Volume 3, Issue 9, September 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Optimizing Range Queries for Handheld Devices

Syed Nazim - Uz- Zama*, R.V.Krishnaiah
Department of CSE & JNTUH
India

Abstract - With the emergence of technologies in mobile computing domain, people are able to access Internet and related services through small hand held devices such as mobile phones, smart phones and PDAs. Mobile devices are resource constraint as they have less memory and processing power. When such devices are used to make range queries on spatial database, the results retrieved by the query are huge number of tuples that can't be easily presented by hand held devices. This is because they have less communication bandwidth and energy. They can't show the results as it is. From this perspective, this paper presents a new way of representing a specified size of results that causes very less information loss. The result of concise range query reduces the cost of communication and also enhances usability of end users. The concise query results are compared with other normal query results. However, finding optimal solution with less information loss is not an easy problem. To achieve this many algorithms were introduced. Experimental results revealed that the proposed system is efficient.

Index Terms – range queries, algorithms, spatial database

I. Introduction

Spatial databases have become popular these days. They have plenty of applications and user can perform range queries on spatial database. The usage of mobile application from which range queries are made have become common practice. However, the range queries result in huge number of records from spatial databases. Such results can't be shown easily by mobile devices as they are resource constraint. For instance they are having less bandwidth, less processing power and less memory. Provided this fact, presenting the results of such queries in mobile devices is a challenging task. As representative skyline points [1], the work done in this paper has same motivation that is to present concise results of range queries. These requirements are not particular to spatial databases only. The same challenges are encountered by huge databases and also dataware houses which are named as OLAP (Online Analytical Processing) systems. In the recent work which is shown in [2], approximate scalable query processing is an essential point. The goal of this is to bestow usable and light representsion of the query results for each and every query being processed. Particularly in mobile devices presenting such huge results is a hard nut to crack. The work done in this area [2] is nothing but random sample of query results. From random sampling, the aim of this paper is different which is meant for approximate representation of query results finally. For this reason the work done in this paper differs random sampling. When the query results are huge in number, the mobile devices can't present them. For this reason light means the representation of query results in small size for many reasons. The reasons include - 1) c/s bandwidth is limited. For embded systems and mobile devices they help that holds true. It is similar to the applications running in the real world.

Provided the case of present scenarios where users need more alternatives, the response time is the critical factors for Google Map vs. Mapquest are best as long response time has negative impact on that. The manipulation of results being rendered in client is essential as the query results are very huge. The results are to be made concise and then presented in mobile devices. Moreover the devices that act as clients may be computationally limited with less memory resource, the results large queries make it difficult if not impossible. When huge number of records are returned by a query, it becomes very difficult in the long run. Fro mobile computing devices and embedded systems it holds true. And when the query result is very large, the I/O burden on server is high. The database community tried to devote lot of time and money to ensure that the structures generated by the query process. The experimental results revealed a fact that the designing of all index structures are used to speed the query processing. The database community also tried to execute space and size. The result of size imposes a lower bound on the dataset. When small representation of whole dataset is returned, it is possible to reduce processing cost on server and get lower bound. Out of the two problems, simple compression techniques can be used. The query results must be usable which does mean that the results should be useful and user friendly in terms of navigation. Presenting lot of information is not usability. Instead it spoils the spirit of such case. Consider a real world example where user given query to GPS device to find spatial data. This query results in huge number of records and the people realize the fact that they are not intended results. This is because, when a subset of records are required, the query results in huge number causes the end user to wait or go through several reulst for knowing his / her results. The results of interactiveness refers to a capability to let other users to give feedback. Very often it is important that users with some idea for large region first provides valuable information to reduce the query to specific regions. This paper presents mechanisms that can be used to calculative attributes to refine further query. The rest of the paper is organized into many sections including related work.

II. Related Work

Recently the query results when they are bulky in case of range queries they are similar to the work presented in [2]. The problem in this paper is also similar to [2] where approximate results are returned by using long-running join queries in relational databases. The approximation is defined and there will be final results. Range queries are having very good geometric characteristics with regard to spatial databases. For this reason, instead of random sampling, it is essential to have very effective model. Therefore the aim of this project is to obtain concise range queries which return query results that are very concise in nature and thereby suitable for presenting in small hand held devices. This has to be done with less information loss. Multidimensional indexing structures are also used for query answering process. However, the ideas presented in this paper are different from that of [2]. Motivation for the work is partly taken from the work of [1] which focuses on skyline queries. Instead of skyline queries we use range queries. To perform the conciseness, many algorithms are considered including [3] and [4]. Especially k-means with variance is used for concise representation of query results. For efficiency reasons R-tree is used. Afterwards, additional operations such as memory heuristics are applied. The work done in this paper is based on a different clustering algorithm altogether. Moreover, an adaptive strategy is designed to make use of R-tree in an efficient way. Another kind of clustering algorithms are studied which are density based. They are namely DBSCAN [5] and CLARANS [6]. Disk based data sets where prepared by extending the work of classic density-based clustering and other variants. However, this paper focuses on an interesting problem known as concise range queries on moving objects [7] and road networks [8].

For privacy preserving, heuristic local recording methods are proposed by [9] which also considers the utility of attributes. The proposed framework in this paper is similar to the IGroup. K-anonymity is the solution [10], and [9] that is different from ours. Summarization techniques are also having relevance with our solution. They are discussed in [11] and [12]. V – optimal histogram in [11] and MinSkew in [13] are in multidimensional spatial data. This paper is aimed at minimizing the loss of information while making concise queries. Constructing histogram and minimizing the information loss are worked out in [12] with acceptable error bound. Our problem is quite opposite to this. Random sampling is another solution studied. Hoever, it is too general and not suitable for this paper. In [14] the reason for using concise range queires has been discussed while more elaboration is given in [14].

III. Problem Description

This paper introduces concise range queries that provide light representation of results which can be presented in resource constrained mobile devices. A point set is represented which is a collection of bounding boxes and corresponding counts. R represents concise query, P represents partitioning. We return R as concise representation of given point set to end user while the P is used by DBMS for calculating the value of R internally.

IV. Algorithms

HGroup Algorithm

This algorithm is meant for computing Hilbert value for each point. The sort P and map it to one dimension. Use dynamic programming to find the partitioning. Build concise representation R for given partitioning and return the result.

IGroup Algorithm

It is an interactive algorithm that finds k groups one at a time. When this algorithm is invoked, for each iteration, a seed value is taken randomly and add points greedily into group one after another. Estiamted total information loss is computed as

$$\tilde{L}(P_i) = \rho(P_i)|P_i| + 2\sqrt{A(U)/(k-i)} \cdot |U|.$$

The algorithm for IGroup functionality is as shown in listing 1.

```
U ← P;
S ← number of seeds to try;
For i=1,……, K – 1 do
L_best = ∞;
U' ← U;
FOR J=1,…….,s do
U ← U' ;
p_s ← randomly chosen seed from U;
P_i' ← { P_s};
U ← U - { P_s};
While true do
```

Let p= arg $\min_p$ Ĺ (P_i’ U{P});
If Ĺ (P_i’ U{P}) < Ĺ (P_i’) then
P_i’ ← P_i’ U {P};
U ← U – {P};
else break;
If Ĺ (P_i’) < $Ĺ_{best}$ then
$Ĺ_{best}$ ← Ĺ (P_i’);
Pi ← P_i’;
U ← U - P_i’;
Output P_i;

Listing 1 – Algorithm for IGroup

To use range queries with budget k from the client, the databse server evaluates query as if it is a range query by using spatial index which is based on the point set P resulting in an R-tree.

V. Experimental Results

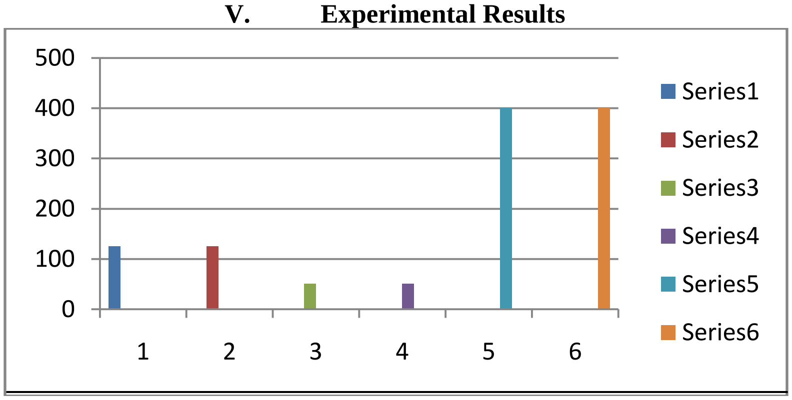

Fig 1 (A) I/O Cost

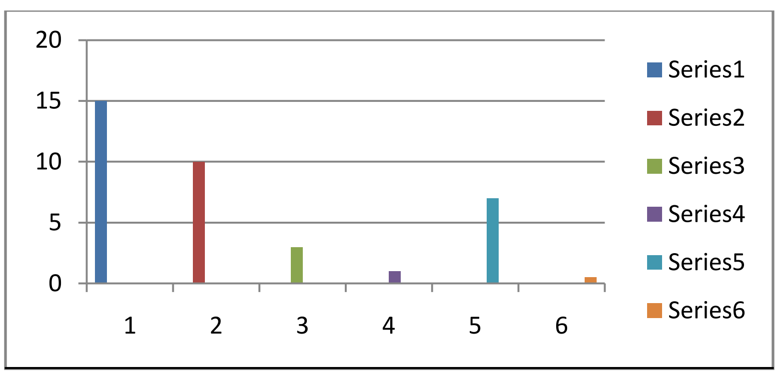

Fig 1 (B) CPU Time

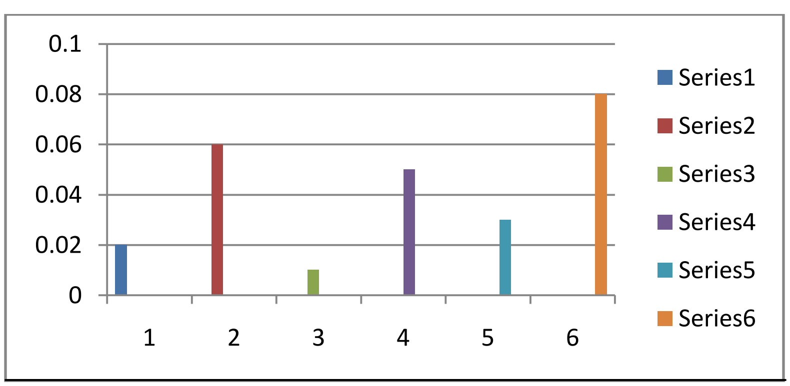

Fig 1 (C) Information Loss

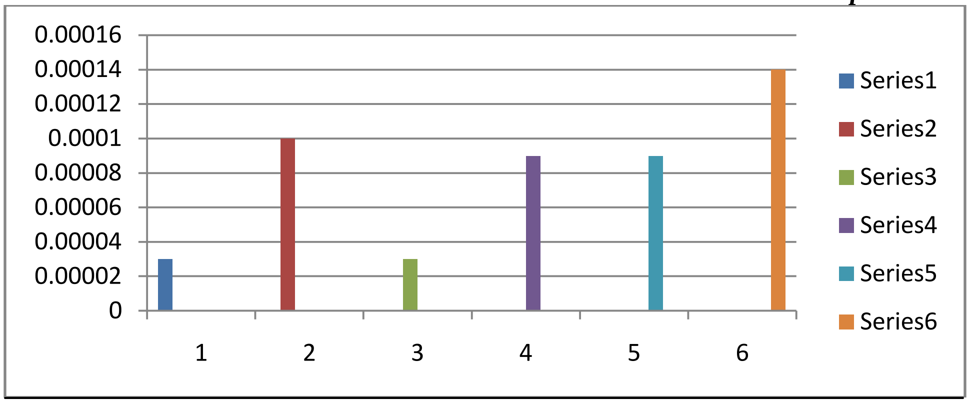

Fig 2 (A)

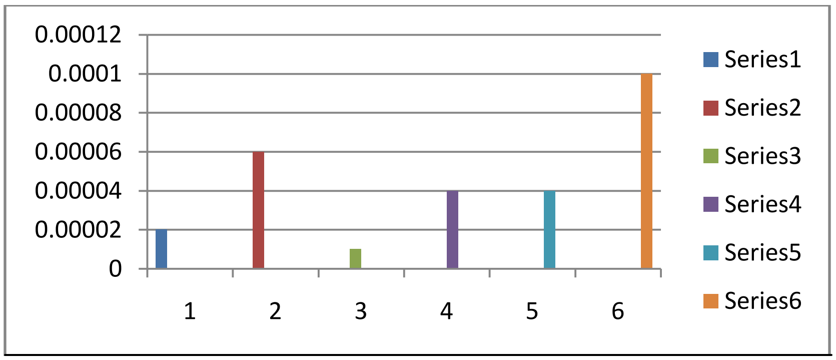

Fig 2 (B)

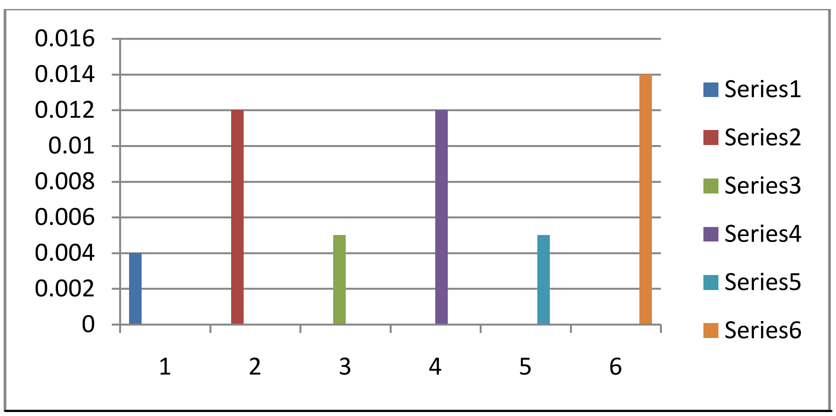

Fig 2(C)

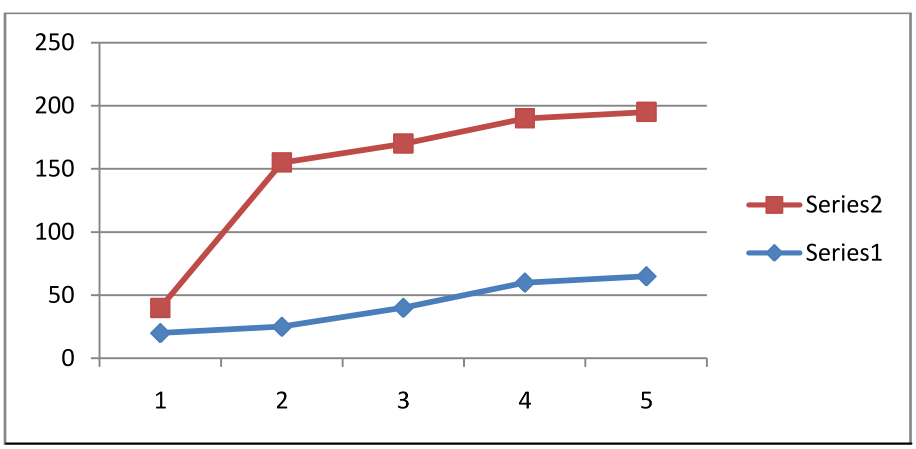

Fig 3 (A)

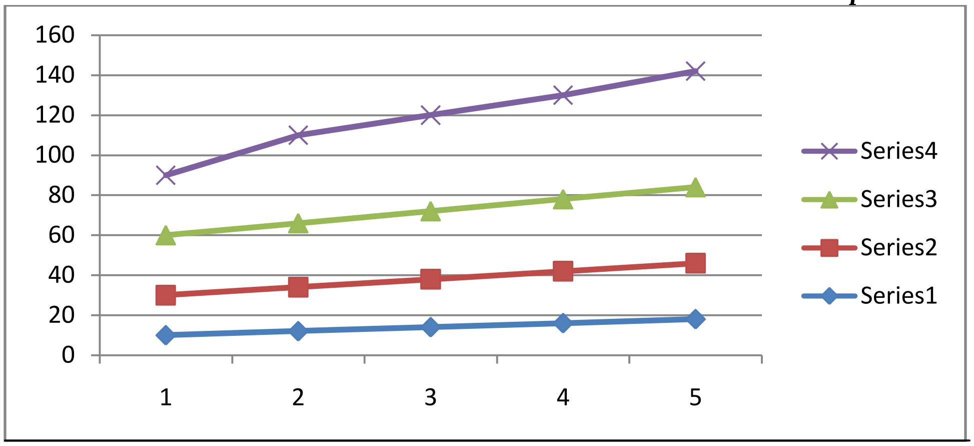

Fig 3 (B)

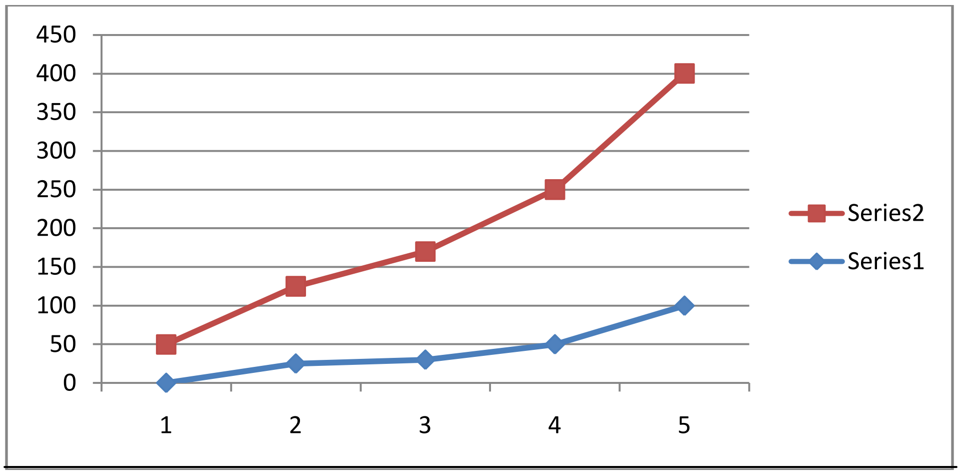

Fig 4 (A)

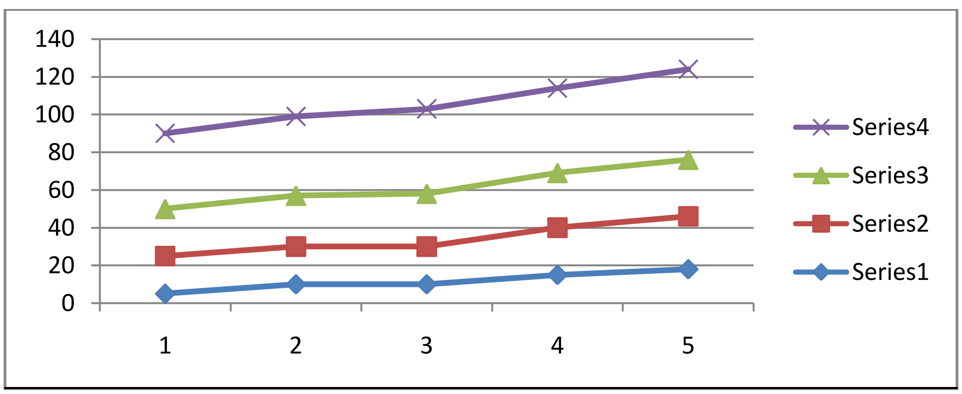

Fig 4 (B)

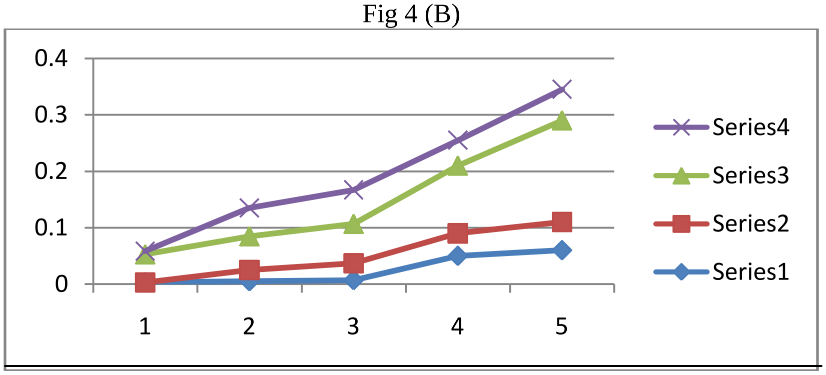

Fig 4 (C)

VI. Conclusion

A new approach in query processing known as concise range query is proposed in this paper which overcomes the three problems of conventional queries. The query results are reduced by the proposed system. The range query which has been made concise form, which saves bandwidth and saves memory usage and computational resources that are very important to hand held devices. The query size gets reduced in the proposed system. However, the concise form of query results is more user friendly and intuitive in nature with respect to spatial database. Then this paper designed T-tree based algorithms for make concise range queries. This and the underlying techniques presented here result in rich user experience. We also developed a prototype application which demonstrates the advantages of concise range queries. The experimental results revealed that the proposed algorithms work fine and capable of providing .

References

[1] X. Lin, Y. Yuan, Q. Zhang, and Y. Zhang, "Selecting Stars: The k Most Representative Skyline Operator," Proc. Int'l Conf. Data Eng. (ICDE), 2007.

[2] C. Jermaine, S. Arumugam, A. Pol, and A. Dobra, "Scalable Approximate Query Processing with the dbo Engine," Proc. ACM SIGMOD, 2007. YI ET AL.: THE WORLD IN A NUTSHELL: CONCISE RANGE QUERIES 153

[3] T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH: An Efficient Data Clustering Method for Very Large Databases," Proc. ACM SIGMOD, 1996.

[4] V. Ganti, R. Ramakrishnan, J. Gehrke, and A. Powell, "Clustering Large Datasets in Arbitrary Metric Spaces," Proc. Int'l Conf. Data Eng. (ICDE), 1999.

[5] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," Proc. Int'l Conf. Knowledge Discovery and Data Mining (KDD), 1996.

[6] R.T. Ng and J. Han, "Efficient and Effective Clustering Methods for Spatial Data Mining," Proc. Int'l Conf. Very Large Data Bases (VLDB), 1994.

[7] C.S. Jensen, D. Lin, B.C. Ooi, and R. Zhang, "Effective Density Queries on Continuously Moving Objects," Proc. Int'l Conf. Data Eng. (ICDE), 2006.

[8] M.L. Yiu and N. Mamoulis, "Clustering Objects on a Spatial Network," Proc. ACM SIGMOD, 2004.

[9] J. Xu, W. Wang, J. Pei, X. Wang, B. Shi, and A.W.-C. Fu, "Utility- Based Anonymization Using Local Recoding," Proc. ACM SIGKDD, 2006.

[10] G. Aggarwal, T. Feder, K. Kenthapadi, S. Khuller, R. Panigrahy, D. Thomas, and A. Zhu, "Achieving Anonymity via Clustering," Proc. Symp. Principles of Database Systems (PODS), 2006.

[11] Y.E. Ioannidis and V. Poosala, "Balancing Histogram Optimalityand Practicality for Query Result Size Estimation," Proc. ACMSIGMOD, 1995.

[12] A.D. Sarma, O. Benjelloun, A. Halevy, and J. Widom, "Working Models for Uncertain Data," Proc. Int'l Conf. Data Eng. (ICDE), 2006.

[13] S. Acharya, V. Poosala, and S. Ramaswamy, "Selectivity Estimation in Spatial Databases," Proc. ACM SIGMOD, 1999.

[14] K. Yi, X. Lian, F. Li, and L. Chen, "The World in a Nutshell: Concise Range Queries," Proc. Int'l Conf. Data Eng. (ICDE), 2009.

AUTHORS

Mr. Syed Nazim-uz-zama, M.Tech from DRK Institute of Science and Technology. His main research interest includes Data Mining, Software Engineering.

Dr.R.V.Krishnaiah (Ph.D) is working as Principal at DRK INSTITUTE OF SCINCE & TECHNOLOGY, Hyderabad, AP, INDIA. He has received M.Tech Degree EIE and CSE. His main research interest includes Data Mining, Software Engineering.