Query Processing by Mining Spatial Database

L. Gowri, R. Seethalakshmi and E. Saranathan

School of Computing, Sastra University, Thanjavur, Tamil Nadu, India
School of Civil Engineering, Sastra University, Thanjavur, Tamil Nadu, India

Abstract: Spatial database is a database which encompasses the spatial data like zonal data, data about population and education. In this paper, we address the issue of creating spatial tables which are mined by spatial queries and produce the statistical output. Four spatial tables are created, namely zonal table, horticulture table, Educational table and Employees table. Here the tables are enabled to join with the state-code id and any query are decomposed appropriately and efficient mining takes place where the data are retrieved accordingly. This data are ported onto the GIS software where the respective area in the map is flashed as a result for the corresponding query.

Key words: Spatial database • GIS • Spatial query • Mining • Trichromatic spatial join

INTRODUCTION

Spatial database is a database that is specially related to spatial data objects. These spatial data objects are connected to a place in the earth. Spatial databases contain the information about location and textual contents. The location based information like point, line, polygon and textual based contents like keyword is present in spatial databases. Spatial data played a major role in many applications like Geo-mapping, Geo-marketing and cartography and so on. The Geographic Information System (GIS) is a tool which has been defined as spatial data that is used for capturing, analyzing, manipulating, storing and displaying Geo-referenced information. For example, GIS [14, 15] can list neighboring state of a given country with the help of political boundaries. The spatial keyword query has become an important research area. The spatial keyword query data are used for searching the spatial data which are stored in spatial databases. The result of the spatial keyword query may be a single value or multiple value based on the search condition passed to that spatial database. The Spatial queries are processed with filter and refine techniques [3]. The filter step is used to produce an approximate representation of spatial objects that are needed to be satisfied by spatial queries. The refine step is very complex rather than filter technique, because the spatial relationship refines spatial objects with join predicates. It requires comparison among every tuple with previously generated spatial objects. There are several spatial keyword query processing methods proposed [7-10]. This paper is mainly focused on spatial join embedded within the trichromatic spatial queries. Spatial joins detect all possible spatial objects with two or more different entities. While performing spatial joins, different perspective of spatial objects can be considered. The query processing encompasses the six steps as shown in figure 1. The spatial query in general may consist of a number of multilevel entities. The block diagram depicts that a Spatial query is taken and is decomposed into simpler components. Then the query engine searches for the correct table and the table is asserted. For each sub-component in the query, the appropriate fields are

Fig. 1: Block diagram of query processing in spatial databases

Corresponding Author: L. Gowri, School of Computing, Sastra University, Thanjavur, Tamil Nadu, India.
chosen and compared with the various spatial objects and values are displayed in the result table. These results are ported onto the GIS and the final results are splashed in the proper area on the map in GIS.

**Related Work:** The several types of spatial queries are discussed by many numbers of authors with different perspectives. The related work of this paper is mainly focused on spatial nearest-neighbor query, range query, density based query and spatial join query. A. Guttman et al., used dynamic index method to find how nearest neighbor queries basically referred by R-tree [3] method. Because of simplicity, the R-tree is very popular indexes to help of Euclidean distance. The R-tree is fully dependent on the heuristic optimization technique that is used for accessing Minimum Rectangle Bounding (MBR) method. Here only one parameter is dealt with namely area. But the variation of R-tree is the R* - tree [4] which was proposed by N. Beckmann, H. P. Kriegel, R. Schneider, B. Seeger et al., that is specially used for different number of parameters, like area, margin and overlap of MBR. The Nearest Neighbor (NN) [5] queries are analyzed using a number of ways which tries to yield the best solution. The NN query was proposed by Roussopoulos et al. That is specially searched for very closest objects in the datasets along a query point with the help of one of the design approach is branch-and-bound algorithm in the R-tree [3] method. The closest pair query [6] was studied by A. Corral, Y. Theodoridis, et al., which considered two different query points and find out the closest value of spatial objects. The All -NN queries [7] proposed by N. Mamoulis, Y. Tao et al., Two approaches were discussed, the Multiple NN which reduced the computational cost of NN queries and Batched NN used hash-based approach which is used only for high dimensional data. The kNN query [5, 8-10, 1] is used to search k number of closest object in a particular area. The value of k is greater than or equal to one. The kNN query algorithm is considered to determine how many numbers of objects are very close to spatial query. The relationship between query point and spatial objects is very important. This relationship might be considered as bidirectional or reversed [9]. Bidirectional means that, the relationship between spatial objects is much strengthened. The next way uses the reverse angle. The Reverse NN [8] gives preference only for spatial objects but not for query point which, means that the relationship between query point and spatial objects are very weak. This problem is partially removed by monochromatic all-k NN [1] which uses spatial objects only of the same type and gives the preference for query point. The bichromatic all-k NN [1] has used two different types of objects. To avoid this type of complexity, the query points are used in the region. The spatial range query is associated with the location and boundary of a given region. The results of spatial range query contain in the form of overlapping and contained data regions. When the numbers of spatial keyword queries are used, these queries should be identified in terms of groups. This way of identification is mentioned by X. Cao, G. Cong, C. S. Jensen, B. C. Ooi, et al., dealt with the name of the collective spatial keyword querying [10]. The result of this approach may be selected either in the most popular area or in the less popular area. This drawback is considered by Li Zhang, Xiaoping Sun, Hai Zhyge, et al., at the minimum level in the density based query [12]. The density based method focuses on high priority which means that highly popular area. This method may be applicable for urban area [13, 16], but not possible for rural area. While comparing rural area, the urban area has high population. In urban areas spatial querying is much easier because the spatial objects are located close to each other. In rural area the technology development is most needed because of its significance. This paper addresses the issue of focusing on not only rural area but also in the urban area. The spatial join query find closest query value pairs from two or more data sets that spatially interact with each other. This work proposed new method namely trichromatic spatial join based query processing method.

**Design of Algorithm:** The spatial query searching is applied in different perspective knowledge of the authors. For example, the spatial query processing applied in the following ways. The first approach is used to identify the relationship between query point and spatial objects. This method gives importance to query point rather than spatial objects. The second approach is applied to the relationship among spatial objects and finally relates to query point and this method says a group of related spatial objects in a particular query point. The third approach is applied with different number of query points and spatial objects in the most popular area say the urban area, but not concerned with the rural area. This paper is focused on three different types of spatial entities, first one is a rural area, second one is spatial objects and finally we consider the query point. In this rural area, we take three views of the projection. The projection of the first view is working status of the particular area. The second projection view is education
status of the particular area. The third projection view is the horticultural status of the particular area. This approach is called Trichromatic Spatial Join (TSJ) query processing in spatial databases. It is an efficient method for analysis in a particular area. First take a whole database to determine the spatial value with respect to rural area.

**Algorithm for Spatial Query Processing:**

Query: Trichromatic Spatial Join

Start ();
Decomposition ();
For fields = 1 to 16
Where the fields are

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>lat_long</td>
<td>Latitude and longitude values</td>
</tr>
<tr>
<td>state_code</td>
<td>State code</td>
</tr>
<tr>
<td>state_name</td>
<td>State name</td>
</tr>
<tr>
<td>rural_pop_percen</td>
<td>Rural population percentage</td>
</tr>
<tr>
<td>urban_pop_percen</td>
<td>Urban population percentage</td>
</tr>
<tr>
<td>tot_males_per</td>
<td>Total males percentage</td>
</tr>
<tr>
<td>main_work_per</td>
<td>Main working percentage</td>
</tr>
<tr>
<td>marginal_work_per</td>
<td>Marginal working percentage</td>
</tr>
<tr>
<td>deumed_univ</td>
<td>Deemed university</td>
</tr>
<tr>
<td>college</td>
<td>College</td>
</tr>
<tr>
<td>fur_area_sqkm</td>
<td>Forest area square kilometer</td>
</tr>
<tr>
<td>veg_area_sqkm</td>
<td>Vegetable area square kilometer</td>
</tr>
<tr>
<td>flow_area_sqkm</td>
<td>Flood area square kilometer</td>
</tr>
<tr>
<td>sqkm</td>
<td>Square kilometer</td>
</tr>
<tr>
<td>tot_area_sqkm</td>
<td>Total area square kilometer</td>
</tr>
</tbody>
</table>

FROM table = population, worker, education, horticulture WHERE JOIN CONDITION object1.
state_code=object2. state_code AND object2.
state_code=object3. state_code AND object3.
state_code=object4. state_code AND object4.
rural_pop_percen;

• Execute_query ();
• Obtain_results ();
• Display_results ();

The spatial query decomposed into a number of tables which consists of the four different spatial entities. These spatial entities are fully dependent on the spatial objects which were stored on spatial database. We used four different spatial entities that are population status, horticulture, employee, education. The population is the main spatial entity which consists of a rural area percentage, urban area percentage, latitude, longitude, total population in each state of India. This information is helpful to find how the state level information could be considered in three different views of spatial entities. The horticultural entity is used to find out the details of cultivation with respect to fruit, flower and vegetable area. The area information is stored in square kilometer representation. The employee entities use two different views such as main working and marginal working status and also consider the depth of female and male working status. This is very useful to compare in terms of main and marginal worker preferences for each state. The education entity considered as three different views, which is number of universities, colleges and schools. The schools might be considered as three different spatial value such as recognized, private and government. The university might be private, state level, central level and deemed. In this paper we focused only on rural area.

**RESULTS**

GIS tool is applied in different technology that assists to analyze various types of geospatial data and also deals with complex data. GIS plays a major role in integrating technology with large quantity of information about the environment. It always supports to analyze the spatial data and display different file formats. The place is always essential to people that dictates much of our life culture. This can be useful to classify human population, education standards, number of employee status in a particular area, the cultivated land area and so on. For these reasons the databases have played a major role. Generally databases used to find out the data with the help of indices. But this is not enough for spatial database. It is needed to represent one more value that is geometric value. With the help of that to identify the relationship between spatial objects very perfectly.

**Experimental Setup:** The experiment is carried out in ArcGIS suite. ArcGIS is a platform for working with maps and spatial entities. ArcGIS allows the user to create different number of layers, view and compile spatial objects, perform spatial analysis in depth view and allow manipulation of shape files and Geo-databases. Here, one should first store the input map into ArcGIS and create the shape file for making the Geo-reference points with the help of latitude and longitude values. It is used to create different views of vector format like point, line and polygon and so on. This work is focused only on the polygon type and displays the output for all states of India with respect to population which is as depicted in figure 2. After making the entire polygon it should be converted into raster format. This is the input to process the Trichromatic spatial join queries. It is used for four different spatial queries and spatial entities. The first step is used to find the rural areas in Indian states. The calculation of all types of area is described in terms of a square kilometer. Next is to calculate the density area with the help of population and area field. We execute the
Fig. 2: State wise population

Fig. 3: State wise rural area

Fig. 4: Maximum employees in rural area

Fig. 5: Best education in rural area
first query to find how much space is placed in terms of rural area for all states. This is the seed value to navigate further steps. This output is shown in figure 3. The states which are obtained as results are depicted in the figure. Very high rural ratio state names are Himachal Pradesh (HP), Bihar (BI), Assam (AS), Orissa (OR), Meghalaya (ME), Manipur (MN), Uttar Pradesh (UP), Arunachal Pradesh (AC), Chhattisgarh (CH), Jharkhand (JH), Haryana (HA), Rajasthan (RJ), Sikkim (SI). The second step is used to find the total number of employment status in rural area. This step is applied in two different views of employees. The first one is main worker status and another one is marginal worker status along with in depth of female and male worker status respectively. This is the input to determine the total number of workers in rural area and execute the proper spatial query. The GIS output is depicted in figure 4. The output list of states is BI, MP, RJ, OR, JH, CH, AS, HA, UP. The third step is used to move forward in another way of angle that is education status in rural area. The methodology of finding the education status is used in three different views that constraints a number of schools, number of colleges and the number of universities in rural area. This query output is projected in figure 5. The list of output states are UP, RJ, BI, OR, AS, MP. Finally the fourth step is considered. This is entirely different from other views that is horticulture places in rural area. This query required only the minimum number of constraints to execute the total number of horticulture places in rural area. This GIS output is shown in figure 6. The output list of states are UP, OR, BI, RJ, MP, CG, AS, HA. At the end, we intersect all the output in one spatial join query namely trichromatic spatial join with the help of unique field of all tables and finally consolidated all output in one projection which is shown in figure 7. The final output states of rural areas are UP being placed in north zone, RJ is placed in west zone, BI, OR, AS all are placed in east zone.

CONCLUSION

In this paper, we focused on the details of the rural area of the Indian states. While comparing with urban area, most of the other areas have a high population, high density and many numbers of industries. The metropolitan areas are well developed areas with respect to socioeconomic connection between these types of area. This is not so in a rural area because it gives the preference only for agriculture. We analyzed the rural area in the views of employee status, education
status and horticulture status. With the help of trichromatic spatial join query processing to implement a different number of constraints and finally produced five output states of India. These states are highly developed in the education level, employee status level and horticultural level. So we can say these states are semi-urban areas. In future work we consider a number of different angles.

ACKNOWLEDGEMENT

We acknowledge our vice chancellor prof. R. Sethuraman for extending support and providing facilities for the current work.

REFERENCES