Approach to calculating spatial similarity degree using map scale change of road networks in multi-scale map spaces

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

1. Department of GIS, Lanzhou Jiaotong University, Lanzhou 730070, China

2. Department of Geography & Environmental Management, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

Abstract: Quantitative relation between spatial similarity degree and map scale change plays an important role in automated map generalization; however, no method has been found to calculate it. Thus, this paper focuses on this issue. It firstly constructs an approach to calculating spatial similarity degrees between road networks on maps at different scales; then it validates the approach and obtains a number of points by a psychological experiment using some representative road networks at multiple scales, taking spatial similarity degree (say, y) and map scale change (say, x) as the coordinates of each point. After this, it forms a formula for calculating y by x. The formula can be used to determine quantitative relations between spatial similarity degree and map scale change in multi-scale representation of road networks on maps.

Keywords: spatial similarity degree; map scale change; map generalization; road networks; multi-scale representation.

1 Introduction

Map generalization, a technique for producing smaller scale maps using larger scale ones (Ruas 2001; Mackaness et al. 2007; Mackaness and Reimer 2014), is a kind of similarity transformation. If the similarity degree between the generalized map and the original ones is calculable, it should be a help to automated map generalization (Yan 2010; Zhang et al. 2013); because if the quantitative similarity is unknown, a map generalization system does not know to what extent an original map should be generalized while the system is executing for generating a resulting map at a given scale, and therefore it does not know when to stop a map generalization procedure (Yan 2014).

Supposed that an object (or a group of objects) A is represented as A_m and A_k respectively at scales m and k. Previous study (Yan 2015; 2016) has given the definition of spatial similarity relations in multi-scale map spaces, presented the concepts of map scale change $C_{m,k} = \frac{m}{k}$ and spatial similarity degree $Sim(\mathrm{A}_m, \mathrm{A}_k)$, and revealed that they depend on each other in map generalization. It has also addressed the four factors taking effects in spatial similarity judgments (i.e. topological relations, direction relations, distance relations, and attributes) and obtained their weights by experiments (Yan 2016), i.e. $w_{topo\log ival} = 0.22$, $w_{direction} = 0.25$, $w_{dis\tan ce} = 0.31$ and $w_{attribute} = 0.22$, respectively. It has also proposed the formulae for calculating spatial similarity degrees among individual linear feature (Yan 2015) and among river basin networks (Yan 2016) on multi-scale maps. Nevertheless, the formulae cannot adapt to the generalization of the other map features such as individual polygonal settlements/islands, clusters of contour lines and road networks. Thus, this paper focuses on road networks and aims at constructing a formula to describe the quantitative relations between $C_{m,k}$ and $Sim(\mathrm{A}_m, \mathrm{A}_k)$.

After the introduction, an approach to calculating the $Sim(\mathrm{A}_m, \mathrm{A}_k)$ between road networks at different scales is addressed, and a number of points taking $< C_{m,k}, Sim(\mathrm{A}_m, \mathrm{A}_k) >$ as coordinates are obtained by means of an experiment using some multi-scale road networks (Section 2); after this the formula for calculating spatial similarity degrees among multi-scale road networks is constructed by the curve fitting methods (Section 3), and a number of insights into the formula are done (Section 4); finally, some conclusions are drawn (Section 5).

2 Approach to Calculating Spatial Similarity Degrees among Road Networks in Multi-scale Map Spaces

Suppose that A_l is a road network consisting of N_l roads on the map at scale l, A_m is a generalized road network of A_l consisting of N_m roads at scale m .Their properties and the weights are $P = \{P_{Topological}, \mathrm{P}_{Direction}, P_{Distance}, P_{Attribute}\}$ and $W = \{\mathrm{w}_{Topological}, \mathrm{w}_{Direction}, \mathrm{w}_{Distance}, \mathrm{w}_{Attribute}\}$, respectively. Because $Sim(A_l, A_m) = \sum_{i=1}^{4} w_i Sim_{A_l,A_m}^{p_i}$, where, $w_i \in W$ and $P_i \in P$; $i = 1, 2, 3, 4$.

The four weights are known. Thus, if $Sim_{A_l,A_m}^{P_{Topological}}$, $Sim_{A_l,A_m}^{P_{Direction}}$, $Sim_{A_l,A_m}^{P_{Distance}}$ and $Sim_{A_l,A_m}^{P_{Attribute}}$ are known, $Sim(A_l, A_m)$ can be easily obtained.

2.1 Calculation of $Sim_{A_l,A_m}^{P_{Topological}}$

To calculate the similarity degree of two road networks in topological relations, it is necessary to know the difference of the topological relations between the two road networks at different scales. There are totally two topological relations between two roads on the map, i.e. topologically disjoint (e.g. R1 and R3 in Fig.1(a)) and topologically intersected (e.g. R2 and R3 in Fig.1(a)). Matrix B with $N_l \times N_l$ integer elements and matrix C with $N_m \times N_m$ integer elements are used for recording the topological relations of the original road network at scale l and that of the generalized road network at scale m. Let $B_{ij} = B_{ji} = 1$ if the i^{th} road and the j^{th} road on the original map are intersected; otherwise, let $B_{ij} = B_{ji} = 0$. Similarly, let $C_{ij} = C_{ji} = 1$ if the i^{th} road and the j^{th} road on the generalized map are intersected; otherwise, let $C_{ij} = C_{ji} = 0$. Based on the two matrix,

$$Sim_{A_l,A_m}^{P_{Topological}} = 1 - \frac{D_{Topological}}{N_l \times N_l} \tag{1}$$

where, $D_{Topological}$ is the topological differences of the road network at scales l and m. It can be calculated using the following method described in computer language C.

Step 1: let $D_{Topological}$ =0;

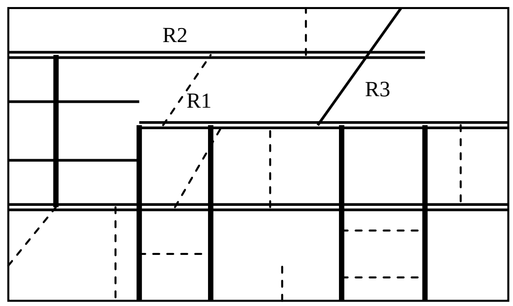
Step 2: take an element C_{ij} from C starting from $i = 0$ and $j = 0$. C_{ij} denotes the topological relations between the i^{th} road and the j^{th} road on the map at scale m.

Step 3: search B for the element B_{pq} that records the topological relations of the i^{th} road and the j^{th} road on the map at scale m .

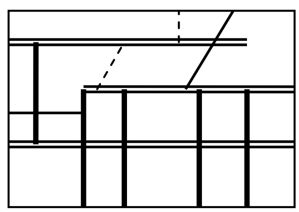
Step 4: If no $B_{pq} = C_{ij}$ can be found, $D_{Topological}$ ++.

Step 5: i++; j++.

Step 6: if $i > N_m$ or $j > N_m$, end the procedure; else go to step 3.



(a) Original city road map at scale l .



(b) Generalized map at scale m

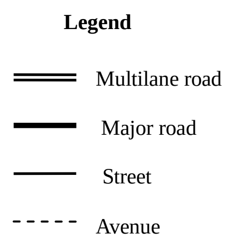


Fig.1 A road network and its generalized

2.2 Calculation of $Sim_{A_l,A_m}^{P_{Direction}}$

Roads on maps are seldom moved before and after map generalization, so the change in direction relations can be viewed as equal to 1 and does not need to be further discussed.

2.3 Calculation of $Sim_{A_l,A_m}^{P_{Distance}}$

Similarity of road networks in distance relations can be evaluated base on road density (D) which is defined as the ratio of the length (L) of the region's total roads to the region's land area (A).

$$D = \frac{L}{A} \qquad (2)$$

Map generalization may lead to the decrease of the number of roads on the map and enlarge the distance among roads, and therefore reduce the road density. Hence,

$$Sim_{A_l,A_m}^{P_{Distance}} = \frac{D_m}{D_l} \qquad (3)$$

It is obvious that the more roads are deleted, the less D_m is, and the less $Sim_{A_l,A_m}^{P_{Distance}}$ is. This means the similarity degree between the original road network and generalized one decrease with the number of roads in map generalization.

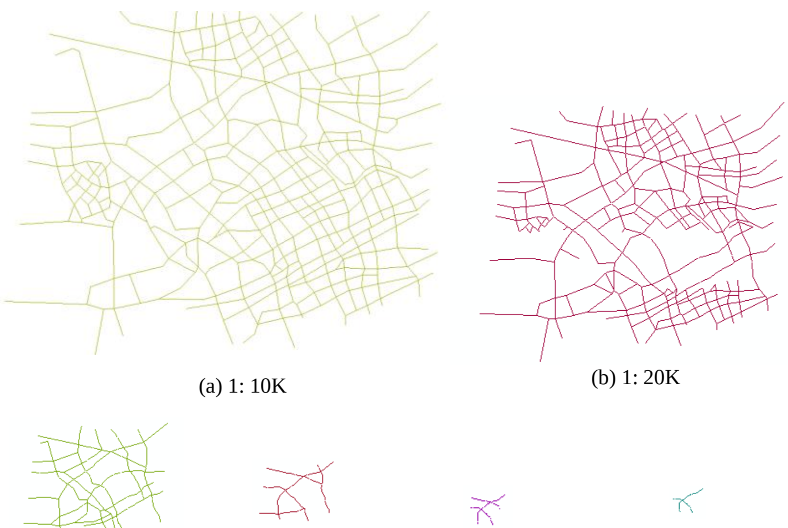
2.4 Calculation $Sim_{A_l,A_m}^{P_{Attribute}}$

Similarity in attributes of road networks depends on a number of attributes such as road type, road class, road condition, etc. To simplify the problem, road class is used to represent the differences of road attributes, which is denoted by the class value.

$$Sim_{A_l,A_m}^{P_{Attribute}} = \frac{\sum_{j=1}^{n_m} L_j^m \times C_j^m}{\sum_{i=1}^{n_l} L_i^l \times C_i^l} \qquad (4)$$

where, L_i^l is the length of the i^{th} road in the road network at scale l; C_i^l is the class value of the i^{th} road in the road network at scale l; L_j^m is the length of the j^{th} road in the road network at scale m; and C_j^m is the class value of the j^{th} road in the road network at scale m.

Here, $\sum_{i=1}^{n_l} L_i^l \times C_i^l$ can be viewed as the total class value of the road network at scale l; $\sum_{j=1}^{n_m} L_j^m \times C_j^m$ is the total class value of the road network at scale m. Thus, $Sim_{A_l,A_m}^{P_{Attribute}}$ represents the percentage of the total class values of the two road networks.



(a) 1: 10K (b) 1: 20K (c) 1: 50K (d) 1: 100K (e) 1: 250K (f) 1: 500K

Fig.2 An ordinary road network at different map scales

3 Construction of the formula

To form the formula, a psychological experiment has been done using three representative road networks at multiple scales as examples. The experiment validates the proposed approach for calculating similarity degrees; on the other hand, it generates a number of points taking $C_{m,k}$ and $Sim(\mathrm{A}_m, \mathrm{A}_k)$ as coordinates. Based on the points, the curve fitting method is employed and the formula is obtained.

3.1 Psychological experiment

Correctness of approaches is often addressed through model validation (Hagen-Zanker 2009; Banks et al. 2010; Sargent 2011). The new approach is validated by a psychological experiment because spatial similarity judgment roots itself in human's cognition.

The experiment was done on October 20, 2013 in Lanzhou Jiaotong University, China. The subjects are 50 students at undergraduate level majoring in geography and each of them has at least six-month work experience in making maps.

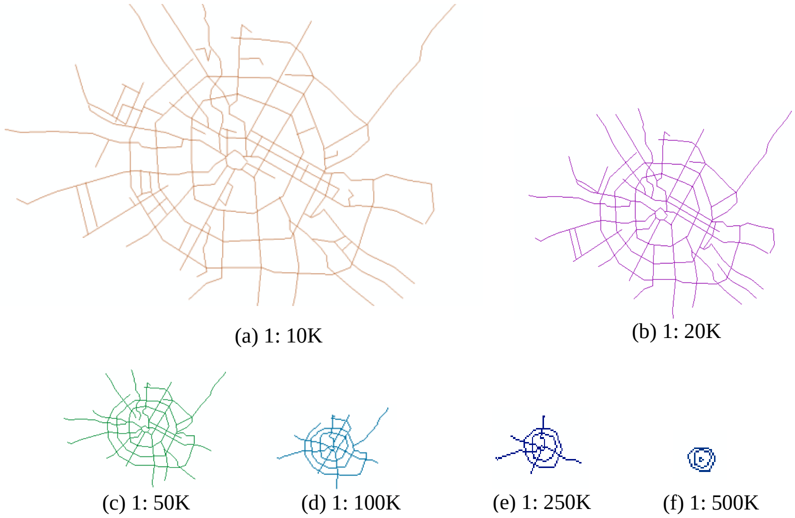


Fig.3 A road network with ring roads at different map scales.

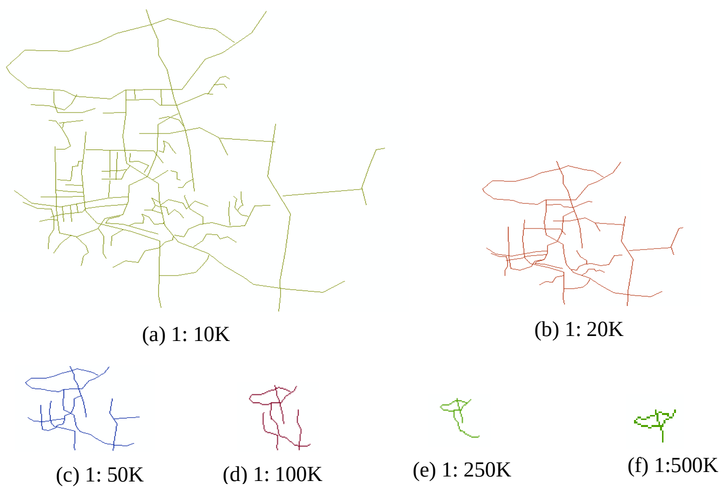


Fig.4 A road network with zigzag roads at different map scales.

Three road networks provided by the Administration of Gansu Geoinformatics Centre, China, are used in the experiment (Fig.2, Fig.3 and Fig.4). Each consists of the original road network at a larger scale and the other five generalized road networks at different smaller scales. In each figure, the spatial similarity degree between the original road network and each of the five other generalized ones is calculated using the approach proposed in Section 2 and listed in Table 1.

In the experiment, the road networks are printed and distributed to the subjects along with the spatial similarity degrees. The subjects are required to answer if they are agree/disagree with the similarity degrees or have no idea about them.

It should be noticed that $Sim(\mathrm{Fig}_{ia}, \mathrm{Fig}_{ib})$ in Table 1 refers to the spatial similarity degree between (a) and (b) in the corresponding figure; $Sim(\mathrm{Fig}_{ia}, \mathrm{Fig}_{ic})$, $Sim(\mathrm{Fig}_{ia}, \mathrm{Fig}_{id})$, $Sim(\mathrm{Fig}_{ia}, \mathrm{Fig}_{ie})$ and $Sim(\mathrm{Fig}_{ia}, \mathrm{Fig}_{if})$ have the similar meaning; $C_{a,b} / C_{\mathrm{a,c}} / C_{a,d} / C_{a,e} / C_{a,f}$ refers to the map scale change from (a) to (b)/(c)/(d)/(e)/(f) in the corresponding figure; $N_{Agree} / N_{Disagree}$ is the number of the subjects that agree/disagree with the spatial similarity degrees; and N_{Noidea} is the number of the subjects that have no idea about the spatial similarity degrees.

In Table 1 the numbers of the subjects that agree with the calculated spatial similarity degrees are 50, 49 and 48 out of 50. Hence, the proposed approach is acceptable, and the coordinate pairs consisting of $Sim(\mathrm{Fig}_{ij}, \mathrm{Fig}_{ik})$ and $C_{j,\mathrm{k}}$ are credible to be used to construct the formula for describing quantitative relations between spatial similarity degree and map scale change.

3.2 Formation of the formula by the curve fitting method

Table 1 Calculated spatial similarity degrees and the subjects' responses

	$Sim(\mathrm{Fig}_{ia}, \mathrm{Fig}_{ib})$, $Sim(\mathrm{Fig}_{ia}, \mathrm{Fig}_{ic})$, $Sim(\mathrm{Fig}_{ia}, \mathrm{Fig}_{id})$, $Sim(\mathrm{Fig}_{ia}, \mathrm{Fig}_{ie})$, $Sim(\mathrm{Fig}_{ia}, \mathrm{Fig}_{if})$	$C_{a,b}$, $C_{a,c}$, $C_{a,d}$, $C_{a,d}$, $C_{a,d}$	N_{Agree}, $N_{Disagree}$, N_{Noidea}
Fig.2	0.77, 0.52, 0.31, 0.22, 0.18	2, 5, 10, 25, 50	50, 0, 0
Fig.3	0.75, 0.55, 0.37, 0.28, 0.19	2, 5, 10, 25, 50	49, 0, 1
Fig.4	0.68, 0.49, 0.34, 0.28, 0.16	2, 5, 10, 25, 50	48, 0, 2

The coordinate pairs $< C_{j,\mathrm{k}}, Sim(\mathrm{Fig}_{ij}, \mathrm{Fig}_{ik}) >$ in Table 1 form 15 points. In addition, a road network is wholly similar to itself, i.e. $C_{j,\mathrm{j}}$ =1 and $Sim(\mathrm{Fig}_{ij}, \mathrm{Fig}_{ij}) = 1.00$. Thus, totally 16 points can be obtained. They are as follows:

(2, 0.77), (5, 0.52), (10, 0.31), (25, 0.22), (50, 0.18),

(2, 0.75), (5, 0.55), (10, 0.37), (25, 0.28), (50, 0.19),

(2, 0.68), (5, 0.49), (10, 0.34), (25, 0.28), (50, 0.16).

(1, 1.00)

Let $x = C_{i,k}$ and $y = Sim(Fig_{ij}, Fig_{ik})$. A general formula for the relation between $C_{j,\mathrm{k}}$ and $Sim(\mathrm{Fig}_{ij}, \mathrm{Fig}_{ik})$ may be $y = f(x)$.

Because curve fitting can capture the trend in the data across the entire range, and can be used as an aid for data visualization to infer values of the function where no data are available and to summarize the relationships among two or more variables (Kolb 1984; Arlinghaus 1994), it is selected to substantiate the formula. Five candidate functions are chosen for the curve fitting. They are (1) $y = a_1 x + a_0$, (2) $y = a_2 x^2 + a_1 x + a_0$, (3) $y = a_2 e^{a_1 x} + a_0$, (4) $y = a_1 \ln(\mathrm{x}) + \mathrm{a}_0$ and (5) $y = x^a$.

The five resulting curves and formulae are shown in Fig.5. R^2 is used to compare the candidate functions. The greater the R^2 , the better its corresponding curve (Lanczos 1988), i.e. the curve with the greatest R^2 among all of the candidates is the best curve fitting the point set. In this curve fitting, the resulting formula is

$$y=1.0022x^{-0.439} \qquad (5)$$

because its $R^2 = 0.9754$ is the greatest in the five R^2 of the candidate curves.

4 Discussion

Some insights can be gained from the formula for describing the quantitative relations between map scale change and spatial similarity degree of road networks in multi-scale map spaces.

·Above all, Formula (5) can be used to calculate the spatial similarity degree (y) if the map scale change (x) of the original and the generalized road networks is known. On the other hand, the inverse function of the formula can also be obtained and used to calculate the map scale change between a map and its generalized result if the spatial similarity degree is given.

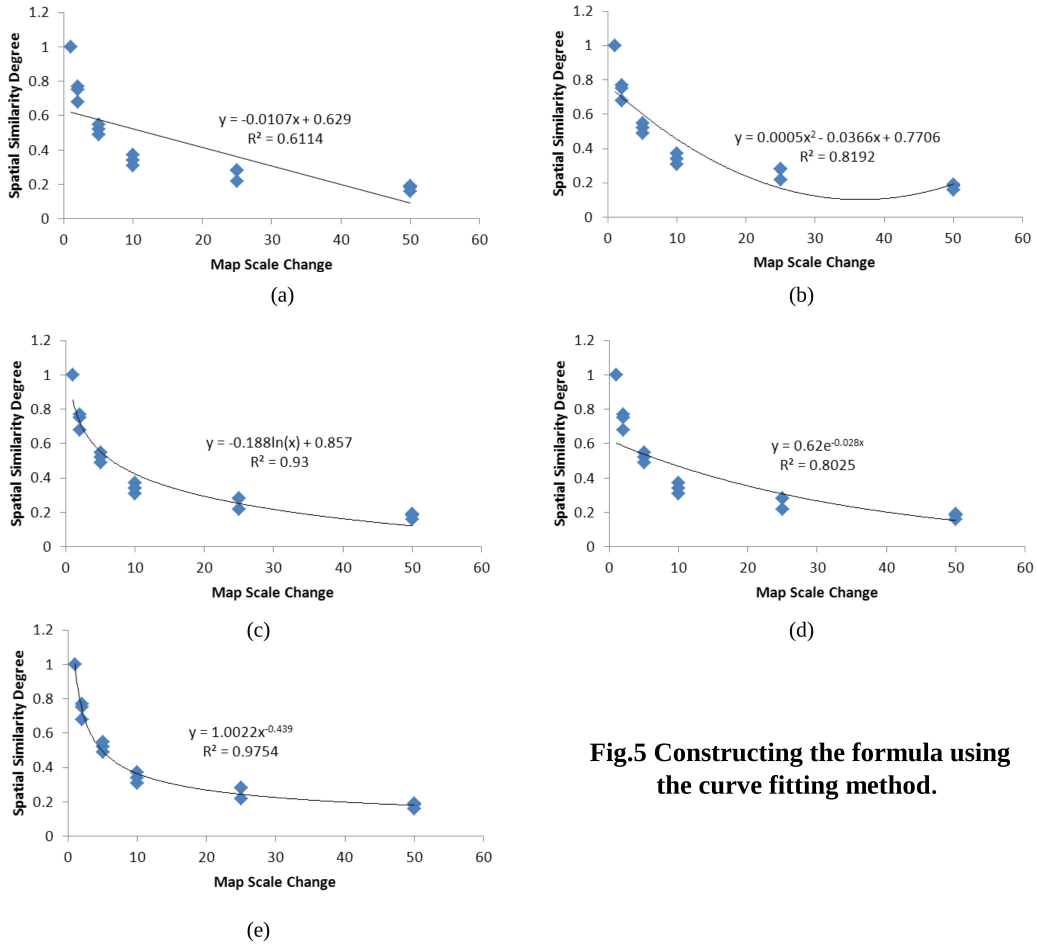


Fig.5 Constructing the formula using the curve fitting method.

• Formula (5) is an empirical function which means the results calculated by it are not strictly accurate. Its accuracy relies on the points used in curve fitting. Generally, the more accurate the original points are and the more points are used in the curve fitting, the more accurate the formula is. Thus, more experiments should be done using more representative road networks and find more subjects with different cultural backgrounds to improve the accuracy and adaptability of the formula.

·The domain of the formula is $x \in [1, +\infty)$, and the range is $y \in [0,1]$. Hence, the formula can be used to interpolate any values belonging to the domain (and belonging to the range if the invers function is used). Nevertheless, $x < 1$ is meaningless and out of the scope of this study, for it denotes the resulting map scale is greater than the original one which is impossible in map generalization.

For example, a road network at scale 1:1 000 is generalized to get a map at scale 1:2 000. The spatial similarity degree between the maps at 1:1 000 and 1:2 000 can be calculated by

$$y=1.0022x^{-0.439}=1.0022 \times 2^{-0.439} \approx 0.739$$

Suppose that the map at 1:1 000 is generalized to produce a map at 1:500, the spatial similarity degree between the maps at 1:1000 and 1:500 can be calculated by

$$y=1.0022x^{-0.439}=1.0022 \times 0.5^{-0.439} \approx 1.359$$

This result is meaningless, because it denotes that the generalized map is more similar than the original map itself if both of them are compared with the original map.

·This formula is a potential tool for improving the automation of map generalization. For example, a 1:1 000 road map is used to produce a map at scale 1:5 000. The spatial similarity degree (say, y_t) between the two maps can be obtained using Formula (5), because the map scale change can be easily calculated. In the process of road network generalization, when an intermediate map is produced it may be compared with the original road network and their spatial similarity degree (say, y_p) can be calculated using the approaches proposed in Section 2. If $\left| y_t - y_p \right|$ is less than a given tolerant value, this means the intermediate map is an acceptable map at scale 1:5 000 and the generalization can be stopped; otherwise, continue with the generalization process. Thus, spatial similarity degrees help the map generalization system to stop the map generalization procedure automatically.

5 Conclusions

Spatial similarity degree between road networks on maps at multiple scales can facilitate the automation of map generalization. This paper focuses on the issue and proposes an approach for calculating spatial similarity degrees between road networks at multiple scales. It validates the approach and constructs a formula for quantitatively describing the relations between map scale change and spatial similarity degree by means of the curve fitting method using the point data from a psychological experiment. The formula can be used to calculate spatial similarity degree if map scale change is given, and vice versa. It would be our future work to integrate the proposed formula and approach into a map generalization system and improve the automation ability of map generalization.

Acknowledgments

The work described in this paper is partially funded by the Natural Science Foundation Committee, China (No. 41371435 and No. 71563025), partially by the Support Plan in Science and Technology, Gansu, China (No. 1304GKCA009), and partially funded by the National Key Technologies R&D Program of China (No. 2013BAB05B01).

References

Arlinghaus S.L., 1994, PHB Practical Handbook of Curve Fitting, CRC Press.

Banks J., Carson J.S., Nelson B.L., Nicol D.M., 2010, Discrete-event System Simulation (Fifth edition), Upper Saddle River: Pearson Education.

Hagen-Zanker A., 2009, An improved Fuzzy Kappa statistic that accounts for spatial autocorrelation, International Journal of Geographical Information Science, 23 (1): 61-73.

Kolb W.M., 1984, Curve fitting for programmable calculators, Syntec Inc.

Lanczos C., 1988, Applied Analysis (Reprint of 1956 Prentice–Hall edition), Dover Publications.

Li B. and Fonseca F.T., 2006, TDD - A Comprehensive Model for Qualitative Spatial Similarity Assessment, Spatial Cognition and Computation, 6(1): 31-62.

Mackaness W. and Reimer A., 2014, Generalisation in the Context of Schematised Maps, In: Burghardt D., Duchêne C., Mackaness W. (eds), Abstracting Geographic Information in a Data Rich World: Methodologies and Applications of Map Generalisation. Lecture Notes in Geoinformation and Cartography, Springer.

Mackaness W., Ruas A. and Sarjakoski T., 2007, Generalization of Geographic Information: Cartographic Modeling and Applications, Elsevier.

Rodríguez A., and Egenhofer M., 2004, Comparing Geospatial Entity Classes: An Asymmetric and Context-Dependent Similarity Measure. International Journal of Geographical Information Science, 18(3):229-256.

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

Sargent R.G., 2011, Verification and validation of simulation models, in: S. Jain, R.R. Creasey, J. Himmelspach, K.P. White, and M. Fu (eds.), Proceedings of the 2011 Winter Simulation Conference, Baltimore, MD, pp. 166-183.

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

Yan H.W., 2015, Quantitative Relations of Spatial Similarity Degree and Map Scale Change of Individual Linear Objects in Multi-scale Map Spaces, GeoCarto International, 30(4): 472-482.

Yan H.W., Shen Y.Z., Li J., 2016, Approach to Calculating Spatial Similarity Degrees of the Same River Basin Networks on Multi-scale Maps, GeoCarto International, http://www.tandfonline.com/eprint/smPNfk8PHV8YZPtAngQm/full

Zhang X., Ai T., Stoter J., Kraak M. and Molenaar M., 2013, Building pattern recognition in topographic data: examples on collinear and curvilinear alignments, GeoInformatica, 17(1): 1-33.