International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307

OPEN ACCESS

Available Online at : www.ijsrcseit.com

doi : https://doi.org/10.32628/CSEIT2410237

A Review on The Spatial Mining (Trends, Techniques & DS Used)

Ghazala Shireen

Assistant Professor, Parul University, Limda, Waghodia, Gujarat, India

ARTICLEINFO

Article History:

Accepted: 15 March 2024

Published: 06 April 2024

Publication Issue

Volume 10, Issue 2

March-April-2024

Page Number

432-438

ABSTRACT

Public and confidential associations have heritage or then again functional spatial information bases or non-spatial information bases, which are likewise some ways or another connected to a spatial data set or a spatial significance. Notwithstanding mission related information bases, these associations either have or access a few information bases containing like registration, financial, security, picture, interactive media, factual data for arranging, insight, choice and strategy making. The lengthy size of these informational indexes makes it hard to look for significant examples or connections among information. Spatial information mining is the course offending fascinating and already obscure, however possibly valuable examples from huge spatial datasets. This paper centers around Procedures and the special highlights that recognize spatial information mining from old style information mining, at long last it recognizes areas of spatial information mining where further examination is required.

Keywords : Spatial Mining, Techniques, Trends Data Structure Used, Latest Technologies, Data Mining, SDM Tools, Process, GIS

I. INTRODUCTION

Spatial data mining is a specialized subfield of data mining that deals with extracting knowledge from spatial data that is data associated to area or geographical region.

Spatial data mining is based on geographical analysis. Spatial data refers to data that is associated with a particular location or geography.

Examples of spatial data include maps, satellite images, GPS data, and other geospatial information. Spatial data mining involves analyzing and discovering patterns, relationships, and trends in this data to gain insights and make informed decisions.

Spatial data mining involves analyzing and discovering patterns, relationship, and trends in this data to gain insights

And make informed decisions.

The use of spatial data mining has become increasingly important in various fields, such as finding specified area, logistics, environmental science, urban planning, transportation, and public health area.

By analyzing spatial data, researchers and data mining professionals can identify correlations, predict future events, and make informed decisions that can have a significant impact

It requires specific techniques and resources to get the geographical data into relevant and useful formats.

For example, any transportation company can observe its delivery routes for faster and more efficient deliveries by using spatial data mining techniques. They can analyze their delivery data along with other spatial data, such as traffic flow, road network, and patterns of weather, to identify the most efficient routes for each delivery to the specified customer.

II. USE OF SPATIAL MINING

A. spatial mining used in:)

NASA Earth Observing System (EOS): Earth science data: Spatial mining prominently use in NASA for observing routes and weather condition while launching any satellites in space.

National Institute of Justice: crime mapping: For any wrong doing planning of any culprit this technique is widely used to chase them and catch them.

Census Bureau, Dept. of Commerce: census data For counting the population this technique used widely used

Dept. of Transportation (DOT): traffic data, shortest routes: This is used to find out the dense traffic area and find out shortest path out of them.

National Inst. of Health (NIH): AIDS clusters: This institute use spatial mining for analyzing data about different type of diseases in different geographical area. Commerce.

Example: stock market, Retail Analysis: This field use spatial mining for analyzing stock, retail, olesale.

Spatial Mining Technology in GIS

The Spatial Data Mining (SDM) technology has emerged as a crucial area for spatial data analysis within Geographical Information Systems (GIS). GIS stores data collected from various sources in geo databases, representing spatial features based on latitude and longitude. These geodatabases are expanding rapidly due to the increasing volume of data from sources like satellite images, providing details on natural resources and other features. GIS is now widely used in traffic monitoring, tourist tracking, health management, and biodiversity conservation.

Computational algorithms play a vital role in extracting valuable information from geodatabases. The document outlines GIS data formats, data representation models, data sources, data mining algorithmic approaches, SDM tools, and the challenges faced in GIS data analysis. It views GIS as a Big Data problem due to the massive amount of data being generated and stored.

The paper provides an overview of GIS data sources, data representations, SDM tasks in various GIS domains, and SDM tools. It also discusses the issues and challenges related to GIS datasets and proposes an architecture to address these challenges. Overall, the document emphasizes the importance of spatial data mining techniques in various domains such as transportation, tourism, soil quality monitoring, water resource monitoring, and deforestation.

B. Concept & Characteristics of Spatial Mining

The concept of Spatial Data Mining, is defined as the extraction of hidden, valid, novel, and potentially useful patterns or knowledge from spatial databases. It emphasizes the interdisciplinary nature of spatial data mining, influenced by disciplines such as spatial database systems, statistics, machine learning, pattern recognition, visualization, and spatial information sciences.

The significance of spatial data mining in understanding spatial data, discovering spatial relationships, exploring the connections between spatial and non-spatial data, structuring spatial knowledge bases, reorganizing spatial databases, and optimizing spatial queries is very crucial. The introduction to the concept of knowledge discovery in databases (KDD) in 1989, which has had implications for spatial data analysis.

The early research on spatial data mining by scholars at Simon Fraser University in Canada and the introduction of the concept of Spatial Data Mining and Knowledge Discovery (SDMKD) was done by **Chinese scholar Li Deren in 1994**. It acknowledges the comprehensive research and theoretical frameworks developed for spatial data mining over nearly two decades.[6]

The characteristics of spatial data are explored, emphasizing its spatial coordinates in 2D, 3D, or higher-dimensional space, along with relevant spatial extent data like latitude and longitude. The unique characteristics of spatial data, such as spatiality, abstractness, and polymorphism, are discussed, along with the complexity of accessing spatial data compared to non-spatial data.

Moreover, the document touches upon the general characteristics of data, including selectivity, reliability, timing, completeness, and detail. It also outlines the specific characteristics of spatial data, such as the non-linear relationships between attribute spaces, multi-scale features, the fuzziness of spatial information, and the increase in space dimension due to advancements in sensor technology.

Overall, it provides a comprehensive overview of Spatial Data Mining, covering its definition, development, characteristics, methods, and significance in the field of spatial information and digital engineering.

C. Spatial Mining Task & Techniques

Fundamental errands of spatial information mining are:

- **Classification** - tracks down a bunch of rules which decide the class of the arranged article concurring to its attributes.
- **Association rules** - find (spatially related) rules from the data set. It portrays designs, which are much of the time in the data set.
- **Characteristic rules**- portray some piece of data set for example "span is an item where a street crosses a stream."
- **Discriminate rules1**- Depict contrasts between two pieces of information base e. g. track down contrasts between urban communities with high and low joblessness rate.
- **Clustering** - Gatherings the item from data set into groups so that item in one bunch is comparable and objects from various groups are disparate.
- **Pattern identification** - Tracks down patterns in data set. A pattern is a transient example in some time series information. A spatial pattern is characterized as an example of progress of a non-spatial property in the neighborhood of a spatial object.

There is no interesting approach to characterizing SDM strategies. Different sorts of examples can be found from data sets and can be introduced in various structures. In light of general information mining, errands can be ordered into two fundamental classes:

a. Expressive information mining and
b. Prescient information mining.

The previous compactly depicts the conduct of datasets and presents fascinating general properties of the information. While the last option endeavors to develop models that will more often than not resist anticipating the conduct of the new datasets. It centers around the association of the specific spatial information mining strategies as

. Clustering and Outlier Detection
. Association and Co-Location
. Classification
. Trend-Detection [5]

D. Data Structure use in Spatial Data Mining

In spatial information mining, a few information structures are generally used to proficiently sort out and break down spatial data. These information structures empower the capacity and recovery of spatial data, working with different spatial examination and mining assignments.

One broadly utilized information structure in spatial information mining is the **Grid based structure**. This approach isolates the spatial space into a customary matrix of cells, with every cell addressing a particular region or region. Grid based structures empower productive spatial ordering and questioning activities by dividing the dataset into more modest, reasonable lumps.

Another well-known information structure for spatial information mining is the **R-tree3**. R-trees are tree-based spatial ordering structures intended to coordinate spatial articles in light of their spatial connections. R-trees take into account effective looking and recovery of spatial items in light of their mathematical properties or location.

In addition, with regards to spatial organization information mining, particular information structures are utilized to deal with spatial organizations like street organizations or informal organizations. These designs consider the availability and connections between spatial elements and empower network-based spatial investigation and mining tasks.

The selection of information structure in spatial information mining relies upon the particular necessities of the application and the attributes of the spatial information being analyzed. Various information structures have different compromises with regards to question execution, capacity proficiency, and backing for various kinds of spatial activities.

E. Spatial Data Mining for GIS

Spatial Data mining in Spatial Data structure generally means to extract the implicit knowledge that is not stored in spatial data structure implicitly

A large amount of spatial data has been collected in various applications, ranging from remote sensing to GIS, computer cartography, environmental assessment and planning. The collected data is in such a huge amount that it’s far of human knowledge to analyze it, new and efficient methods are needed to extract and discover knowledge from large spatial databases. Mining of Spatial informational index is viewed as intricate as the spatial information isn't addressed expressly in geo databases. Examination of Spatial information has become significant for dissecting data regarding area. Along these lines, investigation of spatial information requires planning of spatial characteristics with non-spatial traits for compelling direction. Spatial information is otherwise called geospatial information contains data about an actual item that can be addressed by mathematical qualities in a geographic coordinate framework. Spatial information is multi-layered and auto corresponded. Spatial information incorporates area, shape, size and direction. Non spatial information is likewise called as quality or trademark information which is autonomous of all the mathematical contemplations. Non spatial information incorporates level, mass and age, and so on. The spatial credits are grouped in three significant relations as Distance connection, Heading connection furthermore, Topological connection. Topological connection is generally non spatial information, so it requires spatial planning to change over non spatial to spatial information.[4]

F. Data Types & Data Relation in Spatial Mining:

There are separate data types to represent spatial data types. These are the following three categories:

1. **Raster data type**: Raster data is similar to our digital camera dataset or satellite imagery data, which we see in Google earth. All area is divided into grids (i.e. row and column and its intersection point also called pixel). One purpose of this data type is to know land type whether given area is forest or river.
2. **Vector data type**: point, line and polygon are the example of vector data types. Vector maps are composition of these three data types.
3. **Graph data type**: data is represented as node, edge and path here. For navigation and routing purpose GPS devices are used. Here GPS devices may use road intersection as node, there segment as edge and this complete model as graph.

Data relationship is also an important factor in Spatial Data Mining as here data relationship is different from other traditional data set. This relationship may be categories as following:

i. **Topological**: Intersect, meet, overlap with-in, boundary, etc.

ii. **Metric**: distance, area, volume, perimeter etc.

iii. **Directional**: North, South, east, west, North East, Southeast

iv. **Raster relationship**: a. Local b. Focal c. Zonal These data types and relationships are formally defined as OGC standard and supported by most of the software like Oracle, IBM DB2, PostGIS etc. Still there are many scopes of improvement on direction prediction, three-dimensional data mining, visibility, moving object and many more.[2].

G. Spatial Mining for Marketing and Planning:

Three marketing and planning case studies that use vector data for analysis are presented in this section. The first analysis projects sales at prospective new sites for a commercial enterprise and places emphasis on managing copious volumes of spatial data. The second and third studies, respectively, use visual analytics and subgroup finding for mobile network optimization and client segmentation.

H.1. Sales Forecasting for Retail Location Planning

Any retail business that wants to succeed must select the right location. When assessing possible new locations, the most crucial factor in the decision-making process is the predicted sales at a given location from a microeconomic perspective. Still, planning a retail location still faces significant challenges related to sales forecasting. How can one forecast sales at prospective new locations? And what elements have the biggest effects on sales? One of the top trading firms in Austria is our project partner. To mitigate risk in location selections while maintaining expansion, the organization looked for an automated sales forecasting system to assess potential new locations.

In our project, we determined and measured the key elements influencing sales in active retail sites for three distinct product categories.

To describe the surroundings of specific stores, we first constructed trade regions with socioeconomic, demographic, competitor, and point of interest information Data was combined. The technique of extracting features for every information source is explained in greater depth in the subsequent paragraphs. In general, driving time zones or buffers can be used for aggregation. They indicate, for a fixed location, the region that is reachable in a specific amount of time or that is inside a specified range. However, the effects of location considerations vary on different spatial aggregation levels, and it was not known in advance which levels would have the greatest influence. As an example, if qualities had been considering only time zones that are five minutes away by car, significant positive shopping

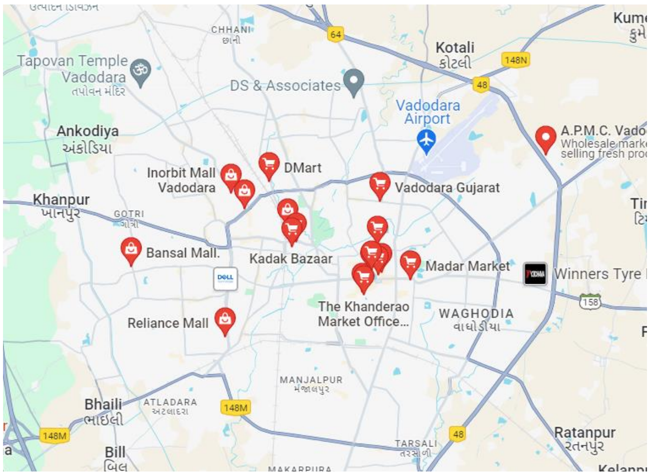
H.2. Customer Segmentation for Marketing Services: Database is offered by one of the top German gas suppliers in the business-to-business sector. Marketing services for regional energy suppliers and power agencies. Providing accurate information on gas users is a major challenge in this field: What are the primary determinants of consumers' interest in natural gas? How can these attributes be used to classify potential clients in a reliable way? Additionally, In what way may this information be applied to automatically assist the choice of advertisements for direct marketing?

The application of spatial data mining and knowledge discovery is thought to be a promising approach to address the aforementioned issues because it entails the creation of models with geographically limited validity and models employing indirect and dependent relationships based on spatial

First, we used visual analytics approaches to explore the data. Afterwards, binomial testing and subgroup mining were employed to assess the statistical significance of the generated hypotheses. Analytical science is called visual analytics.

Visual interfaces that are interactive promote reasoning. The ability to visualize data is crucial, particularly when considering a specific region, as it allows analysts to leverage their creative thinking, previous knowledge, and imagination.

Subgroup identification identifies collections of objects that exhibit a notable departure from the goal value over the whole data set and share common attributes. Subgroups with a noticeably higher response probability to marketing initiatives than the general population were the ones we looked for in our program. A subgroup's quality is determined by two terms: one qualitative and one quantitative. [1]



III.LITERATURE REVIEW

1.Discuss use of spatial mining in shopping and marketing field that how marketing manager analysis the market and demand of customer and as per that they supply their product in that area.[1]
2.Discuss the types and relation in Spatial Mining that are used to extract the relation in the data that are associated with geographical region.[2]
3.Discuss the general goal of geographic data mining in spatial data structures is to retrieve implicit knowledge that is not implicitly contained in spatial data structures. Applications for a vast range of spatial data have been developed, including computer cartography, remote sensing, GIS, environmental evaluation, and planning.[4]

H. CONCLUSION

The review or survey conducted in this paper is an attempt to give an insight to basic meaning of Spatial mining, the prominent area where it is used widely, its tasks, recent trends and techniques related to spatial data mining. Concepts an characteristic, data types and relation used and also cover the statistics used in spatial data mining.

REFERENCES

[1]. Christine KORNER ¨ a,1, Dirk HECKER a, Maike KRAUSE-TRAUDES a , Michael MAY a , Simon SCHEIDER a , Daniel SCHULZ a , Hendrik STANGE a , Stefan WROBEL a,b a Fraunhofer IAIS, Sankt Augustin, Germany b Department of Computer Science III, University of Bonn, Germany .ResearchGate,2010

[2]. Arvind Kumar1 , Aanchal Kakkar2 , Rana Majumdar3 , Anurag Singh Baghel4 J. International Conference on Computer and Computational Sciences (ICCCS) .ResearchGate,2015

[3]. Kamil Raczycki, Marcin Szymański, Yahor Yeliseyenka∗ raczyckikamil@gmail.com, {szymanski7marcin, yahor.yeliseyenka}@gmail.com Department of Artificial Intelligence, Wrocław University of Science and Technology Wrocław, Poland Piotr Szymański, Tomasz Kajdanowicz {piotr. szymanski,tomasz.kajdanowicz}@pwr.edu.pl Department of Artificial Intelligence, Wrocław University of Science and Technology Wrocław, Poland. ResearchGate,accepted to IWCTS 2021

[4]. Mousi Perumal, Bhuvaneswari Velumani, Ananthi Sadhasivam, and Kalpana Ramaswamy Department of Computer Applications, Bharathiar University, Coimbatore, Tamilnadu, India Springer International Publishing Switzerland 2015 .

[5]. N.Sumathi R.Geetha Dr. S. Sathiya Bama Senior Lecturer Senior Lecturer Professor K.S.Rangasamy College K.S.Rangasamy College K.S.Rangasamy College of Technology of Technology of Technology Tiruchengode-637 215 Tiruchengode-637 215 Tiruchengode-637 215 Journal of Computer Applications, Vol – 1, No.4, Oct – Dec 2008

[6]. Du Jiangyi and Hu Cheng Research Center of Spatial Information and Digital Engineering, Wuhan University, Wuhan, China jydu1980@sina.com. Springer International Publishing Switzerland 2015