MATRIX Academic International Online Journal of Engineering and Technology
ISSN: 2348–3326 © MAIOJET
Volume III Issue I, April 2015

Novel Approach for Interactive Spatial Data Mining

Vinodini Katiyar
Professor, Shri Ramswaroop Memorial University, Lucknow, India.
drvinodini@gmail.com

Mohd Haseeb Ahmad
Asst.Professor , Shri Ramswaroop Memorial College Of Engineering And Management, Lucknow, India.
btechfirstyearc@gmail.com

S.Q.Abbas
Professor & Director, Ambalika Institute of Management And Technology, Lucknow, India.
qrat_abbas@yahoo.com

Abstract- Spatial data mining presents novel tasks due to the large magnitude of spatial data, the density of spatial statistics, altitudinal reliance, and the special nature of heterogeneity. Most research in this zone has concentrated on effective query processing by some attractive and affordable means. This paper introduces an active spatial data mining interactive framework that extends the existing spatial data mining algorithms to efficiently support user-interactions on mined spatial data.

Keywords-spatial data mining, spatial database, data mining, Interactive approach, repository, optimizer.

I. INTRODUCTION

SPATIAL data mining, i.e., detection of motivating characteristics and patterns that may indirectly exist in spatial databases in many applications, might become the main feature of interest. There are several desirables for the improved version of this section of mining:

1. How to provide the user an interaction with the spatial data sets?
   - Suppose a farmer is in a location where he is concerned about geographical detailing of the plane, the nature of the sand, the weather forecast etc. or
   - Suppose a politician is in a location where he is concerned about the density of population, religion, sects or
   - Armour deployment in some area can be found via satellite images or
   - Applications need care for a large number of geographically dispersed mobile users collaborating on a domain etc.

2. How to produce the visual interactions to the user?
   - Suppose an airplane is in a location where it is worried about the solidity of air, pollution, altitude etc.

3. How to include user choice along with the interface?
   - It is necessary to include user choice as user wants his required page among the top pages.

4. How to improve the overall accuracy of filtering?
   - Accuracy is the key criteria for improving the performance.

II. RELATED WORK

SPATIAL DATA MINING

A. The Definition of Spatial Data Mining

Spatial data mining (SDM) means finding out wisdom from spatial database, In general, the Data Mining process includes three basic steps:

1) Data Abstraction: This step includes identifying the data to be mined, then choosing appropriate tools required in extracting, manipulating, and examining data from a large data warehouse.

2) Discovery: This step includes using one or more techniques to extract patterns of interest, minimizing data errors and mislead searches.

3) Validation: The extracted information is assessed by means defined by the mining systems.

B. The Characteristics of Spatial Data

Spatial data are generally characterized in that they have an indefinite form and are present in large quantities. There is a confident collaboration between end-to-end objects, so the affiliation between the spatial statistics is further complex. There are noticeable variance between spatial records and other type's data. Spatial statistics possess the following multifaceted characteristics:

1) Flexible data

Spatial statistics are said to be open-ended because the content and assembly of every spatial item are characterized in diverse traditions dependent on the sort of the spatial entity. Further, even two stuffs of the similar sort can fluctuate in the quantity of arguments, sub-objects, etc. that they encompass, and thus have diverse customs and measurements.

2) Ample statistics

Spatial data are commonly large. The size of several categories of spatial statistics can surpass the extent of a DBMS data packing folio. For example, it can take substantial megabytes (MB) of packing planetary to store entirely of the data apply to the coast of a single land.

3) Altitudinal reliance

Spatial statistics is categorized by the presence of spatial reliance. This means that criteria for a particular mutable in an explicit place are interrelated with the standards of another mutable in adjoining location. Altitudinal reliance specifies that the framework has an important influence in the procedure, this complements difficulty to the exploration.

4) Structures instability

There is another vital characteristic of spatial data which is frequently called heterogeneity. Heterogeneity outcomes from the exclusive environment of every place, demonstrating that spatial data very infrequently grants stationary features. This characteristic is also known as nonstationarity. Thus how to recuperate missing data and approximations for the intrinsic distribution of data develops major complications in mining of data tree.

III. THE MAIN METHODS OF SPATIAL DATA MINING

There have been several assistances in this area during recent years [1-14] as:

A research present an dynamic spatial statistics mining tactic that support manager defined triggers on animatedly developing spatial document, which employ a tiered assembly with linked geometric information at several levels. The process spoil user defined trigger into a set of sub triggers connected with cells in the ladder Moreover, this scheme can support incremental query processing as well. [1]

Most of the collected data now a days has certain sort of structural involvement in that, so to hasten data mining of huge databases, there must be some thought to determine directions in operational databases. A research try to discover common substructure within the data, hence interpretation is easy. [2]

A technique to present an option to the real time market competitors to have a grasp on the forecast of the load in future market ,knowledge discovery processes are tried and useful patterns are discovered, hence an exploratory data analysis is done.[3]

because of virtual memory manager tries to project the data base into the memory so system becomes unpredictable sometimes, a research tries to present an associative rule mining which is capable to mine a transcation database without relying on virtual memory manager[11]

A research observe the information of the poised data, including channel position measurements, channel consumption within each individual wireless service. Main findings include significant spectral and spatial correlations which then exploited to develop a 2D frequent pattern mining algorithm that can forecast channel accessibility based on earlier interpretations with significant precision. [14]

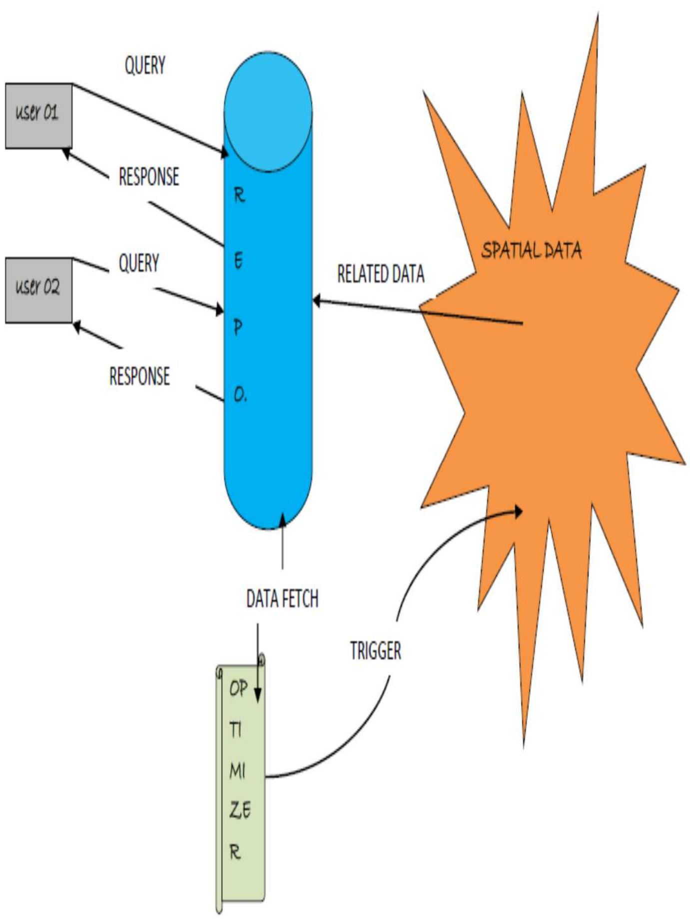
A research propose an innovative joinless tactic for effective colocation shape mining. The joinless colocation mining procedure uses an instance-lookup system instead of an expensive spatial or instance join operation for recognising colocation instances. [7]

OUR APPROACH FOR INTERACTIVE FRAMEWORK

The proposed protocol works as follows:

- User dials the query process held by repo phase of framework.
- Repo tells about the query to optimizer part of framework.
- The optimizer then picks appropriate time sliced algorithm for data fetch.

- The optimizer then fires the appropriate trigger to the spatial data set.
- The related data is then sent to repo set.
- The repo then gives response to the users.
- The repo then stores the data for future reference and the time in overall fetch is tracked by optimizer.



Pros

- user queries are being accommodated in interactive manner
- If the required query finds its association at repository then performance is supposed to be at effective side.

Cons

- The frame work has to maintain repository and optimizer programs.
- Whatever the fetch required the trigger is to be directly written at runtime.
- If repeated queries are referenced then overload will be high.

EXPERIMENTAL RESULTS

REPO

query	Response time
Q1	µ1
Q2	µ2
Q3	µ3
Q2	µ4
Q5	µ1
Q6	µ5
Q2	µ2

OPTIMIZER

query	Data triggers set	Time slices set	selection	To repo
Q1	Đ1 Đ2 Đ3 Đ4	µ1 µ2 µ3 µ4	µ1	Đ4,µ1
Q2	Đ1 Đ2 Đ3 Đ5	µ1 µ2 µ3 µ5	µ2	Đ5,µ2
Q3	Đ5 Đ2 Đ3 Đ6	µ5 µ2 µ3 µ6	µ3	Đ2,µ3
Q4	Đ7 Đ4 Đ5 Đ2	µ7 µ4 µ5 µ2	µ4	Đ4,µ4
Q5	Đ1 Đ2 Đ3	µ1 µ2 µ3	µ1	Đ4,µ1
Q6	Đ2 Đ5 Đ4	µ2 µ5 µ4	µ5	Đ4,µ5
Q7	Đ1 Đ2 Q2	µ1 µ2	µ1	Q2,µ1

VI. CONCLUSION

Given at the set of triggers (Đ1 Đ2 Đ3 Đ4 Đ5 Đ6..)Having time slices of (µ1 µ2 µ3 µ4 µ5 µ6…),if query Q1 may be targeted by set (Đ1 Đ2 Đ3 Đ4) with slices (µ1 µ2 µ3 µ4) selection µ1 is done and repo is tracked by Đ4,µ1.

Query Q2 may be targeted by set (Đ1 Đ2 Đ3 Đ5) with slices (µ1 µ2 µ3 µ5) selection µ2 is done and repo is tracked by Đ5,µ2.

Query Q3 may be targeted by set (Đ5 Đ2 Đ3 Đ6) with slices (µ5 µ2 µ3 µ6) selection µ3 is done and repo is tracked by Đ2,µ3.

Query Q4 may be targeted by set (Đ7 Đ4 Đ5 Đ2) with slices (µ7 µ4 µ5 µ2) selection µ4 is done and repo is tracked by Đ4,µ4.

Query Q5 may be targeted by set (Đ1 Đ2 Đ3) with slices (µ1 µ2 µ3) selection µ1 is done and repo is tracked by Đ4,µ1.

Query Q6 may be targeted by set (Đ2 Đ5 Đ4) with slices (µ2 µ5 µ4) selection µ2 is done and repo is tracked by Đ4,µ5.

Query Q7 may be targeted by set (Đ1 Đ2 Q2) with slices (µ1 µ2) selection µ1 is done and repo is tracked by Q2,µ1 on behalf of q2 repeat.

Thus we may conclude that $\mu4 < \mu2$.

VII. ACKNOWLEDGMENT

This work is supported by Shri Ramswaroop Memorial University (SRMU) Lucknow and Shri Ramswaroop College of Engineering and Management (SRMCEM) Lucknow.

REFERENCES:

[1] Wei Wang; Jiong Yang; Muntz, R., "An approach to active spatial data mining based on statistical information.", IEEE Transactions on Knowledge and Data Engineering ,Volume: 12 , Issue: 5, 715 – 728, 2000 .

[2] Cook, D.J.; Holder, L.B., "Graph-based data mining" , IEEE Transactions on Intelligent Systems and their Applications, Volume: 15, Issue: 2 , 32 - 41, 2000 .

[3] Hung-Chih Wu; Chan-Nan Lu, "A data mining approach for spatial modeling in small area load forecast", IEEE Transactions on Power Systems, Volume: 17, Issue: 2, 516 – 521, 2002 .

[4] Saygin, Y.; Reisman, A.; YunTong Wang, "Value of information gained from data mining in the context of information sharing.", IEEE Transactions on Engineering Management, Volume: 51, Issue: 4, 441 – 450, 2004 .

[5] Yan Huang; Shashi Shekhar; Hui Xiong, "Discovering colocation patterns from spatial data sets: a general approach", IEEE Transactions on Knowledge and Data Engineering, Volume: 16, Issue: 12, 1472 – 1485, 2004.

[6] Shuyu, Zhang; Thongying, Zhu, "Parallelism of spatial data mining based on autocorrelation decision tree", Journal of Systems Engineering and Electronics BIAI, Volume: 16, Issue: 4, 947 – 956, 2005.

[7] Jin Soung Yoo; Shashi Shekhar, "A Joinless Approach for Mining Spatial Colocation Patterns", IEEE Transactions on Knowledge and Data engineering, Volume: 18, Issue: 10, 1323 – 1337, 2006 .

[8] Weili Wu; Xiuzhen Cheng; Ding, M.; Kai Xing; Fang Liu; Ping Deng , "Localized Outlying and Boundary Data Detection in Sensor Networks", IEEE Transactions on Knowledge and Data Engineering, Volume: 19, Issue: 8, 1145 – 1157, 2007.

[9] Sidharta Gautama; Bellens, R.; De Tre, G.; Philips, W, "Relevance Criteria for Spatial Information Retrieval Using Error-Tolerant Graph Matching", IEEE Transactions on Geoscience and Remote Sensing, Volume: 45, Issue: 4, 810 – 817, 2007.

[10] Qin Ding ; Qiang Ding ; Perrizo, W., "PARM—An Efficient Algorithm to Mine Association Rules From Spatial Data", IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, Volume: 38, Issue: 6, 1513 – 1524, 2008.

[11] Adnan, M.; Alhajj, R., "A Bounded and Adaptive Memory-Based Approach to Mine Frequent Patterns From Very Large Databases", IEEE Transactions on Systems, Man, and Cybernetics, Part B: Volume: 41, Issue: 1, 154 – 172, 2011.

[12] Adrienko, N.; Adrienko, G., "Spatial Generalization and Aggregation of Massive Movement Data", IEEE Transactions on Visualization and Computer Graphics, Volume: 17, Issue: 2, 205 – 219, 2011 .

[13] D'Amore, F.; Cinnirella, S.; Pirrone, N., "ICT Methodologies and Spatial Data Infrastructure for Air Quality Information Management", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,Volume: 5, Issue: 6, 1761 – 1771, 2012.

[14] SiXing Yin; Dawei Chen; Qian Zhang; Mingyan Liu; ShuFang Li, "Mining Spectrum Usage Data: A Large-Scale Spectrum Measurement Study", IEEE Transactions on Mobile Computing, Volume: 11, Issue: 6, 1033 – 1046, 2012 .