ISSN:0975-9646

M. Kireet / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 4 (4) , 2013, 546 - 548

A New Approach to Detect Ship Collisions using Spatial & Spatio-Temporal Data mining methods

M.KIREET
Department of Computer Science and Engineering,
Visvesvaraya college of Engineering & Technology
&
Research Scholar in JNTUH

***Abstract*— This paper provides an approach or idea to detect the ship collisions with Icebergs by applying the data mining methods that are combined with Geographic Information Systems (GIS) for carrying out spatial analysis of geographic data. First the basic methods and the tasks related to the spatial data mining and spatio-temporal data mining were explained. Later the methods like clustering, density based clustering algorithms, hotspot analysis and spatio-temporal datamining were taken into consideration and were explained so that it gives the idea that how these methods can be applied for the detection of icebergs in different sea-routes. It concludes with the application of spatio-temporal data mining techniques along with geostatistical approach were explained**

***Keywords*—clustering, spatio-temporal data mining, geostatistical apporoach.**

INTRODUCTION

Icebergs pose a danger to ships traversing in various places in the World like North Atlantic and the waters around Antarctica. The International Ice Patrol uses airplanes and radars to track icebergs that float into major shipping lanes. The U.S. National Ice Center uses satellite data to monitor icebergs near Antarctica. However, it only tracks icebergs larger than 500 square meters (5,400 square feet).This recent technologies like Spatial data mining and their application allows to analyse the data based on the previous Iceberg collisions and provide some advance enhancements in the detection of Icebergs that of which allows to reduce the ship collisions with Icebergs.

Spatial data mining deals with the process concerned with the identification of interesting spatial patterns from data stored in spatial databases. A spatial database is a database that is optimized to store and query data that is related to objects in space, including points, lines and polygons. Spatio-temporal databases are an extension of spatial databases. A spatiotemporal database is a database that manages both space and time information i.e A spatio-temporal database embodies spatial, temporal, and spatio-temporal database concepts, and captures spatial and temporal aspects of data and deals with (i) geometry changing over time and/or (ii) location of objects moving over invariant geometry.

In this paper for the analysis of data Spatial data mining tasks have been taken into consideration , Spatial data mining explains in detail four main output patterns and methods of SDM related to anomalies, clustering, co-location, and prediction. Important tasks in Spatial data mining like spatial outlier detection, co-location pattern discovery, spatial classification and regression modeling, spatial clustering, and spatial hotspot analysis were used.

Now-a-days General purpose data mining tools, such as Enterprise Miner, are used to analyse commercial databases. These tools are used in analysing scientific and engineering data, astronomical data, multi-media data and web data. Extracting interesting and useful patterns from spatial data sets is more difficult than extracting corresponding patterns or data from general purpose databases and categorical data due to the complexity of spatial data types, spatial relationships, and spatial autocorrelation.

Generally Data mining methods are not appropriate to extract spatial data because they do not support location data nor the absolute relationships between objects. Hence, it is necessary to develop new methods which include spatial relationships and spatial data handling. It is time consuming process for calculating these spatial relationships, and also large volume of data is generated by encoding geometric location.

A. Spatial Data mining

Generally, data mining is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both . Spatial data mining is the process concerned with the identification of interesting spatial patterns from data stored in spatial databases. A spatial database is a database that is optimized to store and query data that is related to objects in space, including points, lines and polygons.

B.Spatio Temporal Data Mining:

Taking Spatio-temporal databases into consideration which are an extension of spatial databases. A spatiotemporal database is a database that can organize both space and time information i.e A spatio-temporal database embodies spatial, temporal, and spatio-temporal database concepts, and captures spatial and temporal aspects of data and deals with (i) geometry changing over time and/or (ii) location of objects moving over invariant geometry.

Spatial data mining and spatio temporal data mining plays a important role in

a. Identification of interesting spatial patterns and features.

b. Identifying the relationships between spatial and non spatial data.

c.To reorganize spatial databases as well as to achieve better performance.

C.Using Spatial Clustering for Iceberg detection

Cluster analysis is used in many spatial and spatiotemporal application domains such as remote sensing data analysis

determines the number and distribution of spectral classes, in epidemiology for finding unusual groups of health-related events, and in detection of crime hot spots by police officers. Spatial clustering is a process of grouping a set of spatial objects into clusters so that objects within a cluster have high similarity in comparison to one another, but are dissimilar to objects in other clusters. for finding unusual groups of health-related events, and in detection of crime hot spots by police officers and also can be used to detect the iceberg locations.

The clustering algorithms can be classified into different groups as follows:

(i)Hierarchical Clustering :Hierarchical clustering methods forms all the patterns as a one single cluster and performs splitting or merging until as per some criteria is met. This results in tree of clusters. This tree of clusters can be cut at different levels to get the desired clusters. Taking different routes that ships travels from one source to destination or from one place to another and by using the database related to ship collisions in the past years a hierarchical method is applied. This hierarchical cluster methods can cut the single route to different levels forming a cluster for different locations in one particular route .Comparing the structures of different locations in single route the detection of Icebergs can easily be done.If there are three sea routes from one particular place to another particular place then clustering can be done by splitting each route into different levels. Atlast the complete three routes were analysed and the most active regions of icebergs can be detected.

(ii)Density-based clustering algorithms :These algorithms try to find clusters based on the density of data points in different region and based on this some estimation of Icebergs in different regions can be made.

(iii)Using Spatial Hotspot Analysis for more active regions with Icebergs :

Hotspots are a similar kind of clustered patterns that as in clustered patterns, objects in hotspot regions have high similarity in comparison to one another and are quite dissimilar to all the objects outside the hotspot.The objects in the hotspot area are more active compared with objects in general cluster in terms of density, appearance, etc.Hotspot discovery/detection in Spatial Datamining tasks is a process of identifying spatial regions where more events are likely to happen, or more objects are likely to appear, in comparison to other areas. Hotspot detection is mainly used in the analysis of regions that have more Icebergs in different oceans or sea .They can find areas that have greater than average numbers of Icebergs which can lead to disorderly events, or areas where ships have a higher than average risk of collision. Hotspot analysis also finds applications in detection of cancer disease and several diseases data analysis, hotspots of locations where disease are reported intensively are detected, which may indicate a potential breakout of this disease, or suggest an underlying cause of the disease.By observing the analysis given by the hotspot analysis necessary actions can be taken to avoid ship collisions.

D.Spatio-temporal data mining techniques for detection of icebergs

Spatio-temporal data mining deals with the extraction of unknown and implicit knowledge, structures, spatio-temporal relationships, or patterns not explicitly stored in spatio-temporal databases . Taking one of the Spatio-temporal data mining techniques into consideration and implementing for the detection of Icebergs tasks include , spatio-temporal association rule mining, spatio-temporal sequential patterns mining, spatio-temporal clustering mining relies on how to integrate space and time seamlessly and simultaneously.

The techniques of spatio-temporal mining can be applied for Iceberg detection as follows

A spatio-temporal association rule is an implication of strong association between X and *Y* with the form $X \rightarrow Y$, where *X* and *Y* are sets of spatio-temporal and non-spatio-temporal attributes. If the attributes at *X* take some specific value at a point in time, then with a certain probability, at the same point in time, the attributes at *Y* will take some specific value . A spatio-temporal association rule might find that "If Icebergs are found in a location X with some specific negative temperature at a specific time, at the meanwhile, it can be estimated that it might be quite possible that Icebergs will be there at Y ."Spatio-temporal association rule technique might be able to identify the relationship of locations *X* with *Y* over time change between weather conditions (wind speed, wind direction, temperature, humidity), and geographical conditions (degree of slope, aspect of slope, position of slope).

E.Geostatistical approach

Different types of tools are used for the analysis of spatial data one among them is Geospatial approach. This tool is mainly used for spatial analysis and also for the prediction of spatio-temporal phenomena. It was first used for geological applications.Generally geostatistics techniques are used to analyze and predict the unknown values of variables distributed in space and/or time.It is important to remember that geostastics is limited to point set analysis or polygonal subdivisions and deals with a unique variable or attributes. Under those conditions, it constitutes a good tool for spatial and spatio-temporal trend analysis.

CONCLUSION

This paper provides the different methods that can be used to detect the icebergs by using spatial dataminming and as well as spatio temporal data mining so that the ships collisions with Icebergs can be avoided. Spatial data mining tasks like clustering and Hotspot analysis are used to detect the more active regions of Icebergs .This analysis gives the best to avoid the routes that the ships to be travelled in order to avoid collisions with Icebergs .Spatio temporal data mining is used

By the application of association rule mining technique. Association rule mining technique is used to predict the locations with more Icebergs by comparing it with the different locations. This work constitutes a first step towards a methodology that provides easy detection of Icebergs by using spatial and spatio-temporal data mining so as to reduce the no of ships collisions with Icebergs.

REFERENCES :

[1]. M.Hemalatha.M; Naga Saranya.N. A Recent Survey on Knowledge Discovery in Spatial Data Mining, IJCI International Journal of Computer Science, Vol 8, Issue 3, No.2, may, 2011.

[2]. Identifying patterns in spatial Information: survey of methods by Shashi Shekhar, Michael R. Evans, James M. Kang and Pradeep Mohan

[3]. An Effective Analysis of Spatial Data Mining Methods using Range Queries by Gangireddy Ravi Kumar and Mallireddy Sivareddy,,JGRCS ,Journal of Global Research in Computer Science,vol 3,No 1,Jan 2012.

[4]. 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.

[5]. Lebartetal.,"Statistique exploratoire multidimensionnelle", Editions Dunod, Paris, 439 p., 1997.

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

[7]. Lu, W., Han, J. and Ooi, B.: Discovery of General Knowledge in Large Spatial Databases, in Proc. of 1993 Far East Workshop on Geographic Information Systems (FEGIS'93), Singapore, June 1993, pp. 275-289

[8]. Ester, M., Kriegel ,H.-P., Sander, J., Xu, X.: Density-Connected Sets and their Application for Trend Detection in Spatial Databases, Proc. 3rd Int. Conf. on Knowledge Discovery and Data Mining, Newport Beach, CA, 1997, pp.10-15 .

[9]. Deutsch, S. J. and Ramos, J. A., 1986, Space-time modeling of vector hydrologic sequences, *Water Resource Bulletin*, Vol. 22, pp. 967-981.

[10]. Tsoukatos, I. and Gunopulos, D., 2001, Efficient mining of spatio-temporal patterns, *7th International Symposium on Spatial and Temporal Databases (SSTD),* California, July 12-15, pp. 425-442.

[11]. Verhein, F and Chawla, S., 2005, Mining Spatio-Temporal Association Rules, Sources, Sinks, Stationary Regions and Thoroughfares in Object Mobility Databases, *Technical Report*, *University of Sydney, Number 574*

[12]. Ester, M., Frommelt, A., Kriegel, H.-P., Sander J.: Algorithms for Characterization and Trend Detection in Spatial Databases, Proc. 4th Int. Conf. on Knowledge Discovery and Data Mining, New York, NY, 1998

[13]. Samet H., "Design and Analysis of Spatial Data Structures: Hierarchical (quadtree and octree) data structures ", Addison-Wesley Edition, 1990

[14]. Bédard, Y., Lam, S., Proulx, M.J., Caron, P.Y., Létourneau, F.: Data Warehousing for Spatial Data: Research Issues, Proceedings of the International Symposium Geomatics in the Era of Radarsat (GER'97), Ottawa, May 1997, pp. 25-30

[15]. Fayyad et al., "Advances in Knowledge Discovery and Data Mining", AAAI Press / MIT Press, 1996 .

[16]. Shekhar S, Chawla S, Ravada S, Fetterer A, Liu X, Lu C-T. Spatial Databases - Accomplishments and Research Needs. *Trans. on Knowledge and Data Engineering* 1999, 11:45–55.

[17]. Haining RJ. *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge: Cambridge University Press; 1989.

[18]. Roddick J-F, Spiliopoulou M. A Bibliography of Temporal, Spatial and Spatio-Temporal Data Mining Research. *SIGKDD Explorations* 1999, 1:34–38

[19]. Scally R. *GIS for Environmental Management*. Redlands, CA, USA: ESRI Press; 2006. ISBN: 978-589481429.

[20]. Krugman P. *Development, Geography, and Economic Theory*. Cambridge, MA: MIT Press; 1995.

[21]. Ganguly A, Steinhaeuser K. Data mining for climate change and impacts. In: *IEEE International Conference on Data Mining Workshops, 2008. ICDMW'08* 2008, 385.

[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]. Calkins H. GIS and public policy. *Geographical InformationSystems* 1990, 2:233–245.