Research on Urban Point of Interest Generalization Method Based on Mapping Presentation

Chengming Li, Yong Yin, Peipei Guo, Xiaoli Liu

Abstract—Without taking account of the attribute richness of POI (point of interest) data and spatial distribution limited by roads, a POI generalization method considering both attribute information and spatial distribution has been proposed against the existing point generalization algorithm merely focusing on overall information of point groups. Hierarchical characteristic of urban POI information expression has been firstly analyzed to point out the measurement feature of the corresponding hierarchy. On this basis, an urban POI generalizing strategy has been put forward: POIs urban road network have been divided into three distribution pattern; corresponding generalization methods have been proposed according to the characteristic of POI data in different distribution patterns. Experimental results showed that the method taking into account both attribute information and spatial distribution characteristics of POI can better implement urban POI generalization in the mapping presentation.

Keywords—POI, Road network, spatial information expression, selection method, distribution pattern.

I. INTRODUCTION

POI is short for point of interest. There is no strict definition for POI at present. It refers to all geographic objects that can be abstracted as points in general. POI, in real life, is commonly used for providing map service like auxiliary positioning, surrounding search, location query and path query, etc. Richness of data and definition shown on the map directly influence the quality and availability of a map. A common problem existing in POI applications is overlap generated due to the display of large data volume, seriously affecting the user's cognition of geographic space. Therefore, how to effectively generalize a large number of POI data to satisfy map users' demands has become an important research content. Traditional cartographic generalization is a creative work conducted by cartographers while compiling maps, which is a manual process. The generalization method not only costly in terms of time, a long period length with slow speed and low precision, but also places a high demand on the cartographer's knowledge and experience. With the development of science and technology, automatic and intelligent methods are needed to realize the generalization of POI data.

There are several kinds of generalization methods oriented at point targets, including Voronoi-based algorithm [1], [2],

Kohonen network algorithm [3]; Selection model based on genetic algorithm [4] and the algorithm based on rough-set classification [5], etc. But research objects of these algorithms were targets distributed in point groups like habitation, island and lake in the small-scale topographic map. Such studies were not interested in sizes and shapes of a single point target, but concerned about the spatial structure information implied in group distribution; POI as a special kind of point data, features large volume of data, various types and rich attribute, etc. Others factors should be also considered in the generalization process, besides the spatial feature. Thus, the existing point generalization model cannot be directly applied in the POI generation.

Because of the simplicity of POI data acquisition, early researches on POI were data acquisition processing and multi-source data infusion, etc. And since, research focuses have been steered to POI data mining and application [6]-[9]. Focusing on different aspects, some deficiencies in researches on POI data generalization still exist: Literature [10] put forward a decomposed POI simplification parallel computing method based on road network grid, which partitioned POI data via roads without distinguishing features of POI and ordinary points; the feature simplification algorithm of ordinary points had been still adopted in the generalization algorithm. Literature [11] discussed the differences between POI and ordinary point elements, which proposed a POI dynamic generalization model oriented at the mobile environment. But the generalization algorithm adopted was still the generalization algorithm of ordinary point elements. A POI diluting method based on linear roads has been put forward in literature [12], which adopted distance threshold to dilute POI with consideration of weighted importance of POI data. Discussion was only conducted on road lines in transverse and longitudinal direction without consideration of POI overall distribution features. An online generalization strategy oriented at POI data of spontaneous geographic information has been proposed in literature [13], which realized partial generalization operations on the basis of data pre-processing, multi-hierarchy structural organization, but it only discussed only a single type of POI data with emphasizing efficiency displayed on line rather than a POI generalization method and model. The study by [14] has proposed to adopt the ARCGIS tool to conduct the POI diluting method in a set of illustrations of an electronic map, but the method is belong to a human aided computer mapping , so the automation degree is low.

In conclusion, current researches on POI generalization have deficiencies in the following two aspects: insufficient theoretic researches on generalization, partial generalization methods

Chengming Li is with the Chinese Academy of Surveying and Mapping, Beijing, CO 100830 China (corresponding author, phone: 01063880815; fax: 01063880630; e-mail: cmli@ casm.ac.cn).

Yong Yin, Peipei Guo, and Xiaoli Liu are with the Chinese Academy of Surveying and Mapping, Beijing, CO 100830 China (e-mail: yinyong@casm.ac.cn, peipeiguo@qq.com, 83391860@qq.com).

still adopted the generalization methods of ordinary point elements, and lack of in-depth research on the model of generalization method, display efficiency has been emphasized in the mobile electronic map, but ignores research on the generalization method. Therefore, POI data features and information transfer features have been analyzed in this paper. POI distribution has been divided into three patterns on this basis and three corresponding methods have been proposed according to the characteristic of each distribution pattern, so as to reach the POI generalization effect based on the mapping presentation.

II. FEATURES AND GENERALIZATION REQUIREMENTS OF POI DATA

POI features rich attributes, diverse types, and large volumes of data and updated geographic information, etc. Although belonging to one of point features of the map, POI also differs from point features on ordinary maps, for example, islands, lakes and resident points are distinctly different. POI covers numerous basic information closely related to people's daily lives, such as catering, entertainment, financial institutions, tourist attractions, landmarks, and gas stations and parking lots, etc. POI is often utilized to conduct auxiliary positioning, surrounding search, location query and path query, etc., which requires POI data to have not only geo-spatial coordinates, but also a large amount of attribute information like type, level, address, etc. oriented at user applications; however, ordinary point feature generally expresses the same surface features and reflects a certain geographical phenomenon or geographical features as a whole in collections [11].

Through the research, it found that: although unpredictable non-rigid deviation exists in POI and road network between road networks, there is a certain correlation of geometric pattern between POI and the associated roads. Urban roads are fundamental features of the city map and also essential facilities of a city. The road expresses the urban area as a framework. Road grid is formed through interlaced road, naturally dividing the urban space into different areas. POI distribution has important relevance with roads. Generally speaking, most of the POI is distributed on both sides of urban main roads and internal roads with a part of POI distributing inside the road grid in different modes.

POI generalization can be divided into a generalization based on attribute and a generalization based on mapping presentation. The former mainly involving in data positioning and data query, can filtrate POIs to obtain a certain or a certain type of POI according to attribute conditions assigned by users. Generalization results mainly depend on the richness of POI's attribute information, while the latter is mainly used for map viewing, etc. without input requirements assigned by users, whose purpose of generalization is to provide users with more information about the cognitive environment. If rich types and a large amount of POIs should be processed, the generalization process shall meet the following two requirements: (1) POI data should be generalized in pictures to maintain the spatial distribution features of data; (2) Attribute information like types and grades, etc. of POI points should be considered

during generalization. Such generalization is a generalization based on map presentation, which is a research emphasis of this paper.
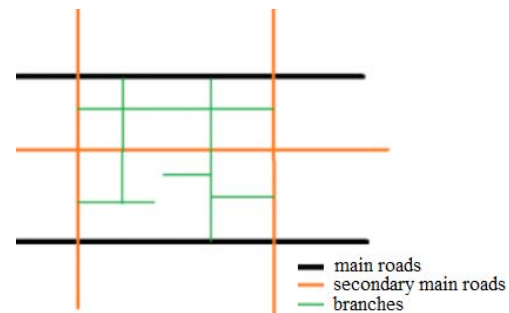

Fig. 1 Urban Road Network

## III. HIERARCHY FEATURES OF POI INFORMATION PRESENTATION

The map as an important presentation tool of spatial information that can transfer rich spatial information to users through various map features The process of the reader reading and analyzing a map is a process to gradually accept the spatial information of the map presentation from the whole to the local and to the individual feature via analysis from the perspective of people's perception on maps. In the process, the map data shows different levels of spatial information to readers. Observing POI data from the holistic view, the first perception obtained is differences in the overall distribution of POI, transferring cluster information of point groups; In the process of further amplification, it can go deeply into each cluster to obtain spatial distribution information that is generated by relationship between points and adjacent points, and then, a single feature is observed further to present the attribute information like position, name, type and grade, etc. of POI. Therefore, spatial information presented by POI has been analyzed from a global clustering hierarchy, local structure hierarchy and single feature hierarchy in this paper.

A. Global Hierarchy

Spatial clustering is one of the important research contents in the areas of data pattern recognition and knowledge mining, etc., which can disclose distribution mode of spatial data to find out relations and regulations existed in the data, thus mining spatial knowledge implied in the data. Spatial clustering divides spatial goals into several clusters with certain meanings, making spatial targets of each cluster to have the maximum similarity, while targets among clusters have the maximum difference [15]. Many researches on concrete methods of clustering analysis have been conducted and will not be covered in the paper.

Upon conducting clustering analysis on POI data inside the road network, POI distribution models in the space of urban road network can be obtained: most POIs are aggregated around the roads, presenting in linear distribution, while part of POIs are aggregated inside the network grid, presenting in cluster distribution, and minor POI are distributed dispersedly

inside the network grid. As POI data have different distribution patterns, the diversity of local structure is a problem that must be considered in the generalization process.
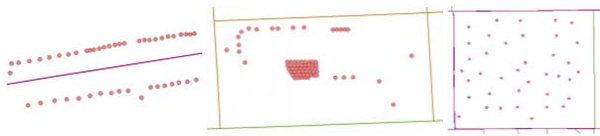
(a) Linear distribution (b) Cluster distribution (c) Discrete distribution

Fig. 2 Distribution patterns of POIs in road network

B. Hierarchy of Local Structure

Diversity of POI distribution patterns represents the diversity of data local structure. POIs in linear distribution and cluster distribution feature large quantity, concentrated distribution and large density, while POIs in discrete distribution feature small quantity, wide distribution and small density. Density is a field expression of space. Each point in space determines local clustering intensity based on the spatial relationship with adjacent points [16]. Distribution density of point group reflects the loose coupling of data. Difference in distribution density is one of the important reasons why local spatial structures of point groups are different. Through analyzing from the geometric angle, distribution density of point groups is relevant to the quantity and distribution range of points, while distribution density of a single point is jointly determined by distances and directions among all adjacent points. Therefore, regarding density measurement, POIs in linear distribution model with one-dimensional feature in space shall start with distances among points, while POIs in cluster distribution start with influencing range of points.

C. Hierarchy of Single Feature

As previously mentioned, POI features rich attributes, diverse types, and large volumes of data and updated geographic information, etc. Uniqueness among POI is one of the important distinctions between POI features and points features of an ordinary map. Each POI has a certain influencing range that is limited by activity space and hobbies, etc., and thus, few people can know all distribution of POIs. Usually only POIs of high significance are known to most citizens [10]. When map users do not have any specific query and filtrating requirements, POI features like well-known landmark building of high significance in the cognitive environment will be presented preferentially, providing users with map guidance and positioning. Location, type and size, etc. of the POI will influence the significance of POI. Therefore, the significance of POI can be regarded as a standard to measure the importance of the single POI spatial information.

IV. Generalization Methods

Generalizing operation of point data includes selection, aggregation, typification and shift, etc. The characteristic of each operation is presented in Table I. In consideration of the particularity of POI data, generalizing operation should meet the following requirements: (1) the generalization result should be a subset of the original point set without generating new data; (2) It should both estimate spatial features and considerate attribute features in the generalization process. The aggregation, typification and shift generalization algorithm fail to meet the first requirement, which are not suitable for POI data, while the selection and simplification fail to considerate both the spatial features and attribute features, which are unable to be used directly in POI generalization. So, POI selection below is a supplement of point selection algorithms, which is also a generalization method considering both spatial feature and attribute feature.

TABLE I
GENERALIZATION OPERATORS FOR POINT FEATURES

Generalization operator	Considering spatial feature	Considering attribute feature	Results with reduced amount of points	New points generating
Selection	No	Yes	Yes	No
Simplification	Yes	No	Yes	No
Aggregation	Yes	No	Yes	Yes
Typification	Yes	No	Yes	No
Shift	Yes	No	No	Yes

As we can see in Fig. 2, distribution of urban POI can be divided into three distribution patterns of different characteristics: linear distribution pattern, cluster distribution pattern and discrete distribution pattern. Characteristics of local structure of diversification of POI data have been reflected. So, corresponding selection methods are proposed as follows in this paper for the three distribution patterns.

A. Significance Selection Method

The amount of POI in discrete distribution is less with low spatial density. Thus, the user concerns more about the information included in a single POI. This type of POI is selected by significance from top to bottom. A small amount of POI such as landscape sites, tourist attractions and residential districts, etc. are often far away from the main roads, which are sparsely distributed within the road grid. When the scale becomes small in a map with only partial POIs, POI with high significance should be reserved. Some scholars studied POI significance. For example, discussion was conducted from public cognition, spatial distribution and individual feature discuss in literature [7]. A measurement model was built from popularity, accessibility and individual feature in literature [17]. Either way, the measurement model can be summarized as the following form:

$$\mathrm{Sig.} = \sum_{i=1}^{n} w_i * Q_i \qquad (1)$$

where, *Sig.* is the significance of POI; *Qi* is value of each influencing factor; *wi* is the influencing weight of each factor.

POI significance as a measuring feature of POI attribute, is an influencing factor that must be considered in the selection algorithm. In the same distribution pattern, the influences of spatial features on the significance are consistent. Thus, the significance can be measured by public cognition of the POI

type and the individual scale. There is no uniform standard for POI classification and coding now. Urban POIs are divided into several class 1 and then divided into class 2 in details according to standard of Map World, Baidu Map and ESRI. Some apps might also divide class 2 into more detailed class 3. POI scale information is also implied in the detailing process of POI classification. For example, class 1 hotels are divided in class 2: star-rated hotels, chain hotels, guest houses and apartments, which imply the individual feature of POIs. Therefore, measurement of POI significance has been transformed into division and rating of POI defining types.

B. Selection Method of Weighted Voronoi

Spatial distribution features of POI should be maintained during selection due to POIs in cluster aggregation with large point density. Voronoi dynamic building sampling method proposed in literature [1] can maintain relative density of points, which is applicable to selection of POI in cluster distribution. Combining with POI characteristics, individual information of POI should also be considered in the selection process. Thus, Voronoi dynamic rebuilding sampling method has been improved with the following steps:

1) Regard the point as the generating element of Voronoi to set up Voronoi figure for the point.
2) Weighted density value of POI is calculated according to

$$\Pr om. = Sig. / Area. \qquad (2)$$

where, $\Pr om.$ is the weighted density value, $Sig.$ is the POI significance and Area is the area of V zone of POI.

3) Sort weighted density of POI and selected a POI target and its adjacent first-order POI from small to large. The point reserved is called the fixed point.
4) Check the amount of fixed points. If the amount reaches the selection requirements, then it will enter into step 5; otherwise, it shall remain in step 3. If the amount of fixed points remains unchanged, it shall return to step 1 with the fixed point as the original point set.
5) Selection is completed.

C. Linear Interval Selection Method

The Characteristic of linear POI data distributed along the road is simple structure, large data volume and abundant types. Selection of the kind of point adopts an interval selection method that is to simulate peasant's farming operation. When field crops grow too dense, it is necessary to remove some of the seedlings at regular intervals to ensure better growth of the remaining seedlings. Similarly, interval selection is to iteratively select POI points of linear distribution according to a certain step length. POI is selected according to the fixed step length and dynamic starting point, so as to guarantee the relative distribution density of POI. At the same time, it shall prioritize reservation of important POIs with consideration of attributes of POIs within a step length. Specific process is as follows:

1) With disordered POI set P {o1...Om} of linear distribution, calculate the projection point from each point to the road

line for setting up the corresponding relation between the projection point and the original point.

2) Calculate the distance of each projection point along the road from the topology line of the road to the line starting point, and carry out ordering of the projection points according to the distance to obtain an ordered projection point set P' {o1'....om'}.
3) Interval selection is conducted.

① Determine to select a number threshold C and initialization selects the number as 0. Calculate according to formula (3) to select the step length T:

$$T = L / C \qquad (3)$$

where T is the step length selected, L is the length of road line and C is the amount of threshold selected.

② An endpoint of the road line is regarded as the starting point to obtain subset Q {q1'....qn'} of POI projection point on the length of step length T; And then, the highest significance M selected from Q is put into the reserved point set R.

③ Check whether it reaches another endpoint of the road line, if yes, then enter into step ④; Otherwise, it shall return to step ② with the position of M point as the starting point for continuing selection.

④ Select the top C projection points of POI with the highest significance to look for the corresponding starting points {o1....on}; that is the selection result.
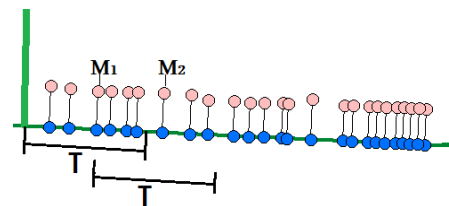

Fig. 3 Interval selection method

V. EXPERIMENT AND ANALYSIS

In this paper, road data and POI data of the main urban area of Weifang, Shandong have been taken as a case study, whose area coverage is about 260 square kilometers. Map data were collected in 2011 with the original scale of 1:500, which contain 40,299 POI points, 425 main roads, 525 secondary main roads, 591 branches and 477 road grid. Experimental steps are designed as follows:

1) Data preparation. Topological structure shall be set up for road data to interrupt road data in the intersection and identify road grid. The significance of each POI is calculated with classification of POI data.
2) Pattern recognition. In order to carry out cluster analysis on POI, the POIs distributing in the road network are divided into three distribution patterns: discrete distribution, cluster distribution and linear distribution.
3) Data generalization. Significance selection method, weighted Voronoi selection method and linear interval selection method have been utilized, respectively, to

conduct data generalization of POIs of the above three distribution patterns.

4) Mapping presentation data results obtained through generalization can be conducted on mapping presentation through symbolization and notes adding.

Based on the WJ-III Map Workstation developed by China Academy of Surveying and Mapping Science, the experiment has been embedded POI parting selection method proposed in the paper to generalize the original data to a measuring scale of 1:2000. Selection proportion can be obtained to be 50% according to root formula. Comparison methods adopted in the experiments include random point selection method, grid point selection method point selection method based on Voronoi area. In order to display the comparison, partial test areas are presented as in Fig. 4.

(a) Original point set   (b) Random selection

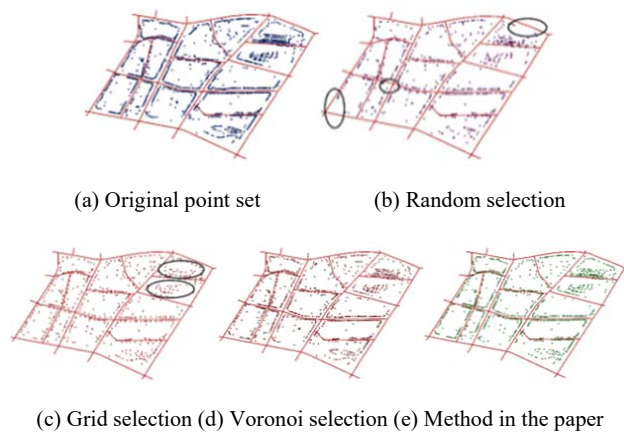(c) Grid selection (d) Voronoi selection (e) Method in the paper

Fig. 4 Generalization Results

Upon comparison of the generalization results, the results of Fig. 4 (b) fail to ensure the spatial distribution feature of data: part of set of points of continuity of linear distribution are deleted, while partial are reserved, and moreover, the characteristic of the selection method determines the uncertainty of the selected results. In Fig. 4 (c), Point density of cluster distribution area is diluted, this is because grid selection method use regular grid to partition data, and it is adverse to maintain relative density among different areas. Fig. 4 (d) presents the selected results based on Voronoi, as selection is conducted on the basis of data density, it can preserve the data spatial distribution feature. Results obtained by the method proposed in this paper are presented in Table II.

In order to compare selection results based on Voronoi and selection results based on the method proposed in this paper, data were added with annotation, as shown in Table II. We can see from the table that, generalization results based on Voronoi miss some points of interest of high significance like hospitals, research institutes and wholesale departments, while the selection process of the method proposed in the paper has considered attribute information while remaining spatial features make up the defect, meeting the generalization requirements.

TABLE II
GENERALIZATION RESULTS WITH ANNOTATION

No	Primary data	Results selected by Voronoi	Results selected by the method proposed in this paper
1	Cai Restaurant	Yes	Yes
2	XUAN Restaurant	No	No
3	Dazhong Noodle Restaurant	Yes	No
4	Flower Shop	No	Yes
5	Shanggong Sewing Machine Shop	Yes	No
6	Werfang Traffic Hospital	Yes	Yes
7	Weifang Ent Specialist Hospital	No	Yes
8	Weifang Ent Research Institutes	No	Yes
9	Shangcheng Cell Phone Store	Yes	No
10	Wenzhou Massage Shop	No	Yes
11	Xiaoxiao Shop	No	No
12	Dazhong Mechanical Wholesale Department	No	Yes

VI. CONCLUSION

Differing from point features of the general map, urban POI data had rich attribute information, whose distribution had a tight relation with the urban road network. The existing researches regarded point feature as point group, the focus of these researches is the spatial structure information of the group distribution, and it is not apply to the urban POI data. Researches on POI mainly centered on multi-source data infusion and data mining as well as application, lacking in-depth research on POI generalization. Given this, the paper has analyzed hierarchical characteristics of spatial information presented by urban POI data and divided POI distribution of urban space into three patterns. Corresponding selection generalization method has also been put forward on the basis of the data characteristic of point sets under each distribution pattern. Compared with the conventional point selection method, the method can consider requirements of spatial distribution features and attributed information of POI data to realize the urban POI data generalization based on mapping presentation.

The method has not considered the influence of symbolization and notes making that will be included in the generalization method in future research. Also, POI shift operation will be added for researching, so as to optimize the generalization results. In addition, correctness and practicality of the generalization methods proposed in this paper have been confirmed in the experiments of small and medium-sized cities. Yet, adaptability of the algorithm should be further verified along with the increasing complexity of data in large and super large cities.

REFERENCES

[1] Ai Tinghua, Liu Yaolin, A Method of Point Cluster Simplification With Spatial Distribution Properties Preserved (J). Acta Geodaetica et Cartographica Sinica, 2002, 31(2):175-181.
[2] Yan Haowen, Wang Jiayao. A Generic Algorithm for Point Cluster Generalization Based on Voronoi Diagrams (J). Journal of Image and Graphics, 2005, 10(5):633-636.
[3] Cai Yongxiang, Guo Qingsheng. Points Group Generalization Based on Kohonen Net (J). Geomatics and Information Science of Wuhan

University, 2007, 32 (7):626-629.

[4] Deng Hongyan, Wu Fang, Qian Haizhong.A Model of Point Cluster Selection Based on Genetic Algorithms (J). Journal of Image and Graphics, 2003, 8(8):970-976

[5] Li Wenjing, Lin Zhiyong, Long Yi. Application of Rough Set Idea to Points Object GIS Generalization (J). Geomatics and Information Science of Wuhan University, 2008, 33(9):896-899.

[6] Xu Zening, Gao Xiaolu. A Novel Method for Identifying the Boundary of Urban Built-Up Areas with POI Data (J). Acta Geographica Sinica, 2016, 71(6):928-939.

[7] Zhao Weifeng, Li Qingquan, Li Bijun. Extracting Hierarchical Landmarks from Urban POI Data (J). Journal of Remote Sensing, 2011, 15(5):973-988.

[8] Yu Wenhao, Ai Tinghua, Liu Pengcheng. Network Kernel Density Estimation for the Analysis of Facility POI Hotspots (J). Acta Geodaetica et Cartographica Sinica, 2015, 44(12):1378-1383.

[9] Chen Weishan, Liu Lin, Liang Yutian. Retail Center Recognition and Spatial Aggregating Feature Analysis of Retail Formats in Guangzhou Based on POI Data (J). Geographical Research, 2016, 35(4):703-716.

[10] Guo Lishuai. Study on Parallel Computing Methods of POI Simplification (D). Nanjing: Nanjing Normal University, 2013

[11] Yu Yanping. Dynamic Generalization Methods of POI for the Representation of Mobile Map (D). Nanjing: Nanjing Normal University, 2012.

[12] Ying Shen, Chen Guiqiu, Cao Xiaohang.POI Selection with the Reference of Linear Roads in Multi-scale Maps (J). Engineering of Surveying and Mapping, 2014, 23(7):6-11.

[13] Yang Min, Ai Tinghua, Lu Wei. A Real-Time Generalization and Multi-Scale Visualization Method for POI Data in Volunteered Geographic Information (J). Acta Geodaetica et Cartographica Sinica, 2015, 44(2):228-234.

[14] Liu Xinhong, Li Tian, Zhang Yongbo. Point Features Thinning Methods Based on ARCGIS (J). Academic Research, 2015(1):31-34.

[15] Li Guangqiang, Deng Min, Liu Qiliang. A Spatial Clustering Method Adaptive to Local Density Change (J). Acta Geodaetica et Cartographica Sinica, 2009, 38(3):255-263.

[16] Yu Wenhao, Ai Tinghua. The Visualization and Analysis of POI Features under Network Space Supported by Kernel Density Estimation (J). Acta Geodaetica et Cartographica Sinica, 2015, 44(1):82-90.

[17] Chen Xiang, Li Xiaoming, Zhan Ran, Xu Weimin. Studying on Extracting Hierarchical Landmarks from Urban POI data (J). Geomatics and Spatial Information Technology, 2015, 38(10):130-133.