

A SURVEY ON SPATIAL DATA MINING

¹Ms. Renu Tiwari, ²Ms. Sandhya Khanna

¹Compucom Institute of Information Technology & Management, Jaipur

²Compucom Institute of Information Technology & Management, Jaipur

Abstract— Data mining is procedure of determining probable valuable and stimulating design secreted in huge data set. Spatial data mining is occurred as per novel era in research communal. Mining in spatial information is dissimilar in several feature relate to mining in traditional information. Spatial data mining use contains then not restricted to common networking, ruin management and even if calculation. The Spatial Data Mining by means of a significant phase of information research procedure means to abstract the indefinite, valued and practicable relationships, designs or movements for judgment provision system from huge spatial databases and data warehouse. a survey of recently accessible spatial data mining movement and methods are accessible in this paper.

Keywords— Spatial Data Mining, KDD, Framework of SDM, Characteristics, Applications.

I. INTRODUCTION

The spatial data mining and the general data mining have not only relation but also difference. The two concepts are data mining (Data Mining, is called DM) and discover the knowledge (Discovery for Spatial Database, is called KDD) usually are confused. Some scholars take DM as a core process of KDD and think KDD process includes data preparation and interpretation and evaluation of discovery result. Some people think that the essence of the two concepts is same, DM only frequently is used in statistics, data analysis and information system, but KDD is widely used in artificial intelligence and machine learning and so on domain. Also some people think the two concepts are separated with difficulty, should take a whole as using, But the terminology "the data mining" compared to "in the database discovered the knowledge" and "the data mining and the knowledge discover" is concision, therefore is widely popular.

The spatial data mining (Spatial Data Mining, is known as SDM), or named the data finding after the spatial database (Knowledge Discovery from Spatial Database), its definition is: in the spatial database and in the data warehouse foundation, comprehensively utilizes statistical method, the pattern recognition technology, the artificial intelligence method, the neural network technology, Rough set, the fuzzy mathematics, the machine learning, the expert systematic, visualization technology and other correlative information technologies taken as the measure, extracts knowledge that is credible, novel, interesting, hidden, unknown, latent, useful and finally comprehensible from the massive spatial data, the management data, the fare data or remote sensing data, thus opens out essence rule containing spatial data, inherence relation and development tendency, realizes knowledge is automatically or semi-automatically gain, provides the suggestion for the management and the fare decision-making.[1]

II. BASIC PRINCIPLE OF SPATIAL DATA MINING

A. The concept of Spatial Data Mining

Spatial Data Mining (SDM), with spatial database or spatial information warehouse for example information causes, integrated the relevant theory and technology, extracted the credible, novel, interesting, hidden, unknown in advance, potentially useful and ultimately understandable knowledge from an avalanche of noisy and uncertain spatial data, revealed the inherent nature, relations and development trends of the objective world behind those datum to achieve the automatic acquisition of knowledge and provide service for information system and operational decision-making.

Spatial data mining is a new development stage of spatial databases, computer management and decision-making. It is the multidisciplinary technology integration, including the artificial intelligence, probability statistics, expert systems, machine learning, spatial databases, fuzzy mathematics, networking, visualization and other correlative disciplines. SDM is a novel exacting information treating technology of geographic data structure.

B. Framework of Spatial Data Mining

So Far, there are a variety of spatial data mining frameworks, such as: multi-component data mining framework put forwarded by Matheus, character architecture advanced by Holsheimer, generic model of data mining proposed by Han. Which should be more mature is multi-component data mining framework, just as shown in Figure 1.

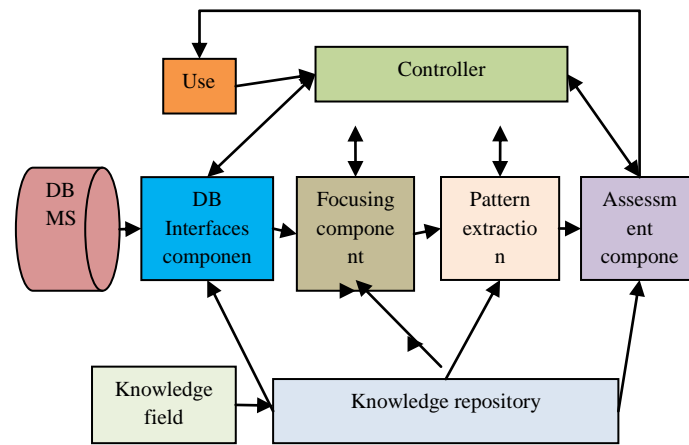


Figure 1. Spatial Data Mining's Framework.

In this system, the user can interactively control the each process of data mining. Knowledge repository stored the background knowledge. Data mining process had completed through four components: DB interface component directly extracted data from the database; focusing component selected data which would determine which type data is useful and 'which type is useless; pattern extraction component would detect knowledge and laws from those useful data; assessment component would evaluate the detected patterns in interesting or importance aspect to remove redundant or useless knowledge. These four components interact with each other through the controller. The final results would deliver to the user. [2]

III. The Characteristics of Spatial Data

Spatial objects have spatial location and distance properties. There is a certain interaction between adjacent objects, so the relationship between the spatial data is more. There are obvious difference between spatial data and other type's data. Spatial data has the following complex characteristics:

1) Massive data

Massive data often make some algorithms cannot be implemented for difficulty or excessive calculating amount, and therefore one of the tasks of spatial data mining is to create a new computing strategy and develop new efficient algorithm to overcome the technical difficulties caused by massive data.

2) Non-linear relationship between the spatial attribute

It is an important symbol of the complexity of space systems, reflects the complex mechanisms of the system inner function, and is 1 of the key responsibilities of spatial data mining.

3) Scale characteristic of spatial data

The following law as well as the embodying characteristic of spatial data is not the same at different observing levels. Scale characteristic is another manifestation of the complexity of spatial data, and can be used to explore gradual change law of the characteristic in the process of simplification and modification of data.

4) Spatial dimension increasing

The properties of spatial objects increases very rapidly, as in the field of remote sensing, due to the rapid development of sensor technology, the number of bands increased from a few to tens or even hundreds, so, how to mine data and discovery knowledge from tens or even hundreds dimensional space becomes another hot spot study.

5) The ambiguity of spatial data

Ambiguity occurs in virtually all forms of spatial data, like the ambiguity of spatial site, the ambiguity of spatial association, along with the attribute values of fuzzy, etc.

6) The lack of spatial data

The lack of data is due to some irresistible external force so that data cannot be acquired or be lost. How to recover lost data and estimates the inherent distribution parameters of data becomes one of the difficulties in solving the complexity of data. [3]

IV. LITERATURE SURVEY

Emine Rumeysa Guler et.al. [2018] In the term of the IOT, vast amount of knowing campaigns gather and/or create several sensual information above time for a extensive variety of fields and uses. Built on the nature of the use, these campaigns effect in great or fast/real time data tributaries. The analytics technique on the subject matter used to discover new information, anticipate future predictions and make decisions on important issues makes IoT technology appreciated for equally the business domain and the value of normal life. In this study, first of all, the concept of IoT and its architecture and relation with big and streaming data are emphasized. Information discovery process applied to the IoT streaming data is investigated and deep learning frameworks covered by this process are described comparatively. Finally, the most commonly used tools for analyzing IoT stream data are introduced and their characteristics are revealed.[4]

Chowdari K.K et.al. [2015] the data mining systems are engaged for effective and real time study of Even if and Weather information. The key objective of studies on Weather is that users example farmers, Scientist, decision & policy maker etc., from altered productions e.g. Agriculture, Scientific, Aerospace etc., is required to understand the importance of various changes in weather and climate parameters like rainfall, humidity, temperature etc. Data unearthing from different sources both in temporal and spatial domains is critical for climate studies and also its impact on different verticals like health, water, energy etc. Yet, using the improvement in technology and accessibility of global Geo – graphical data delivers the data miner’s novel chances all composed. This paper delivers a well accepting of the endure and weather information with spatial – temporal mining. In the current work the advance of new algorithms to study the altered mining systems for endure and weather variation readings will be agreed out using the several case studies like rainfall analysis and simulation, cyclone analysis and simulation and temperature analysis and simulation etc. [5]

Junhuan Peng et.al. [2011] The Ionosphere plays an important role in atmosphere, who’s globally distributed total electronic content (TEC) obtained by GPS technology is the important data source of geographic or earth information system for monitoring global change. This paper put on the equipping process of deionization variable theory to mine the data of big scale of propensity difference and lesser scale of casual deviation, and exposed that the big scale propensity can be displayed as a nine sequences of globe coherence function, and the lesser scale deviation more selects to a zero mean non-stationary random process of symmetrically distributed. Relating the established unit-root check, the lesser scale remaining is recognized using the representative of three sequences of fractional casual walk, and so the residuals after performing 3 orders of difference show the property of white noise process. The common Kriging predication system built on the fractional casual walk model is created to rebuild the spatial procedure exactly. The effect shows that the fractional casual-walk-built check can be used to excavation the auto-correlated arrangement of 0 mean non-stationary mistake function or lesser scale variation, and the constructed general kriging method can improve the prediction result. [6]

Lijie Yan et.al. [2011] the spatial design of primitive expenditure locations in Henan Domain is examined. With examining the relationship b/w place of primitive expenditure places and provincial raise, terrain slope, rivers and further geographic ecological issues, rules of circulation of primitive clearance places are create: More than 70% discarding places are placed at an highness of 50. More than 400 meters, more than 60% of the discarding places are placed in ranges wherever the slope is less than three degrees, and 60 percentage are placed less than four km away the river. With the study of the circulation designs of expenditure places by Geography Information System (GIS) and Spatial Data Mining, well technology is delivered for future removal archeology, and to know the prehistoric human choice approach on the atmosphere. Locations are delivered by examiners. Enlargement of removal structures. [7]

Caixiang Xie et.al. [2010] Climate factors govern the distribution of plant species which is the indicator of the corresponding region climate. Spatial clustering procedures are a main module of spatial data mining. We found dispersal information of more than 100 Chinese open local herb plants to assist as simple information for spatial examine. Spatial clustering algorithm built on spatial contiguity associations in GIS was recycled to set China counties into 7 sections unique in climate and ecology features. The clustering results showed the eco-environment has great effects on geo-authenticities of traditional Chinese medicine materials. [8]

Cheng Jiehai et.al. [2010] to get the innovation of mine initiatives, it is essential to perform the "digital mine" creation. And the main to executing "digital mine" conception is to explain the active storage and administration of my spatial information. The paper first displays the particular features of my spatial information. Before it converses the process of storage and administration of my spatial information built on the Postgres SQL open source database, which is a kind of innovative item relational database models. All of these can beat out a path for mine digital construction. [9]

Yunjun Zhan et.al. [2009] Remote sensing technology has been used in hazardous geology survey and evaluation. But there are many problems such as the difficulty of potential hazards body identification, low automation, and low identification accuracy in the condition of Complex terrain for it, the paper set onward the usage of spatial data mining technology to classify and eliminate the contrary physical body. Examination displays that excavation of multi-source data and data mining can be useful to eliminate dangerous geology spots and possibly contrary physical body. This paper emphases on geo-scientific database formation and verifies its feasibility using an instance of multi-source physical spatial data mining, physical spatial data mining process, mining procedure and mining method, which is built on adverse physical body appreciation. . [10]

Show-Jan Yen et.al. [2001] Mining Sequential Design In huge amounts, most consumers look out the serial buying conduct of consumer communications. In this letter, a data mining language has been obtainable. From data mining languages, operators can identify entities and standards that are wanted for the serial design to be investigated. In calculation, an effective data mining way is suggested to abstract serial designs according to operators needs. [11]

V. APPLICATIONS OF SDM

One of the major applications of spatial data mining is geographic information system in which data is represented in a grid of cells called pixels where each cell resolution depends on its pixel size where more the pixel size more is the resolution.

i. Data structure in SDM: Data structure in spatial data mining also correlate the communication networks in SDM the neighborhood graph and the path leads to the connection of objects. The relation neighbor depicts the topological and the

metric relation in which distance of the objects can be measured to check the conductivity or disjunctively of the earlier relation While non-spatial structures of the entities containing name, population and unemployment degree for city is categorized by the support of non-spatial information.

ii. Spatio-temporal DM: The events and processes representing the change of some kind is represented using spatio-temporal DM where an ongoing phenomenal represent the activities without specifying the endpoints.

iii. Criminology: Every individual has a specific mindset of their rational choice. Spatial process leads to two concepts either heterogeneity or dependency where at local level a change occurs across a defined space where conditions of racial composition vary evenly. In the case of criminals, they have a wide range of variability where crime chose targets within context. Even the criminals think rationally thus making calculative decisions and always commit a crime against the potential benefits while if constrained by time, cognitive ability and the availability of information results in a limited rationality. [12]

VI. CONCLUSION

Spatial data mining is a very young and abroad foreground research domain, at present, gains only some primary fruit, yet still has a lot of theory and method that need intensive research. It mainly includes multi-resource spatial data cleaning and data mining based spatial uncertainty (position, attribute and time and so on), increasing data mining, integration vector with raster data mining, multi-resolution and multi-layer data mining, parallel data mining, research about new algorithm and high effective algorithm, spatial data mining query language, data mining of remote sensing image database, knowledge discovery of multimedia spatial database and network spatial data mining and so on.

In addition, besides development and perfection itself theory and method, spatial data mining needs fully to adopt and assimilate mature theory and method that are data mining and information find, database, machine education, artificial intelligence (AI), mathematic information, conception, remote sensing, grid and portrait , medical treatment and molecule biology and so on.

REFERENCES

- [1] Hu Shengwu, "Method Development about Spatial Data Mining and Its Problem Analysis", 2011 International Conference on Electronics and Optoelectronics (ICEOE 2011).
- [2] LI Gaixiao, P ENG Rencan, ZHENG Yidong and ZHAO Jidong, "Spatial Data Mining and its application in Marine Geographical Information System", 2nd Conference on Environmental Science and Information Application Technology2010.
- [3] Hailiang Jin and Baoliang Miao, "The Research Progress of Spatial Data Mining Technique" 978-1-4244-5540-9/10/\$26.00 ©IEEE 2010.
- [4] Emine Rumeysa Guler and Suat Ozdemir, "Applications of Stream Data Mining on the Internet of Things: A Survey", dx.doi.org/10.1109/IBIGDELFT.8625289. 2018.
- [5] Chowdari K.K, Dr. Girisha R and Dr. K C Gouda, "A study of Rainfall over India With Data Mining", International Conference on Developing Research in Electronics, Computer Science and Technology 2015.
- [6] Junhuan Peng, Kun Qin, Honglei Yang and Lei Cheng, "The consider on the Investigative Spatial Data Mining Method Built on Partial Random Walk and Its use in GPS TEC Examination", 978-1-4244-8351-8/11/\$26.00 ©IEEE 2011.
- [7] Lijie Yan, Yishao Shi, Ruixia Yang, Peng Lu, Li Zhu and Qiang Fu, "The Analysis of the Spatial Pattern of Prehistoric Settlement Sites Based on GIS and Spatial Data Mining", 978-1-61284-848-8/11/\$26.00 ©IEEE 2011.
- [8] Caixiang Xie, Shilin Chen, Fengmei Suo, Dan yang and Chengzhong Sun, "Regionalization of Chinese Medicinal Plants built on Spatial Data Mining", 2010 7th International Conference on Fuzzy Structures and Information Find (FSKD 2010).
- [9] Cheng Jiehai and Li Wei, "Research on the Storage and Administration of Mine Spatial Data built on PostgreSQL" 2010 International Conference on Computer Application and System Modeling (ICCAISM 2010).
- [10] Yunjun Zhan and Haiwei Yang, "Extracting Hazardous Geology Information Based on Spatial Data Mining", Second International Conference on Smart Networks and Smart Methods 2009.
- [11] Show-Jane Yen and Yue-Shi Lee, "An Efficient Data Mining Technique for Discovering Interesting Sequential Patterns", 0-7695-1119-8/01 \$17.00 © IEEE 2001.
- [12] Arvind Kumar, Aanchal Kakkar, Rana Majumdar and Anurag Singh Baghel, "Spatial Data Mining: Latest Movements and Methods", International Conference on Computer and Computational Sciences (ICCCS) 2015.