SPATIAL ANALYSIS IN GEOGRAPHICAL INFORMATION SYSTEMS - A DATA MODEL ORIENTED APPROACH

Kirsi Artimo
Helsinki University of Technology
Department of Surveying
Otakaari 1, 02150 Espoo, Finland

Abstract

This paper presents a data model oriented approach to spatial analysis methods. Both vector and raster oriented data models are dealt with. Spatial analysis methods are classified into computational and visual analysis. Computational methods are further classified into basic operations, analysis functions and applications. This approach is useful in design and implementation of spatial analysis in GIS as well as for educational purposes.

1 Introduction

Analysis in geographical information systems has a significant role. In the traditional function oriented definitions of GIS (9,1) analysis represents a set of necessary functions in a GIS - in order to be a GIS there must be (among others) analysis functions available in the system. It is also quite obvious that geographical analysis - called originally spatial analysis - has been, in a way, the origin of GIS (13).

Spatial analysis is "a type of geographical analysis which seeks to explain patterns of human behaviour and its spatial expressions in terms of mathematics and geometry" (10). The use of mathematics in geography can be classified for example under the following headings (7): 1.sampling, 2.description of spatial distributions, 3.correlation, 4. topology and transformation. This classification outlines a framework for approaching traditional spatial analysis.

Compared to the scope of geographical spatial analysis based mainly on arithmetic, geometric and statistic operations as well as graph theory GIS analysis is a broader subject. In modern GIS analysis methods are utilizing more logic operations (set algebra and Boolean algebra), matrix calculus, optimization and simulation methods. In addition to previous computational methods data management functions and visual analysis possibilities make a strong basis for the modern GIS analysis.

In modern GIS literature we find GIS analysis as a part of the fundamentals. Aronoff (1) gives his classification of GIS functions as well as Laurini (8) and Burrough (3) and many others (2). Aronoff's approach is based on the following classification: analysis functions on spatial data, analysis functions on attribute data, analysis functions on both spatial and attribute data and output functions. Burrough classifies analysis functions in topology oriented, property oriented and topology+property oriented. Like Aronoff he then concentrates on "retrieval", "overlay", "neighbourhood" and "connectivity" functions.

Laurini gives an approach which is more data oriented. He deals with basic operations on points and lines as well as polygons and matrices. Then he treats spatial analysis functions separately in the case of graphs and polygons and grids. The approach is strongly based on the data model in question.

Laurini as well as Aronoff and Burrough refer to "map algebra" developed by Tomlin (14). Tomlin's approach origins from geography and cartography and reflects strongly the procedures of traditional manual map analysis.

GIS software packages provide GIS analysis tools. The "standard" structure of a GIS software is a basic "nucleus" module added by several analysis modules including both vector and raster based analysis functions (Integraph MGE, ESRI ArcINFO). Also the so-called desktop-GIS software packages tend to develop their analysis supply.

An overview on the GIS analysis reveals an unorganized group of methods which seems to require a more systematic approach. In this article a data model oriented approach on computational GIS analysis methods is introduced.

As already mentioned GIS analysis is not only a group of computational methods. Computer based GIS offers an efficient environment for visual analysis. GIS analysis is in the literature approached mainly as computational processing of geographical data. However geographical information systems provide an easy environment of the most efficient analysis which is made by the human eye-brain combination. In this article visual analysis is treated very briefly, however we want to point out the efficiency and important role of it in the practical applications.

2 The goal of this paper

When studying GIS analysis it is important to have a general framework in which all analysis functions can be placed. Also it is important that such a framework is not only an abstract classification of methods but can also be used when designing and implementing analysis applications in GIS software environment. Also for educational purposes a good classification is necessary.

In this presentation a data model oriented approach to GIS analysis is outlined. Various data models are treated. Analysis methods are classified into basic operations, analysis functions and applications. On more general level also the possibilities of visual analysis by using GIS is dealt with and the important role of it is emphasized. The software development viewing point is dominant in this study.

3 Spatial data model

All data processing in GIS environment is based on the data model which is used. A data model oriented approach here means a viewing level where concepts like vector and raster model, topological and "spaghetti" are used. Data model can sometimes be a synonym to "conceptual model" (5) and thus it does not show exactly how the data is stored, on the other hand it is not a pure user-view representation. As Date says a data model is a view of the data "as it really is". A spatial data model can include structures like polygons, graphs, grids, points. In (12) we can find a similar definition of a data model: "an abstraction of the real world which incorporates only those properties thought to be relevant to the application or applications at hand, usually a human conceptualization of reality".

The main types of spatial data models are: vector model and raster model (1). A 2d vector model can be organized by using points, lines or polygons. It can utilize complex concepts like graph, set or area partition. A vecor model can be used in a very simple way, an example of a simple vector model is implemented as a point register consisting of coded points with no other semantics than the quality and location of each point. More advanced vector models are so-called line and polygon models (4) as well as graphs.

In 2d matrix/raster data model only pixels and the matrix can be identified. A non-structured matrix model does not reflect more semantics than the theme of the layer and the quality of each pixel.

In this discussion on spatial analysis we use the following main types of data models: graph, polygon, point and matrix.

4 Hierarchy of GIS analysis functions

In the previous text we already outlined the main types of GIS analysis: computational and visual. Computational includes both data management functions as well as mathematical calculations. GIS analysis functions can be more detailed classified into basic operations and analysis functions as well as applications.

Basic operations are simple generic calculations or data management operations which are not as such making any analysis but which are utilized in more complex analysis functions. Such basic operations are e.g. line and segment intersections, line and polygon intersection, point in polygon problem, intersection and union of polygons, polygon overlay problem, windowing and buffering, length, area and centroid calculation. Basic operations are usually ready made services of a GIS software. For example geometric search is a basic operation in a GIS data management system.

Analysis functions are composite operations which include several basic operations. GIS software packages may include some analysis functions like shortest path in a network or generation and use of a digital terrain model, but sometimes the user must design and program such functions himself by using the basic operations and e.g. the macro language of the software. A good example of a set of implemented analysis functions can be found in the implementations of map algebra (14) in the GIS products of ESRI and Intergraph. Map algebra includes both basic operations and analysis functions as "functions" and the user can create more advanced functions by using the defined language.

Applications is the most complicated level of approaching GIS analysis. In an application we have certain type of source information and a certain process consisting of several analysis functions. Laurini (8) has given two examples as graphical data flow diagrams in his book. Each application consist of several functions and solves a certain problem.

5. Basic operations on data models

Traditionally graphs are coded in matrices (route and distance matrices) (6). For example route matrix shows the connectivity relationships between nodes. In GIS software packages it is more typical to describe graphs by spatial data models expressed by nodes and lines (8). Basic operations on graphs are e.g.: transformations, "rubber-sheeting", point in line/segment, intersection of lines/segments, data model conversions, from geometry to topology, conflation, buffering, generalization.

Polygon can be simple, complex or non-connected (8). Basic operations on polygons are e.g.: centroid, area, point in polygon, intersection of a segment and a polygon, union and intersection of polygons, polygon interpolation, buffering, polygon overlay, transformations (from polygon to point), conflation, topological conversions.

Points are elementary parts of both polygons and graphs (nodes) but sometimes they are used as individual objects. Discrete points are described by individual coordinates. Basic operations on points and sets of points are e.g. (6): distance between two points, Manhattan distance between two points, the nearest neighbour of a given point in a point set, the mean centre in a point set, standard distance in a point set.

Many of the vector based basic operations like distances, area, point in polygon, polygon overlay can be implemented as cell operations. According to Tomlin (14) matrix operations can be classified into operations on individual locations (local), operations within neighbourhoods (focal) and operations within zones (zonal). Examples of them are: local difference, local maximum (between two layers) and reclassification (local); focal maximum (in neighbourhood) and diversity, dominancy (focal); zonal sum, zonal maximum.

6 Analysis functions on data models

Analysis functions are composed by using the basic operations. Several analysis functions can be found ready made in GIS software packages. For some reason, however, e.g. network optimization problems are more frequently programmed from scratch than by using GIS functions. Analysis functions on point sets are typically not very well represented in the GIS software products. Many users claim for this lack. Analysis functions on matrices is quite well covered by the implementations of map algebra.

Examples of analysis functions on graphs: shortest path, travelling salesman problem. Examples of analysis on polygons: map overlay, districting. Examples of analysis on points and point sets: degree of clustering in a point set, triangulation, Thiessen polygons. Examples of analysis on matrices: map overlay, travel costs, visibility.

7 Applications on data models

Application means here a case in a certain planning or decision making situation. Each application must be analysed and solved individually. There are no ready-made solutions for applications. An example of GIS analysis application is described in Laurini (8) where best potential sites (for building) are analysed. The analysis is based on information about soil type, land cover, highways, services like schools and fire stations. This example is describe both on vector and raster data models. Raster-vector-raster conversion possibilities of course makes it possible to combine data from different sources in one application. In practise it is however typical to use only either vector or raster data models in one analysis application. This practical point is one good reason to take a data model oriented view on the analysis techniques.

Visual analysis can be combined with both raster and vector data models. Many problems could be easier to solve if the human eye-brain combination could be utilized. As an example to calculate point-in-polygon problem for five thousand polygons is a nice problem for a traditional single processor computer: several computations of line intersections or sectors and angles. For a human visual perception system it is quite another type of problem: detection, descrimination, identification and recognition of the point as well as the polygon around it. The intuitive reasoning leads to think that the visual analysis might be more easy and light. It is impossible absolutely to compare the complexity of human brain and computer work but at least we can compare them from the practical implementation point of view. The possibilities of visual analysis should always be kept in mind when a practical analysis application is to be designed and implemented.

After making several analysis functions the final result is achieved. If this result is to be used in decision making the decision makers should be very much aware about the data processing methods and their effects on the result. Analysis can be compared to visualization in its many possibilities to tell lies ("How to lie with maps" (11).

8 Conclusions

The study of data models and the analysis methods on them reveals that there are several lacks of functionality in GIS software products. There is only very few analysis possibilities for point sets. Traditionally GIS software packages have been oriented in polygon analysis. The use of raster analysis has been recently growing. But though network analysis modules exist in GIS packages they still are not used very much in practical network optimization problems. In addition to these there are the "underlying" data management problems in spatial analysis. Data management was not dealt with in this paper but it is self evidently one of the topics in the further research. All these facts show that there is a need for studying analysis techniques specially from the viewing point of data models. In this paper an introduction to this direction is given. The approach must still be developed in order to be useful as a stable framework.

9 References

(1) **Aronoff,S.,** Geographic information systems: A management perspective. WDL Publications.1991.

(2) **Bernhardsen,T.,** Geographic information systems, VIAK IT and Norwegian Mapping Authority.1992.

(3) **Burrough, P.A.,** Principles of geographic information systems for land resources assesment. Clarendon Press. Oxford.1990.

(4) **Cromley,R.,** Digital cartography. Prentice Hall. 1992.

(5) **Date,C.J.,** An introduction to database systems.Addison-Wesley Publishing Company.1977.

(6) **Davis,P.,** Data description and presentation.Science in geography 3. Oxford University Press. 1979.

(7) **FitzGerald,B.,P.,** Developments in geographical method.Science in geography 1. Oxford University Press.1982.

(8) **Laurini,R., Thompson,D.,** Fundamentals of spatial information systems. Academic Press.1992.

(9) **Marble,D.F.,** Geographic information systems: an overview, Introductory readings in geographic information systems, edited by Peuquet,D. and Marble,D.,F. Taylor&Francis. 1990.

(10) **Mayhew,S., Penny,A.,** The concise Oxford Dictionary of Geography. Oxford University Press.1992.

(11) **Monmonier,M.,** How to lie with maps. The University of Chicago Press.1991.

(12) **Peuquet,D.,** A conceptual framework and comparison of spatial data models, Cartographica vol. 21, nr.4,1984.

(13) **Steinitz,C.,** GIS: a personal historical perspective. GIS Europe, vol 2, nr 5, June 1993.

(14) **Tomlin,D.,** Geographic information systems and cartographic modeling. Prentice Hall. 1990.