GIS Conceptual Data Model

XIAO Lebin ZHONG Ershun LIU Jiyuan SONG Guanfu

1 Two typical GIS data models

1.1 Topological relation data model

Commercial GIS software in the early age mostly uses the data model based on "node—arc—polygon" topological relation, which is called topological relation data model. On the basis of topological data model, some software stores spatial data and attribute data separately. For example, position coordinates are stored into the file system, but topological attribute and other attributes are stored into two-dimensional tables of the relation database system (RDBMS). Other software stores coordinate and attribute data into various tables of relation database uniformly, and one record responds to one geometric element, such as point, line, etc.

Topological relation data model organizes and stores each geometric element based on topological relation, and their coordinate storage has hierarchy dependence relation of "polygon—arc—node". The advantages of this model are that the data structure is compact, that the topological relation is clear, and that the topological relation stored in a system in advance can increase the efficiency of the system in the aspects of topological query and network analysis. But there still exist some disadvantages in topological relation data model.

1) Operation efficiency is low for single geographic entity. Topological data model faces the entire space and emphasizes the relation among geometric elements, and without paying enough attention to the fact that the entire and independent geographic entity exists in the form of individual. Therefore, when a geographic entity is added, removed, or modified, a series of files and relation database tables is involved, which makes...
system management become complex and decreases the work efficiency of the system.

2) It is difficult to represent complex geographic entity. Because of the requirement of topological relation organization, a complete simple entity needs to be decomposed into many geometric elements in topological relation model sometimes. Complex geographic entity comprises many simple entities, and it is usually broken down naturally. The entire organization character of topological data model determines that it cannot represent this organic aggregation comprising many independent entities efficiently.

3) It is difficult to query fast and make complex spatial analyses. In topological relation data model, geographic entity is decomposed into such basic geometric elements as point, line and polygon, and they are stored in different files and RDBMS tables. Therefore, the operation, query, and analysis involving independent geographic entities will take a lot of CPU time, especially in the complex spatial analysis aspect of large area.

4) Local updating is difficult to proceed, and the system can not be easily maintained and extended. Since the basic geometric element (node, arc, or polygon) is used as the unit for data organization and storage of geographic space, and the complex topological relation stored in system is the data foundation of GIS work. When some entities are updated in local parts, an entire layer of topological relation has to be rebuilt. In such a system, a little change in local parts will result in the whole layer adjustment.

It is worth mentioning that topological relation data model can be realized by the object-oriented methods, but here this object is such a geometric element that cannot be segmented by other elements. And it is usually a part of independent geographic entity, but not an entire independent geographic entity. This is one of the essential different characters between the topological relation data model and entity-oriented data model that will be discussed next.

1.2 Entity-oriented data model

Entity-oriented data model emphasizes that the single, entire and independent geographic entity is used as the basic unit of data organization and storage.

In opposition to the above topological model, this model takes independent and entire entity with geographic meaning as the basic unit to represent geographic space. When executing concrete data organization and storage, coordinate data and attribute data of each entity can be stored in file system and relation database separately. (If a part of topology is established, then topological relation is also stored in RDBMS tables.) Also, they can be stored in relation database uniformly. (Coordinate data and attribute data can be placed in to the same table, and they can be placed into two tables as well.)

Entity-oriented data model is realized generally by using object-oriented software developing method. Each object (independent geographic entity) has its independent attribute (including coordinate data) and its own operation methods. It can finish some operations by itself. Even though the entity-oriented data model can organize data in terms of topological relation, the model discussed by the authors emphasizes the independent character of objects’ coordinate storages (especially the coordinate storage of region and line entity), which is the essential difference from topological relation model. This model can overcome some weakness of topological relation data model, and has some advantages. More importantly, it can conveniently organize and construct any complex geographic entities required by customers, which fits human’s thinking pattern of looking on the real world and can be easily understood and accepted. At the same time, the entity-oriented data model has the advantages of easily maintaining and extending system.

This model is the latest data model used by the currently popular GIS software, but it still has some disadvantages.

1) Topological relation needs to set up temporarily. The entity-oriented data model takes single geographic entity as its basic representation unit, and does not take topological relation as a basis to organize data, store geographic entity, and represent geographic space. Therefore, the topological rela-
tion does not exist at the very beginning, but various topological relations can be deduced temporarily when necessary, which will cost certain system resources.

2) The efficiency of dynamic segment and network analysis is decreased. In the topological relation chain of node—arc—polygon, there are four topological relation tables: node—arc, arc—node, arc—polygon, and polygon—arc. With these four relation tables, we can find out topological attribute of any node, arc, and polygon, and carry on topological analysis, such as dynamic segment, network analysis, etc. GIS based on topological data model can do it easily. Since the space consumption of the system will increase several times if the four topological tables are stored, in some softwares, only two tables of them are stored, or the tables of arc—node and arc—polygon are combined into one. The discarded tables can be educed through the existing tables. Though in this situation, when involving query and analysis based on topological relation the GIS based on topological data model still has pretty high efficiency, and the entity-oriented data model needs to set up topological relation temporarily will decrease the efficiency of topological query and analysis naturally. Of course, well-established topological relation can be stored for future use.

3) Common point and line among entities are stored repeatedly. Since the entity-oriented data model takes the single geographic entity as the basic unit to organize data and represent space, every geographic entity is stored integrally and independently with its point coordinate chain. When storing coordinates, each object is stored independently, not relying on other objects any more, which will result in repeatedly storing common points and lines among entities.

4) It is difficult to execute data management, spatial analysis, and data processing at the geometric element level. Geometric elements are such simple figures as point, arc, polygon, etc. Sometimes, the attribute difference of each geometric element constituting an entity is so great that it needs to be processed in more detailed level than geographic entity level—the level of geometric element. Topological data model can directly process this case. But the entity-oriented data model needs to carry on positioning and decomposing related geographic entities first, thus the efficiency of the system in this aspect will decrease. From the above analysis, we can easily find out that in this model all attributes of geometric elements constituting an entity are the same, ignoring the attribute difference among geometric elements of an entity, which directly results in the difficulty to operate geometric elements when executing processing, query and analysis.

5) It is difficult to execute topological query and analysis across different layers. If this problem is in the topological relation model, it can be solved easily. Since the adjoining elements of each element have existed, topologically-related elements can be tracked down to get their geographic attributes. But for the entity-oriented data model, the problem cannot be solved effectively. When temporarily creating topological relation, geographic attributes across different layers could not be created automatically. It can be solved if further processing is executed. But, obviously, the efficiency of this method is low.

1.3 Common disadvantages of two models

1.3.1 The layers of these two models could not meet the requirement of the holistic character of the real world very well.

Both early GIS and current GIS represent geographic entities as geometric objects with classified attribute, and then organize, store, modify, and display them by using the concept “layer”. Layering almost becomes a necessary basic character of GIS.

The idea of GIS layering gives us great help in the management of geographic object, which has been widely accepted by developers and users in actual application. On the other hand in the real life, the geographic world perceived by users is a geographic entity set, including highway, construction, mountain, tribe, and immigrant areas, etc., but not the data layers. GIS data model should directly reflect this perception\(^{[11]}\). We know that the concept of layering means roughly separating layers accord-
ing to the existing knowledge and experience for the real world. It may meet the short-term need pretty well, but it is uncertain to meet other long-term requirements presented by some new application. Therefore, the layered system for only one purpose hardly meets other purposes, which results in the decrease of system universal capacity. Moreover, the concept of layering enables objects closely connecting with each other before to be stored separately, which results in complex operations and analyses in low efficiency.

In brief, the concept of layering in GIS enables the real world to become a series of coverage layers according to some stiff boundaries, and these classifications cannot reflect the real world completely. Compared with photos, there is large information loss obviously. But layering helps GIS data organization and information selection greatly. It takes time to consider how to solve this contradiction.

1.3.2 Ignoring the semantic relation among geographic entities

Traditional GIS mainly focuses on expressing the geometric composition of geographic character, and its semantic relation and internal relation are neglected usually. This disadvantage affects GIS spatial analysis ability greatly. Therefore, established GIS becomes a system of spatial data storage and management only with simple function, and it is hard to carry out advanced spatial analyses and to make decision or solution directly. The authors believe that ignoring semantic relation means ignoring some laws of geographic phenomena. Since all geographic spatial entities distribute in certain geographic systems, and the correlation of geographic entities comes into being the intension of spatial distribution, form, structure, and laws, etc, ignoring correlation (including semantic relation) is ignoring geographic essential laws of course.

1.3.3 Traditional data model cannot represent some special geographic phenomena sufficiently.

Traditional data model cannot represent some special geographic phenomena sufficiently, and the expressions of fuzzy and uncertain phenomena do not attract much attention (traditional GIS uses binary logic to process most of problems). It uses two opposite geographic data models to simulate natural phenomena: precise object model and continuous field model, or called object model and field model. Precise object model has definite space boundary, topological relation, and clearly defines attribute collection. A continuous field phenomenon is believed as continuous field that is usually expressed by smooth mathematical surface time and space field. These two models are the two kinds of extreme abstract expressions of the real world. Many researches indicate that the internal attribute of mapped soil unit is not always even, and the boundaries between different soil units, geological units, and vegetation units are not always clear. Some researches indicate that mapped continuous field is interrupted by unexpected discontinuities sometimes. All above indicate that traditional precise object model and continuous field model cannot express the geographic space very well, which results in the loss of information.

2 Understanding geographic space

GIS cannot work very well when simulating spatial process and spatial interaction, etc. The main reason is that the existing GIS data model is lack of the genuine description for geographic space. In addition, spatial analysis is the GIS’s unique function different from all other systems, and the basis of GIS is geographic space. Since the understanding for the concept of “geographic space” will affect GIS data model and spatial analysis process, it is necessary to further study and understand the geographic space concept.

2.1 Geographic space

The authors think that since geography is the science of studying spatial distribution law on the Earth’s surface, the space in geography is a relation defined on the object collection of earth surface. There are many kinds of relations between objects. Physical distance is only one of the measurements of these relations. Defining one relation means defining one space naturally. But this space
relates with geometric relation again. Because geometric relation is the basic one of all relations, most of GIS emphasize space position and topological relation. This means that geographic space is a relative space and an object arranging collection (these objects have precise spatial position), and that it focuses on macroscopical spatial distribution and correlation among objects (each geographic object being a node or carrier of connection). Topological relation is one of these relations. At the same time, if geographic space needs to be positioned precisely on the Earth, Euler space and absolute position in earth coordinate system must be introduced into geographic space. By uniforming geographic space and Euler space together, the macroscopical character of geographic phenomena and precise character of space position will be connected closely. In the geographic space, macroscopical character is mainly represented by the topological and semantic relation among geographic objects (represented by data model of geographic space), and the carriers are those single geographic objects (represented by data structure of single object) with precise position and connection function acting as nodes.

2.2 Three-dimensional character of geographic space

Geographic space is a three-dimensional space in nature. In the past several decades, the rapid development and wide application of two-dimensional mapping and GIS enables people in different areas to accept the conceptual data model of the two-dimensional project simplified from the three-dimensional actual world and geographic space. With the further development of application and the practical requirement, the defects of two-dimensional GIS simplifying world and space are revealed gradually. At present researchers and developers on GIS have to re-consider the essential three-dimensional character of geographic space and a series of processing methods under the three-dimentional space conceptual data model. Geographic space should have the following three-dimensional features different from two-dimensional space:

1) The third dimension information, i.e., vertical coordinate information, is added into geometric coordinates.

2) The addition of vertical coordinate information results in complication of spatial topological relation, and especially no matter whether it is zero dimensional, one-dimensional, two-dimensional, or the three-dimensional object, there exists complex topological relation in vertical direction; if two-dimensional topological relation means the round radiation extension in plane, then the three dimensional topological relation means the global unlimited dimension extension in the three-dimensional space.

3) The three dimensional object in the three-dimensional geographic space has rich internal information (such as attribute distribution, and structure form, etc.).

2.3 multi-scales and temporal characters of geographic space

In fact, the above geographic space contains the concept of space scale. If the geographic space involved in GIS focuses on the study of macro-holistic structure and rough topological relation, then the scale is large. Conversely, if the application purpose focuses on absolute position and precise topological relation, then the scale is small. So is the geographic time scale. Geographic space scale, scope, time scale, and time scope all relate to the studied problems and geographic area. Different geoscientific problems have different geographic space and geographic time[7]. Therefore, maybe two kinds of coordinate spaces are needed in GIS. One is the space having relatively precise position and topological information (simply called precise space), and the other is space focusing on macro-geographic phenomena (simply called rough space). This means that the geographic database should act as a precise basis layer and have the ability to derive many different kinds of scale data, which is the foundation for GIS to process multi-scales analyses according to different needs.

3 Object-oriented holistic data model

The geographic space concept discussed above
tells us that only is the geographic world represented by the object-oriented method, the established GIS can fit human's thinking habit. Here, the holistic data model does not deny the above "entity-oriented data model", but is the further development based on that model. These two kinds of data models are unitary. In addition to the advantages of entity-oriented model, the object-oriented holistic data model can overcome the defects of topological relation model very well, and it has its own special characters. It requires us to organize the geographic world by the unit of entity, and look upon the real world as a whole. That is to say, each entity not only has spatial position attribute and topological relation in geographic space, but also, more importantly, has logical semantic relation with other entities. Besides, it has time attribute. Concretely speaking, the entity-oriented holistic data model has the following points:

1) According to the human thinking habit to understand the geographic space as a relative object-based space. The geographic space is a relation defined on object collection of the Earth's surface.

2) To consider geographic space from the viewpoint of holistic theory, in addition to studying geometric location and topological relation of objects, semantic relation among objects should be emphasized as well. Time, attribute and space, these three attributes occupy the same important positions. We can call geographic entity as "feature" in the holistic data model. Here, the concept of feature emphasizes the organic connection with other features (in aspects of time, space, and logic relation), and a feature is an element of a whole.

3) Even though holistic data model requires us to look upon the real world as a whole, when detailed data organization is executed, numerous geographic entities need to be layered. Layering is necessary, but a layered system for only one purpose can hardly satisfy other requirements. Therefore, supporting the efficient operation of maintaining layer, such as adding and removing objects conveniently from layers, is more important than providing a universal layered system. Compound layers have the meaning of layering not according to object’s dimension, and it can represent the holistic character of the real world very well. It offers an independent space of converged storage and interactive operation for those related objects in different layers or those different types of objects that users are interested in. Holistic data model should enforce the function of compound layers particularly, enable users to add, remove, modify, and query any kinds of geographic entities (point, line, region, and complex entity) freely, and carry on powerful space analyses simultaneously.

4) Even though traditional GIS usually represents the object model with vector structure, and represents the field model with raster structure[^4,^8], in fact, the two kinds of models, which seem opposite to each other, can be united into object-oriented frame, and the uncertain and fuzzy information can also be attached into the attribute collection. Object-oriented method as a kind of frame can describe both the object model and the field model[^8,^9]. The authors suggest that field objects

![Diagram](image-url)
should be established when developing a GIS software, and the internal structure of field object is quite different from common entity objects. A field is an object, in which the elements have the same attribute, and are related to each other closely.

5) A space object is in the three dimensional space, and has multi-scales character.

Summarizing the above analyses, we have designed a simple object-oriented holistic GIS data model (Fig. 1).

4 Conclusion

Data model is the spirit and key of GIS. In order to enable the developed GIS to reflect geographic space precisely and satisfy GIS users’ requirements in various areas completely, this paper analyzes and summarizes the advantages and defects of several early and current GIS data models, gives a new understanding for the concept of geographic space, presents several key points of designing holistic GIS data model, and designs a simple data model.

References

1 Usery E L (1996) A concept framework and fuzzy set implementation for geographic features. In: Burrough P A, Frank A U, eds. Geographic Objects with Indeterminate Boundaries. New York: Taylor & Francis, 71-85
2 Tang A Y, Adams T, Usery E L (1996) A spatial data model design for feature-based geographical information systems. International Journal of Geographical Information Systems, 10(5): 643-659
3 Li T J (1997) The research on distributed geographic spatial object model. [Dissertation]. Beijing: Peking University. (in Chinese)
4 Burrough P A (1996) Natural objects with indeterminate boundaries. In: Burrough P A, Frank A U, eds. Geographic Objects with Indeterminate Boundaries. New York: Taylor & Francis, 3-28
5 He J Z (1997) The research on data model of object-oriented GIS. [Dissertation]. Beijing: Peking University. (in Chinese)
6 Zhang X S (1997) The research on city spatial kinetic model and its application. [Dissertation]. Beijing: Institute of Geography, Chinese Academy of Sciences. (in Chinese)
7 Gong J H (1997) Geoscientific visualization—theory, technique and application. Postdoctoral Research Report. (in Chinese)
8 Worboys M F (1994) Object-oriented approaches to georeferenced information. International Journal of Geographical Information Systems, 8(4): 385-399
9 Worboys M F (1995) GIS: a computing perspective. New York: Taylor & Francis, 177-178
10 Freksa C, Barkowsky T (1996) On the relations between spatial concepts and geographic objects. In: Burrough P A, Frank A U, eds. Geographic Objects with Indeterminate Boundaries. New York: Taylor & Francis, 109-121