International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 2 Issue 4, April - 2013

Spatial Data Mining Used In Marine Geographical Information System

Girish V. Patil
Computer Engineering Dept.
Pdm. Dr. V. B. Kolte College of Engineering,
Malkapur, India

Mahesh V. Shastri
Computer Engineering Dept.
Pdm. Dr. V. B. Kolte College of Engineering,
Malkapur, India

Bharat K. Chaudhari
Computer Engineering Dept.
Pdm. Dr. V. B. Kolte College of Engineering,
Malkapur, India

Abstract

Spatial data mining (SDM) is to extract the credible, innovative, interesting, hidden, unknown, potentially useful and ultimately understandable knowledge from the vast spatial production data, management data or remote sensing data on the basis of the spatial database, and in which the statistical methods, pattern recognition technology, artificial intelligence methods, neural networks, rough sets, fuzzy math, machine learning, expert systems and related information, technology, etc are synthetically used[1]. With the rapid development of marine surveying and mapping technology and computer storage devices, massive amounts spatial data appears in the Marine Geographical Information System. Traditional data processing could not discover the potential knowledge. How to make full use of the Spatial Data Mining technology to obtain the implicit rules, trends and relevance in marine spatial data becomes the most important issue on hydrographic and cartographic field. This paper introduced the basic principle of Spatial Data Mining technique, analyzed the potential knowledge which could be obtained by Spatial Data Mining, and studied the application which spatial Data Mining in marine geographic information system[2].

Keyword- Spatial Data Mining, hydrographic and cartographic, Marine Geographical Information System.

I. INTRODUCTION

Data mining is the process of abstracting unaware, potential and useful information and knowledge from plentiful, incomplete, noisy, fuzzy and stochastic data. Simply stated, data mining refers to extracting or mining knowledge from large amounts of data [3].

With the development of satellite technology, communication technology, computer technology, data processing technology and other modem science and technology, hydrographic and cartographic has already broken through the traditional spatio-temporal limitations, come into an new epoch which has digital surveying and mapping means as main body, computer technology as support, 3S technology as the representative. Using the new equipment and technology, we had obtained a large number of marine surveying and mapping spatial data, which has been processed and backed up day after day. The vast amounts data, indicted ample and valuable knowledge, was foundation of Marine Geographical Information System. Those multi-sources, multi-dimensional, multi-temporal states data had far beyond the analytical capacity of traditional data processing technology. As the integration of statistics, database technology, machine learning method, artificial intelligence and other expert system; only Spatial Data Mining technology is the effective means.

II. BASIC OF KEYWORD

1. Spatial Data Mining

Spatial data mining is based on spatial database, and on the comprehensive utilization of statistical method, pattern recognition technology, artificial intelligence method, neural network technology, rough set, fuzzy mathematics, machine learning, expert system and related information

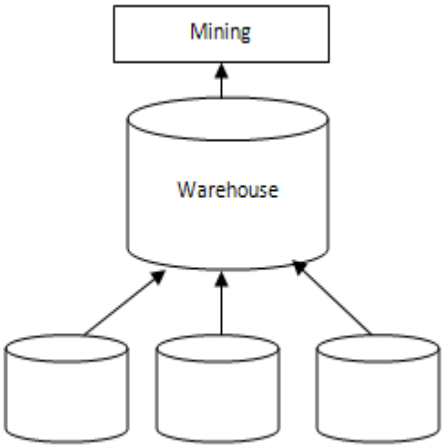

Figure 1 Mining for Data Warehouse

technology, etc. It disjunctives credible, novel, interested, hidden, prior unknown, potential useful and ultimately comprehensible knowledge from a lot of spatial data such as geographic data production data, or remote sensing data [1]. Simple figure given below shows the Spatial Data Mining (SDM) in which spatial data is store in Warehouse.

Aim of the SDM is to reveal the objective world essential routine, internal relations and development trends contained in the data, and then to realize the knowledge automatic acquisition and provide the basis of technical and management decision making. As the main target of the SDM is the spatial database, and in which not only the geometric data, attribute data of the spatial things or object are stored, the graphical space relation between the spatial things or objects is also stored, so its dealing approach is different from the general data mining, and it is a new research branch of the data mining.

2. Hydrographic and Cartographic

Hydrographic refers to the mapping or charting of water's topographic features. It involves measuring the depths, tides, and currents of a body of water and establishing the topography and morphology of seas, rivers, and lake beds. Normally and historically the purpose of charting a body of water is for the safety of shipping navigation. Such charting includes the positioning and identification of things such as wrecks, reefs, structures, navigational lights, marks and buoys and coastline characteristics.

Cartographic refers to study and practice of making maps. Combining science, aesthetics, and technique, cartography builds on the premise that reality can be modeled in ways that communicate spatial information effectively. Cartographic organizations and data publishers must be able to produce professional, standard maps from their libraries of data.

3. Geographical Information Systems

Geographical Information Systems (GIS) are computer based systems that enable users to collect, store, process, analyze and present spatial data. It provides an electronic representation of information, called spatial data, about the Earth's natural and man-made features. A GIS references these real-world spatial data elements to a coordinate system. These features can be separated into different layers. A GIS system stores each category of information in a separate "layer" for ease of maintenance, analysis, and visualization. For example, layers can represent terrain characteristics, census data, demographics information, environmental and ecological data, roads, land use, river drainage and flood plains, and rare wildlife habitats. Different applications create and use different layers. A GIS can also store attribute data, which is descriptive information of the map features. This attribute information is placed in a database separate from the graphics data but is linked to them. A GIS allows the examination of both spatial and attribute data at the same time. Also, a GIS lets users search the attribute data and relate it to the spatial data. Therefore, a GIS can combine geographic and other types of data to generate maps and reports, enabling users to collect, manage, and interpret location-based information in a planned and systematic way. In short, a GIS can be defined as a computer system capable of assembling, storing, manipulating, and displaying geographically referenced information.

Marine GIS enables study of sea level change, marine population, sea surface temperature, and coral reef ecosystem [4].

III. SDM ARCHITECTURE

The SDM system can generally be divided into the three layers, as shown in figure 2. The first layer is data source, which refers to use the index, query optimization and other functions the spatial database or data warehouse management system provided to obtain and abstract the data related to the problem areas, or directly use the data stored in the spatial data cube, the

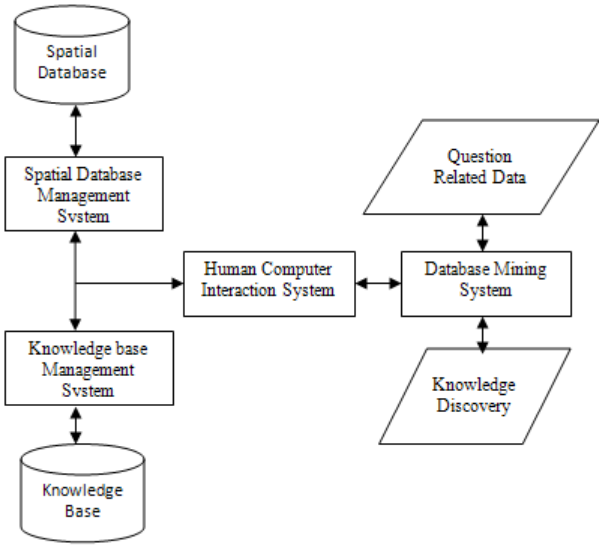

Figure 2 SDM System Structure

data can be called the data source or information base of the data mining. In this process, users mutually collect the data related to the tasks directly through the spatial data warehouse management tools, and carry out necessary visual analysis of the query and retrieve results, repeatedly, to abstract the data related to the problem areas; or take out the related data through the analysis operations such as the spatial data cube gathering, drill, cut and rotation, and then begin the data mining and knowledge discovery process. The second layer is mining, to analyze the abstracted data by using the various data mining methods in the SDM system, interactive form is generally used, which refers to that user will select appropriate data mining method according to the question type and the data type and scale, but to certain specialized data mining system, the system can automatically select the mining approach. The third layer is the user interface, in which the obtained information and knowledge will be reflected to users with the way facilitate to be understood and observed by them, and the knowledge will be provided to the spatial decision making for use or stored in the field knowledge base. In the entire data mining process, users can control each step. Generally speaking, a number of steps of the data mining and knowledge discovery are interconnected, so repeated human-computer interaction is needed to achieve final satisfied results. It is clear that in the entire data mining process, a good human computer interaction user interface is the basis of favoring data mining and achieving satisfactory results [13].

IV. RULES FOR SPATIAL DATA MINING

As we know that Spatial Data Mining is different from Spatial data processing, Spatial Data Mining could collect the recessive, valuable important, hidden knowledge, spatial relations, or other meaningful patterns, in order to discovery all kinds of unknown and known important knowledge. For getting this knowledge we use some Rules. They are as follows.

A. Spatial Association Rule

Spatial association rule is a rule indicating certain association relationship among a set of spatial and possibly some non spatial predicates. A strong rule indicates that the patterns in the rule have relatively frequent occurrences in the database and strong implication relationships. The extraction and comprehension of the knowledge implied by the huge amount of spatial data, though highly desirable, pose great challenges to currently available spatial database technologies [5]. A spatial association rule is expressed in the form A=>B [s%, c%], where A and B are sets of spatial or non-spatial predicates, s% is the support of the rule, and c% is the confidence of the rule.

Also a spatial association rule is a rule which describes the implication of one or a set of features by another set of features in spatial databases and space entities, such as adjacent rule, connected rule, symbiotic rule and including rule. Discovered knowledge was expressed in logical rule. Spatial association rule is an important kind of knowledge in Spatial Data Mining. When the data is the spatio-temporal data or the historical data in same area at different times, we would discover the time sequence rule of space entities.

B. Spatial Characteristic Rule

A spatial characteristic rule is a general description of a set of spatial-related data which include type, size, shape or other general characteristic. For example, the description of the general weather patterns in a set of geographic regions is a spatial characteristic rule. The characterization of a selected part of the database has been defined in as the description of properties that are typical for the part in question but not for the whole database. In the case of a spatial database, it takes

account not only of the properties of objects, but also of the properties of their neighborhood up to a given level.

C. Classification Rule

An object can be classified using its attributes. Each classified object is assigned a class. Classification is the process of finding a set of rules to determine the class of an object. By the spatial classification rule, the data of spatial data set could be mapped to a certain class. It mainly used for data prediction. Space classification rule is the universal knowledge. Its essence is abstracting or generalizing to the given data set.

D. Spatial Dicriminant Rule

A spatial discriminant rule is the general description of the contrasting or discriminating features of a class of spatial related data from other class(es). For example, the comparison of the weather patterns in two geographic regions is a spatial discriminant rule.

E. Spatial Clustering Rule

Spatial clustering rule is widely used for data analysis, which organizes a set of data items into groups (or clusters) so that items in the same group are similar to each other and different from those in other groups. Many different clustering methods have been developed in various research fields such as statistics, pattern recognition, data mining, machine learning, and spatial analysis. It can be used for generalization or comprehensive information. Different with spatial classification rule, it did not know the group's number and the type ahead of the schedule.

V. SPATIAL DATA MINING APPLICATION IN MARINE GIS

As you know Hydrographic and cartographic data is the important part of marine geographic information system, which has the general characteristics of GIS data, so you can effectively use spatial data mining methods to process the marine survey and mapping information. There are different types of applications in marine GIS such as for decision making in navigation.

A. Submarine Topography Data Mining

Chart is a kind of thematic map to describe the areas of ocean and its adjacent land. It mainly represents submarine topography by soundings-note. But because of the limited sheet, soundings note on chart existed in isolated and discrete way, which arrangement form greatly affected the interpretation of the submarine topography and the security channel choice to chart user.

Using the statistical methods of spatial data mining, can dig out the intrinsic rule of submarine topography distribution from the discrete and abstract soundings note, and then build a sounding depth grid model, combined with Visualization technology, we can obtain a specific representation of the true three-dimensional submarine topography.

By spatial data mining analysis tools, can discover the target in spatial connected, neighboring and symbiotic association rules etc., can detect the shortest path and the optimal path between the objectives, and other knowledge to support decision-making .Under the circumstance of human-computer interaction, by setting the different weights for different factors: choosing the navigable area, the shortest navigating time, at least vessel consumption, safety course and so on, users can obtain several recommend sea routes, then the command personnel can select channel, map out the navigational plan to meet the needs of navigation in the shortest time.

B. Marine Sediment Data Mining

Marine sediment also is an important composing factor in submarine topography, at the same time it is the magnitude effect factor to vessel navigation, anchoring and the submergible action.

Exerting the data mining tools, such as statistics, rough set theory and clustering method, the distribution spectrum and spatio-temporal changing rule of the marine sediment would be discovered. The safety of ship navigation would be enhanced and the utilization of waterways would be increased: It would provide the best and comprehensive decision-making for the national defense and economic construction.

C. Navigational Obstruction Data Mining

Navigational obstructions, also known as navigation obstacles, refer to reefs, wrecks, obstructions, etc. Although most of navigational obstructions' location is fixed, due to water flow and other reasons, some type's obstacles' position was often shifted. Navigation obstruction data includes not only

its spatial location and attribute feature, but also implied its incidence, developing trend and other spatio temporal information.

In support of the theory of spatial data mining, MGIS could automatically map the obstruct-water-region, point the different safety index in different navigation condition. When vessels maybe enter the danger zone, system can intelligently issue a warning to mariners as much as possible to ensure the safety navigation of ships.

D. Navigational Aid Data Mining

Navigational aid, also known as navigation mark, is to show reefs or other hazards, or for mooring. It is the important artificial device for vessel at sea to mensurate position, guide course and avoid obstacle.

Navigational mark data stores a large number of navigation marks' attribute feature and spatial location information. That knowledge not only reflects the selected channel and course information, but also implies much horizontal and vertical information such as navigable waters coverage and navigation capability. Using the spatial statistical analysis, spatial clustering, genetic algorithms and other SDM techniques, this information system can reclassify navigational aids in cording with their attribute and grade. Finally mariner could obtain the optimization program, realize automatic selection of navigational marks, and enhance system capacity to support decision-making.

E. Hydrological Data Mining

Hydrological data, refers the water power, the physical properties or the chemical properties of seawater, including the tides, tides stream, currents, rapids, whirlpools, waves, icebergs, sea ice and sea water temperature, salinity, density, water color and transparency elements, etc. On the one hand, hydrological thematic data describes the explicit spatial location, attributes, strength, direction and other information; on the other hand it implies a number of unknown, time-varying and ambiguous information. For example, currents in different water depth with different speed and direction, furthermore when the terrain changed, it would translate into tide rip or whirlpool.

Using SDM, system not only can automatically discovery clustering rule of hydrological elements, but also can provide the optimum analysis and forecast for the navigation decision-making.

VI. CONCLUSION

Now a day's study of marine space environment has become a hot topic. The hydrographic and cartographic data is an important part for digital marine environment. By means of spatial data mining technology, it could be re-induce and re-classify the discrete and disordered information to discover the hidden and common rule, to rootle the new law. It would drastically remodel traditional database's flaws data redundancy, low utilization rate, nonconnection and so on. For establishing the digital marine spatial environment enhancing the decision-making support capability of marine geographic information system, and, it is necessary to study the marine spatial data processing.

VII. REFERENCES

[1] Wang Jinlin, Chen Xi, Zhou Kefa, Zhang Haibo, Wang Wei "Research of GIS-based spatial data mining model" Xinjiang Institute of Geography and Ecology Chinese Academy of Sciences, Urumqi 830011, China.

[2] LI Gaixiao, P ENG Rencan, ZHENG Yidong, ZHAO Jidong "Spatial Data Mining and its application in Marine Geographical Information System" 2nd Conference on Environmental Science and Information Application Technology 2010, 978-1-4244-7388-5/10/©201O IEEE.

[3] Wenliang Cao, Xuanzi Hu, Chongjie Dong "Research of The Mining Algorithm Based on Distributed Database" Department of Computer Engineering DongGuan Polytechnic Dongguan, Guangdong, China.

[4] Arul Prakash "Geographical Information Systems . An Overview" Indian Institute of Information Technology.

[5] Subhasmita Mahalik "SPATIAL DATA MINING TECHNIQUES".