Topic Series 17

GIS Elementary Analysis Operations and Classification Functions

Up until now, we've be chiefly concerned about getting information into the database for our applications. The data selection and design of the database had to be driven in large part by the analysis operations that would be used in our work. Now we will start looking at some of the operations that got us into GIS in the first place; namely, the analysis functions.

There are a number of simple operations used in GIS for **database exploration (display and list operations), simple measurement and simple query.** Other functions that often involve analysis take on the characteristics of **maintenance functions** (basic editing, updating, tiling, recoding) in that for many of these functions, analysis is the first step in the maintenance procedures. You've already experienced some of these functions first-hand in using ArcView in lab exercises.

Database Exploration

The most basic of functions in GIS are those associated with showing us the unaltered data in the database. These functions take the form of the exploration capability we looked at in ArcView to simply **display a theme**. This, combined with use of the legend editor, gives us immediate information of the distribution of features in each theme and their relative relationships to features in other themes. We could **list the attributes** of the active theme by use of the **information button**.

The **steps involved in list operations** that are key components of database exploration capabilities:

- the list command is issued (1)
- list the coverages available is given (2)
- elect a coverage (theme) (3)
- ask to see a the attribute table (4)
- list records (5)
- list fields (items) (6)
- all of these things can be done like any other database operations without viewing the graphic features (7)
- display the graphic features in the coverage (8)
- zoom in or out to get the area of interest on the screen (9 & 10)
- simple query may be used by pointing at the feature and getting data back (11& 12)
- this represents the surface of the powerful link (**relational**) between the graphic and non-graphic components of the GIS. (13)

Much of what we typically do day-to-day in GIS involves these simple point and list operations. The various characteristics of **point**, **line and area features** can be examined. The more robust abilities of GIS revolve around relational database structures that let us look for entities or groups of entities based on their characteristics.

More on Relational Databases

Lets revisit some of the concepts of relational databases: the concepts of original field data and other information that can be attached to it (top) and also the ease of extraction of information (bottom). A relational database, unlike simple tabular data, allows us to formulate questions to narrow down our choices based on a specific set of conditions that are of interest.

We may select features based on relational queries to get at information that may be in error. This gives us editing tools that are not available with simple tables.
Examples of error checking would be to find problems and either automatically correct them or give the operator a field to fill in with the new correct value. (**overhead**)

- find text records that have numeric data (these need to be changed)
- find all text records that have misspelled entries
- search for a specific record, if not there, create
- search for an attribute (field), if not there, create and prompt to input
- find and display records with numeric attributes that are outside of a specific range
- find numeric records with text data

Database-Graphics Links (ArcView overhead)

GIS's provide for interactive links between the graphics and the attributes in the database. We have seen this first hand in our exercises in ArcView where we could either **select an object on the screen and ask for its attributes**, or we could build a query to find geographic features based on attributes then **display or highlight those selections** on the computer screen. A large portion of our daily use of GIS revolves around these types of capabilities.

Graphic Selection

Another useful tool in simple exploration is graphic selection. This is done in ArcView by **drawing a graphic** and selecting everything that is within it or possibly anything that touches it. Few GIS's provide for the second example of reducing area based on the bisecting line on a screen. This is both computationally difficult, and as we have seen on our exercises, imprecise because of the limitations placed on coordinate location based on the screen resolution. We can also drag and draw a **window** to select features.

Boolean Query

This involves properties comparisons reduced to answering questions that have (yes/no) answers only. The standard operators demonstrated in this example are: **AND, OR, and NOT.** In each case we ask (yes/no) of the attribute on each side of the Boolean operator then subject that result to the operator. In the first example, we need to know (yes/no) what blocks are NOT HI? (if it is HI then we are NOT interested in it, if it isn't HI then we are). Two sided comparisons work the same way but the condition is tested for each side and must be true for both sides to satisfy the selection criteria.

Measurements

One way we can do this is with the **ruler button** in ArcView to measure simple distances on the screen. Measurements along existing lines (**interconnected arcs**) rely on topological relationships of the arcs to determine distance along a **selected path** or one computer based on **network analysis**. Arc/Info also automatically keeps the measurement information for each arc (**length**) and the **area/perimeter** information for each polygon.

Because **length, polygon perimeter, and area** are created in Arc/Info and ArcView with topology, these **can be automatically accumulated through summaries** derived from queries that identify features of interest. For instance, we can get the length of all roads along a path to travel (**network**) or the area of all stand polygons that are pine plantations. We'll discuss more complex distance measurement constraints later in the course under the topic of networks and modeling.

An interesting point is that the default length of features in GIS is the **horizontal distance** and does not account for **changes in elevation**. This can be taken care of with some of the analytic features of GIS. **(overhead)**

Statistical Reports

These are a useful tool in GIS. We can use these to summarize data with a common attribute: for example, getting simple totals for all records (e.g. **length of roads**), or getting percentage distribution of a features within primary classes (example of Districts, bottom of graph).

Classification

Much of the functionality of GIS is rooted in the basic human desire to explain the organization of the world through **categorizing things that exhibit common attributes**, a process we call **classification**. For many, the ordered world (except for my daughter's room), is a much easier place to understand.

Many of you have dealt with classification systems already. Examples include: **(overhead)**

- soil classification (NRCS)
- forest cover types (SAF)
- general land-cover assessments (Anderson et al., 1976)
- vegetation (UNESCO)
- wetlands (Cowardin et al., 1979)

An example classification system for **vegetation** is the one developed by the **United Nations Educational, Scientific, and Cultural Organization (UNESCO)** for field description of vegetation worldwide. A commonly cited system used in remote sensing for **land-use/land-cover classification** was developed by **Anderson and others (1976)**. In this topic, we will cover basic operations associated with classification processes.

Reclassification (Recode)

These operations are used to:

1) reorganize data by logical groupings or
2) redefine attributes based on other information.

We might group soils in the same general families (A or B) by reassigning the attributes to a smaller number of groups. The results make these maps much easier to interpret. This procedure is often used to **prepare for coverage overlay operations or buffer zone development**. This is an attribute manipulation function that may be later reflected in the graphic database by subsequent **dissolve** or recombination of the graphic features based on the newly recoded attributes.

> **Dissolve operation** The way this is done in Arc/Info is to re-attribute first then the dissolve operation automatically eliminates lines between polygons with the same attribute. This may be done to simplify the data for display purposes or, more importantly, to restructure the data for more efficient use in decision making processes.

Reclassification can be performed on both **nominal** and **numeric** attributes. Nominal reclassification is as implied, a renaming function (i.e. **Corn and Beans** reclassified to a single **Crop** class). **Numeric reclassification** can be done through **math operators** (e.g. add/subtract/multiply/divide by a scalar constant: could also be considered modeling), **range grading** (assign **ranges of values** to a new value).

Neighborhood Functions

One way to reclassify vector polygons or cells in a raster database is through proximity or neighborhood analysis. The entities in the database can be reattributed based on a common attribute of immediately **adjacent entities** (**immediate neighborhoods**) or a **region (extended neighborhoods).** For example: the parcels in a single city block may be associated by a block code while all parcels in a subdivision (multiple blocks) may be attributed with the subdivision name. **(overhead)**

In this example, **immediate neighborhoods** are assumed to be either **entities that touch** or are closely associated by being **within a threshold distance** of each other. The result of our example may be as simple as attributing each parcel with the **common block letter** for future reference. These **closely associated blocks** could be fields or stands in a natural resource management compartment.

Note that neighborhood assignment, in some cases, **could rely on overlay functions** of GIS. For example, we may use a GIS **coverage of stands** with each stand considered as a neighborhood to determine the **number of RCW colonies** in each stand (the entities in the neighborhood) and output the **result as a stand (neighborhood) attribute**. To do this, we have to use an overlay operation to examine the stands and colonies coverages at the same time.

We might use **extended neighborhood** assessments to **group blocks within a region** or to determine the **density of critical habitat** for wildlife management.

Some functions that work in the raster domain are designed to look a groups of cells (**filter kernel**) while incrementing through the raster data set. A common filter kernel identifies the majority value within the kernel and outputs the result for the middle cell of the kernel into the output dataset. This, in essence, aggregates the rasters. (**overhead)**

Buffer Operations

Typically, buffer zones are used to extract and group features based on a distance function as applied to entities on which the buffer zone is based. Buffer zones can also be considered a **tool for reclassification** of other entities and as such, are also a type of **neighborhood function**.

Buffer zones can be defined around point, line and polygon features. Consider the examples: Bbuffer zones can be used to **mask information** of specific interest (land use near a railroad). They can also be used to aid in **proximity searches** such as flood zones. In realistic application, we would generate **variable buffers** based on runoff, stream flow, and terrain (all can vary along a stream corridor).

Note that buffers can be used to either select features within or outside of the buffer zone **like clipping/masking** operations that you'll see again in our discussion of overlay operations. Include can be modified to take the features strictly within the boundaries or include all features that cross through the boundary.

If we have different distance zones around the feature of interest, these are represented as concentric zones.