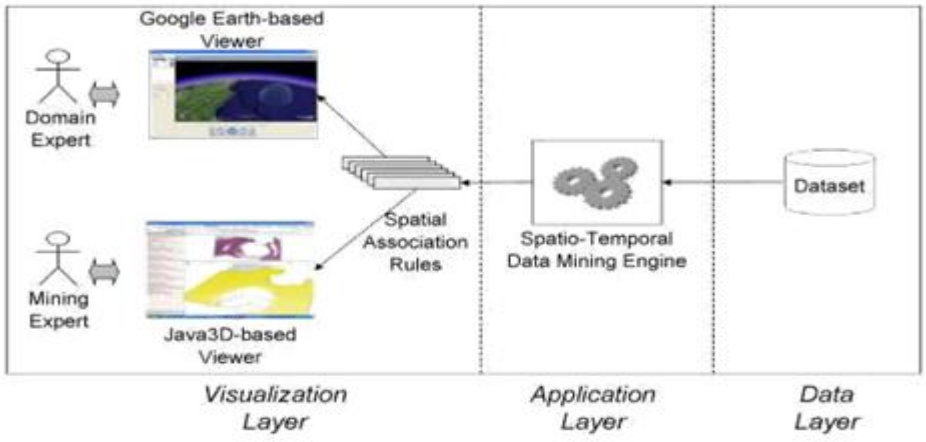

Fig. 1. System architecture.

Since several different application domains can be considered the application server must include domain-specific wrappers that transform raw data into the input format required by the mining engine. These wrappers implement the data set type models described.

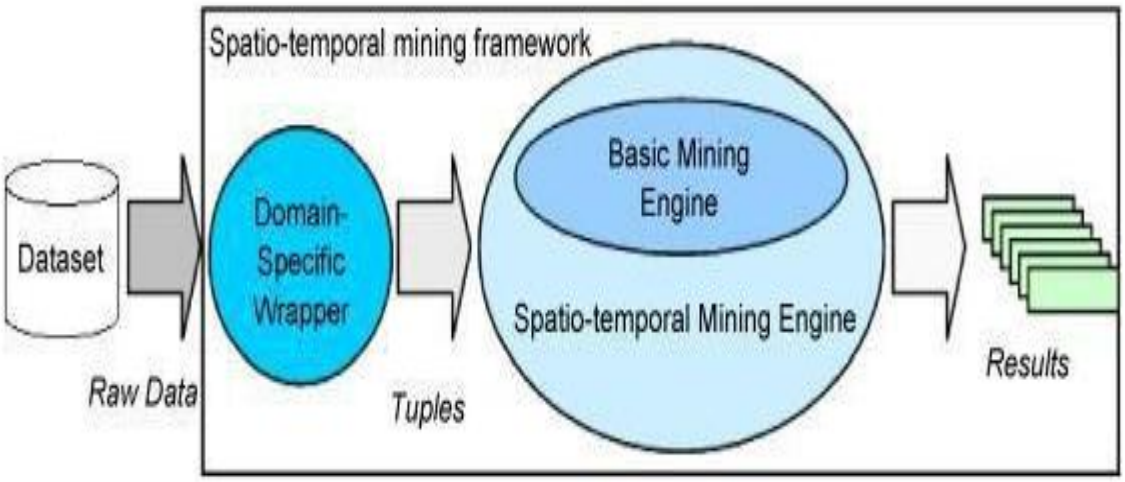

Fig. 2: Architecture of the spatio-temporal data-mining engine.

III.THE DATA-MINING PROCESS

In this section we describe our data- mining approach that deals with spatio-temporal data sets. We start with a review of the current state-of the art in this field, and then we present our solution.

Related work on spatial data mining

Numerous research projects on spatial data mining have been conducted in the last two decades. Some attentions have been dedicated to the application of existing as well as the development of new mechanisms to extract relevant information from large data repositories. However, due to the huge volume and diverse nature of this kind of data, traditional techniques such as statistical methods have high computational burdens and seem often inadequate to elicit complex spatial and temporal relationships among data. Association rules have also been used successfully on special data sets. The main idea is to design spatial association rules that not only can find local correlations between patterns ,but also global ones .Spatial association. rules constitute an improvement to generalization-based spatial data-mining methods, as they cannot discover rules reflecting spatial pattern structures.

Most efforts have been spent in trying to adapt, modify or improve conventional techniques, relying on a solid knowledge discovery in database (KDD) experience to design new suitable mining models. Some good attempts have been made in proposing a better-more meaningful-data format to highlight spatio -temporal relationships prior to elaboration; this can be achieved either by preprocessing the database or by imposing a level of meta-data to properly access information within the whole data set.

The proposed approach

In this paper we propose a new approach for spatio-temporal data mining, whose conceptual schema is depicted in Fig.3. The approach consists of two main components; localizer and miner. The localizer deals with the data attributes and especially with spatial and temporal dimensions. The miner process the data based on the spatio-temporal relationships provided by the localizer.

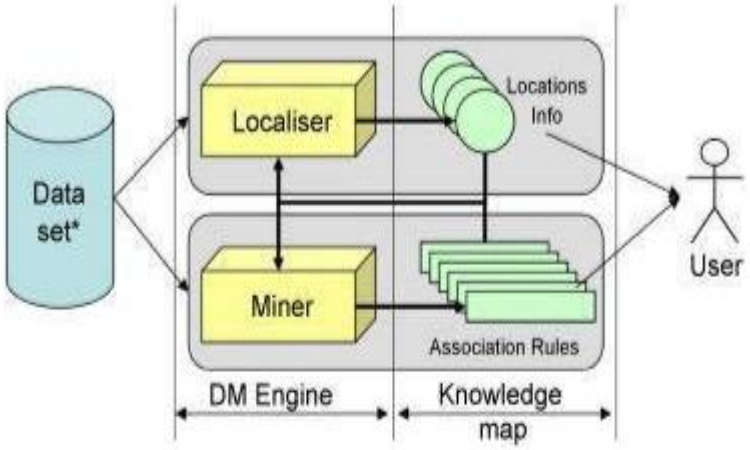

Fig.3: A schematic view of the proposed approach for spatial data mining. In the following we will focus on the techniques used in each phase.

Spatio-temporal data set model

It was already mentioned above that spatial data sets more complex than conventional data.

This complexity is not only in processing and interpreting the data but is also present at the data- mining process inputs. Spatial data is usually characterized by two different types of attributes: spatial and non-spatial attributes. The former identifies the spatial locations of spatial items. These include 3D space coordinates, item shape, temporal, geometry, etc.

The latter is usually the same as in conventional data sets such as item name , item key ,type ,rate ,size ,etc. .The main difference between these two types of attributes is that the relationships between spatial patterns/items are often implicit ,while they are usually explicit in non-spatial objects. Some methods of how to optimize the categories have been developed and, mainly based on heuristic methods and clustering analysis.

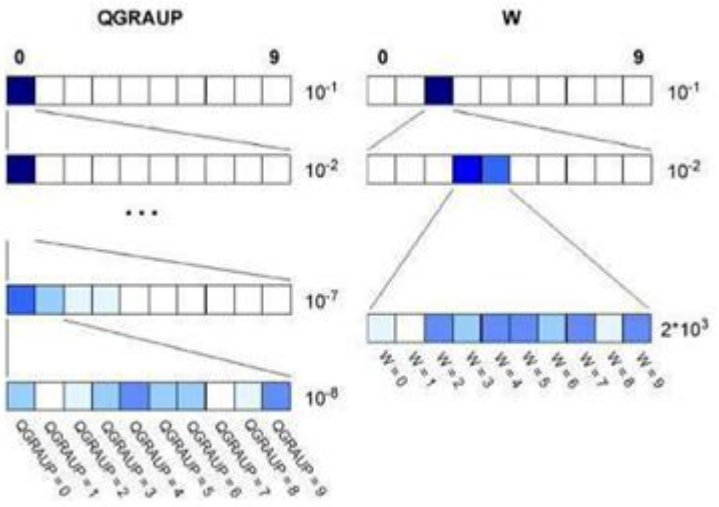

Fig.4.Example of categorization for the two variables W and QGRAUP. Dark color=many points are in that range of values; light color=few points.

The model used here consists of mapping the spatial data sets onto a virtual partitioned space. This can be seen as a layer in which original data are aggregated into virtual points representing the minimal spatial unit that can be occupied

by a spatio-temporal entity. Each virtual point is defined by a set of attributes including coordinates, size, neighborhood, etc. For instance, traditional geographical databases have two or three dimensions, while in spatio-temporal data sets the number of dimensions can range from two to N.

Spatial association rules

In the proposed system we focus our attention in developing a technique based on association rules to discover relationships between spatial patterns. A spatial association rule is of the form "A B(s %,c%)"Where the pattern A is called antecedent and B consequent, and the percentages s and c are called the support and the confidence of the rule. The problem of discovering association rules consists of identifying all rules, within the data set, satisfying minimum support s and confidence c. This usually requires a solution to the following two sub-problems :1) find frequent(large)spatial patterns;2)extract strong spatial association rules. In the first problem the rules should satisfy a minimum support (support>s) and in the second a spatial association is said to be strong if it satisfies a minimum confidence (confidence>c).

IV. VISUAL TECHNIQUES FOR ADVANCED SPATIAL ANALYSIS

Visual data mining refers to methods, approaches and tools for the exploration of large data sets by allowing users to directly interact with visual representations of data and dynamically modify parameters to see how they affect the visualized data.

The Google Earth-Based Tool

The first tool has been meant for domain experts, i.e. users that study the specific phenomenon but are not (necessarily) experts in data mining.

Google Earth (shortly GE) is a virtual globe, currently freely available for personal use on PC running on Windows and Mac OS, while the Linux version is expected shortly. For commercial and professional use, many purchasing options are available, ranging from basic licenses to enterprise services. The original project was developed by *Keyhole*, which was bought by Google in 2004.

Google Earth combines satellite raster imagery, with vector maps and layers, in a single and integrated tool, which allows users to interactively fly in 3D from outer space to street level views. Most places of the world are available at (at least) 1 km of resolution, while many large cities are available at high enough resolution to see individual buildings, houses, and even cars. A very wide set of geographical features (streets, borders, rivers airports, etc.), as well as commercial points of interest (restaurants, bars, lodging, shopping malls, fuel stations, etc.), can be overlaid onto the map. The application uses data from NASA databases to render 3D terrain models, thus providing also Digital Elevation Model features.

This application turns out to be very flexible, being able to deal with a large variety of spatio-temporal phenomena, ranging from worldwide (e.g. weather, pollution, epidemic diffusions, etc.) to local ones (e.g. local health, traffic, economics, etc.). The tool we developed embeds GE and presents the same ease of use, resulting very suitable for domain-expert user.

***Dimensional panel*:** This panel allows the user to move in four dimensions, namely the 3D permitted by GE (by exploiting six DoF), and the time dimension, through a sliding bar. To this aim, the horizontal panel, located at the bottom of the window, realizes a unique control panel/set of commands to follow the data painted on screen.

Technical Aspects

The development of an environment exploiting Google earth technologies to render mining outcomes posed two main challenges:

How to arrange the information of the data set and/or the rules in a way that it could be displayed by Google Earth.

How to improve the Google Earth user interface to provide the tools to carry out the exploratory spatio-temporal data mining and visualization.

The first issue we exploited the ad hoc language provided by GE, named keyhole markup language (or KML), which is an XML grammar and file format suited to model one or more spatial features to be displayed in GE. In relation to the second issue, there are two main ways to programmatically interact with GE. This approach is straightforward, but does not provide an effective management of the user interactions. The alternative solution is to use the set of API provided by GE.

CONCLUSION

In this paper we have described the system for exploratory spatio-temporal data mining we have developed. This system includes a mining engine based on an adapted version of the well-known A-priori algorithms. Since results of a mining algorithm require interpretation, we have focused on visual techniques. To this aim we have developed two independent visualization tools for viewing and interacting with the results of the mining process, meant, respectively, for the domain experts and data mining experts. The Google Earth application allows to relate the phenomenon being studied to the specific geographic area and associated features. The second visualization tool presents more sophisticated interactively. Our system has been tested on a large real-world data set and has produced interesting results. However we plan to perform more extensive testing with domain experts. The system offers much scope for enhancements and further developments. We also intend to integrate the two visualization tools allowing to switch in a continuous fashion between them maintaining the same perspective.

REFERENCES

[1] R.Agarawal, T.Lmielinski, A.Swami, Mining association rules between sets of items in large databases, in: ACM SIGMOD Conference,1983.
[2] U.M.Fayyed,G.G.Grinstein,Introduction in Information Visualization in Data Mining and Knowledge Discovery ,Morgan Kaufmann,Los Altos,CA,2001,PP.1-17.
[3] Y.Bedard,T.Merrett,J.Han,fundaments of spatial data warehousing for geographic knowledge discovery,Geographic Data Mining and Knowledge discovery,Taylor&Francis,London,2001,PP.53-73.
[4] M..F.Constabile,D.Malerba(Eds),Special issue on visual data mining Journal of Visual languages and Computing14(2003)
[5] W.L.Johnston ,Model Visualization,in Information Visualization in data Mining and knowledge discovery,Morgan Kaufmann,los Altos,CA,2001,PP.223-227.
[6] Stolorz, P.; Nakamura, H.; Mesrobian, E.; Muntz, R.; Shek, E.; Santos, J.; Yi, J.; Ng, K.; Chien, S.; Mechoso, R.; et al. Fast Spatio-Temporal Data Mining of Large Geophysical Datasets; AAAI Press: Palo Alto, CA, USA, 1995.
[7] Guting, R. An introduction to spatial database systems. VLDB J. 1994, 3, 357–399.
[8] Shekhar, S.; Chawla, S. Spatial Databases: A Tour; Prentice Hall: Upper Saddle River, NJ, USA, 2003.
[9] Shekhar, S.; Chawla, S.; Ravada, S.; Fetterer, A.; Liu, X.; Lu, C.T. Spatial databases— Accomplishments and research needs. Trans. Knowl. Data Eng. 1999, 11, 45–55.
[10] Worboys, M. GIS: A Computing Perspective; Taylor and Francis: London, UK, 1995.
[11] Hot Spots. Available online: http://www.ncjrs.gov/pdffiles1/nij/209393.pdf (accessed on 10 May 2015) .
[12] ssaks, E.H.; Svivastava, RM. Applied Geostatistics; Oxford University Press: Oxford, UK, 1989.
[13] Haining, R.J. Spatial Data Analysis in the Social and Environmental Sciences; Cambridge University Press: Cambridge, UK, 1989.
[14] Roddick, J.F.; Spiliopoulou, M. A bibliography of temporal, spatial and spatio-temporal data mining research. SIGKDD Explor. 1999, 1, 34–38 .
[15] Scally, R. GIS for Environmental Management; ESRI Press: Redlands, CA, USA, 2006.
[16] Leipnik, M.R.; Albert, D.P. GIS in Law Enforcement: Implementation Issues and Case Studies;
[17] CRC Press: Sacramento, CA, USA, 2002.
[18] Lang, L. Transportation GIS; ESRI Press: Redlands, CA, USA, 1999.
[19] Elliott, P.; Wakefield, J.; Best, N.; Briggs, D. Spatial Epidemiology: Methods and Applications;
[20] Oxford University Press: Oxford, UK, 2000.
[21] Hohn, M.; A.E. Liebhold, L.G. A Geostatistical model for forecasting the spatial dynamics of defoliation caused by the Gypsy Moth, Lymantria dispar (Lepidoptera:Lymantriidae). Environ. Entomol. 1993, 22, 1066–1075.
[22] Yasui, Y.; Lele, S. A regression method for spatial disease rates: An estimating function approach. J. Am. Stat. Assoc. 1997, 94, 21–32.
[23] Ruß, G.; Brenning, A. Data mining in precision agriculture: Management of spatial information.
[24] In Computational Intelligence for Knowledge-Based Systems Design; Springer: Berlin, Germany, 2010; pp. 350–359.
[25] Gubbi, J.; Buyya, R.; Marusic, S.; Palaniswami, M. Internet of Things (IoT): A vision, architectural elements, and future directions. Future Gener. Comput. Syst. 2013, 29, 1645–1660.
[26] Marcus, G.; Davis, E. Eight (no, nine!) problems with big data. N. Y. Times 2014, 6, 2014.
[27] Caldwell, P.M.; Bretherton, C.S.; Zelinka, M.D.; Klein, S.A.; Santer, B.D.; Sanderson, B.M. Statistical significance of climate sensitivity predictors obtained by data mining. Geophys. Res. Lett. 2014, 41, 1803–1808.
[28] Shekhar, S.; Zhang, P.; Huang, Y.; Vatsavai, R.R. Trends in spatial data mining. In Data Mining: Next Generation Challenges and Future Directions; AAAI Press: Palo Alto, CA, USA, 2003; pp. 357–380.
[29] Worboys, M.; Duckham, M. GIS: A Computing Perspective, 2nd ed.; CRC Press: Sacramento, CA, USA, 2004.

[30] Li, Z.; Chen, J.; Baltsavias, E. Advances in Photogrammetry, Remote Sensing and Spatial Information Sciences: 2008 ISPRS Congress Book; CRC Press: Sacramento, CA, USA, 2008.
[31] Yuan, M. Temporal GIS and spatio-temporal modeling. In Proceedings of the Third International Conference Workshop on Integrating GIS and Environment Modeling, Santa Fe, NM, USA, 21–26 January 1996.
[32] Allen, J.F. Towards a general theory of action and time. Artif. Intell. 1984, 23, 123–154.
[33] George, B.; Kim, S.; Shekhar, S. Spatio-temporal Network Databases and Routing Algorithms: A Summary of Results. In Proceedings of the 10th International Symposium on Spatial and Temporal Databases (SSTD'07), Boston, MA, USA, 16–18 July 2007.
[34] George, B.; Shekhar, S. Time Aggregated Graphs: A model for spatio-temporal network. In Proceedings of the Workshops (CoMoGIS) at the 25th International Conference on Conceptual Modeling (ER2006), Tucson, AZ, USA, 6–9 November 2006.
[35] Gelfand, A.E.; Diggle, P.; Guttorp, P.; Fuentes, M. Handbook of Spatial Statistics; CRC Press: Sacramento, CA, USA,.