Spatial Similarity Relations in Multi-scale Map Spaces: Basic Concepts and Potential Research Issues

Haowen Yan[1,2], Tao Liu[2], Weifang Yang[1,2], Jonathan Li[1]

1. Department of Geography & Environmental Management, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada.
Email: (h24yan, w38yang, junli)@uwaterloo.ca
2. Department of GIS, Lanzhou Jiaotong University, Lanzhou 730070, China.
Email: dlkc@mail.lzjtu.cn

1. Introduction

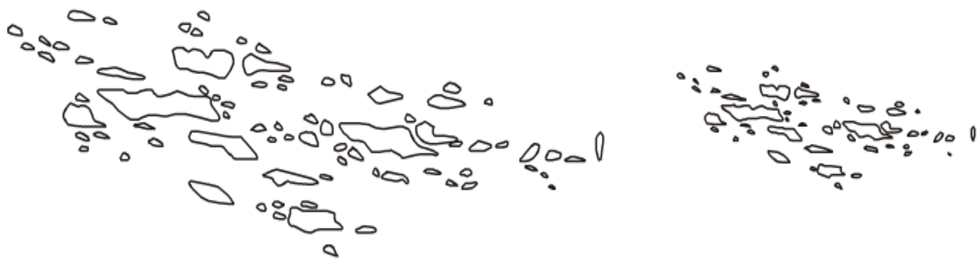
Multi-scale map databases fundamentally support the construction of cities', regions' and nations' spatial data infrastructure. They provide geographically spatial positioning bases for various location-based services in politics, economy, military, environment, traffic, transportation, and telecommunication, etc. Traditionally, multi-scale map databases are built manually or semi-automatically by means of the "multiple-version method". For example, to build digital map databases containing maps at scales 1:10K, 1:50K, 1:250K and 1:1M using the multiple-version method, cartographers firstly need to digitize and compile the maps at the four scales, respectively, and save the map data in individual map databases. The sum of the four databases constitutes a multi-scale database. Past research and practical applications have revealed that this method has at least the following disadvantages (Wang, 1993; Ruas, 2001):

(1) repeated storage of map data at different scales generate redundant data in multi-scale map databases and leads to the waste of computer memory spaces;

(2) repeatedly storage of the maps of a same region greatly increases the quantity of the databases, therefore makes the map data transmission via the Internet become difficult;

(3) consistency of map data can not be ensured due to repeated compilation and digitization of the maps at different scale of the same region; and

(4) renewal of the databases is time-consuming and uneconomical.



1:100K 1:250K

Fig.1 Similarity transformation in map generalization

A most prospective method that can overcome the above disadvantages is automated map generalization. In essence, map generalization is a kind of similarity

transformation in graphics and semantics. Take Fig.1 as an example: the island map at scale 1:250K is generated from the one at scale 1:100K. Although the graphics have been simplified in the process of scale change, the two maps are intuitively similar.
It is obvious that the similarity degree between a generalized map and its original one and the scale of the generalized map are functionally dependent on each other. The more the original map is generalized, the less similar the original map and the generalized result is, and the larger the scale change between the original map and the generalized result is. Nevertheless, no achievement has been made on the calculation of the similarity degree between a map and its generalized version (Yan, 2010). This hampers the automation of map generalization; because if similarity degrees are unknown, a map generalization system/software does not know to what extent an original map should be generalized if the resulting map scale is specified, and therefore it does not know when to stop the map generalization procedure.

2 Concepts

The fundamental concepts of spatial similarity relation, including its definition, features and classification are essential for researching on other relative issues; hence, they are discussed in this paper above all.

2.1 Definition

Seemingly, similarity is a very simple concept. People encounter and use similarity almost every second in daily life, for example, people can recognize familiar persons by their faces if they meet.
Similarity plays a crucial role in many fields in science (Bronstein et al., 2009). A typical example in geometry is "similar triangles": two triangles are similar if the three pairs of corresponding sides are proportional or two pairs of corresponding angles are congruent. In computer science, the definition of similarity, in many cases, is closely relative to character processing (e.g. comparing similarity of character strings). In pattern recognition, with a slight exaggeration, it may be true that all pattern recognition problems are based on finding methods for giving a quantitative interpretation of similarity, or equivalently, dissimilarity between objects (Bronstein et al., 2008). It is not easy to find a unique definition of similarity from existing literatures. Every field has its criterion to define similarity for the purpose of solving a group of problems.

> ***In geometry**, two objects are called similar if both of them have the same **shape**.*
>
> ***In computer sciences,** there are two important concepts that are closely related to similarity: similarity metrics and semantic similarity. Similarity metrics (also called string metrics) are a class of metrics that are used for measuring similarity (closeness) and dissimilarity (distance) between two character strings for approximate matching or comparison in fuzzy string searching. Semantic similarity (it is also known as semantic relatedness) is a concept used for assessing the likeness of the meaning/semantic content of a set of documents or terms within term lists by means of defining a metric (Budanitsky and Hirst, 2001).*
>
> ***In engineering**, similitude is a concept used for testing the similarity between two engineering models. An engineering model can be defined as "having similitude" with a real application on condition that they both share geometric similarity, kinematic similarity and dynamic similarity (Hubert, 2009). Similarity and similitude are interchangeably used in this context.*
>
> *Similarity in **psychology** refers to the psychological nearness or proximity of two mental representations.*

*Similarity does exist in **music**. There are a number of types of musical similarity that has been research (Toussaint, 2006), such as metrical structure similarity, rhythmic pattern similarity, section structure similarity, modality structure similarity, etc.*

Similarity is one of the basic research issues in Geo-Sciences (Nedas and Egenhofer, 2003); it is also called spatial similarity relation. Here, a definition of spatial similarity relations between objects in the map space is given.

Suppose that A_1 and A_2 are two objects in the geographic space. Their property sets are C_1 and C_2, and $C_1 \neq \Phi$ and $C_2 \neq \Phi$. If $C_1 \cap C_2 = C_\cap \neq \Phi$, $C_\cap$ is called the spatial similarity relations of object A_1 and object A_2.

The definition of spatial similarity relation in multi-scale map spaces can be:

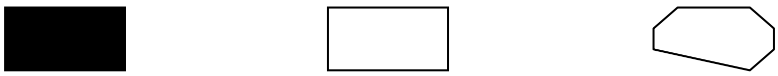
Suppose that A is an object in the geographic space. It is symbolized as A_1、A_2……A_k separately on the maps at scales S_1、S_2……S_k. The property sets of A_i (i=1, 2, ..., k) are C_1、C_2……C_k, and $C_i \neq \Phi$ (i=1, 2, ..., k). If $C_1 \cap C_2 …… \cap C_k = C_\cap \neq \Phi$, $C_\cap$ is called the spatial similarity relations of the multiple representations of object A in multi-scale map spaces.

2.2 Features

Similarity relations in multi-scale map spaces have the following five features (Tversky, 1977; Rodriguez and Egenhofer, 2004).

(1) Reflexivity: any object has similarity relations with itself.

(2)Symmetry: if object A has similarity relation with object B, B has the same similarity relation with A.



（a）settlement （b）vegetable land （c）vegetable land

Figure 2 Taking {shape, land type} as the properties to test similarity relations, the properties of (a), (b), and (c) are C_a={ rectangle, settlement}, C_b={rectangle, vegetable land}, and C_c={irregular polygon, vegetable land }. $C_a \cap C_b$={rectangle} and $C_b \cap C_c$={vegetable land} do not mean the objects in (a) and (c) have similarity relations, because $C_a \cap C_c = \Phi$.

(3)Non-transitivity: object A does not definitely have similarity relations with object C, though A has similarity relations with B and B has similarity relations with C (e.g. Figure 2).

S_1 S_2 S_3

Figure 3 the settlements at scales S_1, S_2 and S_3 obviously have similarity relations.

(4) Self-similarity in multi-scale map spaces: multi-scale representations of an object on maps have spatial similarity relations (e.g. Figure 3).

(5) Scale-dependence in multi-scale map spaces: suppose that original map scale is S_o, and target map scale is S_t, and spatial similarity degree is D_s, the relations among the

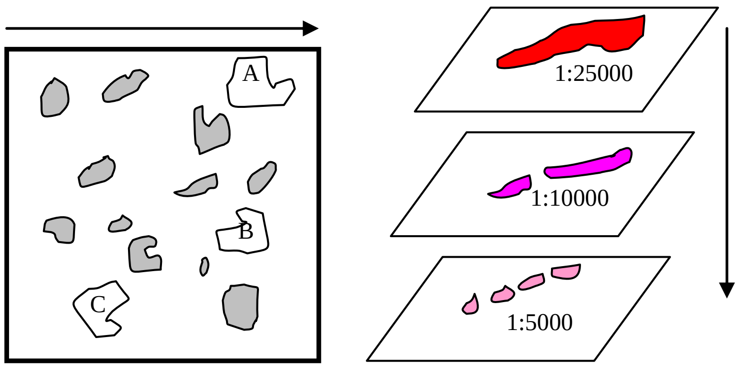
three variables may be:

$$D_s = f(S_c) \tag{1}$$

where, $S_c = S_o / S_t$.

This function states that the greater the scale change from the original map to a target map (S_c) is, the less the similarity degree (D_s) between the two maps should be. In other words, D_s monotonously increases with S_c. This relation is evident but not strictly proved.

2.3 Classification



(a) horizontal similarity relations; (b) perpendicular similarity relations.

Figure 4 A scale-based classification of spatial similarity relations.

Two rules should generally be obeyed in any classification, i.e. completeness and exclusiveness. Completeness means the union of all subsets of the sub-categories equals to the whole set; while exclusiveness means the intersection of every two subsets is empty. If map scale is taken as the classification criterion, two categories can be differentiated:

(1) horizontal similarity relation, and

(2) perpendicular similarity relations.

If objects are at same scale, their similarity relations are called horizontal similarity relations; whereas if objects are at different scales, their similarity relations are called perpendicular similarity relations (Figure 4).

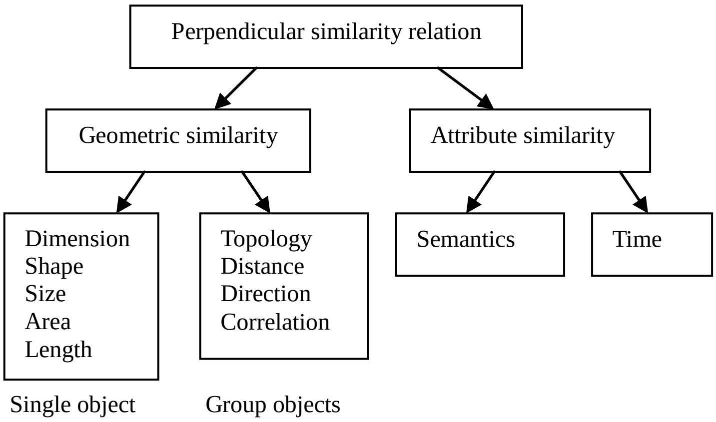


Figure 5 Classification of perpendicular similarity relation.

Perpendicular similarity relation can be further classified, taking geometric and attribute of objects as the criterion (Figure 5).

3 Potential research issues

At least the following topics related to spatial similarity relations need to be explored in multi-scale representation of map features.

• What affects human's judgment of similarity in geographic space? To be exact, the factors that take effects in human's judgment of spatial similarity relations need to be obtained, which lays the foundation for constructing models of spatial similarity degree calculation.

We will select appropriate factors based on our experiences and make questionnaire surveys by means of a lot of typical examples. The conclusion about the factor will be drawn from the calculations and analyses of the statistical data from the surveys.

• Calculation models of spatial similarity degrees between two individual spatial objects or two object groups on maps in multi-scale spaces is the most important and difficult task in spatial similarity relations. Here, spatial objects and object groups include individual linear/areal objects, point clusters and connected/separated polygon groups. The models should consider two types of similarity, i.e. geometric similarity and attribute similarity.

• The relations between similarity degree and map scale change in multi-scale representation of map features need to be quantitatively expressed. Formula (1) is only a qualitative description of their relations.

• It is necessary to compute the threshold values of some algorithms for map generalization, e.g. the distance tolerance in the Douglas-Peucker Algorithm, using the relations between similarity degree and map scale change, because the threshold values obviously have quantitative relations with the similarity degrees and map scale change in map generalization. This helps to reduce human interception so that the algorithms become fully automated.

4 Conclusion

This paper, after discussing the definitions of similarity in many other fields, presents a definition of spatial similarity relations in multi-scale map spaces using set theory; then it proposes five features and gives a classification system of spatial similarity relations; finally, it shows four potential research issues related to spatial similarity relations in multi-scale map spaces.

Spatial similarity relation is not a new concept in the family of spatial relations (i.e. topological, distance, direction, correlative and similarity relations), and it is of great important in multi-scale representation of map features; however, it has not been explored systematically by far. Our future research will focus on constructing the models for calculating similarity degrees among objects at multi-scale maps.

Acknowledgements

The work described in this paper is partially funded by the NSERC, Canada, and partially funded by the National Support Plan in Science and Technology (No. 2013BAB05B01), China.

References

Bronstein A.M., Bronstein M.M., and Kimmel R., 2008, *Numerical geometry of non-rigid shapes*, Berlin: Springer.

Bronstein A.M., Bronstein M.M., Bruckstein A.M., and Kimmel R., 2009, Partial Similarity of Objects, or How to Compare a Centaur to a Horse, *International Journal of Computer Vision*, 84: 163-183.

Budanitsky A. and Hirst G., 2001, Semantic distance in WordNet: An experimental, application-oriented evaluation of five measures, *Workshop on WordNet and Other Lexical Resources, Second meeting of the North American Chapter of the Association for Computational Linguistics*. Pittsburgh

Hubert C., 2009, Turbulent Air-water Flows in Hydraulic Structures: Dynamic Similarity and Scale Effects. *Environmental Fluid Mechanics*, 9(2):125–142.

Nedas K.A. and Egenhofer M.J., 2003, Spatial Similarity Queries with Logical Operators. *Proceedings of the 8th International Symposium, SSTD 2003*.

Ruas A. 2001, Automating the generalization of geographical data. *Proceedings of the 20th International Cartographic Conference*. Beijing, China, 1943–1953.

Toussaint G.T., 2006, A comparison of rhythmic dissimilarity measures. *FORMA*, 21(2): 129–149.

Tversky A., 1977, Features of similarity. *Psychological Review*, 84 (4): 327–352.

Wang J Y, 1993. *Principles of general map generalization*. Beijing: Surveying and Mapping Press. (in Chinese)

Yan H.W., 2010, Fundamental Theories of Spatial Similarity Relations in Multi-scale Map Spaces. *Chinese Geographical Science*, 20(1):18-22.

Rodriguez M.A., and Egenhofer M.J., 2004, Comparing geospatial entity classes: an asymmetric and context-dependent similarity measure. *International Journal of Geographic Information Science*, 18(3): 229–256.