TOWARDS AN ENUMERATION AND CLASSIFICATION OF GIS FUNCTIONS

Michael F. Goodchild

University of Western Ontario

London, Ontario, Canada N6A 5C2

ABSTRACT

The data structure used in a geographic information system can be seen as a model of spatial data. Primitive GIS data structures represent features and their attributes; more advanced data structures include explicit representation of the complex spatial relationships which may exist among objects. The analytic functions of a GIS can be seen as operations on the underlying data model. This raises the question addressed in this paper: is it possible to enumerate the set of possible GIS functions in a rigorous manner, and to organize them within some consistent conceptual framework? The paper proposes such a framework based on a relational model. A comprehensive set of GIS functions results, which can be compared to the functionalities of existing vendor products. It is also possible to describe the minimum requirements of a data base which would be capable of supporting the full range of functions.

1. INTRODUCTION

The ability of a Geographic Information System to analyze spatial data is frequently seen as a key element in its definition, and has often been used as the characteristic which distinguishes the GIS from systems whose primary objective is map production. It is widely accepted that systems for analysis require more completely developed data structures than those intended for map production; hence the distinction between "topological" and "cartographic" data structures. On the other hand there is very little agreement on the nature of such analysis; it is not clear what degree of analytic functionality is necessary to identify a system as a GIS, or what constitutes the totality of possible forms of analysis. The Canada Geographic Information System was clearly designed with two specific capabilities in mind - measurement of area and topological overlay of polygons - while more recent systems have begun to include more sophisticated forms of spatial analysis. In the past few years we have seen the first instances of location-allocation algorithms executed within a GIS as components of a spatial search application, and demand is clearly growing for more capabilities in other areas of spatial analysis.

Spatial analysis can be defined broadly as that set of analytical methods which requires access both to the attributes of the objects under study and to their locational information. It appears to be more extensive than statistical analysis, which we define for the purposes of this paper as requiring only attribute information, but the literature of spatial analysis provides no readily available principles of organization of the field, or any clear view of its extent. It seems clear that the GIS field will have to develop such principles itself if it requires them as a basis for designing appropriate functionality in spatial analysis.

Besides allowing more comprehensive GIS design, a discussion of functionality in spatial analysis has a more immediate purpose. Goodchild and Rizzo (1) have discussed benchmarking and workload estimation for GIS acquisitions using a model which decomposes workload into individual GIS operations, and Goodchild (2) has described an application of the model. The list of operations used in this work was based on familiarity with the user interfaces of current vendor products, rather than on any comprehensive view of spatial analysis. A more general list based on clearly defined principles would be more internally consistent and less dependent on currently offered capabilities.

The present paper therefore aims to provide a comprehensive framework for the organization of spatial analysis techniques, in the context of GIS. It is hoped that this will serve two immediate purposes: first, to serve as a framework for substantial expansion of spatial analysis capabilities, and second to provide the basis for organizing a generic list of GIS functions.

2. GIS FUNCTIONALITY

Much of the workload of a production GIS is concerned with input and output of data; a large proportion of the costs of operation are input-related, at least in the first few years of operation. Moreover these costs have proven to be relatively fixed, and there is little prospect of dramatic reductions in costs in the immediate future. However the focus of the present paper is on analysis, and it is analytic capabilities which fundamentally distinguish a GIS. So although input and output functions are overwhelmingly important cost components in many systems, the present discussion will focus exclusively on manipulation and analysis.

The demands placed on a GIS appear to fall into two modes, although the distinction between them is often blurred. We define the <u>product</u> mode as a series of GIS operations designed to generate an information product, usually hard copy, in the form of a map, table or list, from which the user will later extract selected information. The product is usually more comprehensive than the uses to which it will be put; it might, for example, contain information for an entire map sheet, of which the user will later examine only a small area. There is usually some time period over which the product remains useful, allowing the generation of products to be scheduled on a regular basis independent of the actual pattern of use. So for example a forest resource manager might request quarterly updates of the forest resource map on which most of his or her decisions will be based, but would likely not require direct interaction with the system in the periods between arrival of new quarterly updates.

The <u>query</u> mode assumes that the user is able to formulate a specific question, for which the answer should be provided immediately, or within some reasonable response time, and is willing either to interact directly with the system, or to give the request to an interactive operator. The question will usually be designed to provide no more than the information actually needed. Thus two dimensions distinguish product from query modes:

- immediacy, in that the query mode requires interaction with the data base whereas the product mode assumes that queries are directed to some previously extracted hard copy, and

- domain, in that the query mode assumes that the limits of the required information are known, whereas the product mode assumes only that the product will be sufficiently broadly defined to answer all likely queries.

We have already argued that spatial analysis is more general than statistical analysis, because it requires access not only to attributes but also to associated locational information. Well-known software for statistical analysis offers both query and product modes. Packages such as SAS and SPSS have developed from a product orientation, partly, one suspects, because both extend back to the days of batch processing, and partly because the complexity and volume of output of many forms of analysis requires it. It would be impractical, for example, to conduct a factor analysis in query mode. On the other hand limited statistical analysis, in the form of totals, means or bivariate displays, are available in query mode in many packages, including the more recent versions of SAS and SPSS.

It seems that the most important factor dictating the choice of mode is the complexity of the analysis. Query seems more suitable for simple analyses; more complex ones are likely to require more voluminous output, to be more difficult to specify through the user interface, to involve substantial and perhaps unacceptable response times, and perhaps to require the use of more than one package. All of these factors seem to argue for a product mode of operation for spatial analysis using a GIS. It follows that demands for more sophisticated spatial analysis are likely to be satisfied by product-oriented systems, while query mode will continue to be associated with rather limited analytic functionality. For this reason the remainder of the paper adopts a product orientation. At the present time it seems that both product- and query-oriented systems are available from vendors; the relative importance of the two modes in the vendor's perception of the marketplace dictates both the breadth of functionality and those design characteristics associated with speed of response.

3. DATA MODEL

In order to build an organizing framework for spatial analytic methods it is necessary first to establish an underlying data model. This may or may not correspond to the structures actually employed in storing spatial data in the GIS; however it will be assumed that it represents a possible view of the data from the spatial-analytic user's perspective.

The primitive elements of spatial data, representing map features, are referred to as objects. We assume that the data model includes point, line and area primitives. In addition the model recognizes the raster form, an ordered regular tesselation with location implied by position in the scan order. A single map coverage may have several classes of objects of each type; churches and windmills may both be represented by point object classes.

All object classes may have unlimited numbers of attributes. Each attribute for a given class must be homogeneous, meaning that the value of the attribute for each object in the class must be drawn from the same set of permissible values. It may be useful to classify attributes into interval, ratio, ordinal or nominal scales; we assume that the conventional methods for describing and manipulating statistical data are available for the attribute table of each object class, for example through SPSS or SAS, or to a more limited extent through data base management systems such as ORACLE or INFO.

Associated with each type of object are a number of reserved fields, which have fixed meanings. These may be geometrical; for example for an area data type there might be reserved fields for measures of area and perimeter length. Other reserved fields have topological significance: each edge or arc of a line data type in a network might carry pointers to the nodes or point objects at each end, and to the polygons or area objects on each side; polygon objects carry pointers which allow retrieval of the line objects from which they are constructed. We assume that the data base codes topology at least to the extent of allowing retrieval of adjacencies between pairs of both area and line objects.

Note that the structure of fixed attribute tables implies that the data base can code 1:1 and n:1 relationships with ease, but not

1:n. For example it is easy to carry pointers from each edge to adjacent nodes in a network, but not from nodes to adjacent edges. However the data model being described is merely the user's view; the actual data structures used internally to represent topological relationships may well include 1:n mappings.

The model must also include the object-pair as a recognized data type, because of its fundamental importance in spatial analysis. It must be possible to take two classes of objects, which may or may not be of the same type, and create a new "object" of each unique pair. Object-pairs may combine any types, although pairs which include raster types seem unlikely to have much practical application. The reserved fields of object-pairs include pointers to the objects which created each pair, together with distance between the objects. Other attributes might include flows of goods, numbers of migrants or trips, or details of transportation links. Although object-pairs have associated distances, these must be evaluated using a metric as the object-pair has no physical existence, unlike a line object.

Higher levels of combination are possible, beginning with the object-triple and including pairs of object-pairs, but there appear to be few applications in spatial analysis; the triangular inequality seems to be the only well-known property of objects (or object-pairs) taken three at a time.

The object/attribute model provides a simple view of spatial data, and appears to be relatively comprehensive when augmented with the concept of object-pair. It allows conventional statistical packages to be seen as processors of single, rectangular attribute tables; the more complex hierarchical, network and relationship models are supported by numerous data base management systems but seem to have had insignificant influence on statistical analysis.

However there are examples of data structures which occur in certain areas of spatial analysis and which are not included in any obvious way in the model. One is the path, which results, for example, from the solution of a shortest path problem, and can be viewed as an ordered set of network links or edges. A path might be represented in several ways in the model: as a binary attribute of each link, indicating whether the link was or was not included in the path; as an attribute of each link interpreted as a pointer to next link; or as a binary attribute of an object-pair composed of links paired with links, indicating whether each pair of links is adjacent in the path. The first of the three alternatives would require a search to recover the correct order of links, whereas order is coded directly in the other two. The tree structure is another example; it results when the shortest path problem is solved from one origin to all possible destination nodes in a network; it is easy to represent paths from destination to origin as link or node attributes interpreted as pointers, but the reverse requires 1:n capabilities.

4. GIS OPERATIONS

Since spatial analysis is seen in this discussion as including statistical analysis, the focus of this section is on the additional operations which are necessary to support the former; the ability to carry out statistical operations is assumed to exist in the form of standard statistical package capabilities.

The proposed system of organization of GIS operations has two levels. At the highest, operations are classified according to whether they:

a) analyze the attributes of a single class of objects (statistical analysis);

b) analyze one class of objects using both locational and attribute information;

c) analyze the attributes of object-pairs;

d) analyze more than one class of objects;

e) create new object-pairs for one or two existing classes of objects; or

f) create a new class of objects from one or more existing classes of objects.

In cases of new object creation it is assumed that reserved fields will be filled appropriately; in all cases operations are assumed to generate additional attributes for the new or existing objects as appropriate.

The lower of the two levels of organization concerns the types of objects being processed in each class. Thus the creation of an area object type from a point type is distinct from creation of a line type from a point type, but both are examples of group (a) operations above.

The definitions derive from operations on the data model, rather than from the cartographic meaning of those operations. For example, the pairing of points in a class of objects representing churches might generate an object-pair in a group (e) operation, a reserved field recording the distance between each pair of churches, or might generate a line object class of straight line segments between each pair in a group (f) operation, with a reserved field of line length. The effects on the data model are quite different, although the operations are conceptually similar.

The next section of the discussion analyzes a number of example operations in terms of this framework, in increasing order of complexity. This is followed by a more extensive discussion of one particularly important operation, polygon overlay.

5. EXAMPLE OPERATIONS

5.1 Boolean Selection

The selection of a subset of objects from a class based on their attributes requires no access to locational information; it falls into class (a) above.

5.2 Nearest Neighbour Analysis

Various nearest neighbour statistics can be computed for a point object class by processing their locational information in a group (b) operation. Alternatively the same statistics can be computed from the distance reserved field of point-point object-pairs in a group (c) operation.

5.3 Spatial Autocorrelation

Calculation of spatial autocorrelation for a class of areas, using either the Geary or Moran coefficients (3), would be structured as a group (d) operation if weights are to be based on polygon adjacencies. The operation requires access to the area attribute for which spatial autocorrelation is to be computed, together with the adjacent polygons for each arc in the line object class representing polygon boundaries; two object types, areas and lines, are therefore required, together with the necessary topological links between them in the form of reserved fields.

5.4 Thiessen Polygons

Many GISs include the capability of creating Thiessen or Voronoi polygons from point objects, and the literature describes various generalizations (4), as well as numerous applications to spatial data handling. In vector mode the operation falls into group (f); the new area objects created would have standard reserved fields, and would likely include a pointer to the generating point as attribute.

5.5 Spatial Interaction Modelling

Openshaw (5) has commented on the growing gap between the level of theoretical development in many areas of spatial analysis, and the corresponding level of operationalization and practical application. One of the difficulties contributing to this appears to be the inability of standard statistical packages to handle the complex data models of many forms of spatial analysis. The calibration of a spatial interaction model is a case in point; it requires attributes of point objects representing destinations, area objects representing origins and point-area object-pairs representing trips. The need to handle multiple types of objects places this operation in group (d). The results would be returned as summary statistics or as new attributes of the existing objects; for example predicted numbers of trips would be a new attribute of the point-area object-pair.

5.6 Transportation Network Analysis

The large number of techniques for analysis of transportation networks provides a good example of the rapidity with which spatial analysis capabilities can expand when combined with a spatial data base. The origins and destinations of the system can be represented as point objects, and the connections between them either as lines with appropriate pointers or as point-point object-pairs, depending on whether flow is constrained to follow the links of a finite network. Goodchild and Lee (6) have recently programmed the following four classes of transportation algorithms as group (d) operations for the Electronic Atlas of Canada:

- shortest path analysis;
- optimum tour routing;
- location-allocation;
- transportation and transshipment problems.

Each algorithm receives node and link attributes from the data base and returns the results of the analysis in the same form, as new link or node attributes, allowing existing capabilities of the system to provide input and output support and other functionality.

5.7 Shortest Path Analysis

The finding of the shortest path through a network from an origin node to a destination falls into group (d), as it requires access to both point and line objects, representing nodes and links respectively. Results can be returned in several forms; the total distance from origin could be returned as a node attribute, and the shortest path could be returned either as pointers from node to next node, or from node to next link (link to next node would likely be available as a reserved field). A more developed version of the shortest path problem allows the addition of impedances in turning from one link to another at each node; this "turntable" might represent delays due to having to slow a vehicle, or might be used to prohibit illegal turns. The turntable can be represented as an attribute of a line-line object-pair.

6. POLYGON OVERLAY

The case of polygon overlay seems sufficiently rich and important to merit special attention. Superficially this is a group (f) operation, involving the creation of a new class of area objects from two (or more) existing classes. However the results of the operation can be returned in a number of ways:

- new area objects with reserved fields, and either the concatenated attributes of the generating areas, or pointers back to them;

- new point objects at the intersections of the existing area objects, each with four pointers to the areas whose common boundaries intersected;

- new line objects, being the old line objects broken at intersections and having either the attributes of the old line objects or pointers back to them;

- new attributes for the existing areas pointing to the new areas formed from them;

- a binary attribute of a new area-area object pair indicating whether or not the pair of areas intersect.

It is clear that not all of these will be relevant in any application, suggesting that the system designer might provide a list of appropriate options. However the set of information required to some extent dictates the algorithm to be used. This leads to a potential conflict between the need for consistency on the one hand, which places all polygon overlay operations together, and the need for operational efficiency on the other. A case in point is the window operation; the overlay of an area object (the window) on a set of area objects. Since the user is not interested in those objects falling outside the window, the system designer is able to provide a special algorithm for windowing which is generally faster than the full polygon overlay. Consistency with the organizing framework is lost if this algorithm is invoked by a different command, but not if the system is able to identify which algorithm to use in each case.

7. CONCLUSIONS

The true potential value of Geographic Information Systems lies in their ability to analyze spatial data using the techniques of spatial analysis. Although an enormous amount of development has taken place in spatial analysis in the past two or three decades, much of that work remains largely unapplied due to the inherent difficulties of processing spatial data, with its complexity of relationships. GIS provides the key to this problem through a consistent and relatively complete model of representation, and through the economies of scale inherent in the sharing of many input, output and housekeeping functions. GIS development now seems to have reached the stage where rapid expansion of spatial analytic capabilities is possible, in such fields as spatial search, spatial decision making and spatial interaction modelling. So for example the potential exists to apply many of the theoretical developments of the 1970s in spatial interaction modelling to retail site selection.

Since the set of techniques in spatial analysis is so large, a modular approach to development is clearly necessary, a pattern seen already in the TIN and NETWORK additions to the ARC/INFO data base. To be fully effective, modular development requires that the base be complete, and it appears from the analysis presented in this paper that that point has not yet been reached by any available system. When it is, we can expect very rapid expansion of capabilities, since there will be a major incentive for any theoretician or software developer working in the field of spatial analysis to provide implementations consistent with the data model; this pattern is evident in the history of SAS and more generally in third-party software development for the IBM-PC.

References

1. *GOODCHILD, M.F. & B.R. RIZZO*, 1987: Performance Evaluation and Workload Estimation for Geographic Information Systems. International Journal of Geographical Information Systems, 1: 67-76.

2. *GOODCHILD, M.F.*, 1987: Application of a GIS Benchmarking and Workload Estimation Model. Papers and Proceedings of Applied Geography Conferences, 10.

3. *CLIFF, A.D. & J.K. ORD*, 1981: Spatial Processes: Models and Applications. Pion, London.

4. *BOOTS, B.N.*, 1987: Modifying Thiessen Polygons. The Canadian Geographer 31: 160-8.

5. *OPENSHAW, S.*, 1987: Guest Editorial: An Automated Geographical Analysis System. Environment and Planning A 19: 431-6.

6. *GOODCHILD, M.F. & J. LEE*, 1987: Network Algorithms for the Electronic Atlas. Energy, Mines and Resources Canada, Ottawa: Final Report of Contract 2ER.23242-6-2385.