International Journal of Information & Computation Technology.
ISSN 0974-2239 Volume 4, Number 1 (2014), pp. 55-61

http://www. irphouse.com

Processing Queries on Road Networks in Spatial Data Base Perspective for Selectivity Estimation

K. Bhanu Prakash* and N. Bala Krishna

Computer Science and Engineering,
Madanapalle Institute of Technology & Science,
Jawaharlal Nehru Technological University,
bargav.prakash@gmail.com, balu1203@gmail.com

ABSTRACT

This work mainly focuses on building a framework that is capable of analyzing spatial approximate substring queries, for mainly to solve the selectivity estimation problem of range queries which belongs to road networks represented in spatial databases. The selectivity estimation is nothing but estimating the size of the results i.e., estimating the number of points that presents in a graph which is nothing but result, as per user request. Here for that we propose an algorithm which is nothing but the Selectivity Estimation Algorithm. This is very useful for the effective processing of queries by the query optimizer. This algorithm generally uses the frame work for establishing the spatial network and there by using the spatial graphs for estimating the size of those graphs. The Selectivity Estimation algorithm uses range queries and based on that the result will generate, after that estimation of results takes place by building the selectivity estimator.

Keywords— Selectivity Estimation, range queries, spatial approximate string, road networks, spatial databases, MHR tree, spatial graph, min-wise signatures.

I. INTRODUCTION

The typical databases support integer type of data. To support geometric type of data like polygons and so on, databases of spatial type are introduced. Here such type of structures may leads to support 3D objects which may provide storing objects of geometric space type in databases. The spatial databases [1] also support objects those are nothing but geometric objects such as polygons, points and lines. Spatial database is the database which was designed in an optimized manner for the purpose of storing and querying the data that represents objects defined in geometric space.

In generally the system was implemented such that even though an approximate string [2] was entered by the user then he can able to get exact results without the need of exact query. There by the efficiency of query matching has been improved even though misspell of words has taken part such that the user can get exact results for his respective query. The networks which are of spatial type are generally represented in type of graphs, such that the road junctions are designed as vertices/nodes which are crossroads and the connections which are used to connect those junctions are represented as graph edges. Those graphs which were treated as spatial graph may be directed, weighted or undirected. Based on this procedure spatial query can able to represent on the actual network for executing on the graph G specified as either Directed/Un-Directed. There are so many processing techniques that have been pro-posed especially for queries which are of fundamental type in spatial networks. Such types of techniques are applicable especially for range queries. For implementing the process of combining those queries the techniques followed are called appropriate query optimization techniques which are used for increasing the efficiency. Therefore, the process of estimation generally affects the query performances which are very important for optimization of queries. Generally such type of data can be able to retrieve from the spatial database. Therefore the region of relative interest was denoted by starting vertex v_0 and the distance can be measured on network-base.

There are several methods that are developed to examine the Selectivity Estimation with respect to spatial or spatiotemporal queries [3] over spatial networks considering non-Euclidean spaces. 1) Multi Dimensional Scaling method: The MDS consists group of tools which can able to transform data objects to Euclidean space from a non-Euclidean space, allowing effective visualization. MDS [3] can also able to generate representations of data objects in a space of low-dimensional, which can be desirable frequently. The main advantage of this MDS method is it can able to utilize previous results for the selectivity estimation of multi-dimensional objects based on Euclidean distance. 2) Global Parameters Estimation [3] method: this method was mainly dependent on global parameters. Global Parameters Estimation method is very efficient in accordance of estimation accuracy. It can able to estimate the result size based on global parameters. The global parameters are average weight of an edge and average degree of a node. The global parameters method can be able to apply for some special types of graphs which can able to obey some specified properties. Generally those graphs are called as regular uniform graphs. Those graphs are formally defined as: A graph G which is connected one is said to be regular uniform (RU) graph [3] if and only if it obeys some properties which are mentioned below: i)The degree of all the nodes should be same. ii) The weights of all edges should be identical, and iii) Select a random node named it as v and thereby gradually increasing the distance around v which is multiples of the edge weight (constant), then the number of the reachable nodes which are new can be increased by following an arithmetic-growth. 3) Binary Encoding Estimation method: This method generally uses particular graph transformations, and also specific binary encoding technique. The next and efficient method is Binary Encoding Estimation method [3]. The binary method efficiency is not dependent on the parameters edges e and vertices v.

Generally this methods rationale depends on the below mentioned observations: Total local densities pre-computations requires complete distances of shortest path that present in between the nodes of the graph. As based on that the process could encode complete distances of shortest path in to labels of node with respective to pre-processing step which can able to avoid total computations of the network distances. After that, assigning the labels to nodes of a particular graph must be done such that the distance between two nodes of a network can be calculated by using their labels only. After the process of encoding the node labels as binary numbers, the network distance between two nodes can be approximately calculated. Here this process needs a unitary weighted graph i.e., the weights of all edges should be equal to 1.

The main disadvantage of these methods is: none of these methods are capable of estimating the number of points that present between nodes in the specified range. Those methods are only capable of estimating the number of nodes in the specific range. And some methods can work efficiently in Euclidean spaces only but not in road networks. To overcome the above problem we propose an efficient mechanism which is nothing but selectivity estimator which clearly estimates the size of the result.

II. RELATED WORK

The selectivity estimation problem has been analysed in several sections as 1) selectivity estimation which belongs to spatial networks [4] 2) Estimation of nodes and edges specific to road networks [3] are examined. 3) The selectivity estimation of with respect to spatial databases [5] is determined. The methods has been analysed about approximate string search those are 1) for fast processing of approximate string for producing accurate results [6]. And another method is analysed for determining the query processing based on spatial keyword [7].

III. SELECTIVITY ESTIMATION

The proposed system is an extension to the existing system which focuses on building a framework that is capable of analysing spatial approximate substring queries, for mainly to solve the selectivity estimation problem of range queries which belongs to road networks represented by spatial database. The selectivity problem mainly concentrates on a range query for which it is mainly used to detect the number of points/elements in the respective dataset that presents in a given range. As the data stored in database is very bulky so that for traversing those data it might take very huge time, so that the procedure is to implement such that the estimation of the result should takes place. Clearly the selectivity estimation is developed because it is very helpful for query optimizers such that size of the result should be estimate. By using the estimated result the query optimizer can able to run the query in very effective way such that we can able to get better results than existing system. Generally after estimating the size of the result it has to generate a feedback for the user about the size of the result for particular query as estimated by selectivity estimator.

A. **System Architecture:**

Query optimizer is an inbuilt component of database management system such that it can able to useful to run query in an efficient way. Generally the Query Optimizer examines complete expressions of algebra which are equivalent to the given relevant query. The query optimizer implements the process of execution of a SQL statement in a most effective way by considering so many factors which generally related to the objects that are referenced, and also conditions which are specified in the query. The clear importance of query optimizer: it is a critical component of spatial database management system which analyses spatial queries and it clearly useful to determining the mechanisms for efficient execution of queries.

The proposed system generally implemented in the way as shown in Fig. 1 such that there is a technique implemented for estimating the size of the result for a respective query. For implementing that the process is as follows: At first the user enters the spatial approximate query, for example as “Ameerpet”. Then after that the construction of MHR-Tree [2] has to be takes place based on user query. The data should get from Spatial Data-Base which presents in Data-Base server as shown in Fig.1.Next step involves the estimation of result that is generated in Selectivity Estimation step. Then after that the filtering (i.e., first primary filtering and then secondary filtering) have to be takes place. After completion of filtering successfully the pruning has to be takes place. At last the processing of final candidates takes place. There after the query result is displayed to user with a feedback of estimated points.

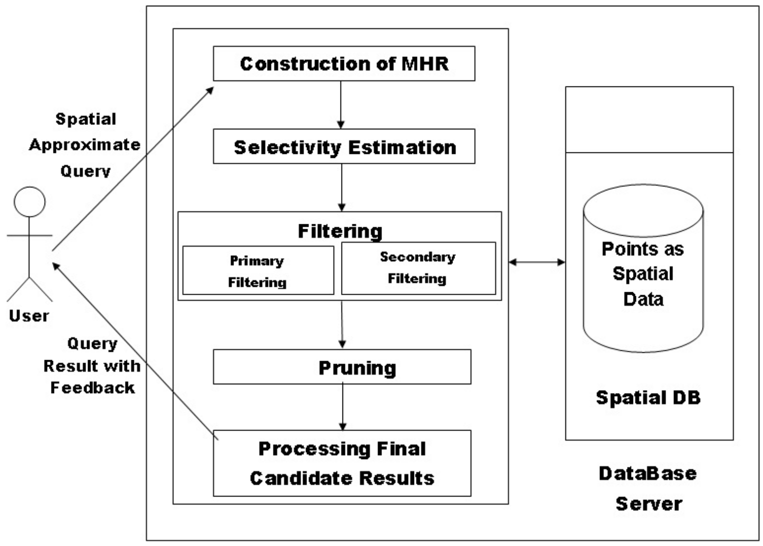

Fig 1: Selectivity Estimator

B. **Algorithm:** Selectivity Estimator Processing Algorithm

Purpose: Selectivity estimation and processing range queries on road networks represented in spatial databases

Inputs: Spatial approximate query q, spatial database SDB
STEP 1: CONSTRUCTION OF MHR TREE
Initialize an R-tree denoted as ***rt***
Initialize MHR tree ***mt***
Compute q-grams ***QG***
Compute min-wise signature s(***QG***) from ***QG***
mt = embed min-wise signatures into ***rt***
STEP 2: SELECTIVITY ESTIMATION
Convert the ***q*** into range query ***rq***
Divide ***mt*** into buckets ***B***
FOR each bucket ***b*** in ***B*** that intersects ***rq***
Use min-wise signatures of ***QG*** to build selectivity estimator ***se***
END
STEP 3: FILTERING AND PRUNING
Use se for primary filtering ***pf***
Then use ***pf*** results for secondary filtering ***sf*** based on spatial predicates
Apply pruning ***p*** on results of ***sf***
After processing ***p*** produce candidate results ***CR***
STEP 4: PROCESSING CANDIDATE RESULTS
Initialize ***CR`***
FOR each candidate result cr of ***CR***
Compute distance d of ***cr*** from query point
Find ***cr`*** that satisfies distance ***d***
Add ***cr`*** to ***CR`***
END
return ***CR`***

The algorithm procedure is: At first the technique was taken from existing system which was construction of MHR tree [2]. In that the nodes are divided in to a tree format with respective indexes assigned to that particular node. After indexing the nodes the computation of q-grams for particular node is implemented. And then the min-wise signatures of respective q-grams is calculated. The second step is the technique involves in the way such that selectivity estimation implementation mainly starts from here. The procedure is: the query has to be converted in to range query and afterwards min-wise signatures are divided into buckets. After that range queries are to be intersect with min-wise signatures which are stored in the buckets. Then the result should be built to estimate the number of points that lies within the given range. Such that a count should be established for counting the number of points that lie between nodes in a particular spatial network of spatial graph. The Google API is taken for the concept of respective implementation of displaying places which are near to particular area related to query. The present method overcomes the drawback of existing methods which is implemented for estimating the number of nodes within the given range as follows: here to determine particular region Google API has taken as medium and from that an area based on latitude and longitude values is selected with respect to the query entered by the user. There after the results are analysed for estimating the points in respective region for that a count is implemented. The points are generally retrieved from the Spatial Database which presents in the Data Base Server as mentioned in the architecture as shown in Fig.1. The third step involves filtering in which the results mainly based on primary filtering and secondary filtering. In the first phase of primary filtering the records are filtered from a bulk of records to some extent. The secondary filtering needs in the phase where the filtering takes place on the records that obtained from primary filtering. The fourth step involves here is pruning. The pruning is the process of getting rid of unwanted data by selecting the respective relevant data. Here based on that, the data can be clearly analysed for respective query. At last step the procedure is of processing final candidate results should takes place. Here based on the pruning results the final results are processed and displayed to the user.

C. Results:

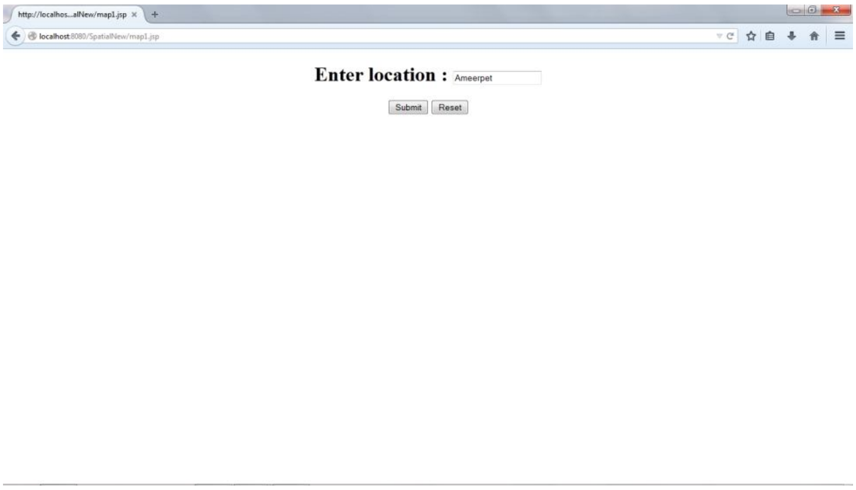

Fig 2: Entering Query as location for estimating the Result

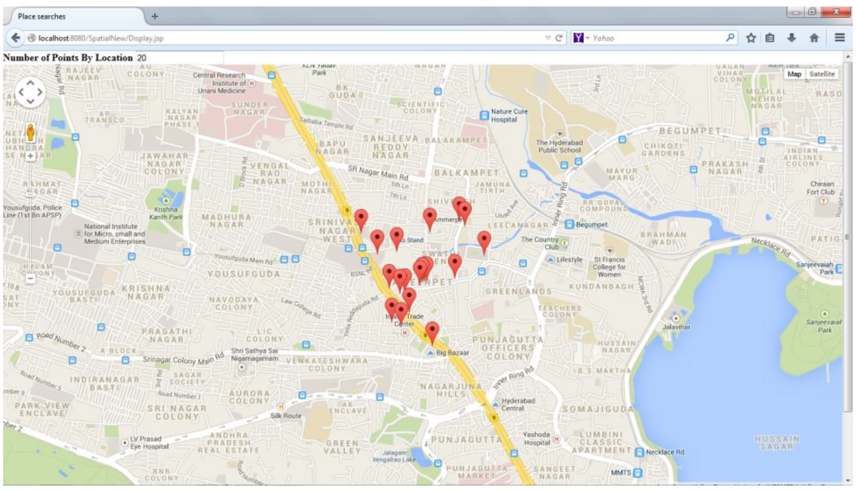

Fig 3: Estimating the result for the area Ameerpet

Here for representation of the results the algorithm is implemented successfully and those results are analysed for the area of Ameerpet which is located in Hyderabad and after that the procedure implemented as shown in Fig. 1. Here for that the query has to be given as “Ameerpet” in the text field of “Enter Location”. After that the estimation of result shown as number of locations as shown in the following Fig. 3. In Fig. 3 the count of locations are shown as 20 and those counted places are also highlighted by marking them.

IV. CONCLUSION

In this proposed work the selectivity estimation for range queries on road networks is implemented. Here we use the selectivity estimator as the technique for implementing the proposed work. The selectivity estimator mainly estimates the size of the result. The size of the result is nothing but counting number of points and nodes within the range. This is successfully implemented by using the selectivity estimator. Here there was another problem addressed which is integration of string selectivity estimator with spatial selectivity estimator in an efficient way. This problem is not correctly rectified and left as future work.

V. REFERENCES

[1] Ralf Hartmut Güting, Praktische Informatik IV, FernUniversität Hagen, (1994), “An Introduction to Spatial Database Systems”, VLDB Journal Vol. 3.
[2] Feifei Li, Bin Yao, Mingwang Tang, and Marios Hadjieleftheriou, (2013), “Spatial Approximate String Search”, IEEE Transactions On Knowledge And Data Engineering, vol. 25.
[3] E. Tiakas, A.N. Papadopoulos, A. Nanopoulos, and Y. Manolopoulos, (2009), “Node and Edge Selectivity Estimation for Range Queries in Spatial Networks”, Information Systems, vol. 34.
[4] E. Tiakas A.N. Papadopoulos A. Nanopoulos Y. Manolopoulos, (2008), “Selectivity Estimation in Spatial Networks”
[5] Swarup Acharya Viswanath Poosala Sridhar Ramaswamy, “Selectivity Estimation in Spatial Databases”.
[6] ZiqiWang, Gu Xu, Hang Li, Ming Zhang, “A Fast and Accurate Method for Approximate String Search”.
[7] Lisi Chen, Gao Cong, Christian S. Jensen, Dingming Wu, “Spatial Keyword Query Processing: An Experimental Evaluation”.