Research on the Method of Feature-Based Multi-scale Vector Data Model

Yibing Dong, Jianyu Yang, Chao Zhang, Dehai Zhu, Xingyue Tu, Xianzhe Qiao

HAL Id: hal-01348242
https://inria.hal.science/hal-01348242

Submitted on 22 Jul 2016

Research on the Method of Feature-based Multi-scale Vector Data Model

Yibing Dong[1] , Jianyu Yang[1,*], Chao Zhang[1], Dehai Zhu[1] ,Xinyue Tu[1], Xianzhe Qiao[1]

[1] College of Information and Electrical Engineering,
China Agricultural University, 100083, Beijing, China；* Author for correspondence
{ycjyyang@cau.edu.cn；294726701@qq.com}

Abstract. Multi-scale representation of spatial data is a research focus in GIS, while building multi-scale data model is a key to implementing multi-scale representation of vector data. In view of the shortcomings of existing multi-scale data model in geographical cognition and special analysis, this paper puts forward a method of feature-based, and studies on it qualitatively from definition, description and extraction. Compared to conventional single-scale E-R model, in this paper, the key strategies of building multi-scale conceptual model are put forward. Deeply study and analysis are applied on abstraction and expression of multiple geometric characteristics, of multi-attribution, and of semantic relation among different scales，and the design of feature-based multi-scale conceptual model is realized. Finally, the object-oriented multi-scale logic model is researched, which lays a theoretical foundation for building the feature-based multi-scale vector data model.

Key words: scale, feature-based, multi-scale representation, extended E-R model, object-oriented

1 Introduction

Multi-scale representation of spatial data refers to different expression forms of one geographical entity in geometry, topology and properties, which arise from different scales (resolutions) of geographical entity in the computer storage, analysis and description [1]. With the continuing expansion of application fields in GIS, there is a growing demand, not only to observe, understand and describe the geographic phenomena in different scales, but also to analyze, manage and express multiple geographic representations.

In recent years, as the development of wavelet compression and image pyramid model, the problems of multi-scale raster data representation and visualization have been solved. However, the relationships among spatial elements are complex, the multi-scale representation and management of vector data have been an open problem [2]. At present, the research of multi-scale vector data representation mainly focus on four aspects, including formal description of the multi-scale data model, design and

implementation of data structure, link and update mechanism between different levels and consistency maintenance of different expression versions. Specialists and scholars have long committed to improve previous algorithms and data structures, and made considerable achievements in reducing data redundancy, improving the efficiency of data access and maintaining consistency of spatial data [3-7]. In addition, a series of concept were proposed, such as "critical scale", "incremental", providing new ideas and methods for the representation of multi-scale vector data [8-10]. The existing method of multi-scale representation is broadly divided into four categories: multi-scale explicit storage, initial scale changes accumulation, Key scale function evolution and Senior scale automatic generalization. Their basic ideas, advantages and disadvantages are shown in Table 1.

Table 1. Methods and technical strategies of multi-scale representation

Technical Strategy	Basic Idea	Advantage and Disadvantage
Multi-scale explicit storage	Store the pretreated data in different scales explicitly, and when is needed, the data at a certain scale can be extracted immediately	*Advantage:* data structure is simple and easy to access data *Disadvantage:* data redundancy large and scale change discrete, cannot get data at any scale
Initial scale changes accumulation	Store the coarsest data representation and the change data between adjacent scales, therefore, data in any scale can be obtained by accumulated the change on initial data	*Advantage:* save storage spaces and improve access efficiency *Disadvantage:* scale change discrete, cannot get any scales of data and present the difficulty of maintaining consistency within the multiple versions of the maps
Key scale function evolution	Store a series of "key scale" data representation in hierarchical structure, and can derive any data representation between adjacent scales by interpolation or Morphing function	*Advantage:* data redundancy less and can get data at any scale *Disadvantage:* key scale of the different objects is inconsistent and difficult to select
Senior scale automatic generalization	Store the most detailed data representation and using on-line automatic generalization, achieve the export of data representation at any scale	*Advantage:* data redundancy at least and can get data at any scale *Disadvantage:* subject to the automatic generalization, this method cannot be achieved right now

As current GIS lack methods for on-line automated map generalization, the senior scale automatic generalization (the last method in Table 1) remains a challenging area of research, so we cannot get data at any scale [11]. On the other hand, the multi-scale explicit storage (the first method in Table 1) is widely used with simple structure, but the data storage is too large. Therefore, the initial scale changes accumulation (the second method in Table 1) and key scale function evolution (the third method in Table 1) are used as compromise methods for multi-scale vector data representation. Found by previous studies, specialists discuss multiple representation issues from a data-modeling point of view, with emphasis placed hierarchical structure to organize and store data at different scales. But the hierarchical structure has introduced additional problems, including ignore the essential characteristics and intrinsic links of geographical phenomena, educe the efficiency of spatial analysis and complex operation, and bring a certain of error [12, 13].

Based on the above, this paper introduces a "feature-based" thought into the progress of multi-scale modeling, and studies on it qualitatively from definition,

description and extraction. It also includes the design of multi-scale conceptual model, which not only overcome the insufficient of previous model, but also achieve the cognitive and description of the geographical world.

2 Qualitative Research of Multi-Scale Feature

2.1 Scale Definition

Typically, the scale is a measure of study object, including two parts of space scales and time scales. The spatial scale refers to the size of study scope (geographical scope) or the level of detail (size and level of geographic resolution). The time scale indicates the duration and frequency of study.

In different study areas, the meaning and application of scale is various. In cartography and GIS, scale is often defined as a ratio of the distance on the map and the actual length of a certain object. Among them, the distance is fixed, the relationship between "space scales" in the actual geographical space and "scale" is same as the relationship between denominator and business [10]. In this paper, large-scale refers to larger spatial areas, corresponding to low resolution. On the contrary, small-scale corresponds to smaller spatial areas and high resolution, which can reflect complex phenomena.

2.2 Feature Description

Traditionally, feature was a graphical representation of the real world. By the application of object-oriented methods in GIS, the concept of feature was re-defined as an abstract of real-world phenomena; using these features, all the geographical phenomena can be represented [14]. Geographic feature is an abstract of geographic entity, which includes geographic entity and further descriptions of them.

According to the definition, the feature F in the horizontal is divided into three parts: entity feature EF, attribute feature AF and relation feature RF. Thus, the feature can be shown as:

$$F=<EF, AF, RF> \tag{1}$$

Where EF is a description of the spatial position and geometry of geographical entities, AF describes the composition and characteristics of them, such as name, type, while RF is to describe the relationship between entities, including topological and semantic relations.

Through considering the domain of information representation, the spatial positions and distributions of entities can be represented in the geometric domain, as well as the semantic domain can explain the logic pattern of non-spatial information descriptions. Thus, in vertical, feature F is divided into geometric domain feature GDF and semantic domain feature SDF. Combine of two definitions above, features F can be complete expressed as:

$$GDF=< GEF, GAF, GRF> \qquad SDF=<SEF, SAF, SRF> \qquad \textbf{(2)}$$

Where, respectively, GEF, GAF, SRF represents the geometric characteristics of entities, attributes and relationships, and SEF, SAF, SRF is the semantic characteristics of entities, attributes and relationships.

2.3 Multi-Scale Feature Extraction

In different scales, a geographic entity will show different features. In this paper, the extraction method of multi-scale features is proposed, which can lay foundations for the establishment of multi-scale vector data model.

According to the definition and description of feature, some features at a variety of scales can be extracted, such as location, geometry, attribute, semantic relationship (shown in Figure 1). The extracted multi-scale features mainly show as the following three aspects:

(1) In different scales, a certain entity shows different geometries or abstract geometric types. For example, with increasing scale, the abstract geometry type of a building will change from surface to point.

(2) The topologic relations of a certain entity in different scales must be consistent, but the semantic relations between different geographical entities will change. For example, as the scale increases, two school buildings and a track field aggregate into a school.

(3) In different scales, the attributes expressed of one certain geometry will be different.
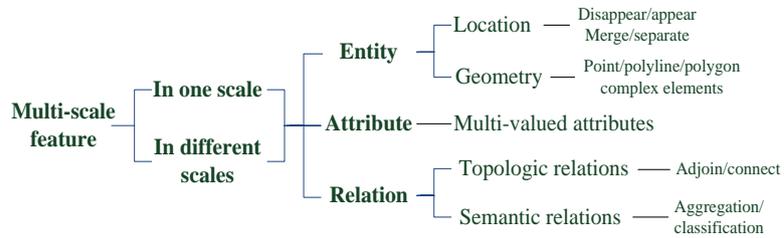
Fig. 1. Extraction of multi-scale features

3 Feature-Based Multi-scales Conceptual Modeling

In GIS, the process of building data model is an abstract to the real world. Like single-scale models，multi-scale models consist of three levels: conceptual model, logical model and physical model. The abstract level of three models is gradually deepened. Now, based on the multiplicity of geometry, attribute and relations in different scales, how to properly abstract, definite and describe the geographic entities

and their relations is still a problem, which must face in multi-scale conceptual modeling [15].

3.1 Choice of Conceptual Modeling Tool

The conceptual model is defined as a description of entities in the real world, and there are a lot of methods for conceptual modeling. Entity-Relation (E-R) model is the most famous method, and it can use diagram to represent the relationships between entities and attributes. A basic E-R model, consisted of entities, attributes and relations, can be used to represent geographical features. However, the traditional GIS conceptual model only supported single-level expression that cannot express complex changes between different scales [15].

In this study, the traditional E-R model is been expended to meet the demand of multi-scale representation and application. This paper also introduces a series of abstract concept, like generalization, aggregation, classification, to support multi-valued description of attributes, relations and others.

3.2 Multi-Scale Extended E-R Modeling

Multi-scale extended E-R model contains three elements: the entities, attributes and relations. According to the composition and law of multi-scale features, this paper is applied to study the multiple geometric features, multi-valued attributes and the change of semantic relations in different scales.

3.2.1 Abstraction and Expression of Multiple Geometrical Features

Borges said, the focus of building a conceptual model was to make a clear expression of geographical entity, and the type of abstract geometry would affect the spatial relationships between different entities [1]. In this paper, geographic entities were abstracted as point, polyline, polygon and complex elements formed by these simple elements. In any scale, entity can be represented by these elements. In addition, the concept of "abstract feature class" is proposed. So the geometric features of a certain entity in different scales can be unified in Figure 2.
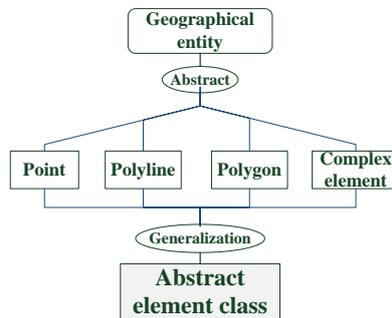

Fig. 2. Abstraction and expression of multiple geometrical features

3.2.2 Extraction and Description of Multi-valued Attributes

In order to reflect the abstract information of entities, in the stage of conceptual modeling, it is very important to extract the relationships between attribute domains or attribute value domains. With the abstraction level dropping in different scales, the attributes of a same entity is considered to be redefined, and the attribute values is consider to be refined, Parent said [16].

(a)   (b)   (c)
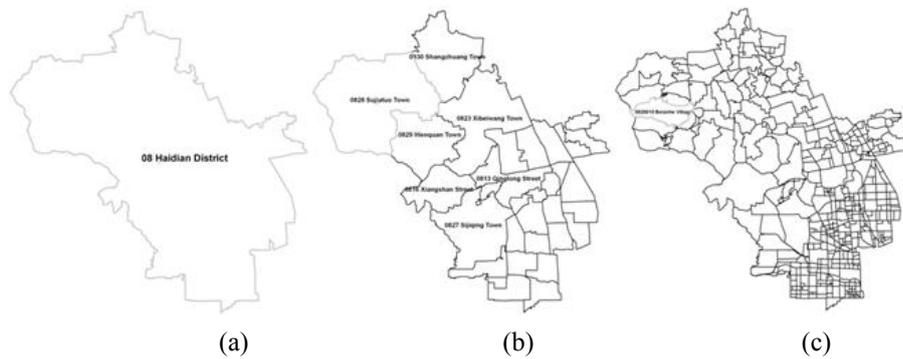
Fig. 3. Multi-scale representation of Haidian District in Beijing. The highlighted part in (a) is Haidian district, the highlighted part in (b) is Sujiatuo Town, and the highlighted part in (c) is Beianhe Village.

For example, in Beijing's Handian District, the gathered characteristics of geographic elements abstracted at different scales is described clearly (shown in Figure 3). Simultaneously, the attribute domains and attribute value domains of highlighted elements in each layer are extracted (Figure 4). Therefore, the fields of attribute domains (XZQM, JDMC, and JFMC) are different and unrelated, which indicates the geographic elements exhibit different attributes in a variety of scales.

Fig. 4. Extraction and description of multi-valued attributes. XZQM, JDMC, JFMC is the name field of highlighted element in (a), (b) and (c) respectively; and QHDM, JDDM, JFDM is the code field of highlighted element in them

However, the refining fields of attribute value domains (QHDM、JDDM、JFDM), not only reflect the different attributes of geographic elements represented in different scales, but also link multi-scale data together to establish multi-scale hierarchical structure. Besides, the attribute value domains can be used to extract multiple attributes.

3.2.3 Extraction and Representation Semantic Relations

Although the geometric shapes and details in different scales are different, the topologic relations between geographic entities should be consistent. Therefore, the semantic relations should be extracted to express the scaling relations between different scales. The four basic concepts of semantic relations are described as follows:

(1) Aggregation. Aggregation can combine primary objects with different characteristics into a higher level object—composite object. The composited object can derive some new characteristics from primary objects, but cannot own the characteristics inherited from primary objects. The relationships between composite objects and primary objects mainly represent as "part of". For example, the school is a composite object, and school buildings, roads, playgrounds are the parts of it.

(2) Generalization. Similar to aggregation, generalization is an operation that can extract general attributes or characteristics from common class to establish a high-level class. So the relationship between high-level class and common class can be represented as "is a".

(3) Classification. Classification is a method which can form a hierarchic class structure. In this structure, every class is consisting of objects with the same characteristics, and there is a "is instance of" relationship between object and its class.

(4) Association. In one scale, some objects can be combined into an abstract object. The relationship between objects and abstracted object is represented as "is member of".

(a)   (b)   (c)
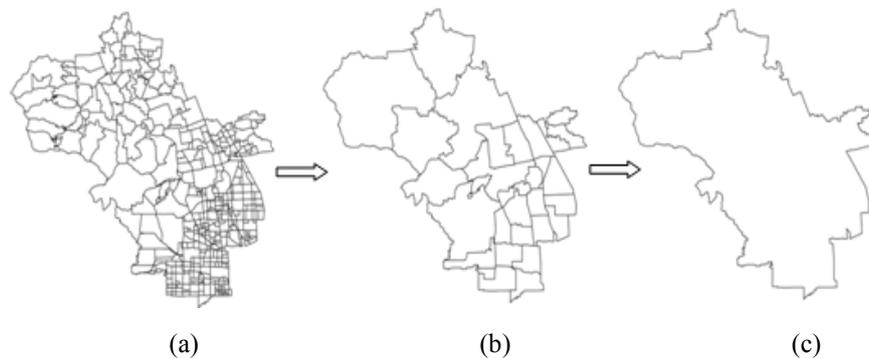
Fig. 5. Semantic relation of multiple geographical elements of Haidian District in Beijing

In this paper, some methods are applied to extract multiple semantic relations of Haidian District in Beijing (shown in Figure 5). For example, the elements in neighborhood layer (Figure a) are classified by street code (JDDM). Then the street layer (Figure b) can be achieved by generalizing all classes. In the same way, through classifying and generalizing the elements by zoning code in street layer (Figure b), the block layer (Figure c) can also be gained. Each element in the layer can be simple element or composite element aggregated from geographical entities with different characteristics. Using the four basic semantic relations and making a certain

expansion of them, it is possible to establish scaling relation between different scales and thus to connect data representations together.

4 Feature-based Object-Oriented Multi-Scale Logic Modeling

In object-oriented method, the concept of "class" is a collection of objects with common attributes and methods, where the object is a package of its stages and behaviors. For an object, the stage is a collection of attribute values and the behavior is a collection of methods. In this paper, the concept of feature and class are similar, through class the geographical features can described formally. Shown in Figure 6, the object is an instantiation of class, and the feature is an abstraction of entity. Therefore, feature-based multi-scale logic modeling can be achieved by object-oriented data model.
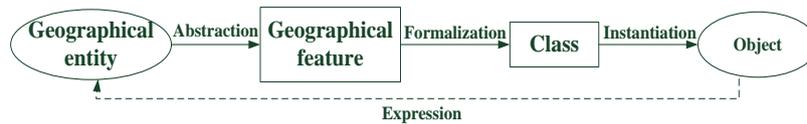

Fig. 6. Formal description of geographical feature

In multi-scale logic modeling, feature is defined as a unity of geographical data and functions, corresponding to the "class" of the object-oriented data model. The "object" described by "feature class" is a geographical entity. Combining the design ideas of feature-based, as well as the method of object-oriented, a unified multi-scale data model can be build from conceptual modeling to logic modeling.

In this study, the feature object FO is divided into geometry object GO and semantic object SO. GO is composed of the spatial locations and relations of geographic entities, which indicates the geometric features. Similar to GO, SO represents the semantic features, which include attributes, non-spatial relations and others (shown in Figure 7). The GO and SO are not independent of each other, and they are encapsulated into FO. Through this method, the features of geographical entities can be represented completely.
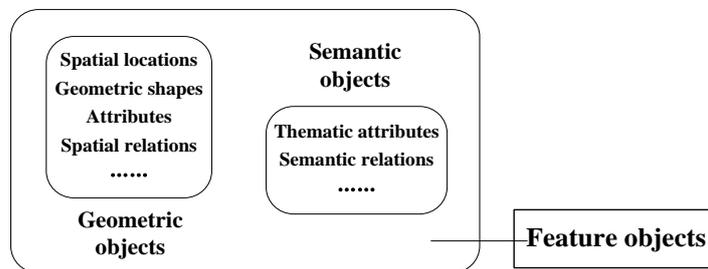

Fig.7. Composition of feature object

Therefore, the SO are extracted to organize spatial representations in different scales. The semantic relation in SO can be used as the basic description object. Through these semantic relations explained above, we can link a series of spatial representations together, and establish the spatial indexing mechanism efficiently by semantic features. Thus, the multi-scale logic model can be achieved preliminary.

5 Conclusions and Outlooks

The multi-scale vector data model is an important part of multi-scale spatial data representation. However, in terms of geographical cognition and spatial analysis, the existing multi-scale hierarchical data models had some lacks. For the requirements and applications of multi-scale data representation, this paper puts forward the thought of feature-based, and studies on it qualitatively from definition, description and extraction. The design of feature-based multi-scale conceptual model was realized by extended E-R model. Finally, the feature-based object-oriented multi-scale logic model was researched, which lays a theoretical foundation to building the feature-based multi-scale vector data model. From this study, we can draw the following conclusions:

(1) Qualitatively study and deeply analysis of multi-scale features, not only realize the cognition and representation of the real world, but also obtain a more integral abstraction and description of geographical entity to maintain the consistency of topological relations between different scales.

(2) The thought of feature-based and object-oriented is consistent. Using object-oriented method, the cognitive paradigm of geographical space and the representation of geographical data can be unified together naturally. Therefore, a unified multi-scale data model can be build from conceptual modeling to logic modeling.

Next step, the major works are: (1) Combined with experimental data, the quantitative research will be carried out, which will be extract multi-scale features and change rules of geographical entities in different scales. (2) Guided by the conceptual model in this paper, we will deeply study the implement of feature-based object-oriented logic model, including the establishment of object-oriented data structure and spatial indexing mechanism, the maintenance of level connectivity relation and consistency. The establishment of feature-based multi-scale data model will lay foundations for the representation of multi-scale vector data.

References

1. Lin Li, Fan Wu: Multi-scale Spatial Data Representation Model and Its Visualization. Science Press, Beijing (2005).
2. Changxiu Cheng, Feng Lu: A Bi-hieerarchical Multi-scale Model and Indexing Approach for Large Vector Data Representation. Journal of Image and Graphics. 14(6): 1012-1017 (2009).
3. Tinghua Ai, Oosterom P.V.: GAP-tree Extensions Based on Skeletons. Advances in Spatial Data Handling. 501-514 (2002).
4. Min Zhou, Michela B.: Efficiently Generating Multiple Representations for Web Mapping. Lecture Notes in Computer Science. 3833: 54-65 (2005).
5. Haunert J.H., Dilo A., Van Oosterom P.: Constrained set-up of the tGAP structure for progressive vector data transfer. Computers & Geosciences. 35(11): 2191-2203 (2009).
6. Bisheng Yang, Qingquan Li: An Algorithm for Progressive Transmission Vector Map Data over the WWW. Acta Geodaetica Et Cartographic Sinica. 34(4): 355-360 (2005).
7. Yingchao Ren: The Research of Vector Data Progressive Transmission. Institute of Remote Sensing Applications Chinese Academy of Sciences, Beijing (2007).
8. Sheng Z., Christopher J.: Design and Implementation of Multi-scale Databases. Lecture Notes in Computer Science. 2121: 365-384 (2001).
9. Jean-Michel F., Alain B., Fr éd éric B.,*et al.* : An Increment Based Model for Multi-resolution Geodata Management in a Mobile System. Lecture Notes in Computer Science. 3833: 42-53 (2005).
10. Tinghua Ai, Jianguo Chen: Key Issues of Multi-Scale Representation of Spatial Data. Geomatics and Information Science of Wuhan University. 30(4): 1-9 (2005).
11. Tinghua Ai, Jingzhong Li: The Lifespan Model of GIS Data Representation over Scale Space. Geomatics and Information Science of Wuhan University. 35(7): 757-781 (2010).
12. Sang Hun L., Kunwoo L.: Simultaneous and incremental feature-based multiresolution modeling with feature operations in part design. Computer Aided Design. 44(5): 457-483 (2012).
13. Wenjuan Li, Hongwei Li, Rupeng Liang,*et al.*: Research on Feature-based Spatial Data Model. Geomatics&Spatial Information Technology. 01, 42-45 (2010).
14. Zhangcai Yin: Featurre-based Data Models in GIS. Management Geological Science and Technology. 02, 50-53 (2002).
15. Yanhui Wang, Jun Chen, Jie Jian: Multi-Scale Conceptual Model for GIS Geographical Feature. Journal of China University of Mining & Technology. 32(4): 376-382 (2003).
16. Parent C.: Modeling Spatial Data in the MADS Conceptual Model. In the Proceedings of the International Symposium on Spatial Data Handling. Spring-Verlag, Berlin (1998).