Vector Data Model: A Literature Review

Author: Qianni Wang
Student number: 2015112956

1. INTRODUCTION

The vector data model is a strategy for describing distinct features in GIS. In order to have a better understanding of the vector data model, we read the literature and find the introduction on the website. Through the integration of these materials, we can easily understand what the vector data model is and the usage of the vector data model, the development and the features of the vector data model. The vector data model is an important concept in the GIS system so when we truly understand it, we will have an opportunity to improve the weakness of current models and to promote the development of Geographic data processing.

2. REVIEW

2.1 Data Model in GIS

The vector data model is one of the two main types of GIS data models, so we should learn data model before deeply understand of the vector data model.

2.1.1 The definition of data model

A data model in GIS system is a mathematical construct to represent geographic objects or surfaces as data. It contains two types of information which are the attribute information and spatial information that constitute a geologic map. A data model may be highly conceptual, or it may describe the data structure for managing information within a specific hardware or software platform. (1) In either case, it is a central construct because it addresses the database design for geologic maps in GIS format.

There are two types of data models which are widely used in GIS software, the raster data model and the vector data model (2). We can zoom the pictures to distinct two types of model, if the edge is still crisp, it might be the vector data model.

2.1.2 Raster data model

The raster data model represents data using a grid of cells. And the rectangles and squares cells are usually used as raster girds (3). There is only one value in every cell which can be used to interpret to mean several different things. The raster models are usual for storing continuous da (4) ta, so it is widely used to describe the ocean, vegetation and animal distribution (5).

2.1.3 Vector data model

The vector data model represents data using the collection of points, lines, and polygons (6). Through 0 dimensional, 1 dimensional, 2 dimensional methods, objects located on the ground can be easily described. Although there are many relationships between different components to describe how those are joined, all vector data fundamentally consists of lists of co-ordinates. The vector models are useful for storing data that has discrete boundaries, such as country borders, land parcels, and streets (7).

2.1.4 Vector-Advantages and disadvantages

There are many differences between the vector data model and the raster data model. So they have both advantages and disadvantages when applying to a real problem.

Advantages:

- Vector data can better represent topographic features than the raster data model. Vector data models can represent all types of features with accuracy. Points, lines, and polygons, are accurate when defining the location and size of all topographic features.
- Vector data can contain information about topology which underlies a large number of GIS operations. And it can make the task more efficient which requires topological information.
- Graphic output is usually more aesthetically pleasing.
- Vector Data can be represented at its original resolution and form without generalization.

Disadvantages:

- Compared to a raster data model, vector requires more time and technology to input the data. (8) Errors are likely to occur when inputting vector data. For effective analysis, vector data must be converted into a topological structure. This is often processing intensive and usually requires extensive data cleaning.
- Continuous data, such as elevation data, is not effectively represented in vector form, usually substantial data generalization or interpolation is required for these data layers.
- Spatial analysis and filtering within polygons is impossible.

So when performing GIS analysis, it's important to think about the most appropriate data model to suit your needs.

2.2 The Development of Vector Data Model

The vector data model is only a general strategy for representing objects. There are dozens of physical data structures (file format) that organize vector geometry and attributes in different ways, with unique capabilities. So the development of vector data model can be reflected by the development data format. The following will introduce several mainstream data formats.

2.2.1 Coverage

Coverage is a vector data format introduced with *ArcInfo* in 1981. It is an older type of GIS geospatial format. It has no extension, just a set of folders. *Coverage* use feature classes, stored as points, arcs, polygons or annotation. Tics are geographic control points and help define the extent of the coverage.

It is a topological vector data model, but compare to the *Geodatabase*, it is simple and original.

2.2.2 Shapefile

Shapefile is a vector data format introduced with ArcView in 1993. The *Shapefile* is the most common geospatial file type we will encounter. All open source and commercial accept *Shapefile* as GIS formats. So it has become the industry standard. But when we make up a *Shapefile*, it will comprise at least 3 physical disk files and all of which must be present. These 3 physical disk files must exist. SHP is the feature geometry, SHX is the shape index position and DBF is the attribute data (9).

It is a non-topological vector data model.

2.2.3 Geodatabase

Geodatabase is a new format introduced with *Arcgis 8.0* in 2000. *Esri* created the file *Geodatabase* to be a container for storing multiple attribute tables, vector and raster data sets. Within a *Geodatabase*, *Shapefile* are referred to as feature classes. It can store more complex data such as networks, raster mosaics and feature data sets (10).

So we can find that the development of vector data model, from Coverage to many modern data formats. More information is integrated. With the development of database technology, the vector data model also evolved from relationship-oriented to object-oriented.

2.3 The Features of Different Vector Data Models and their strength and weakness

2.3.1 Coverage

Coverage is a topological data model. And it is a set which can include one or more feature classes. The advantages of *Coverage* are:

Spatial data can be associated with attribute data. The spatial data can be saved as binary file by a kind of encoding method. And the attribute data can be saved as many datasheet in the DBMS. They can be associated by a common Id code.

Moreover the topological relation between vector data can be preserved.

2.3.2 Shapefile

Shapefile is a non-topological data model. The advantages of *Shapefile* are:

It has characteristics of display at a high speed and easy to operate. This may be why the *Shapefile* data structure was developed for *ArcView* , a software program that was originally designed for data viewing rather than analysis.

Polygon features in *Shapefile* format can contain one or more parts, so that disjunct and overlapping features can be represented (11).

Moreover, *Shapefile* can easily be copied and do not require importing or exporting as other format files.

Also, The *Shapefile* specification is readily available, and a number of other software packages support it.

2.3.3 Geodatabase

With the development of IT technology, the database has been changed from the traditional file management to relational database management. So *Geodatabase* comes up to reflect the objects in real world. *Geodatabase* is the core data model in *Arcgis* which not only has the function to show but also to store data. It can achieve the physical storage of geographic data and support extended cross-platform storage models. In addition, *Geodatabase* is a transactional model for managing GIS workflows and a set of COM components that access data.

Geodatabase has 3 types. They are *Personal Geodatabase*, *File Geodatabase* and *ArcSDE Geodatabase.* Different types have different storage format, storage ability, users amount and supported platforms.

2.3.4 KML

KML stands for Keyhole Markup Language. *KML* format is XML-based and is primarily used for *Google Earth*. (12)*KML* was developed by *Keyhole Inc* which was later acquired by *Google*. *KMZ* replaced *KML* and became an international standard of the Open Geospatial Consortium in 2008 because it is a compression version of the file.

KMZ has the advantages. It allows any images you use – say custom icons, or images in your descriptions – to be zipped up within the *KMZ* file. That way you can share these details without having to reference the files through some link to the Internet.

2.3.5 TIN

TIN is the abbreviation of triangulated irregular network. It is a digital data structure used in GIS to represent the surface. A TIN is made up of irregularly distributed nodes and lines with three-dimensional coordinates (x, y, and z) that are arranged in a network of nonoverlapping triangles (13).

The first TIN program for GIS was written by W. Randolph Franklin, under the direction of David Douglas and Thomas Peucker (Poiker), at Simon Fraser University in 1973.

Like other vector data model, TIN has its advantages and disadvantages. Firstly, compare to DEM (a kind of raster data model called Digital Elevation Models) in mapping and analysis, the points of a TIN are distributed variably based on an algorithm that determines which points are most necessary to an accurate representation of the terrain. Secondly, in three dimensions, TIN has the advantage of being able to portray terrain. Thirdly, data input in TIN is flexible and fewer points need to be stored than in DEM with regularly distributed points. But sometimes TIN has its disadvantages that not suitable for all applications. When using TINs, details on the actual triangles may be lost. In addition, because TIN is linear, many edges will appear jagged, distorting the image.

2.4 The Future Trends of Vector Data Model

Different data models have different advantages, so we should choose the appropriate data model based on the real problem.

I think the vector data model will be changed in these areas, and those may be the future trends of vector data model.

- Use less space or use easier way to store the data.
- Reflect more information about the real world and will be a integration of more information.
- Focus more on 3D instead of 2D.
- There must be faster calculation speed and more vivid display method.

3. REFERENCE

[1]. **David R. Soller, Thomas M. Berg.** *The National Geologic Map Database Project: Overview and Progress.* s.l. : Digital Mapping Techniques '03 — Workshop Proceedings, 2003.

[2]. Wiki.gis. Wikipedia. http://wiki.gis.com/wiki/index.php/Vector_data_model#cite_note-3.

[3]. wiki.gis. Wikipedia. http://wiki.gis.com/wiki/index.php/Raster_data_model.

[4].jniemuth.hubns.https://jniemuth.hubns.net/gis520/basics-of-raster-data-using-raster-data-for-site-selection/.

[5]. slideplayer. http://slideplayer.com/slide/7674817/.

[6]. **Bolstad, Paul.** *GIS Fundamentals.* White Bear Lake : Minnesota: Eider Press, 2008. ISBN978-0-9717647-2-9.

[7]. slideplaye-GIS data structure. http://slideplayer.com/slide/5114248/.

[8]. **LoK. W. YeungC.P.,Albert.** Concepts and Techniques of Geographic Information Systems.Prentice Hall, 2002.

[9]. **Esri.** Shapefile Technical Description White Paper. Environmental Systems Research Institute, 1998.

[10]. wenku.baidu. https://wenku.baidu.com/view/c3d0e8c608a1284ac850430d.html.

[11]. wiki-shapefile. wikipedia. https://en.wikipedia.org/wiki/Shapefile.
[12]. gearthblog. https://www.gearthblog.com/blog/archives/2005/09/google_earth_fi.html.
[13]. wikipedia-TIN. https://en.wikipedia.org/wiki/Triangulated_irregular_network.
[14]. gisgeography. http://gisgeography.com/gis-formats/.