International Journal of Multidisciplinary Research and Modern Education (IJMRME)
ISSN (Online): 2454 - 6119
(www.rdmodernresearch.org) Volume I, Issue I, 2015

AN INTERACTIVE TWO TIER FRAMEWORK FOR SPATIAL DATA MINING

Mohd Haseeb Ahmad*, Vinodini Katiyar & S. Q. Abbas*****
* Assistant Professor, Department of Computer Science and Engineering, Shri Ramswaroop Memorial College of Engineering and Management, Lucknow, India.
** Professor, Department of Computer Science and Engineering, Shri Ramswaroop Memorial University, Lucknow, India.
*** Professor & Director, Department of Computer Science and Engineering, Ambalika Institute of Management and Technology, Lucknow, India.

Abstract:
Though spatial data possess certain complex characteristics, still vast utilization of spatial data emphasizes on techniques to be developed to efficiently mine and produce excellence in presentation of it. Typically the spatial data spectrum tree retain into the memory of fetching program of the system but this sometimes result in an unpredictable behavior of the system, This problem is tackled in this paper by presenting an interactive two phased framework capable of mining spatial data spectrum tree as well as a simple database.

Index Terms: Spatial Data Mining, Spatial Database, Data Mining, Interactive Approach, Analyzer Shield, Analyzer Program & Repo Sets.

1.Introduction:

Spatial data bases contain large amount of data which may be space related, spectrum related or even people related. Because of size being too large to be handed by traditional data base system, mining becomes typical in this section. May be the case with a politician concerned about religious sections of a plane or a jet being concerned about the air coordinates or a farmer be concerned about the typographical description of a location ,the spatial data management method become the key for spatial data mining.

2. Related Work:

Spatial Data Mining:

A. The Definition of Spatial Data Mining:

Spatial data mining(SDM)means finding out wisdom from spatial database, it require data base technology together with appropriate abstraction of spatial data base, isolating the encouraging patterns and validating it with the specified frame work of the system.

B. The Characteristics of Spatial Data:

Spatial data are generally characterized with capaciousness, diversity, fluctuating attributes and heterogeneity. Following are the common noticeable differences between traditional and spatial data:

Feature	Traditional data	Spatial data
Size	Normal.	Very huge.
Definiteness	Definite.	Flexible (reference frame has a significant impact).
Influences	Does not have dependency on another data at all times.	Has a significant influence on another spatial object in adjoining location.
Heterogeneity	Not heterogeneous.	Heterogeneous or non-stationary as it may vary over distance and time etc.

Rapidness	Does not change rapidly.	Changes rapidly (as in the case of remote sensing).
Shortage	Termed inconsistent if it can't be acquired.	It may be lost or not acquired due to some constraints.

3. The Main Methods of Spatial Data Mining:

It is being approximated that approaches achieve three zones for developments, while approaches produce the same result without or with any delay.

i) Data Mining Based on Statistical Information

The various approaches used are as follows:

- Employing an ordered structure with associated geometric information at the different levels of the hierarchy and spoiling the user-defined trigger into a set of sub triggers. Updates are deferred in the ladder until their aggregate consequence might cause the trigger to fire.[1]
- Targeting the documentation of outlying sensors (that is, outlying interpretation sensors) and the detection of the reach of events in sensor networks. Typical applications include the detection of the transportation front line of some vegetation or animalcule's growth over a certain geographical region.[8]
- Using a Game-theoretic framework to suggest the fair value for information extracted via data mining and shared between two retail-market competitors.[2]
- On a collection of Boolean spatial features, using collocation pattern discovery process which finds the subsets of features frequently located together.[5]
- using error-tolerant graph matching to find correspondences between the detected image features and the geospatial vector data.[9]
- Proposing a well-organized tactic to derive association rules from spatial data using Peano Count Tree (P-tree) structure. P-tree structure offers a lossless and crushed representation of spatial statistics.[10]
- Offering a model skilled of mining a transactional list, irrespective of its magnitude and lacking memory manager.[11]

ii) Graph-Based Data Mining

The various approaches exclusively using this technique are as follows:

- Developing systems for determining perceptions in databases the collocation design detection process finds the subsets of features often situated organized. [2,5,7]
- Using error-tolerant graph alike to find communications between the distinguished image and the geospatial trajectory statistics.[9]
- Suggesting a process for spatial generalization and aggregation of association data, which renovates curves into aggregate streams between zones.[12]
- Offering a comprehensive spectrum measurement reading, with data collected in the 20 MHz to 3 GHz spectrum group and at four sites, scrutinizing the information of the poised data, including channel vacancy statistics. [4].

iii) Spatial Modeling in zonal sets

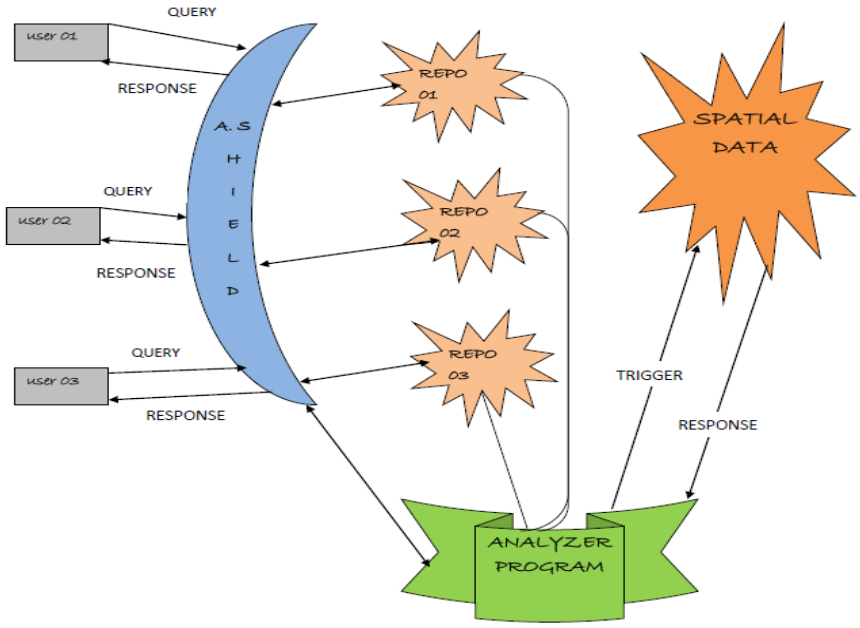
The various approaches exclusively using this technique are as follows:

- Recommending spatial modeling approach is an investigative data exploration, trying to notice useful patterns in spatial data that are not noticeable to the data user.[3]
- Proving a join less algorithm to be accurate and thorough in finding collocation rules. Labeling a partial join approach for spatial data which are grouped in neighborhood areas.[7]

4. Our Approach for Interactive Framework:

The proposed protocol works as follows:

- User dials the query process held by analyzer shield of framework.
- Analyzer shield tells about the query to analyzer program part of framework.
- The analyzer program then picks appropriate time sliced algorithm for data fetch.
- The analyzer program then fires the appropriate trigger to the spatial data set.
- The related data is then sent to appropriate repo set.
- The concerned repo then gives response to the analyzer shield.
- The analyzer shield then stores information about the data repo referred for future reference and the time in overall fetch is tracked by analyzer program.
- The analyzer shield then responds to the users.



Pros:

- Choice of user acceptance/many users do not degrade performance.
- If repeated queries are generated then performance will be high.
- Repo sets might use standard DBMS to get managed.

Cons:

- The framework has to come up with Analyzer shield and repo management.
- If size of repo sets increase then query would be needed.
- Refresh process will restart the algorithm at level 0.

5. Experimental Results:

Analyzer Shield:

Query	Response Set	Responding Repo
Q1	Ð4,µ1,N1	01

Q2	Ð5,µ2,N2	02
Q3	Ð2,µ3,N3	03
Q2	Ð5,µ2,N2/N2,Q2	02
Q4	Ð4,µ4,N4	03
Q5	Ð4,µ1,N1	01
Q6	Ð4,µ5,N5	02
Q2	Ð5,µ2,N2/N2,Q2(Fr)	02

Analyzer Program:

Query	Data Triggers Set	Time Slices Set	Selection	To Repo
Q1	Ð1 Ð2 Ð3 Ð4	µ1 µ2 µ3 µ4	n1	Ð4,µ1,n1
Q2	Ð1 Ð2 Ð3 Ð5	µ1 µ2 µ3 µ5	n2	Ð5,µ2,n2
Q3	Ð5 Ð2 Ð3 Ð6	µ5 µ2 µ3 µ6	n3	Ð2,µ3,n3
Q4	Ð7 Ð4 Ð5 Ð2	µ7 µ4 µ5 µ2	n4	Ð4,µ4,n4
Q5	Ð1 Ð2 Ð3	µ1 µ2 µ3	n1	Ð4,µ1,n1
Q6	Ð2 Ð5 Ð4	µ2 µ5 µ4	n5	Ð4,µ5,n5
Q7	Ð1 Ð2 Q2	µ1 µ2	n1	Q2,µ1,n1

6. Conclusion:

Given at the set of triggers (Ð1 Ð2 Ð3 Ð4 Ð5 Ð6...) Having time slices of (µ1 µ2 µ3 µ4 µ5 µ6…), if query Q1 may be targeted by set (Ð1 Ð2 Ð3 Ð4) with slices (µ1 µ2 µ3 µ4) selection n1 is made and Analyzer shield is mirrored by Ð4, µ1,n1 and repo 01.

Query Q2 may be targeted by set (Ð1 Ð2 Ð3 Ð5) with slices (µ1 µ2 µ3 µ5) selection n2 is made and Analyzer shield is mirrored by Ð5, µ2, n2 and repo 02.

Query Q3 may be targeted by set (Ð5 Ð2 Ð3 Ð6) with slices (µ5 µ2 µ3 µ6) selection n3 is made and Analyzer shield is mirrored by Ð2, µ3, n3 and repo 03.

Query Q4 may be targeted by set (Ð7 Ð4 Ð5 Ð2) with slices (µ7 µ4 µ5 µ2) selection n4 is made and Analyzer shield is mirrored by Ð4, µ4, n4 and repo 03.

Query Q5 may be targeted by set (Ð1 Ð2 Ð3) with slices (µ1 µ2 µ3) selection n1 is made and Analyzer shield is mirrored by Ð4, µ1, n1 and repo 01.

Query Q6 may be targeted by set (Ð2 Ð5 Ð4) with slices (µ2 µ5 µ4) selection n5 is made and Analyzer shield is mirrored by Ð4, µ5, n5 and repo 02.

Query Q7 may be targeted by set (Ð1 Ð2 Q2) with slices (µ1 µ2) selection µ1 is done keeping tracked by Q2,µ1 on behalf of q2 repeat and selection n1 is made and Analyzer shield is mirrored by q2,µ1,n1 and repo 02 noting that this query is frequent(Fr). Thus we may conclude that next two references to Q2 will have shorter response time.

7. Acknowledgment: This work is supported by Shri Ramswaroop Memorial University (SRMU) Lucknow, India and Shri Ramswaroop College of Engineering and Management (SRMCEM) Lucknow, India**.**

8. 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. Vinodini Katiyar; Mohd Haseeb Ahmad S. Q. Abbas "Novel Approach for Interactive Spatial Data Mining", MATRIX Academic International Online Journal of Engineering and Technology, ISSN: 2348–3326 © MAIOJET Volume III Issue I, April 2015.
5. Yan Huang; Shashi Shekhar; HuiXiong, "Discovering collocation 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.