

Quantitative Evaluation Method of Elements Priority of Cartographic Generalization Based on Taxi Trajectory Data

Zhiguan Long^a, Nai Yang^{a,*}, Yuanlin Huang^b, Yi Chao^a, Lin Wan^a

^aFaculty of Information Engineering, China University of Geosciences, 430074 Wuhan, China (long, yang, chao, wan)
longzhiguan@163.com

^bCollege of Resource and Environment, Qinzhou University, 535011 Qinzhou, China 1732305940@qq.com

Commission IV, WG IV/4
Commission III, WG III/4
Commission V, WG V/4

KEY WORDS: Cartographic Generalization, Taxi GPS data, Density Analysis, Elements Selection, Quantitative Research

ABSTRACT:

Considering the lack of quantitative criteria for the selection of elements in cartographic generalization, this study divided the hotspot areas of passengers into parts at three levels, gave them different weights, and then classified the elements from the different hotspots. On this basis, a method was proposed to quantify the priority of elements selection. Subsequently, the quantitative priority of different cartographic elements was summarized based on this method. In cartographic generalization, the method can be preferred to select the significant elements and discard those that are relatively non-significant.

1. INTRODUCTION

Cartographic generalization combines selection and generalization (He,2016). Selection is carried out based on the significance of the cartographic elements. Therefore, how the significance of elements quantitatively is measured is essential for cartographic generalization. To date, many experts and scholars have conducted research from different perspectives. For instance, Guo proposed a location selection method based on the structure of a point-like feature group and the automatic generalization method which changed the former spatial clustering method (Guo,2002). Bjrke established the location selection model to optimize the point group classification mechanism based on the theory of communication (Bjrke,1997). Kreveld proposed the structured selection algorithm of a point-like feature group based on the circular growth feature, which realized the automatic selection of this group (Kreveld,1997). Li proposed 11 kinds of cartographic generalization operators, which form the qualitative description of cartographic generalization with accurate mathematical characteristics (Li,2002). Steiniger attempted to quantify the horizontal relationship among map objects, such as geometry, topology, and statistics, to provide a mathematical support for the generalization of map objects in thematic and topographic maps (Steiniger,2007). Ai considered the spatial distribution and statistical laws of interaction between points, proposed the group simplification algorithm, and then extracted the point group center via the image gray scale to represent the distribution density (Ai,2002). Jiangnan divided the map into several areas according to density, and then established the grading system of basic geographic elements for qualitative analysis and survey (Jiang,2013). Li Z and Pei Z explored quantitative measures for spatial information of maps (Li,2002). Steiniger S and Weibel R proposed several quantitative methods for standardizing relations and structures in categorical maps(Steiniger,2007). Sylvain B. made some research about quality assessment of generalized geographical data accuracy. (Sylvain,2002). Cristina C F and De N, F made quality evaluation and non-uniform compression of

geometrically distorted images using the quad-tree distortion Map. (Cristina C F,2004) Bereuter P focused on generalisation of point data for mobile devices. (Bereuter,2010). Okabe A and et al developed a kernel density estimation method for networks, its computational method and a GIS-based tool. This method would be applied in estimating the density of points on a network and implement the method in the GIS environment. (Okabe,2009).

The significance of the location of elements is strongly related to the aggregation of human activities. Studies on the priority of elements selection in cartographic generalization from the perspective of human travel is limited. However, the advent of the big data age provides an opportunity for us. Today, we get a large quantitative of geospatial data, via satellite, unmanned aerial vehicles, mobile measurement vehicles. which are traditional mapping technology to bring the massive basic mapping data. However much real-time data via monitoring sensors, mobile terminal data and even a variety of UGC data constitute geographical big data. This paper presents a method for calculating the priority of elements selection quantitatively based on GPS data through the qualitative analysis of human activity hotspots and the quantitative analysis of elements that fall into hotspots.

2. DATA PROCESS

The research data in this study included the random sampling of a taxi's one-day GPS data in Wuhan and the POI data. The GPS data contained the taxi operation information of more than 10,000 vehicles within 24 hours. The total amount of data was approximately 109,600,000. Raw data were stored in text format and contained attribute information, such as taxi ID, latitude and longitude, GPS recording time, passenger status, speed, and direction.

The following is a detailed description of the process of data processing. This change in passenger status between 0 and 1 was

* Corresponding author

used to gather the on and off points, that is, passenger status 0 corresponded to off and 1 to on. The results consisted of 614,209 on and off points. Then, density analysis of both points was conducted using ArcGIS. The density was divided into five grades using the natural discontinuity classification method, and three regions of relatively large densities were selected as dense, medium, and general area hotspots. Then, the vector layer of the hotspot regions was extracted through vectorization, as shown in Figure 1.

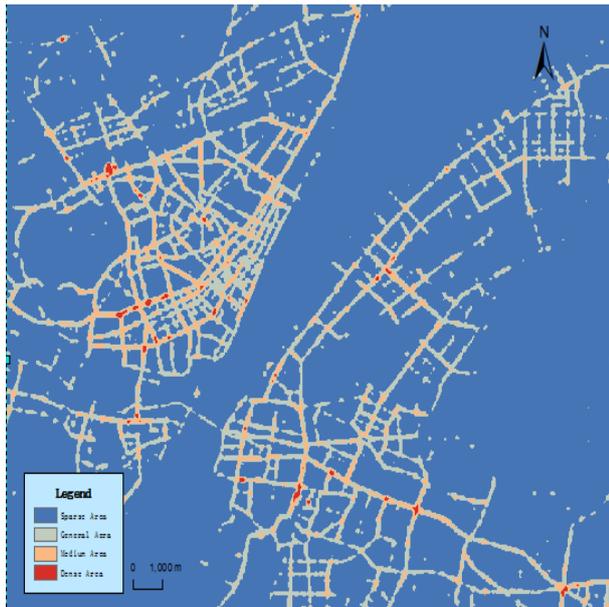


Figure 1. Hotspot regions with various grades

The POI data that fall into the different areas were extracted by buffering the hotspots within the 200m range.

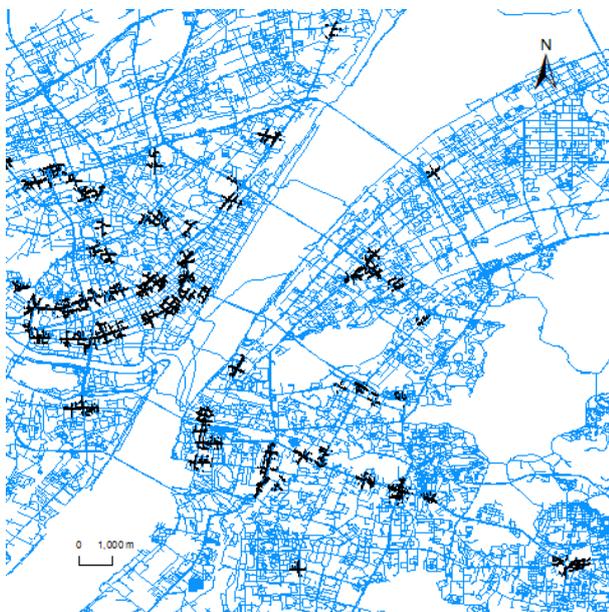


Figure 2. POIs in Dense Area

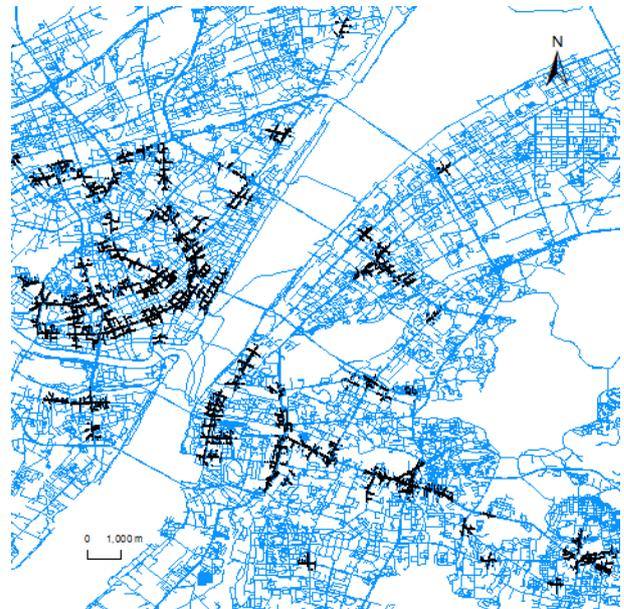


Figure 3. POIs in Medium Area



Figure 4. POIs in General Area

3. STATISTICS ANALYSIS

According to the POI classification system in the electronic map and the GB/T13923-2006 basic geographic information element classification and code among others, the POI data was divided into 235 secondary classes under 8 comprehensive classes, which included commercial retail, food consumption, financial and business service, medical and health security, public sports, public facility, research and education/culture, and public utility management. The POI data from the hotspots were counted according to the classification system mentioned above, and the analysis is shown in Table 1.

Comprehensive Class	amounts	T1 Area	Proportion	T2 Area	Proportion	T3 Area	Proportion
Food Consumption	18075	2138	11.83%	4233	23.42%	5013	27.73%
Commercial Retail	48727	7247	14.87%	10421	21.39%	12457	25.56%
Research and Education/culture	9548	516	5.40%	1136	11.90%	1383	14.48%
Medical and Health security	4171	380	9.11%	736	17.65%	838	20.09%
Public Sports	8215	571	6.95%	1300	15.82%	1740	21.18%
Public Utility Management	7062	346	4.90%	766	10.85%	934	13.23%
Financial and Business Service	25141	2695	10.72%	5358	21.31%	6863	27.30%
Public Facility	24400	1671	6.85%	3724	15.26%	5025	20.59%

Table 1. Proportion of comprehensive classes in three areas

(In order to avoid confusion due to the large number of parameters above. The following describes the parameters in detail)

In the Table above, “amounts” represents the total amount of each comprehensive class and “proportion” represents the percentage of each class in each region. For the statistics of the secondary classes, considering that the number of secondary classes was too large, this study took food and consumption as an example. The statistical results are shown in Table 2.

Comprehensive Class	Secondary Class	T1 Area	Proportion	T2 Area	Proportion	T3 Area
Food and Consumption	Chinese-Restaurant	967	7.56%	1995	7.52%	2525
	Fast Food	588	4.60%	1136	4.28%	1309
	Cold Store	250	1.95%	438	1.65%	411
	Exotic	183	1.43%	315	1.19%	295
	Cafe	78	0.61%	162	0.61%	214
	Teahouse	35	0.27%	74	0.28%	92
	Bar	35	0.27%	111	0.42%	165
	Local-Cuisine	1	0.01%	2	0.01%	2

Table 2. Proportion of Food and Consumption class in three areas

4. QUANTITATIVE CALCULATION METHOD OF ELEMENTS SELECTION

4.1 Set weights of hotspots

Establishing the significance of elements is the basic part of the adaptive expression of map elements. Under the premise of reasonable load capacity, the high-significance objects in a small-scale map can be preferentially displayed and the low-significance objects are discarded, making the map clearer and more usable.

The distribution of passenger locations based on the on and off points reflects the importance of these areas at different levels when people travel. It also reflects that the more concentrated the distribution points are, the more frequently people move, thus making their location more significant for people to travel. The three hotspots mentioned previously were given different weights, as shown in Table 3.

Area ID	Area Name	Weight
T1	Dense area	0.6
T2	Medium area	0.3
T3	General area	0.1

Table 3. Weights of hotspots

In this paper, subjective weights assignments are set to three-level hotspots. For example, weight (max: 0.6) is set to a region with a large human collection. which means that in the subsequent calculations. When it comes to elements generalization, the POIs in the dense area have the possibility of more retention than the other area.

4.2 Calculation of significance index

The significance index of different POI categories was calculated as follows:

$$P = \sum_{i=1}^3 W_i * K_i ; K_i = N_{[T_i]} / \Delta \quad (1)$$

K_i : For the calculation of the comprehensive class, this represents the percentage of the total number of a certain comprehensive class in the T_i area, which accounted for the entire city, $i \in [1, 3]$. For the calculation of the secondary class, K_i represents the percentage of the total number of a certain secondary class in the T_i area that accounts for the entire city of Wuhan.

W_i : the weight of the different hotspots, $i \in [1, 3]$.

$N_{[T_i]}$: For the comprehensive class, it denotes the total number of a certain comprehensive class in the city; and for the secondary class, it denotes the total number of a certain secondary class.

Δ : For the comprehensive class, it represents the total number of a certain comprehensive class; for the secondary class, it represents the total POI number in a certain area.

(1) Calculation of the significance index of the comprehensive class. The significance index of the comprehensive class was for the comparison between different comprehensive classes. With the food consumption class as an example,

the significance index of formula (1) is calculated as follows:

$$P = \sum_{i=1}^3 W_i * K_i = 0.6 * 11.83\% + 0.3 * 23.42\% + 0.1 * 27.73\% = 0.17$$

According to the calculation method above, the results of the calculation of the significance index for all categories are shown in Table 4.

Comprehensive Class	Significance Index
Commercial Retail	0.18
Food Consumption	0.17
Financial and Business Service	0.16
Medical and Health security	0.13
Public Sports	0.11
Public Facility	0.11
Research and Education/culture	0.08
Public Utility Management	0.08

Table 4. Significance ranking of POI comprehensive class

Through the calculation above, all classes of elements were allocated with different significance indices, and the priority order of the cartographic elements was as follows: commercial retail (0.18) > food consumption (0.17) > financial and business service (0.16) > medical and health security (0.13) > public sports (0.11) > public facility (0.11) > research and education/culture (0.08) > public utility management (0.08).

(2) Calculation of the significance index of the secondary class. The significance index of the secondary class was for the comparison between different secondary classes. Taking the Chinese restaurant class under the food consumption class as an example, the significance index of formula (1) is calculated as follows:

$$P = \sum_{i=1}^3 W_i * K_i = 0.6 * 7.59\% + 0.3 * 7.60\% + 0.1 * 7.46\% = 0.076$$

Based on the formula above, the ranking order was as follows: Chinese-restaurant>fast-food> Cold Store >exotic>cafe>teahouse>bar> local cuisine, as shown in Table 5.

Comprehensive Class	Secondary Class	Significance Index
Food consumption	Chinese Restaurant	0.076
	Fast Food	0.045
	Cold Store	0.010
	Exotic	0.009
	Cafe	0.007
	Teahouse	0.003
	Bar	0.003
	Local Cuisine	0.001

Table 5. Significance ranking of Food Consumption

Using the calculation above, all classes of elements were allocated with different significance indices, and the priority order of the cartographic elements was the same as that above.

4.3 Experiments Results

At a certain scale, the display priority was given to those cartographic elements with high comprehensive significance indices rather than those with low values when conflicts come between cartographic elements under the same comprehensive class. In addition, priority was given to those cartographic elements with a high secondary significance-indices for the same conflict mentioned previously.

This paper took the expression of POIs in Optical Valley as an example based on the quantitative priority calculation of the cartographic elements. The expression at the 1:5,000 scale is shown in Figure 5, and 17 comprehensive classes were displayed, including the food consumption and commercial retail classes. When the scale changed from 1:5,000 to 1:10,000, conflict occurred between the comprehensive classes. For instance, in the display conflict that occurred between the Agricultural Bank and the malls, the significance index of the commercial retail class was higher than that of the commercial and financial class in view of the significance index ranking of POIs. Therefore, the mall was displayed prior to other POIs in area at a 1:10,000 scale. When the scale changed from 1:10,000 to 1:15,000, conflicts occurred between the malls and the surrounding government agencies and residential areas. Considering the significance of the index ranking of POIs, the malls were displayed preferentially in the screen.

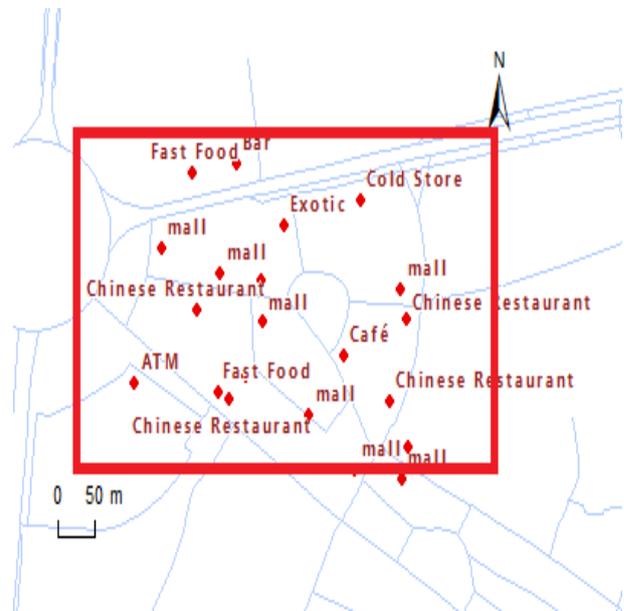


Figure 5. POI expression at 1:5,000 scale

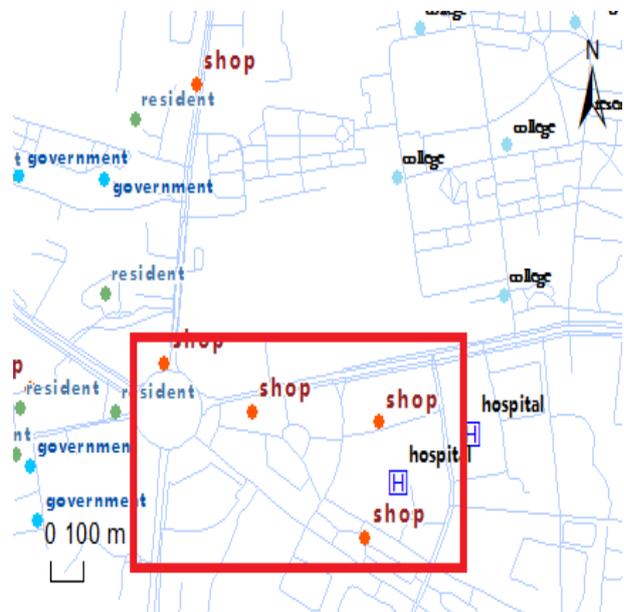


Figure 6. POI expression at 1:10,000 scale

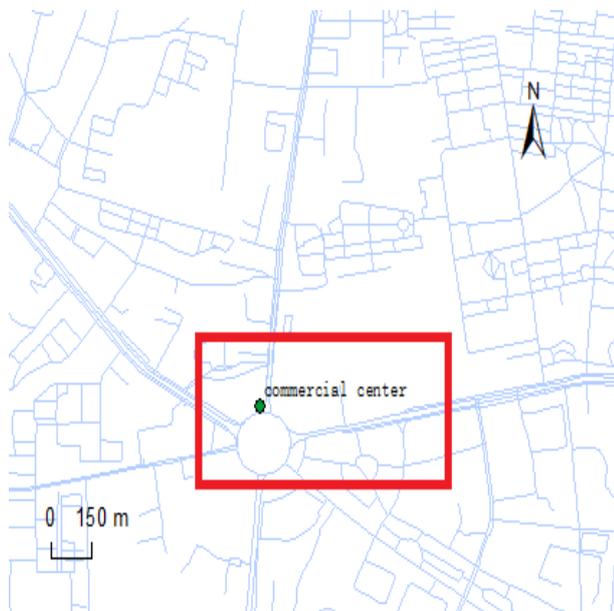


Figure 7. POI expression at 1:15,000 scale

5. SUMMARY

This paper presents the method for evaluating the significance indices of cartographic elements by quantifying them. On this basis, a set of reference systems for the priority sorting of cartographic elements selection was established via the GPS data from a taxi. One-day data was randomly selected for research, and a possibility of urban hotspot formation existed.

Moreover, the research objects of this paper were POI elements that were closely related to human activities. Other elements, such as roads and rivers, were not considered. Relevant research on diversified cartographic needs and demand-driven users should be conducted in future works.

ACKNOWLEDGEMENTS

This research was financially supported by State-Sponsored Scholarship Program for Visiting Scholars, the Fundamental Research Funds for National University, China University of Geosciences (Wuhan) (G1323541712, CUGL130256), National Natural Science Foundation of China (41401449), the Open Foundation of Key Laboratory of Precise Engineering and Industry Surveying of National Administration of Surveying, Mapping and Geo-information (Grant No. PF2015-16), Hubei Education Science "Twelfth Five-Year" Planning Project (NO. 2013A010)

REFERENCES

- He, Z Y., Song, Y., Li, L Y., 2016. Principles and methods of cartography. Wuhan University Press. 211.
- Guo, Q S., 2002. Principles and theory of cartographic generalization. Beijing: Surveying and Mapping Press.
- Bjrke, J T., 1997. Map Generalization: Information Theoretic Approach to Feature Elimination. 18th ICA.
- Kreveld, 1997. Efficient settlement selection for interactive display. In Proc. Auto-Carto 13: ACSM/ASPRS Annual Convention Technical Papers: 287-296.

- Li, Z., and P, H., 2002a. Quantitative measures for spatial information of maps. *International Journal of Geographical Information Science*, 16 (7), pp. 699-709.
- Steiniger, S., and R. Weibel., 2007a. Relations among Map Objects in Cartographic Generalization. *Cartography and Geographic Information Science*, 34(3), pp. 175-197.
- Ai, T H., 2002a. A Method of Point Cluster Simplification with Spatial Distribution Properties Preserved. *Journal of Surveying and Mapping*, (02), pp. 175-181.
- Jiang, N., Bai, X S., & Cao, Y. N., et al., 2013a. Establishment and application of multi-scale display model in basic electric map. *Geomatics and Information Science of Wuhan University*.35(7), pp.768-772.
- Li, Z., & Pei, Z., 2002a. Quantitative measures for spatial information of maps. *International Journal of Geographical Information Science*.
- Cristina, C F., & De N F. 2004. Quality Evaluation and Nonuniform Compression of Geometrically Distorted Images Using the Quad-tree Distortion Map. *EURASIP Journal on Advances in Signal Processing*.
- Steiniger, S., & Weibel, R., 2005. Relations and Structures in Categorical Maps. 8th ICA Workshop on Generalization and Multiple Representation.
- Bereuter, P., & Weibel, R., 2010. Generalisation of point data for mobile devices. A problem-oriented approach:13th Workshop of the ICA commission on Generalisation and Multiple Representation.
- Okabe, A., Satoh, T., & Sugihara, K. 2009. A kernel density estimation method for networks, its computational method and a GIS-based tool. *International Journal of Geographical Information Science*, 23(1), pp. 7-32.
- Yu, Y., Zhu, H., Yang, L., & Wang, C X., 2016. Spatial Indexing for Effective Visualization of Vector-Based Electronic Nautical Chart. International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration(pp.323-326). IEEE Computer Society.
- Wang, Z L., Weng, Q., Jiang, & et al., 2008a. Information organization and visualization mechanism of electronic map. *Geo-spatial Information Science*, 11(4), pp. 262-268.
- Wang, J Y., 1992. Principles and methods of cartography. Zhengzhou: Surveying and Mapping Press.
- Standardization Administration of the People's Republic of China. Specifications for feature classification and codes of fundamental geographic information: GB/T 13923-2006 [S]. Beijing: Standards Press of China.2006:1-30.