A Summary of analysis and application research on the spatial distribution of POI data based on urban service industry

Li Zhu¹, Lu Wang*, Xiaoling Li² and Lei Zhang³
¹Department of Computer Technology and application, Qinghai University, Xining, Qinghai, 810016, China
²School of Civil Engineering, Qinghai University, Xining, Qinghai, 810016, China
³College of Computer Science, Information Management Center, Sichuan University, Chengdu, Sichuan, 610065, China
*Corresponding author’s e-mail: wlgql@126.com

Abstract. Data mining and data analysis based on POI data of urban service industry has gradually become the research trend of urban spatial layout. In the present study, many scholars use different research methods to analyze the agglomeration characteristics and development trend of urban service industry from multiple perspectives. This article combines POI data analysis methods and application fields to summarize the relevant research of urban service industry POI data analysis and application and summarizes the representative research methods of POI data analysis, and compares the characteristics of each method, and then the development trend of POI data mining, analysis and application in urban service industry is prospected.

1. Introduction

In the era of big data, the rapid development of data mining technology is an important force to promote the research on the spatial layout characteristics of POI data in the urban service industry [1]. Studying the formation mechanism of urban service industry, the development model of agglomeration areas and the rationality of spatial layout, exploring the agglomeration trends, agglomeration motives and agglomeration characteristics of the service industry will help promote the development of urban service industry and accelerate the urbanization process [2-4]. POI stands for "Point of Information". POI data of urban service industry represents the spatial data of service industry entities in geographic information system. Compared with traditional geospatial data, POI data is easy to obtain and has high practical significance [5]. Mining the potential value behind POI data in the study of urban geography is an important content of the spatial layout of urban service industry. At present, there are many kinds of methods for POI data analysis at home and abroad, and each method has its own advantages and limitations. Therefore, this paper focuses on the urban service industry POI data, based on the five types of data analysis algorithms such as kernel density analysis, summarizes the basic principles of each algorithm in detail, and then briefly describes the city service industry POI data research case under this method. Finally, taking the POI data of the entertainment industry in Chengxi District of Xining City as an example, the characteristics, research contents and application fields of the five algorithms are compared, and the development trends of the spatial distribution characteristics of urban service industry are discussed.
2. POI data analysis method

2.1. Kernel density estimation

Kernel density estimation is currently one of the most commonly used methods in POI data analysis. Mainly by calculating the density of point or line features in its neighborhood, to analyze the distribution characteristics of the data and discuss the distribution rules of the data. In the spatial agglomeration analysis of POI data, the kernel density estimation needs to select an appropriate distance as the search radius. The kernel density value represents the degree of aggregation of POI \(^{[6-7]}\).

The kernel density calculation formula is defined as follows.

\[
f(x) = \frac{1}{nh^d} \sum_{i=1}^{n} k\left(\frac{x-x_i}{h}\right)
\]

Where \(f(x)\) is the kernel density calculation function at point \(x\), \(k\) is the spatial weight function; \(h\) is the search radius, \(d\) is the data dimension, \(n\) is the number of POI within the search radius.

Aiming at this method, domestic scholars analyzed the POI data of urban service industry and obtained the aggregation characteristics of POI data. Ran Zhao et al. proceeded from the urban service industry as a whole to discuss the characteristics of the spatial pattern of life service industry \(^{[8]}\). In the field of the retail industry, Bai Yongping et al. analyzed the agglomeration trend of retail pharmacies in Lanzhou city to explore their differentiation characteristics and accessibility level, so as to reasonably allocate urban service industry resources \(^{[9]}\). In the field of express logistics, Li Gang et al. took Xi'an express industry as an example, based on the two types of perspectives of dependence type and object type, and studied the spatial layout characteristics of express pickup points from the macro and micro characteristics through this method \(^{[10]}\). In the field of public facilities and services, Cheng Chezhi et al. took Hefei as the research area, excavated the POI data of leisure function areas, and combined the research results to study the spatial network system \(^{[11]}\). It can be seen that the research content of the kernel density estimation method in the spatial layout of the urban service industry is relatively extensive.

2.2. Standard deviation ellipse

The standard deviation ellipse is an algorithm used to analyze the directional characteristics of the spatial distribution of POI data. The result is an ellipse that can reflect the dominant direction and distribution profile of geospatial data as a whole \(^{[12]}\). The size of the ellipse reflects the concentration degree of the overall elements of the urban spatial pattern. The area of the ellipse is the main spatial distribution range of the POI data. The direction of the long axis of the ellipse indicates the main trend direction of the POI data distribution. The main steps of the standard deviation ellipse method are the determination of the center \(o\), the rotation angle \(\alpha\) and the length of the X and Y axis (Figure 1).

![Figure 1 Standard deviation ellipse](image)

Based on the standard deviation ellipse method, He Weichun et al. used Kaifeng City as the research area to classify the types of retail commercial outlets and study their retail commercial spatial structure and influencing factors \(^{[13]}\). Tu Jianjun and others took the overall geographical structure of the city as a starting point and analyzed the spatial heterogeneity and layout direction characteristics of the location selection of the catering industry from the perspective of the mountainous urban pattern in the main urban area of Chongqing \(^{[14]}\). Taking the cultural facilities as the research object, Zhao Hongbo et al. used the standard deviation ellipse method and the kernel density estimation method to
quantitatively analyze the evolution of the spatial layout of cultural facilities in Zhengzhou in 2007 and 2017\cite{15}.

2.3. **DBSCAN**

DBSCAN is an unsupervised learning algorithm\cite{16}, which is sensitive to the parameters Eps and Minpoints. It can find noise points and generate clusters of arbitrary shapes. It has broad application prospects in data mining, geospatial data processing, machine learning and other fields. The process of DBSCAN clustering algorithm is as follow:

1. Scan the data set D and determine all core objects according to the given neighborhood parameters Eps and Minpoints.
2. Select an unprocessed core object, find a sample whose density is up to generate a cluster
3. Repeat the above process until all points in the data set D are classified as a cluster or marked as noise points.

The advantage of DBSCAN is that it can distinguish the differences among clusters. Based on this feature, Li Xinyan et al. used this method to process POI data in the public service industry, and proposed a targeted development strategy by analyzing the clustering results and distribution patterns\cite{17}. Yang Fan et al. took the urban catering industry as a research perspective, analyzed its hierarchical structure, explored the characteristics of the catering industry cluster under the spatial pattern, and provided scientific basis and guidance for the layout planning of the urban catering industry\cite{18}. Regarding the problem of urban traffic congestion, Liu Chang et al. based on GPS trajectory data, predicted the next time of traffic congestion, and used the perspective of traffic planning as the starting point to provide research directions for relevant departments to deal with traffic problems\cite{19}. In the planning of the spatial pattern of the service industry in the new urban area, Li Jiangsu et al. took Zhengdong New Area as a research case to perform cluster analysis on the service industry of Zhengdong New Area and provide development suggestions for its functional area planning\cite{20}.

2.4. **Complex network**

A complex network is composed of multiple nodes, and there are certain connections between the nodes. When using complex network analysis method to process POI data, by comparing the distance between any two POI points of interest and the parameter radius R, the matrix and the network topology are obtained. This method mainly analyzes the spatial layout characteristics of POI data through five parameters (Table 1) of network degree, network density, K-core, average path length, and network clustering coefficient in complex networks\cite{21}.  

| Parameter       | Concept                          | Characteristic                                      |
|-----------------|----------------------------------|-----------------------------------------------------|
| Degree          | Number of neighboring nodes connected to nodes | Describe the strength of node connectivity           |
| Density         | The ratio of the actual number of edges to the number of edges that can be accommodated | Describe the density and sparseness of the network |
| K-core          | Any node has at least K neighbors | Describe the level of network cohesion              |
| Characteristic path length | The average of all node distances | Indicators for measuring network convenience |
| Clustering coefficient | The average value of the clustering coefficient of all contacts | Represents the degree of network node aggregation |

The complex network has obvious industry characteristics in analyzing the urban service industry. In related research, the layout characteristics of bank outlets are mostly analyzed. Zhen Maocheng et al. studied the layout characteristics and spatial differences of bank outlets under two different service...
radii from the perspective of a complex network by constructing a network topology diagram [22]. Li Feng et al. implemented an optimization model for the layout of bank outlets and analyzed the reasons for the unreasonable layout of bank outlets [23]. Wan Ma provided scientific basis and guidance for bank location planning by exploring the law of spatial agglomeration of bank outlets [24]. In addition, Shi Jiayi et al. used this method to study the unreasonable characteristics of the network and layout of gas stations in Shanghai, and to analyze the differences between different agglomeration areas [25].

2.5. Getis-Ord $G'_i$

In the study of the distribution hotspot of POI data of urban service industry, local characteristics are generally used as indicators [26]. Getis-Ord $G'_i$ can be used to measure the correlation of a certain geographical phenomenon or a certain attribute value in the adjacent location, and can identify POI spatial clustering with statistical significance, so as to analyze the hot spots of urban service POI clustering.

The method based on Getis-Ord $G'_i$ is the most commonly used method in the analysis of hotspot characteristics of urban service industry and can accurately excavate the hotspots of clustering. Chen Weishan et al. used this method to analyze the POI data of the retail industry in Guangzhou and proposed a method to identify hot spots in the distribution of the retail industry, so as to improve the scientific basis for rationally optimizing the allocation of urban retail resources [27]. Li Weiwei et al. took Xi’an tourism as the research object and analyzed the formation mechanism of its hot spot area [28]. Xue Bing et al. summarized the characteristics of the distribution structure of the retail industry in Shenyang and compared the differences in the spatial layout of different retail industries [29]. Cao Fangjie et al. combined Getis-Ord $G'_i$ with kernel density estimation to study the commercial spatial pattern of the central city of Beijing [30].

3. Algorithm comparison analysis

Based on the spatial data mining technology, this paper crawls the POI data of entertainment industry in Chengxi District, Xining City, Qinghai Province through online electronic map, and realizes the visualization of POI original data map of Chengxi district in Figure 2 (a). At the same time, based on the data set, five algorithms are implemented and compared.

The Kernel density estimation method is easy to realize and has a wide range of research contents, and can visualize the differences of urban service industry agglomeration. Figure 2 (b) uses this method to process the POI data in Chengxi district to analyze the agglomeration degree of the entertainment industry in Chengxi district, which better reflects the diffusion trend and distance attenuation effect of the entertainment industry in Chengxi district in the spatial distribution. However, the lack of quantitative statistical indicators to evaluate the spatial distribution characteristics of the entertainment industry in Chengxi district makes it impossible to accurately identify the distribution hot spots of the entertainment industry from the statistical significance level.
Based on the standard deviation ellipse method, Figure 2(c) selects three standard deviation levels to study the urban entertainment industry. This method has certain advantages in exploring the differences in the extent and directionality of the service industry. Commonly used in the determination of the spatial distribution profile and main development direction of the urban service industry, if the POI data of different time periods in the region are combined, the spatial distribution evolution process of the urban service industry in the region can be clearly found.

It can be seen from Figure 3 (c) that DBSCAN can identify noise points and generate clusters of any shape. Compared with the Kernel density estimation, which can only analyze the overall structure characteristics, DBSCAN can explore the local characteristics and details of urban service industry, cluster various industries of urban service industry, and can deeply study the scale, spatial form and agglomeration degree of the various services industries. However, when the density of POI data distribution is not uniform, its clustering effect is poor.

Figure (e) is based on the research of complex network method, which is easy to find the degree of aggregation and connection of POI. The research content of complex network method mainly focuses on the service industry with special service nature, such as bank outlets and California sites. Regarding the analysis of the complex network characteristics of the city service industry POI, it is mainly used in the location selection of the city service industry outlets and the judgment of the rationality of the network outlets in order to optimize the spatial layout of the city service industry.

Figure (f) shows the clustering results of POI data in Chengxi district under different confidence levels. It can be seen that the Getis-Ord $G^*_i$ method is a kind of local spatial autocorrelation analysis with high accuracy, which can mine out the hot areas of urban service industry through the statistical
significance parameters of hot spots, and optimize the resource allocation of urban service industry in zoning planning. The analysis of hot spot characteristics in retail, finance, transportation service, entertainment and other industries is the focus of identification of hotspot centers in urban service industry.

4 Conclusion and discussion

Although various methods based on POI data to study the spatial distribution characteristics of urban service industry have made good achievements in urban planning, site selection of service industry outlets, resource allocation and other aspects, there are still deficiencies. In the research, the method of combining spatial data mining and analysis technology with urban geography theory should be adopted to qualitatively and quantitatively analyze the formation mechanism and development status of the urban service industry spatial layout, so as to better solve the problems encountered in the research. The prospects for the research and analysis methods and applications of POI data are as follows:

(1) Before the POI data analysis, the quality of the dataset has a decisive effect on the results of the data analysis. Therefore, the mining and preprocessing of the POI data in the early stage is very important.

(2) Urban geography research based on POI data is an important research content in the era of big data. In the study of the spatial layout of the urban service industry, the POI data in different regions will have the characteristics of complexity and diversity. In data processing and actual analysis, specific analysis methods suitable for the region should be selected according to the research needs.

(3) Each analysis method has some limitations when processing and analyzing POI data. The combination of multiple algorithms for analysis can more comprehensively study the spatial layout characteristics of the urban service industry.

(4) In the study of the agglomeration characteristics of the urban service industry, the research perspective should be expanded, data models should be continuously explored, innovative research methods should be explored, and the research content and application fields should be enriched.

Acknowledgments

This paper was supported by the Youth-natural Science Foundation Project of Science and Technology Department of Qinghai Province (NO. 2019-ZJ-985Q)

References

[1] Zhou C G.. (2016) The value of spatial data in the era of big data -- a review of the theory and application of spatial data mining [J]. Journal of geography, 71 (07): 1281.
[2] Chen X., Lv B.. (2014) Study on the development mode of service industry cluster in small and medium-sized cities [J]. Economic geography, 34 (04): 105-111.
[3] Fang Y P., Yan X P.. (2004) A Comparative Study on the Characteristics and Trends of Service Industry in Coastal Central Cities of my country Since the 1990s: Taking Beijing, Shanghai and Guangzhou as Examples [J]. Economic Geography, (05): 614-619+637.
[4] Li H J.. (2013) A review of the research on service industry agglomeration abroad [J]. Cooperative economy and technology, (13): 18-20.
[5] Zhang T Y ., Li H W., Xu D H., et al. (2016) Visualization method of interest point data using density clustering algorithm [J]. Surveying and Mapping Science, 41 (5): 157-162.
[6] Gao Z Y., Zhang H F.. (2019) Spatial pattern of retail industry in Xining City Based on POI data [J]. Geography of arid area, 42 (05): 1195-1204.
[7] Ran Z., Zhou G H., Wu J M., et al. (2019) Spatial pattern of life service industry in Changsha Based on POI data [J]. World geographic research, 28 (03): 163-172.
[8] Bai Y P., Zhang W X., Wang Z G.. (2020) Distribution characteristics and accessibility of pharmaceutical retail stores based on POI data -- Taking Lanzhou as an example [J]. Journal of Shaanxi University of science and Technology (NATURAL SCIENCE EDITION), 36
[9] Li G., Yang L., he J X., et al. (2018) Spatial pattern and spatial relationship of Xi'an express delivery self delivery point based on POI data -- Taking Cainiao post as an example [J]. Geosciences, 38 (12): 2024-2030.

[10] Li G Q., Jin F J., Chen Y., et al. (2017) Location characteristics and differentiation mechanism of Beijing Logistics Industry Based on POI [J]. Journal of geography, 72 (06): 1091-1103.

[11] Cheng C Z., Liu Z T., Zhang Q.. (2019) Study on spatial distribution and evaluation of Hefei Urban Leisure functional area based on POI data [J]. Journal of western human settlements, 34 (02): 74-79.

[12] Chen H X.. (2018) Evolution of spatial distribution and agglomeration characteristics of producer services in Beijing [J]. Economic geography, 38 (05): 108-116.

[13] He W C., Li E L., Cui Z Z., et al. (2018) Spatial distribution of retail business in the main urban area of Kaifeng City and its influencing factors [J]. Economic geography, 38 (09): 158-167.

[14] Tu J J, Tang S Q., Zhang Q., et al. (2019) Spatial heterogeneity of influence of mountain city pattern on location selection of catering industry [J]. Journal of geography, 74 (06): 1163-1177.

[15] Zhao H B., Yu D F., Miao C H., et al. (2018) Study on the location layout characteristics and influencing factors of Zhengzhou cultural facilities based on POI data [J]. Geosciences, 38 (09): 1525-1534.

[16] Ester M., Kriegel H P., Sander J., et al. (1996) A density-based algorithm for discovering clusters in large spatial databases with noise[C]. Knowledge Discovery and Data Mining, 226-231.

[17] Li X Y., Li D R.. (2005) DBSCAN spatial clustering algorithm and its application in urban planning [J]. Surveying and Mapping Science, 03: 51-53 + 5.

[18] Yang F., Xu J G., Zhou L.. (2016) Catering cluster recognition and spatial feature analysis in Guangzhou based on DBSCAN spatial clustering [J]. Economic geography, 36 (10): 110-116.

[19] Liu C., Li Z J., Jiang S X.. (2015) Discovery of urban traffic congestion area based on DBSCAN algorithm [J]. Intelligent computer and application, 03: 69-71.

[20] Li J S., Liang Y., Wang X R.. (2018) Spatial clustering of service industry in Zhengdong New Area Based on POI data [J]. Geographic research, 37 (01): 145-157.

[21] He Y B., Zhou Y D., Zhang H.. (2018) Study on the layout characteristics of urban residential areas based on complex network theory [J]. Journal of Wuhan University of Technology (information and Management Engineering Edition), 40 (01): 55-60.

[22] Zhen M C., Zhang J Q., Yang G L.. (2013) Characteristics of outlets layout of commercial banks based on complex networks: a case study of Bank of China in Beijing [J]. Progress in Geosciences, 32 (12): 1732-1741.

[23] Li F., Xue Y W.. (2011) Research on bank network layout based on complex network [J]. Complex system and complexity science, 8 (02): 64-70.

[24] Wan M.. (2015) Analysis on the layout of domestic commercial banks' sub branches in Southwest China based on complex network -- Taking Nanning China Construction Bank as an example [J]. Knowledge economy, 12: 52-53

[25] Shi J Y., Guo J L.. (2019) Study on the layout characteristics of gas stations in Shanghai from the perspective of complex network -- Taking CNPC as an example [J]. Software guide, 18 (09): 146-149.

[26] Yu W H., Ai T H., Yang M., et al. (2016) Hot spot detection of urban facility interest distribution using nuclear density and spatial autocorrelation [J]. Journal of Wuhan University (Information Science Edition), 41 (02): 221-227.

[27] Chen W S., Liu L., Liang Y C.. (2016) Hot spot identification and business cluster characteristics of Guangzhou retail business center based on POI data [J]. Geografie research, 35 (04): 703-716.

[28] Li W W., Chen T., Ma X L.. (2020) Hot spot characteristics and formation mechanism of Xi'an
urban tourism and leisure industry space [J]. Geosciences, 40 (03): 437-446.
[29] Xue B., Xiao X X., Li J Z., et al. (2018) Spatial hot spot analysis of urban retail based on POI big data -- Taking Shenyang City, Liaoning Province as an example [J]. Economic geography, 38 (05): 36-43.
[30] Cao F J., Xing H F., Hou D Y., et al. (2019) Identification and spatial pattern of Beijing business center based on POI data [J]. Geographic information world, 26 (01): 66-71.