

Arbitrary Polygon Query Handling Algorithm on GIS based on three Value Logic- An Approach

Ferdousi Khatun
Department of Computer
Science and Engineering
Sikkim Manipal Institute of Technology

Pratikshya Sharma
(Assistant Professor)
Department of Computer
Science and Engineering
Sikkim Manipal Institute of Technology

ABSTRACT

Recent year the GIS (Geographic Information System) is used widely in various field such as planning and management, environment and natural resource management, medical science, facility and utility management, land information system, street network management. GIS mainly deals with two type of data spatial data and tabular or attribute data. The spatial analysis mainly conducted through the classical point-in-polygon algorithm such as the ray casting algorithm and winding algorithm. The above mentioned algorithm have linear complexity with the no of point available in the map. An approach to handle arbitrary polygonal queries on a GIS based on the three value logic has been put up. The above said approach can find whether a point is inside, outside, or on the boundary of a given polygon by utilizing a very simple overlay mechanisms in a GIS.

General Terms

Spatial analysis in GIS.

Keywords

GIS, Spatial Analysis, Polygon Query, Point-in- Polygon, Spatial query, Overlay, Three value logic.

1. INTRODUCTION

A geographic information system or geographical information system (GIS) is a system which is designed in such a way that geospatial data or geo-reference data can be input, capture, store, manipulate, analyses, manage, output effectively in order to decision making for planning and management of land use, natural resources, environment, transportation, urban facilities and various administrative records. The GIS deals with mainly two types of data those are spatial data and tabular or attribute data. The spatial data are of two type vector data which is represented in point, line or polygon and the raster data which are represented in pixel. Spatial data are used to provide the visual representation of a geographic space and are stored as raster and vector types. This data is a combination of location data and value data to render a map. The spatial data are usually stored as co-ordinates. The vector data are in the form of point, line or polygon. The spatial data is very important in GIS. As GIS has a huge range of application field in recent year, there may be cases where the organization, their infrastructure or assets and their customers are spatially distributed in several locations. Without the knowledge of where customers and assets are spatially located, it is difficult for organization to conduct their operational activities. The GIS always represent any geographic objects in terms of location (where is it?), shape (what form is it?), size (how big is it?) and the orientation (in which direction it is facing?). The above geographic objects are represented in point, lines and area in the GIS system. The points are used to represent the shape and size of one

dimensional feature. Lines are used to represent one dimension objects such as road, rivers, canals, railroads, power lines etc. The area are represented in polygons such as lake, forest, water body etc.

2. SPATIAL QUERY

The heart of GIS is the analytical capabilities of the system. The analysis function use geospatial and non-spatial attributes in the database to answer about the real world. To analyze the GIS perform various spatial query. Generally query is a way of acquiring the information stored in a database. The GIS has the ability to ask complex questions to the database. In GIS spatial query is defined as special type of database query which is supported by geodatabases and spatial databases. For example, a spatial query might find which points are contained within a polygon or set of polygons, find features within a specified distance of a feature, or find features that are adjacent to each other. The GIS system can carry out a no of queries, such as selection by cursor, selecting geographic objects, and selection by spatial queries. Spatial queries differ from non-spatial SQL queries in several important ways. Two of the most important are that they allow for the use of geometry data types such as points, lines and polygons and that queries consider the spatial relationship between these geometries. For any type of display, processing and analysis of GIS spatial data the user select data from a GIS via asking queries. The complex queries require the buffering or overlay technique. The overlay operation involves point in polygon overlay, line –in-polygon overlay and polygon –in –polygon overlay. Today one of the most common queries is the 2-D polygonal query. The polygonal query allows user to draw a plane or polygon to search for objects on information attribute to the objects within the plane. This interaction required a GIS to effectively determine the relation between objects in the map and the polygonal plane. For example, to determine on which lots the medical shop are located, whether it lays inside, outside, or on the boundary of a given polygon. The spatial data are mainly represented in 2D plane. There are two main approaches in recent year for point-in-polygon operation those are the even-odd rule or the ray-crossing and the winding number or the angle summation algorithm. The ray casting algorithm tests how many times a ray (line), starting from the point and going to any fixed direction, intersects the edges of the polygon. The winding number algorithm computes the given point's winding number with respect to the polygon. The above two algorithms consist of a pre-processing, a sorting method, and a method for querying points for polygon inclusion. The ray casting or the winding algorithms are applied to all the points in the map. Thus, those search methods have run times on the order of $O(m * \log(n))$, with m is the number of query points and n is the number of edges of the given polygon. In order to achieve optimal performance, the spatial indices are employed in practical

such as those in a GIS, especially for handling large dataset .One of common indices algorithm is the R-tree. The R-tree indices algorithm are having good performances especially if the indices are well design. The current GIS using R-tree that has at least two main disadvantages: 1) the polygon drawn as the query should be available in the R-tree indices, it means that the GIS users cannot query or drawn arbitrary polygon; 2) The point-on-polygon algorithm has to be tested to every points in the map, thus the computation time is very high especially if there are many (thousands) points in the map.

3. GIS AND SPATIAL DATA

In GIS the spatial data are generally represented in 2D plane. In GIS as an example of a fire station at 32 MG road, beside State bank of India and 10 KM away from hospital, can be represented in spatial form as a point, line and polygon. The region of the location can be represented as polygon. Someone may ask a query from the GIS map such that how far is the fire station from the hospital. This spatial analysis is done on 2D display map. The location are stored in latitude and longitude value. The spatial data are stored in GIS in various form such as spaghetti model, vertex dictionary model, topological model, point model, contour model etc. In fig.1 representation of a Spaghetti model is depicted.

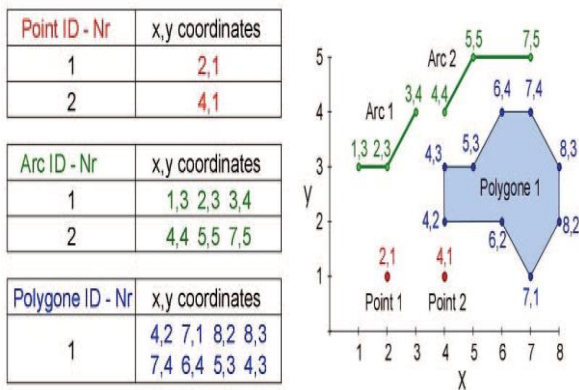


Figure 1: Vector data model representation

[<http://www.geogra.uah.es/patxi/gisweb/GISModule/images/Image121.gif>]

One of the major advantage of GIS is its analytical capability. The GIS has the capability of carrying out any no of analysis related to any discipline oriented problem. The GIS spatialanalysis can be made in various way such as database query,geospatial measurement, overlay operation, network analysis,surface analysis, geovisualization etc. The overlay analysis performs the spatial data and attribute integration by combining information from one GIS layer with another GIS layer to derive or infer an attribute for one of the layers.

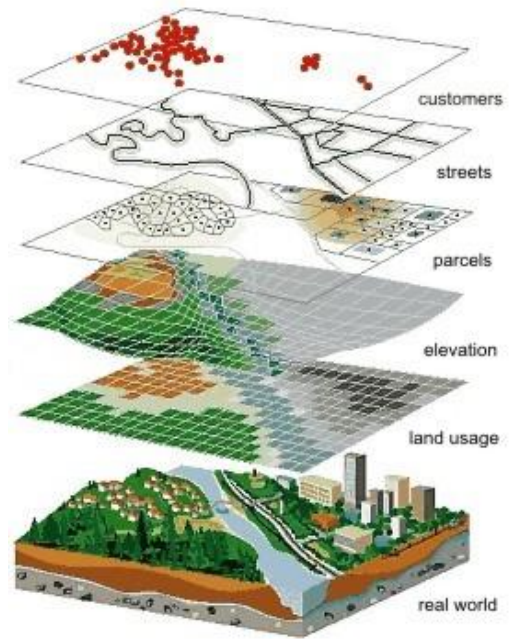


Figure 2: Example of overlay analysis in GIS

[<http://www.ukurtanah.com/wp-content/uploads/2013/01/gis-overlay-analysis-tile.jpg>]

3.1 Winding number algorithm

The overlay operation involves point in polygon overlay, line –in-polygon overlay and polygon –in –polygon overlay. The vector overlay is achieved by creation of a new topologicaltheme from two or more existing themes. Counting the number of points in a polygon is a very common overlay operation. But unless you're aware of what happens when points fall on polygon boundaries, or when points fall just outside the coverage of your polygons, you may not be getting the results you expect. This analysis involves points in one data layer and polygon in another data layers. In this process, a GIS determines which point falls in which polygon. The answer can be made by applying the ray casting algorithms or the winding algorithms. The Winding number counts the total number of times that a curve of polygon travels counterclockwise around the point. If the number is non-zero, the point lies inside the polygon otherwise the point is outside the polygon. One way to compute the winding number is to sum all the angles subtended by each side of the polygon. However, since this involves costly inverse trigonometric functions, this algorithm is slower than the ray casting algorithm. Another simple way of finding whether the point is inside or outside a simple polygon is to test how many times a ray, starting from the point and going in any fixed direction, intersects the edges of the polygon. If the point is on the outside of the polygon the ray will intersect its edge an even number of times. If the point is on the inside of the polygon then it will intersect the edge an odd number of times. Unfortunately, this method won't work if the point is on the edge of the polygon. This algorithm is sometimes also known as the crossing number algorithm or the even-odd rule algorithm. Assume that winding number is $wn(P, C)$ of any closed continuous curve C around a point P in the 2D plane. Let the continuous 2D curve C be defined by the points $(u) = C(x(u), y(v))$, for $0 \leq u \leq 1$ with $C(0) = C(1)$. And let P be a point not on C . The vector can be defined as $c(P, u) = C(u) - P$ from P to $C(u)$, and the unit vector $w(P, u) = c(P, u) / |c(P, u)|$ which gives a continuous function $W(P): C \rightarrow S^1$ which is

mapping the point $C(u)$ on C to the point $w(P,u)$ on the unit circle $S^1 = \{(x,y) \mid x^2+y^2=1\}$. This map also can be represented in polar coordinates as $W(P)(u) = (\cos\theta(u), \sin\theta(u))$ where $\theta(u)$ is a positive counter clockwise angle in radians. The winding number $wn(P,C)$ is equals to the integer value ie the no of times that wraps C around S^1 . The winding no can be computed by the integral:

$$wn(P,C) = \frac{1}{2\pi} \oint_{w(P)} d\theta = \frac{1}{2\pi} \int_{\alpha=0}^1 \theta(u) du$$

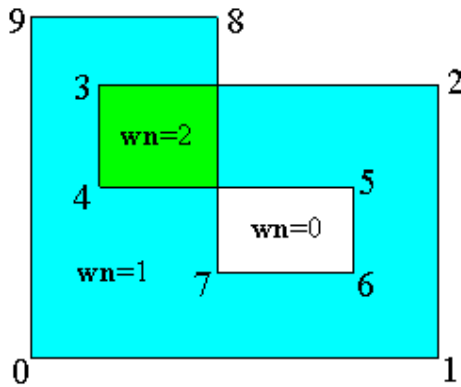


Figure 3: Example of winding algorithm

3.2 R-Tree indexing

The performance of GIS can be increased by applying the R-Tree indexing method. R-Tree is a tree data structure, similar to a B-tree that is used for indexing spatial data within a database. In an R-tree structure, data is sorted into a set of hierarchical nodes that may overlap. Each node has a variable number of entries, each of which includes an identifier for child nodes or actual data elements and a bounding box for all entries within the child node or the data elements. Searching algorithms check the bounding boxes before searching within a child node, thus avoiding extensive searches.

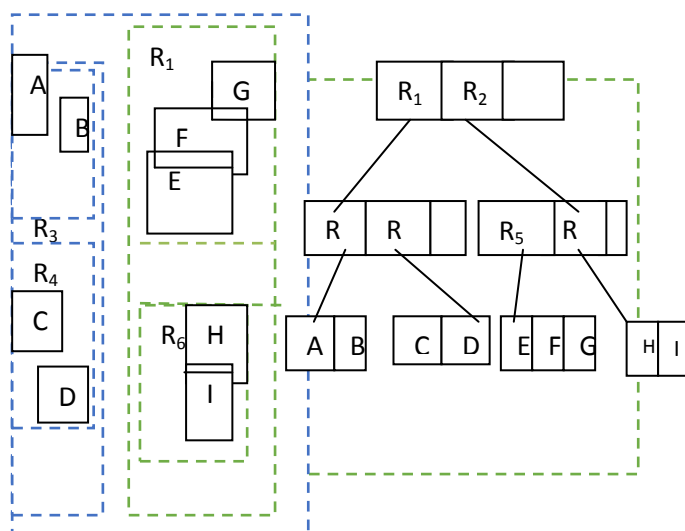


Figure 4: Example of R-TREE indexing

While building the R-tree indices, we need to minimize the coverage of tree nodes to reduce the amount of empty space in the map and to minimize the overlap of tree nodes to reduce the search paths to the leaves. Coverage is the entire area covered by tree nodes, while overlap is the area covered by two or more nodes. Minimizing coverage reduces the amount

of empty area and minimal overlap reduces the set of search paths to the leaves. R-tree indices, point-polygon relationship query algorithm consists of the following steps:

1. For each point P , query the polygon index to find polygons whose envelopes contain P .
2. Query whether P is inside the polygon using ray casting or winding algorithms.

3.3 Arbitrary polygonal query handling

A recent method of handling arbitrary polygon query is proposed by IpingSuprianaSuwardi, DessiPuji Lestari, Dicky Prima Satya in IEEE where overlay analysis based on color or Boolean value assigned to the given polygon to handle the arbitrary polygonal query. This proposed method works on Boolean value. The above mention three algorithms works to find all points exist in the given polygon by checking whether the point coordinate in the point layer has "true" value or not. If the answer is true it implies that the point is inside or on the boundary of the given polygon. False value means the point is out side the given polygon. By using a temporary table that stores all Points that are shown on the map. The table has all attributes that contains in the points database and additional attribute called the point-in-polygon that is used to store the Boolean value of each point for indicating whether the point lays inside a given polygon or not.

4. THREE VALUE LOGIC FOR ARBITRARY POLYGON QUERY

The ray casting algorithm, winding number or the R-Tree index algorithm has some disadvantages like computationally expensive across trigonometric function is used in case of winding number which is difficult and it is very slow in case of computation. All the above mentioned algorithm determine the points is either inside or outside a given polygon. But if a particular point is in the boundary the ray crossing, winding no fails to determine them. In case of Boolean overlay it assumes that if point is in the boundary then gives result as true that the point is either inside or on the boundary but can not exactly differentiate. The three value logic is based on Boolean logic. The Boolean logic takes only the value true and false but the three value logic has true, false and unknown value ie neither true nor false state. By using three value the three state can be identified separately. If the value is true it implies that the points inside the polygon, if value is false then it is outside the polygon and for unknown value it is on boundary of the polygon. Like Boolean logic AND, OR, NOT, XOR operation are allowed in three value logic. The three value logic can be implemented using the following steps: At first create a temporary table and attached it with spatial database. Then Copy all the data points that will be visualized on the map from the point database into a temporary table. Visualize the map by rendering spatial data in the temporary table using layering. If the map contains only the point data then there will be only a point layer. If the user want to made an arbitrary polygon query they can draw a polygon on the map. These polygon is separately captured in separate layer. Now Colour all surfaces inside and in the boundary of the given polygon query with different colour let blue in the boundary and grey inside the polygon which not used in the map.

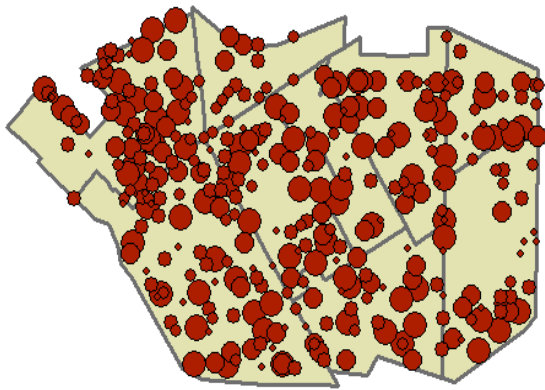


Figure 4: Example of R-TREE indexing

The colour has to different form the colour which is not present in the map. In the temporary table assign the true value for grey and unknown value for blue for all the points in the polygon layer. By zooming the map it shows the grey values are inside the polygon. In Figure 5 a point layer of one place has been taken and the polygon layer of same place is represented in Figure 6. After the overlay the output using three value logic depicted in Figure 7. The blue points are on the boundary of the polygon for a query may be how many hospital are present on that area which name starts with alphabet A.

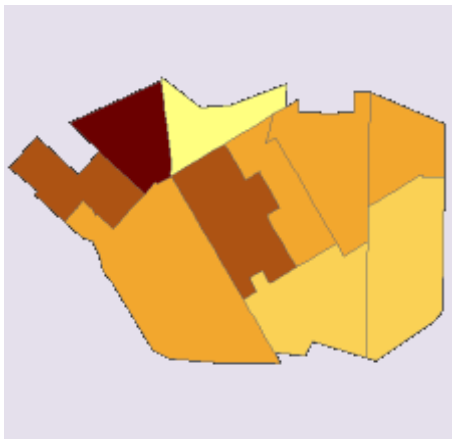


Figure 6: Polygon layer data representation

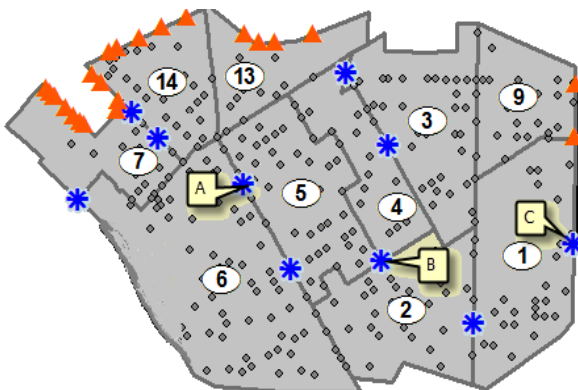


Figure 7: Overlay output

5. CONCLUSION

An approach to handle arbitrary polygonal queries on a Geographical Information System based on the three value Overlay has been described. The algorithm is able to find whether Points in map are inside, outside, or on the boundary of a given polygon by utilizing overlay mechanisms in a GIS. Three value logic is extension of Boolean logic. This algorithm distinctly found the three feature, inside, outside or on border point. This approach will works very well for conducting spatial analysis between object on a map in a point shape and polygon query.

6. ACKNOWLEDGMENTS

Our thanks to the CSE department of SMIT .

7. REFERENCES

- [1] IpingSuprianaSuwardi, DessiPuji Lestari, Dicky Prima Satya” Handling Arbitrary Polygon Query based on the Boolean Overlay on a Geographical Information System”, Advanced Informatics: Concepts, Theory and Applications (ICAICTA), 2015 2nd International Conference on, 978-1-4673- 8142-0, 19-22 Aug. 2015
- [2] T. Zhou, H. Wei, H. Zhang, et. Al, “Point-Polygon Topological Relationship Query using Hierarchical Indices,” SIGSPATIAL 2013.
- [3] I. Supriatna, D.P. Satya, D.P. Lestari D.P, “Spatial Data Model for Corporate Based on Google Maps Platform,” in Proc. International Conference on Data and Software Engineering, 2014
- De Smith, Goodchild and Longley, ”Geospatial Analysis”, 5th edition, 2015
- [4] Point in polygon” https://en.wikipedia.org/wiki/Point_in_polygon” 15.03.2016.
- [5] Three value Logic” https://en.wikipedia.org/wiki/Three-valued_logic”
- [6] 16.03.2016.
- [7] Three value logic” <http://c2.com/cgi/wiki?ThreeValuedLogic>” 16.03.16.
- [8] E. Clementini, ”Computational geometry Part V”, 13.10.2014
- [9] Ying Hu, Siva Ravada, and Richard Anderson” Geodetic Point-In-Polygon Query Processing in Oracle Spatial” 2011
- [10] Category:GeometricAlgorithm, ”https://en.wikipedia.org/wiki/Category:Geometric_algorithms”, 16.03.2016