Spatial Distribution Analysis of Urban Retail Industry Using POI Big Data

Hongwei Su [1,3] and Maria Amelia E. Damian [2]
[1]Graduate School, University of the East, Manila, Philippines
[2]Faculty College of Engineering,University of the East, Manila, Philippines
[3]China Urban Construction Desig&Research Institute CO.,LTD, Beijing,China
mariaamelia.damian@ue.edu.ph

***Abstract*—Identifying urban functional areas and understanding their spatial distribution characteristics are crucial for urban planning and government decision-making. The spatial distribution of the urban retail industry plays a significant role in promoting economic development, optimizing urban structure, and meeting residents' consumption needs. This study analyzes the spatial distribution characteristics and formation mechanism of different retail industry clusters in Nanning, China, using Point of Interest (POI) data, Gross Domestic Product (GDP), population, and transportation data. The methods employed include kernel density estimation, spatial correlation analysis, Getis-Ord G* index, and average nearest neighbor distance. The results reveal that (1) the retail industry is concentrated in Nanning's central urban area, spreading from the center to the surrounding areas; (2) retail hotspots exhibit a multi-center development and surrounding spread pattern; and (3) the spatial distribution of the retail industry is highly correlated with main traffic lines, economy, and population. This research provides valuable insights for optimizing urban commercial facility layout and decision support for urban commercial space optimization.**

***Index Terms*—POI data, Business center identification, Retail industry layout, Data analysis**

I. Introduction

Urban retail industry is one of the important factors to evaluate the overall development level of a city. It has a positive effect on increasing residents' income, solving population employment, and stimulating domestic demand. Since China's reform and opening up in 1978, the retail industry has developed rapidly. In 1993, foreign-funded retail enterprises entered the country, and the retail industry developed further, forming a variety of retail formats such as shopping centers, supermarkets, and department stores. The rapid development of e-commerce such as Alibaba Mall, Jingdong Mall, and Suning eBay has had a huge impact on the traditional retail industry in our city. According to the "2015 China Retail Industry Development Report", the growth rate of total retail sales in 2015 was only 4.3% (the growth rate in 2010 was 21%), and there were 31 retail companies with negative sales growth. It is particularly important to analyze the main factors that affect the spatial layout of the retail industry, explore new retail models, increase the number of outlets, adjust business formats, and promote the development of the retail industry. Several issues need to be considered in the transformation and development of urban retail industry: Is the spatial layout of retail industry reasonable? What is the formation mechanism of retail space layout? What is the spatial correlation between retail trade and factors such as transportation, population, economy, etc.?

This study is based on the hot POI data of commercial institutions in Nanning City, combined with GDP, population, transportation and other data, and analyzes the spatial distribution pattern of different retail industry clusters through kernel functions and other methods. Density method, spatial correlation analysis, Getis Ord Gi* index statistics, average nearest neighbor distance, in order to provide reference for the rational allocation of commercial resources in the city. Further study the formation mechanism of the spatial layout of the retail industry, analyze the spatial relationship between the retail industry and population, economy, transportation, etc., in order to enrich the research content at the micro level of the urban retail industry location, and provide a certain basis for the urban retail industry location. Decision support for urban commercial space optimization.

II. Overview of Relevant Literature

In recent years, Point of Interest (POI) data, one of urban spatial geographic data, has emerged. POI data has the characteristics of rich data volume, strong temporality, and high accuracy. Statistical learning methods can be applied to analyze the spatial distribution patterns and influencing factors of urban retail industries using large-scale POI data in the era of big data [1]. It has certain advantages in urban infrastructure research, retail agglomeration analysis, commercial spatial pattern, city center identification, population and economic distribution, and housing layout. In the retail industry, Chen Weishan et al. used POI data to study the characteristics of hotspot identification and format accumulation in Guangzhou's retail industry [2]. Xue Bing et al. analyzed the spatial hotspots of Shenyang's retail industry based on POI big data [3]. Gao Ziyi et al. used POI data to explore the spatial pattern of retail industry in Xining [4]. To sum up, it is a new trend to use POI big data to quantify commercial agglomeration space. However, based on the spatial correlation of various influencing factors such as transportation, population, and economy, there are few empirical studies on the spatial layout of different retail formats, analysis of agglomeration areas, and identification of hot spots.

III. Theoretical framework

A. Basic Theory of Spatial analysis

GIS Spatial analysis refers to the realization of spatial data analysis in GIS (geographic information system), that is, obtaining and processing the spatial location, distribution characteristics, formation and evolution of the geographical objects studied from the spatial data. This article will analyze the spatial distribution of stores and their interaction with commercial districts based on actual geographical factors, in order to broaden the perspective for exploring the issues between retail industry and urban commercial districts.

B. POI

POI is the abbreviation for "Point of Interest", which can be translated as "Point of Interest" in Chinese. It refers to a facility point (set) with latitude and longitude information. POI data is a commonly used data type in spatial data analysis. People can understand the location of a city's central area through the aggregation of POI data. Whether it is retail or catering, it can be completely presented through POI Data and information visualization.

C. Business district (commercial center)

A business district (commercial center) refers to the radiation range that a store extends in a certain direction and distance, centered around its location, to attract customers. Simply put, it refers to the area where customers come to the store to live. Whether it is a large shopping mall or a small store, their sales always have a certain geographical range, which is centered around the mall and radiates to the location where potential consumers who may come to the store to purchase reside.

IV. Methodology/Research Design

A. Nuclear density estimation

Kernel Density Estimation is a type of nonparametric estimation proposed by Rosenblatt (1955) and Emanuel Parzen (1962), also known as Parzen window. Ruppert and Cline proposed a revised kernel density estimation method based on the dataset density function clustering algorithm. The nuclear density estimation formula (1) is obtained:

$$p(x) = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{h}K(\frac{x-x_k}{h}) \tag{1}$$

Where the parameter h is called bandwidth, $K(x)$ is called kernel function, and $K(x)$ satisfies the following conditions:

$K(x) \geq 0, \int K(x)dx = 1$

$\int xK(x)dx = 0$

$\int x^2K(x)dx > 0$

Correlation analysis

The correlation coefficient represents the correlation between two spatial variables, and the calculation method is shown in formula (2).

$$Corr_{ij} = \frac{\sum_{k=1}^{n}(z_{ik}-u_i)(z_{jk}-u_j)}{(m-1)\delta_i\delta_j} \tag{2}$$

Among them: $Corr_{ij}$ is the correlation coefficient, i and j represent the i-th and j-th grid layers respectively, u represents the average pixel value of this layer, m represents the number of pixels, z_{ik} represents the k-th pixel value of the i-th layer, and z_{jk} represents the k-th pixel value of the j-th layer, δ represents the standard deviation.

Hotspot Analysis of Local Getis Ord G_i^* Index Method

This method is used to analyze the aggregation degree of attribute values at the local spatial level, which can be expressed as formula (3).

$$G_i^* = \frac{\sum_{j=1}^{n}W_{ij}X_j}{\sum_{j=1}^{n}X_j} \tag{3}$$

Among them: X_j represents the attribute value of the j-th spatial element, n represents the total number of elements, and W_{ij} represents the spatial weight matrix. If the distance between the i-th and j-th spatial elements is within a given critical distance d, they are considered neighbors and the element in the spatial weight matrix is 1; otherwise, the element is 0. The test of the local Getis-Ord G_i^* statistic can be expressed according to the corresponding standardized form (Z-value), such as formula (4) and formula (5).

$$Z(G_i) = \frac{G_i - E(G_i)}{\sqrt{Var(G_i)}} = \frac{\sum_{j=1}^{n}W_{ij}X_j - \overline{X}\sum_{j=1}^{n}W_{ij}}{S\sqrt{\frac{n\sum_{j=1}^{n}W_{ij}^2-(\sum_{j=1}^{n}W_{ij})^2}{n-1}}} \tag{4}$$

$$\delta_j = \frac{m_j}{P_j} \tag{5}$$

Among them: $\overline{x}$ represents the average value of the attribute values of all spatial units, δ_j represents the commercial network density of the j-th street unit, m_j represents the number of commercial network points in the j-th street, and P_j represents the area of the j-th street.

Average nearest neighbor distance analysis method

This method uses the ratio R of the closest measured value to the expected value to judge the clustering characteristics of POI distribution. The calculation method of R is shown in formula (6).

$$R = \frac{\overline{d_i}}{\overline{d_c}} \tag{6}$$

Among them: d_i represents the measured distance value of the shortest distance of a certain type of POI, and d_c is the expected value of the shortest distance. R standard deviation Z The expression is shown in formula (7).

$$Z_R = \frac{\overline{d_i}-\overline{d_c}}{S_c} = \frac{\overline{d_i}-0.5\sqrt{\frac{A}{N}}}{0.26136/\sqrt{N^2/A}} \tag{7}$$

Among them: A represents the area of the region, and N represents the total number of hotspots. The Average Nearest

TABLE I
POPULATION, ECONOMY, GEOGRAPHY, AND BUSINESS PROFILE OF THE STUDY AREA

AREA	AREA COVERED/KM^2	GDP/RMB100 MILLION	POPULATION/10000	Retail POI data		
				Supermarket	Convenience Store	Large shopping malls
WuMing	3389	321	56	71	1260	16
ZiNing	1231	76	28	25	310	2
BinYang	2298	199	81	57	766	11
Heng	3348	270	90	38	589	14

Neighbor calculation returns a Z-score and a P-value. The smaller the Z-value for a node, the closer the significance P-value is to or equal to 0, and the higher the degree of clustering.

Data source

In this paper, Geosharp1.0 is used to collect the POI data of the retail industry in 2018 published on the Internet, including attribute information such as convenience stores, supermarkets, shopping malls, and shopping malls, as shown in Table I. The economic statistical data of population and GDP are obtained from the public data of the Statistics Bureau of Guangxi Zhuang Autonomous Region, and the urban traffic and road data are vectorized using GIS software.

V. Presentation and Discussion of Results

A. Spatial distribution pattern of retail industry

Divide the study area by an average of 1 km × A grid of 1 km in size; Thereby calculating the core density of the retail industry, and connecting spatial information to the grid; And overlay the administrative zoning map to identify the height of the concentration area of retail distribution within each region, and define it as a ”commercial center.”. According to the nuclear density estimation method, we obtained a nuclear density map of retail venues, supermarkets, convenience stores, and large shopping malls, as shown in Figure. 1,2,3,4. We further learned about the spatial layout and commercial agglomeration of the retail industry in the study area. Different types of commercial venues also differ in terms of spatial distribution and agglomeration.

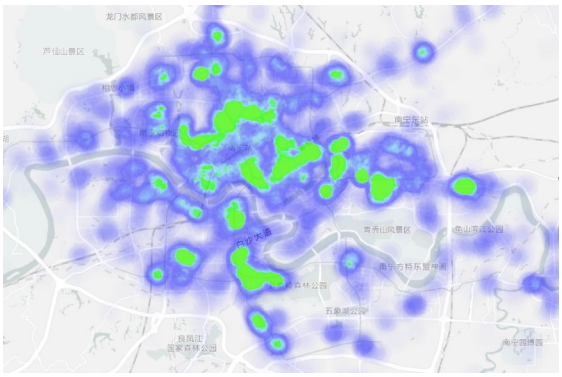

Fig. 1. Retail Site Nuclear Density - NanNing area

(1) Nanning retail outlets are mainly concentrated in the core urban area, as shown in Figure 1, distributed from the inside to the outside, from the middle to the periphery, and

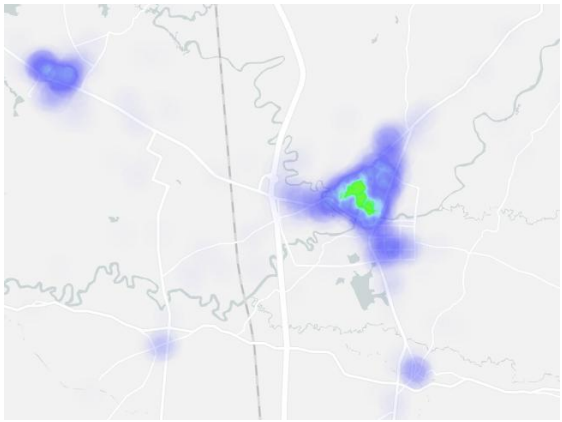

Fig. 2. Retail Site Nuclear Density - WuMing area

Fig. 3. Retail Site Nuclear Density - BinYang area

there are individual high-density hotspots outside the urban area; there are significant differences in the kernel density values of various regions in the urban area. The areas with high density values are Wuming, Binyang, and Hengxian, indicating that the retail industry in these areas is relatively developed.

(2) The characteristics of supermarket core density values are relatively obvious, indicating that supermarkets are mainly concentrated in the surrounding areas of the core urban area, as shown in Figure 5,6,7,8. The main supermarkets are: Wal-Mart, Grandbay, Sam’s Club, China Resources Vanguard, Hualian Supermarket, Li Kelong, Nancheng Department Store, etc. The target market of these supermarkets is to meet the daily consumption of people’s families.

(3) The core density value of convenience stores is the highest, and convenience stores are mainly distributed in commercial centers, transportation places, and population gathering places, as shown in Figure 9,10,11,12. Generally, the coverage

Fig. 4. Retail Site Nuclear Density - Heng area

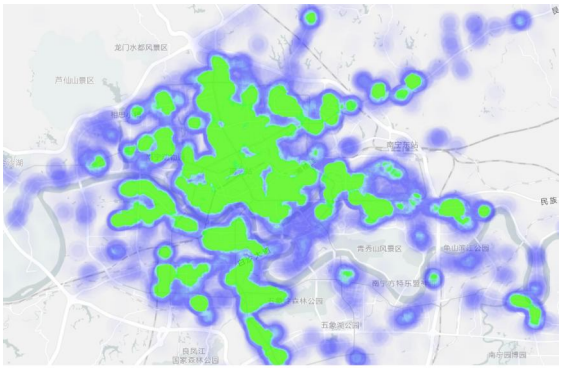

Fig. 5. Supermarket Core Density - NanNing area

area of convenience stores is within 500 meters, which is greatly affected by passenger flow. The core density value of convenience stores ranks first in the city center, because there are the most people gathered here; the second is schools, where students are the main consumers; the third is that they are close to residential areas, which is convenient for consumers to shop nearby. Therefore, population agglomeration is the most important factor affecting the layout of convenience stores.

(4) The core density of large-scale shopping centers is the lowest, mainly distributed in the core urban area, and there are also layouts in surrounding districts and counties, as shown in Figure 13,14,15,16. The area with a higher core

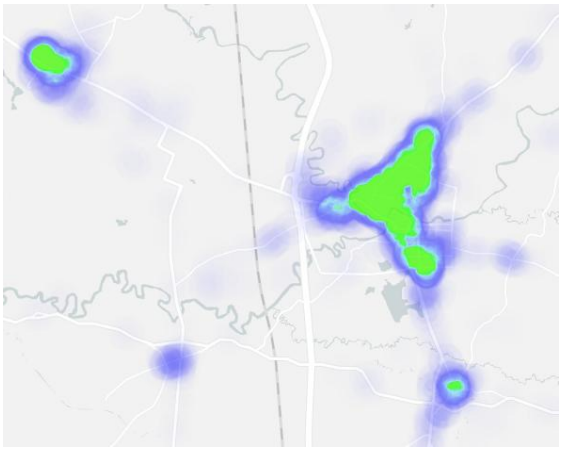

Fig. 6. Supermarket Core Density - WuMing area

Fig. 7. Supermarket Core Density - BinYang area

Fig. 8. Supermarket Core Density - Heng area

density of large shopping centers is the city center, such as: MixC, Hangyang Convention and Exhibition City, Neverland Department Store, Wanda Department Store, INCITY, Greenland Central Plaza, Wuyue Plaza, Logan Commercial Plaza, Parkson, etc. Commercial shopping center. These large shopping centers are usually located in areas with a large population, convenient transportation, complete supporting facilities, and strong commercial atmosphere. These large shopping malls can usually promote the opening of catering, fitness, entertainment and other places, forming a comprehensive comprehensive shopping mall.

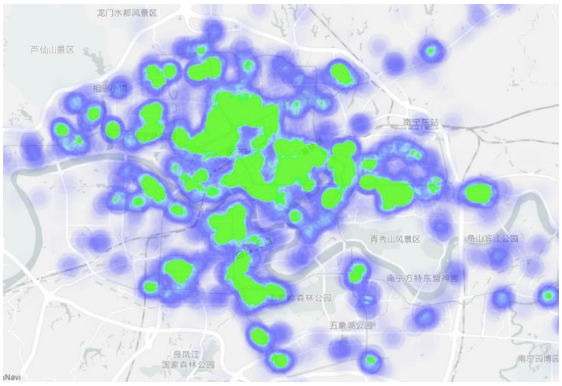

Fig. 9. Core Density Of Convenience Stores - NanNing area

Fig. 10. Core Density Of Convenience Stores - WuMing area

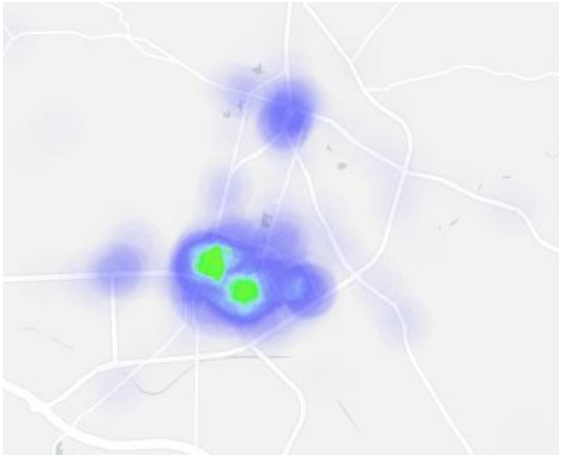

Fig. 11. Core Density Of Convenience Stores - BinYang area

B. Analysis of hot spots in the retail industry

Using the local Getis Ord G_i^* index method to analyze the retail industry hotspots in Nanning, as shown in Figure V-B. layout. Large shopping malls and supermarkets are concentrated in the core area, forming a concentrated and contiguous "commercial center"; the commercial layout in the urban fringe area is mainly retail, with scattered layout and small scale. From the perspective of hotspot values, there are significant differences between regions, with high hotspot values in core areas and low hotspot values in suburban counties.

Fig. 12. Core Density Of Convenience Stores - Heng area

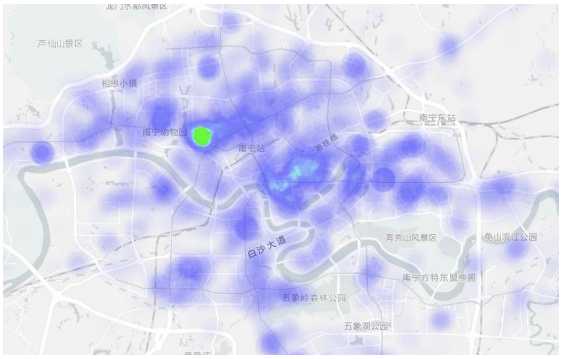

Fig. 13. Core Density of Large Shopping Malls - NanNing area

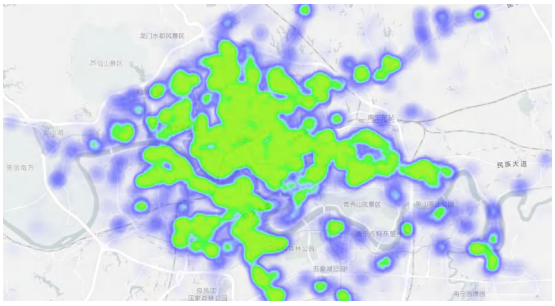

Fig. 17. POI hotspot analysis

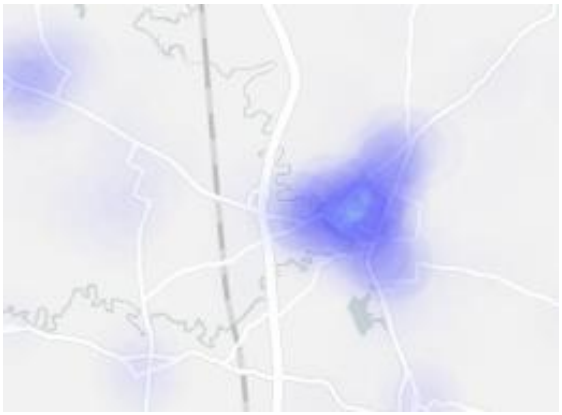

Fig. 14. Core Density of Large Shopping Malls - WuMing area

TABLE II
SPATIAL LAYOUT AND DATA CATEGORY CHARACTERISTICS OF THE RETAIL INDUSTRY IN NANNING REGION

POI Data category	quantity	Proportion	R	P	Z	Average adjacent	Layout Features
convenience stores	**2100**	**94**	**0.63**	**0**	**135**	**89**	**General aggregation**
Supermarkets	**1100**	**5**	**0.52**	**0**	**53**	**53**	**comparative aggregation**
shopping malls	**200**	**1**	**0.42**	**0**	**35**	**35**	**strong aggregation**

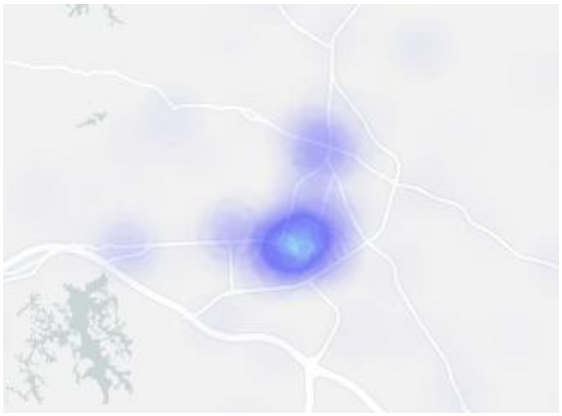

Fig. 15. Core Density of Large Shopping Malls - BinYang area

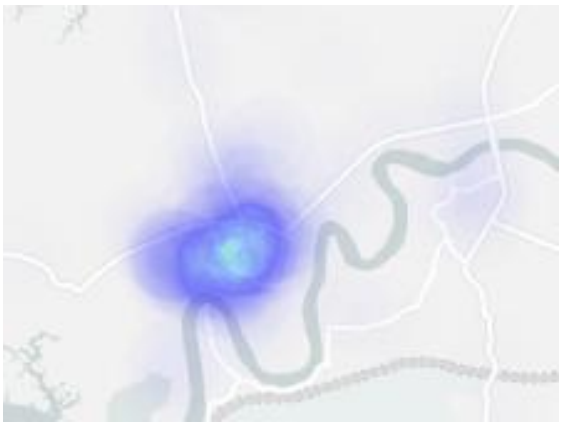

Fig. 16. Core Density of Large Shopping Malls - Heng area

The spatial layout and data category characteristics of Nanning's retail industry can be obtained through the nearest distance analysis method, as shown in Table II. From the perspective of the number and proportion of POIs, convenience stores accounted for the largest proportion, reaching 94%, accounting for most of the retail industry; supermarkets ranked second, accounting for 5%; large shopping malls accounted for the smallest proportion, only 1%. From the perspective of layout characteristics, large shopping centers have the highest R value and the highest degree of aggregation, which belongs to the strong aggregation type; supermarkets are next, relatively clustered; convenience stores have the lowest R value, and are generally distributed in clusters.

C. Analysis on the formation mechanism of spatial layout

1) Traffic Layout Impact Analysis: From the analysis chart of hot spots in the retail industry in Nanning, it can be seen that transportation accessibility has a significant impact on the layout of the retail industry, as shown in Figure 5. The transportation in the core urban area is convenient, and the core density of retail industry is the highest, especially near the main urban roads; As marginal areas extend outward along major roads, the retail industry also expands outward.

2) Economic level impact analysis: By comparing and analyzing the POI data of districts and counties with the GDP data of the region, the results show that there is a strong correlation between the two. The POI data of the retail industry is closely related to the GDP data of the secondary and tertiary industries, reaching 99%. Therefore, the retail industry is most concentrated in regions with developed secondary and tertiary industries.

3) Analysis of the impact of population agglomeration: Statistics show that the population density ranks first in Xixiangtang, second in Qingxiu District, third in Xingning District, and last in Long'an County. The population density is highly consistent with the layout of the retail industry, and the correlation between the population density and the spatial distribution of the catering industry reaches 90

VI. CONCLUSIONS

In this paper, by studying the retail POI data of Nanning City and its districts and counties in the target area, combined with the population, GDP indicators, transportation infrastructure and other data in the government statistical report, quantitative and qualitative research methods are used to analyze local hotspots and obtain retail places, supermarkets, Convenience stores, large shopping malls and other retail spatial layout results, study the formation mechanism of this layout, the relationship between retail spatial distribution and regional population, and obtain factors such as GDP and transportation facilities. The detailed conclusions are as follows:

(1) On the whole, retail outlets in Nanning are mainly concentrated in the core urban area, spreading from the inside to the outside, and from the center to the periphery. The core density values of different regions in the urban area vary greatly, and the areas with high density values are Wuming, Binyang, and Hengxian; the core density values of convenience stores are the highest, mainly distributed in commercial centers, transportation places, and crowded places; the core density value of supermarkets The characteristics are relatively obvious, mainly concentrated in the surrounding areas of the core urban area; the core density of large shopping malls is the lowest, mainly concentrated in the core urban area.

(2) According to POI data, the retail industry in Nanning presents a spatial distribution of multi-center development and band-like expansion. From the number and proportion of POIs, it can be seen that convenience stores accounted for the largest

proportion, reaching 94%, accounting for most of the retail industry; supermarkets ranked second, accounting for 5%; large shopping malls accounted for the smallest proportion, only 1% . From the perspective of layout characteristics, large shopping centers have the highest R value and the highest degree of aggregation, which belongs to the strong aggregation type; supermarkets are next, relatively clustered; convenience stores have the lowest R value, and are generally distributed in clusters. The distribution of retail hotspots generally forms a spatial layout along the main traffic road, concentrated in the middle, and surrounded by strips.

(3) From the perspective of retail agglomeration and spatial layout, traffic accessibility has a significant impact on retail layout. The transportation in the core urban area is convenient, and the core density of the retail industry is the highest, especially close to the main road in the urban area; the layout of the retail industry is closely related to the level of regional GDP. The layout of the retail industry is also most concentrated in areas with developed secondary and tertiary industries; the population density is highly consistent with the layout of the retail industry, and the spatial correlation between population density and retail industry reaches 90%.

Finally, the research on the spatial layout and formation reasons of retail industry based on POI big data is highly consistent with the real activities and agglomeration status of urban retail hotspots. The relationship between the spatial layout of the retail industry and GDP, population, and transportation infrastructure obtained through the study can provide a reference for optimizing the layout of urban commercial facilities in the future. However, only using POI data cannot fully support the simulation and optimization of the city's overall commercial space structure, and more aspects such as urban planning, government policies, population flow, market changes, and urban humanities need to be considered. In the future, if POI data can be loaded onto the smart city platform, it will provide better support for urban management.

VII. Acknowledgements

At this moment, looking back on the fulfilling, busy, and happy learning time of the past year, I have a lot of emotions. This unforgettable learning journey has been profound and enriching for my growth and refinement. This kind of growth and refinement comes from the tireless teaching of teachers, the passionate debates of classmates, and the strong support of family. I am particularly grateful to my wife for giving birth to a lovely baby during my research period, which led to the birth of two lovely sons. I would like to sincerely thank all those who have given me help and care.

References

[1] Z. Wang, X. Liu, W. Zhang, Y. Zhi and S. Cheng (2022). The statistical analysis in the era of big data. International Journal of Modelling, Identification and Control, 40(2), 151-157.

[2] W. Chen, L. Liu, and Y. Liang, "Hot spot identification and industry agglomeration characteristics analysis of Guangzhou retail business center based on POI data," *Geographical Research*, vol. 35, no. 4, pp. 703–716, 2016.

[3] B. Xue, X. Xiao, J. Li, et al., "Spatial association analysis of residential and retail industries in Shenyang based on POI big data," *Geographic Sciences*, vol. 39, no. 3, pp. 442–449, 2019.

[4] Z. Gao and H. Zhang, "Research on the spatial pattern of Xining's retail industry based on POI data," *Geography of Arid Areas*, vol. 42, no. 5, pp. 1195–1204, 2019.

[5] P. J. Clark and F. C. Evans, "Distance to nearest neighbor as a measure of spatial relationships in populations," *Ecology*, vol. 35, no. 4, pp. 445–453, 1954.

[6] J. Ge, "Research on retail industry promoting urban development: A case study of Jinan City," *Journal of Shandong University of Finance*, no. 3, pp. 17–21, 2013.

[7] A. Getis and J. K. Ord, "The analysis of spatial association by use of distance statistics," *Geographical Analysis*, vol. 24, no. 3, pp. 189–206, 1992.

[8] E. Parzen, "On estimation of a probability density function and mode," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.

[9] M. Rosenblatt, "Remarks on some nonparametric estimates of a density function," *The Annals of Mathematical Statistics*, vol. 27, no. 3, pp. 832–837, 1956.

[10] C. Xiu and C. Xia, "Comparison of urban commercial distribution patterns between China and the West," *Geography and Territorial Research*, no. 1, pp. 31–35, 1999.

[11] F. Yan, "Challenges and strategic choices of traditional retail formats in China under the network background," Ph.D. dissertation, Henan University, Zhengzhou, China, 2015.

[12] J. Zhu, "Research on innovation in China's retail industry," *Heilongjiang Foreign Trade and Economic Cooperation*, no. 6, pp. 36–37, 2011.