ISSN (ONLINE): 2395-695X
ISSN (PRINT): 2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Volume 2, Special Issue 19, October 2016

FOCAL AND LOCAL TEST BASED SPATIAL DECISION TREE

S.Chandrakumari[1] and P.Narendran[2]

[1]*Research Scholar, Department of Computer Science, Bharathiar University, Gobi Arts and Science College, Gobi, India.*

[2]*Head of the Department, Department of Computer Science, Bharathiar University, Gobi Arts and Science College, Gobi, India.*

[1]Chandra.psg.10@gmail.com,[2]narendranp@gmail.com

Abstract— **The Group Nearest Neighbor (GNN) question uses cluster functions to supply the most effective account for the highest cluster purpose among datasets. The novel type of abstraction keyword question noted as Group Nearest Group (GNG) question area unit getting to optimize the question. Associate information purpose set D, an query purpose set character associated degree and associate number k, the Group Nearest Group question finds a collection of purposes from the desired overall distance from all functions in character to the highest point in ω, isn't any larger than the opposite set of points in D. Each nearest purpose obtained matches minimum of one all question keywords. For processing this question several algorithms area units are projected. The method of GNG question consists of Complete Stratified Combination formula and Set Stratified Refinement formula. Group Nearest Neighbor (GNN) question returns the situation of a gathering place that minimizes the mixture distance from ramification out cluster of users. The duplicates among information set is thought to spice up the search question from the given knowledge. The knowledge set is analyzed for looking out the duplicates among knowledge set. The applications of cluster question come from location-based services.**

Keywords— **GNN, GNG, RNN, Hybrid.**

I.INTRODUCTION

Spatial data mining, i.e., mining data from massive amounts of spatial information, could be a hard field since vast amounts of spatial information are collected in numerous applications, starting from re mote sensing to Geographical Information Systems (GIS), PC devising, environmental assessment designing. The collected information so much exceed people's ability to investigate it. Thus, new and economical strategies are needed to measure the required to find data from massive spatial databases. A spatial association rule describes the implication of one or a group of options by another set of options in spatial databases. A spatial classification method could be a method that assigns a group of spatial objects into variety of given categories supported a group of spatial and non-spatial options.

II. EXISTING SYSTEM

Most ancient spatial queries on spatial databases like nearest neighbor queries, vary queries using CLARANS (**C**lustering **L**arge **A**pplications based upon **RAN**domized **S**earch) of GNG ends up in gap of few proportion points lost. The present system, takes long question time interval and information accuracy issues were known.

In nearest neighbor queries, an optimization drawback is evaluated for locating the highest points in metric areas .Given a set S of points in a metric space M and a query point q ∈ M, finding the highest purpose in S to letter. The informal observation is sometimes noted, because the curse of spatiality states that there's no general purpose precise answer error for NNS in high-dimensional metric Euclidean space using polynomial pre process and poly power search time. This system is unable to look at the placement of the spot in spatial information once new website is further.

In follow, using local search heuristics for GNG question ends up in some proportion points between the obtained answer and therefore the world optimum. Within the worst case, the local search heuristics are tested to realize at the most five times of the world optimum. The present system reduced the cluster quality.

A.Demerits

- It has the classification errors.
- Existing system not provide the largest information gain in one tree node test.
- Thus if all the candidate tests have poor spatial autocorrelation, this type of decision tree will still select one of them.
- Less Accuracy

III. PROPOSED SYSTEM

The proposed system uses two algorithms. They are FTSDT algorithm and Subset Hierarchical Refinement (SHR) algorithm. Use hierarchical blocks instead of data points to optimize the number of subsets evaluated. This technique aims at minimizing the I/O accesses to the object and feature data sets.

A. Merits

- Optimized version provides more efficient technique for computing the scores of the objects. It develops solutions for the spatial preference query based on the temporal data.

- It minimizes the access and reduces search space. In this work, database techniques are explored to boost the GNG query processing of local search heuristics without any loss on clustering quality.
- To refine the solution, the search space in lower hierarchical level is minimized. In FTSDT, every set of blocks is evaluated in high hierarchical level and the set with the current best value (i.e., the minimum total distance) are refined by visiting their children in next level FTSDT is capable to provide the optimal solution.

B.System Overview

The spatial object p is a pair in the form (p.l, p.t), where p.list a location descriptor in the multidimensional space, and p.t is the textual description represented by sets of keywords. Let D be the universe of all objects in the database. Given a group of query points Q = {q1, q2, . . .qn} and a set of m query keywords Q_w= {w1, w2, . . . , wm}. A top keyword query retrieves query points from Q with the minimum sum of distances ∀ q ,Q, the nearest keyword w of q is a point p_i (p_i ,D) which contains keyword w such that ∀ pj ,D/p_i and w p_j.t and dist(pi, q) ≤ dist (p_j, q).The function dist(q, p_i) is the Euclidean distance between q and p_i. The function near key (q, w) present the distance between q and its nearest keyword w. Then the summed distance of q is defined as $\sum_{i=1}^{m}$ near key (q,w_i), where w_i,Q_w. The GNG query returns the nearest query points in Q with the minimum summed distance. Here each query point in Q only contains the spatial information.

Given a set of data points D which contains keyword information, a group of query objects Q and m query keywords, a GNG query retrieves objects in Q with the minimum sum of distances to its nearest points in D such that each nearest point matches at least one of query keywords. It can be widely utilized in various decision support systems and multiple domains like service recommendation, investment planning, etc. For example, consider a spatial database D which manages facilities such as schools, restaurants and hospitals, represented by sets of keywords.

A user wants to GNG the locations with respect to the sum of distances to nearest interested facilities. The user may issue a set of locations and multiple query keywords representing his/her interested facilities, the result returns best locations that minimize the summed distance to these facilities.

V. MODULE DESCRIPTION

The project is divided into three modules.

A. Data Group

B. Group Combination

C. Subset Refinement

A. Data Group

A real data set of points are collected which consists of the place with the longitude and latitude of the metropolitan city. The synthetic data points were obtained containing the uniformly distributed points around the city. These data sets are unified into a unit region. Q is distributed in an area whose Minimum Bound Rectangle is a percentage of the whole data space, denoted as M. All the data sets are indexed by R-trees for FTSDT and SHR.

B. Group Combination

FTSDT algorithm minimizes the access and evaluation of potential subsets. The data points in FTSDT are hierarchically represented by data blocks, e.g., using R-tree. The algorithm process GNG query by treating the blocks as points to find an intermediate solution in higher hierarchical level. To refine the solution, the search space in lower hierarchical level is minimized by following the guided search direction.

C. Subset Refinement

Subset Hierarchical Algorithm is a local search heuristic with support of the database techniques. In higher hierarchical level, each block is treated as a point by SHR to replace every element in the subset, and the resultant subset with the current best value is refined by visiting the children of the block. The solution of SHR is usually close to the global optimum and guaranteed to be within a factor of at most close to the global optimum.

D. FTSDT Classifier From Training Samples

The algorithm retrieves the query result by computing the summed distance of every query point in Q. Initially, the data's are fetched from the database. In the front end, the data (nearest features) corresponding to the input query object is fetched from the database. Next, the distance calculation takes place for the interested neighbors of the selected data. The minimum distance of the interested neighbor with respect to the input object is obtained. Then the summed distance of the neighbors are calculated, which is done by the sum of distances of the three nearest neighbors. It is given by,

$$\sum (q1) = q1.a + q1.b + q1.c$$
$$\sum (q2) = q2.a + q2.b + q2.c$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\sum (qn) = qn.a + qn.b + qn.c$$

Where,

$\sum (q1)$ = summed distance of the input object q1.
$\sum (q2)$ = summed distance of the input object q2.
$\sum (qn)$ = summed distance of the input object qn.
q.a = distance of 1st nearest feature of q.
q.b = distance of 2nd nearest feature of q.
q.c = distance of 3rd nearest feature of q.

E. Subset Hierarchical Refinement Algorithm

The algorithmic program computes the boundary of the summed distance,that considerably reduces the amount of question objects and therefore the data points to be examined. The primary step involves the method of clustering, i.e. grouping of similar data objects. The clustered data is developed to create a tree referred to as hierarchical tree that is then followed by fetching of data in the database. Within the forepart, supported the index, the information (nearest features) resembling the input question object is fetched from

the information. Next, the distance calculation takes place for the interested neighbors of the chosen knowledge. The distance between every interested neighbor with relevance to the input object is obtained. Then the summed distance of the neighbors are calculated that is completed by the sum of distances of the three nearest neighbors.

The summed knowledge is then sorted so on show the results of the item within the ascending order. The map overlay are often obtained with the input object, premeditated to its nearest key options. The directed line within the map links the item to its key options, that is displayed in a very little parallelogram of differentiated colours.

Algorithm SHA (Dataset D, Query set Q, Integer K)

1. begin
2. ω_{cur}= find ω_{ini}
3. γ =Compute sum based on ω_{ini}
4. N= root of R-tree on D
5. H=∅
6. for each entry E in N
7. for each p$\epsilon\omega_{cur}$
8. compute sum when E replaces p $\epsilon\omega_{ini}$
9. if sum < γ , H ←{sum,p,E}
10. if H=∅ , SHR terminates by returning ω_{ini}
11. removeh ϵ H where h.$\sum_{lb}^{\omega}$ is minimum sum
12. p=h.p
13. N=Node referred by h.E
14. While N is a non-leaf node
15. for each entry E in N
16. Compute sum when E replaces p$\epsilon\omega_{cur}$
17. if sum < γ , H ←{sum,p,E}
18. if H=∅ , returnω_{cur}
19. removeh ϵ H where h has the minimum sum
20. p=h.p
21. N=Node referred by h.E
22. ω_{cur}= replacep$\epsilon\omega_{cur}$ by N
23. γ=h.sumgoto line 3
24. end

VI. COMPARISION SCREENSHOTS

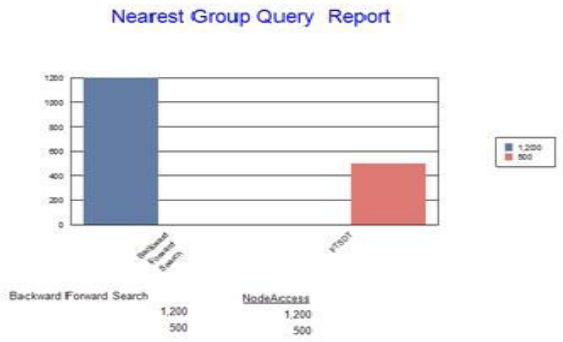

Fig. 1 Node Access Report

A. Node Comparison

In the Proposed System, Node Access can actually compare up to 1200 nodes. Then It can Cover the full data up to 1200 but the existing method could cover the data maximum of 500 Nodes .So the Proposed System can give the exact results from database.

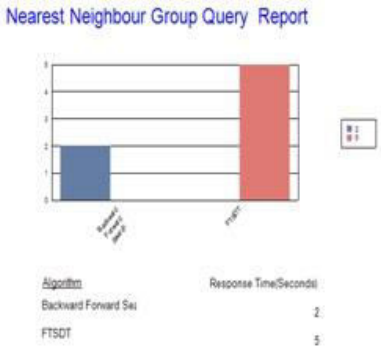

Fig. 2 Response Time Report

B. Time Comparison

In the existing system time taken for the data search from database is up to 5 seconds but the proposed system time taken reduced to 2 seconds. During search of large databases this proposed system reduces the time consumption.

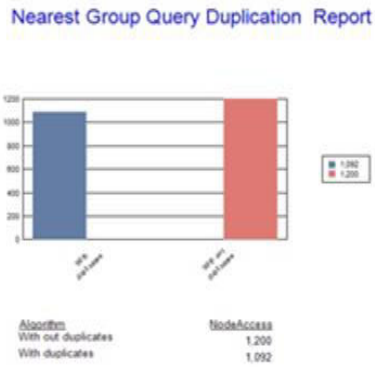

Fig. 3 Node Access Report for Duplicates Search Operation

C. Data Duplication

In the existing system the data duplication can't be found from 1092 nodes but the proposed system data duplication is not allowed in 1200 nodes. During search of large databases this proposed system reduces the data duplication.

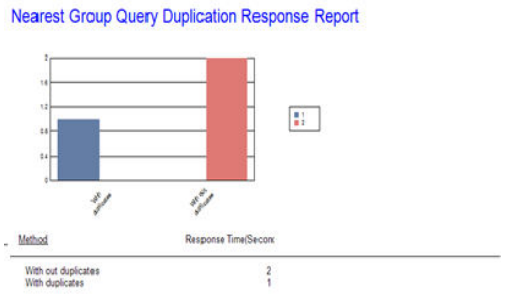

Fig. 4 Response Time Report For Duplicates Search Operation

VII. CONCLUSION

The Group Nearest Group question retrieves different objects from question keyword Q character with minimum total of distances to its nearest information points, Complete Stratified combination and Set Stratified Refinement rule, prunes the question objects and eventually the diminished summed distance is calculated. The amount of node accesses is in addition that reduces the time period interval,that exhibits sensible quality with the question objects and additionally the variability of question keywords.

REFERENCES

[1] V. Arya, N. Gary, R. Khandekar, A. Mayerson, K. Munagala and V. Pandit, " Local Search Heuristics for k-Median and Facility Location Problems", Proceedings 33rd ACM Symposium on Theory of Computing, 2001.

[2] Baihua Zheng, Jianliang Xu, Wang-Chien Lee "Data management in location dependent information Services", IEEE Pervasive Computing, Vol. 1, PP. 65–72, 2002.

[3] C. Bohm, S. Berchtold and D. Keim, "Searching in High Dimensional Spaces Index Structures for
Improving the Performance of Multimedia Databases", ACM Computing Surveys, Vol.33, PP. 322-373, May 2001.

[4] K. Cheung and A.W.C. Fu, "Enhanced Nearest Neighbor Search on the R-Tree", ACM SIGMOD Record, Vol.27, PP. 16-21, 1998.

[5] A. Civilis, C.S. Jensen and S. Pakalnis, "Techniques for efficient road-network-based tracking of moving objects" , IEEE Transactions on Knowledge and Data Engineering, Vol. 17, PP. 698–712, Feb 2005.

[6] K. Deng, H. Xu, S. Sadiq, Y. Lu, G. Fung, and H. Shen, "Processing Group Nearest Group Query",Proceedings 25th IEEE International Conference on Data Engineering, Apr 2009.

[7] K. Deng, X. Zhou and H. Shen, "Multi-Source Skyline Query Processing in Road Networks", Proceedings 23rd IEEE International Conference on Data Engineering Mar 2007.

[8] G. Hjaltason and H. Samet, "Distance Browsing in Spatial Databases", ACM Transactions on Database Systems, Vol. 24, PP. 265-318, 1999.

[9] K.Mouratidis,D. Papadias, and S. Papadimitriou, "Tree-Based Partition Querying: A Methodology for Computing Medoids in Large Spatial Datasets", The Very Large Database Journal, Vol. 17, PP. 923- 945, 2008.

[10] R. Ng and J. Han, "Efficient and Effective Clustering Method for Spatial Data Mining", Proceedings 20th Very Large Data Bases Conference, 1994.

[11] D. Papadias, Y. Tao, K. Mourstidis and C.K. Hui, "Aggregate Nearest Neighbor Queries in Spatial Databases ", ACM Transactions on Database Systems, Vol. 30, PP. 529-576, 2005.

[12] K.E. Rosing, "An Empirical Investigation of the Effectiveness of a vertex substitution Heuristic", Environment and Planning B: planning and design, Vol. 24, PP. 59- 67, Jun 1997.

[13] M. Yiu, N. Manoulis, and D. Papadias, "Aggregate Nearest Neighbor Queries in Road Networks", IEEE Transactions on Knowledge and Data Engineering, Vol. 17, PP. 820-833, Mar 2005.