

# Spatial Association characteristics of facilities around scenic spots considering distance and orientation: A case study of 3A and above scenic spots in Beijing

Hansong Cao<sup>1,2,3,4</sup>, Jiahui Qin<sup>1,2,3,4</sup>, Disheng Yi<sup>1,2,3,4</sup>, Yusi Liu<sup>1,2,3,4</sup> and Jing Zhang<sup>1,2,3,4,\*</sup>

<sup>1</sup>College of Resources Environment and Tourism, Capital Normal University, Beijing, 100048, China

<sup>2</sup>Beijing Laboratory of Water Resources Security, Capital Normal University, Beijing, 100048, China

<sup>3</sup>3D Information Collection and Application Key Lab of Education Ministry, Capital Normal University, Beijing, 100048, China

<sup>4</sup>Beijing State Key Laboratory Incubation Base of Urban Environmental Processes and Digital Simulation, Capital Normal University, Beijing, 100048, China

**Abstract.** In the urban tourism and service industry, the POI data with coordinate and attribute information of the major map platforms constitute one of the important data sources of the urban tourism and service industry. In this paper, the spatial data transaction database under four distances was established based on the gate buffer of 3A and above scenic spots in Beijing. The Apriori algorithm was used to calculate the lifting degree to obtain the distance for mining the best association features of 3A, 4A and 5A scenic spots, and then the association features of the three scenic spots in different directions were analysed.

## 1. Introduction

With the acceleration of modernization process and the development of mobile portable devices, LBS acquired by mobile devices provides massive data, which provides a strong theoretical basis and data support for the mining of spatial knowledge and spatial relations hidden in spatial data. Spatial association rules (SAR, Spatial Association Rules) is an important part of Spatial data mining. Many researchers have carried out a lot of urban calculation work based on big data, using POI data, floating car track data and microblog check-in data to conduct research on the division of urban functional areas [1], travel characteristics of residents [2], urban vitality [3] and other aspects.

After the concept and algorithm of association rules [4] were proposed, scholars conducted a large number of researches, which mainly focused on the application field and the mining algorithm. From the perspective of application fields, studies are mainly focused on emergency events [5], environment [6], and urban functions [7]. In terms of mining algorithm, researches mainly focus on optimizing algorithm efficiency by using vertical data format [8], binary constraint [9], removable window [10] and other conditions. However, there are few research examples of spatial association analysis of urban facilities, and the current research is mainly based on the fuzzy perspective to study the association characteristics of urban services. In this paper, using POI point data of

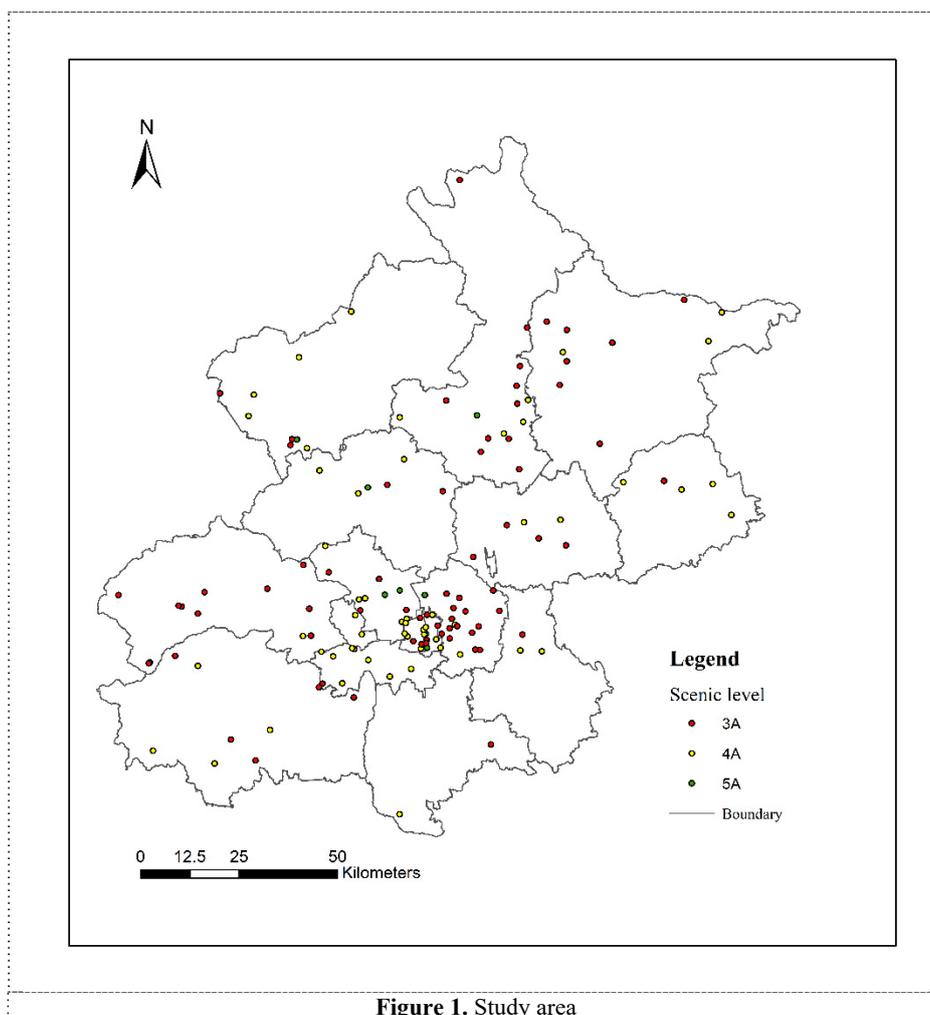
urban public facilities and considering the influence of distance and orientation, the spatial association characteristics of facilities in urban scenic spots are analysed. The main research contents are as follows: (1) Calculate the average promotion degree and analyse the optimal distance for mining the optimal association characteristics in different levels of scenic spots. (2) Calculate the similarities and differences of the association features of different levels of scenic spots in different directions and corresponding distances.

## 2. Materials and Methods

### 2.1. Overview of the study area and data acquisition

In this paper, the spatial association characteristics of 3A and above scenic spots in Beijing were analysed. As shown in the figure1, as of June 26, 2020, Beijing has a total of 149 scenic spots, including 72 3A-level scenic spots, 56 4A-level scenic spots and 8 5A-level scenic spots. In addition, the POI data of different operators offer abundance, there is a difference, to the POI classification system is not the same, Scott map as one of the most popular maps usage, this article from Beijing university open data platform for the scope of gold map of Beijing in 2018, a total of about 850000 POI data, selected from 29 class POI data contains the public service facilities, And about 420,000 taxi OD data on a rest day for experimental analysis.

\*Corresponding author's e-mail: [zhangjing@mail.cnu.edu.cn](mailto:zhangjing@mail.cnu.edu.cn)



**Figure 1.** Study area

## 2.2. Overview of the study area and data acquisition

Association analysis is one of the core technologies of data mining. Association rule model and data mining algorithm were first proposed by Agrawal in 1993 [11]. The goal is to find interesting associations or interrelationships between projects in large amounts of data. The most classic case of association rule data mining is the story of "Beer and Diaper" in Wal-Mart. By mining the association of various goods in customers' "shopping basket", customers' shopping habits can be analysed to help merchants make better sales strategies. In basket mining, an item is an item, a collection of multiple items is called an item set, and each purchase record is called a transaction. Association rules are generally written in the form of  $X \Rightarrow Y$ , where the left item set  $X$  is the prerequisite and the right item set  $Y$  is the corresponding association result, which is used to represent the implied association in the data. For example,  $\{\text{scenic spot}\} \Rightarrow \{\text{parking lot}\}$ , indicating that the two spatial entities, scenic spot and parking lot, have a certain association within the distance threshold. For the intensity of association rules, whether scenic spots and parking lots are more concentrated or social security agencies and bus stations are more

concentrated. The evaluation is mainly carried out by the concepts of support, confidence and lift [12].

Support refers to the possibility of appearing in all item sets  $\{X, Y\}$ , namely the probability of containing both  $X$  and  $Y$ , as shown in Equation (1).

$$\text{Support}(X \rightarrow Y) = P(X, Y) \quad (1)$$

Where,  $\text{Support}(X \rightarrow Y)$  is the degree of Support, and the probability that  $P(X, Y)$  item set contains both  $X$  and  $Y$ . Support is the first requirement for mining strong association rules. Its significance lies in eliminating meaningless rules with low probability by setting minimum threshold, and reserving items with relatively frequent occurrence as spatial frequent item sets. The above example data is taken as an example. Assuming that the minimum support is 0.05,  $\{\text{parking lot, Chinese restaurant}\} = 80/1000 = 0.08$ ,  $\{\text{parking lot, convenience store}\} = 10/1000 = 0.01$ . Since  $\{\text{parking lot, Chinese restaurant}\}$  meets the minimum support requirements, it is retained as the frequent item set. At the same time, the spatial association rule  $\text{parking lot} \Rightarrow \text{convenience store}$  is retained, while the rule corresponding to  $\{\text{parking lot, convenience store}\}$  is eliminated.

Confidence refers to the probability of occurrence of association result  $Y$  under the condition of occurrence of prerequisite  $X$  of association rule, that is, the item set containing  $X$  contains the possibility of both containing  $Y$ ,

as shown in Equation (2).

$$Confidence(X \rightarrow Y) = P(Y | X) / P(X) \quad (2)$$

Where, Confidence( $X \rightarrow Y$ ) is Confidence;  $P(Y | X)$  is the probability of occurrence of the corresponding association result  $Y$  under the condition that the precondition  $X$  of the association rule appears;  $P(X)$  is the probability of  $Y$ . Confidence is the second requirement for generating strong association rules, which measures the reliability of association rules. It is also necessary to set the lowest confidence level to continue filtering. In the above example, when the minimum confidence is set as 0.6, the confidence of parking lot  $\Rightarrow$  Chinese restaurant is  $80/100=0.8$ , and this rule is retained; the confidence of parking lot  $\Rightarrow$  convenience store is  $10/100=0.1$ , and the rule is eliminated.

The promotion degree refers to the ratio of the possibility of both containing  $Y$  under the condition of containing  $X$  to the possibility of the item set containing  $Y$  without such a condition, namely, on the basis of the occurrence probability  $P(Y)$  of  $Y$  itself, the promotion degree of the occurrence probability  $P(Y | X)$  of  $Y$  by the appearance of  $X$ , as shown in Equation (3).

$$Lift(X \rightarrow Y) = P(Y | X) / P(Y) \quad (3)$$

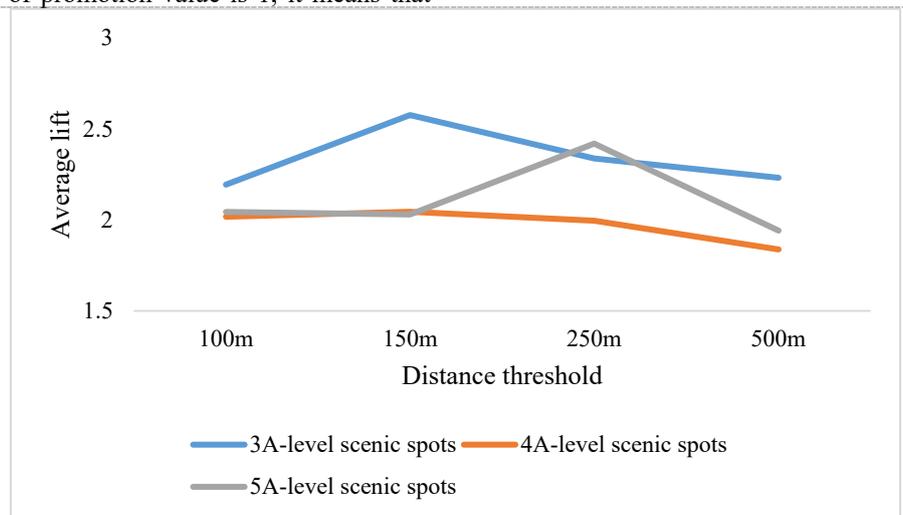
In the formula, Lift( $X \rightarrow Y$ ) is the lifting degree; Lifting degree can make up for the deficiency of confidence and can be regarded as a complementary index of confidence. When the degree of promotion value is 1, it means that

instance  $X$  and  $Y$  are independent, and the occurrence of  $X$  has no effect on the possibility of occurrence of  $Y$ . The larger the degree of promotion value (confidence  $> 1$ ) is, the greater the degree of promotion of  $X$  to  $Y$  is, that is, the stronger the spatial association is.

The association analysis of tourist attractions and public service facilities can be similar to the analysis of shopping basket. Each facility type is equivalent to the goods in the shopping basket, and the facility type in each unit is equivalent to a shopping record.

### 3. Results & Discussion

In this paper, the Apriori algorithm is used to analyse the tourist attractions in Beijing at four distances of 100, 150, 250 and 500 meters. The minimum support is set as 0.2 to calculate the average promotion degree of association rules of the three types of scenic spots at different distances, as shown in Figure 1. It can be found that the 3A and 4A scenic spots have the highest average lift degree at a distance of 150 meters, and the 5A scenic spots have the highest average lift degree at a distance of 250 meters. Therefore, 150 meters is the best distance to excavate the association characteristics of the 3A and 4A scenic spots, and 250 meters is the best distance to excavate the 5A scenic spots.



**Figure 2.** Trend of average promotion degree of association rules in scenic spots at all levels with distance

In the distance, on the basis of in this paper, three kinds of scenic area according to the north and south, east and west four different bearing characteristics of the spatial association analysis, through many experiments, the 3 level scenic spot in the 150 - meter distance, the select 0.1

**Table 1.** The top five rules of 3A-level scenic spots are in descending order of support.

Orientation	Association rules	Support	Confidence	Lift
East	{Social security agency} $\Rightarrow$ {parking lot}	0.22	1	3.85
	{parking lot} $\Rightarrow$ {social security agencies}	0.22	0.85	3.85
	{social security agencies, 3A-level scenic spots} $\Rightarrow$ {parking lot}	0.22	1	3.85

as the minimum support degree of the east and the south, 0.2 as a scenic spot in west and north minimum support degree, minimum confidence level of 0.6. Due to space constraints, only the first five items in each direction are listed in the table.

	{parking lot, 3A-level scenic spots} => {social security agencies}	0.22	0.85	3.85
	{Medical facilities} => {parking lot}	0.18	1	3.85
West	{Medical facilities} => {Chinese restaurants}	0.35	1	2.12
	{Chinese restaurants} => {Medical facilities}	0.35	0.75	2.12
	{auto service} => {Chinese restaurants}	0.35	1	2.12
	{Chinese restaurants} => {auto service}	0.35	0.75	2.12
	{illegal parking sign} => {parking lot}	0.35	1	2.83
South	{Chinese restaurants} => {parking lot}	0.36	1	2.27
	{parking lot} => {Chinese restaurants}	0.36	0.83	2.27
	{Chinese restaurants, 3A-level scenic spots} => {parking lot}	0.36	1	2.27
	{parking lot, 3A-level scenic spots} => {Chinese restaurants}	0.36	0.83	2.27
	{Medical facilities} => {parking lot}	0.31	0.86	1.97
North	{pick-up points} => {drop-off points}	0.4	1	1.52
	{drop-off points} => {pick-up points}	0.48	0.73	1.52
	{3A-level scenic spots, pick-up points} => {drop-off points}	0.48	1	1.52
	{3A-level scenic spots, drop-off points} => {pick-up points}	0.48	0.73	1.52
	{parking lot} => {drop-off points}	0.42	0.93	1.42

In the distance of 150 meters, the overall support degree of association rules is at a low level. Strong association rules related to parking lots are found in 3A-Level scenic spots in all four directions, indicating that their confidence degree is above 0.9. In the east direction, {social security agencies, 3A-level scenic spots} => {parking lot} is taken as an example. The probability of parking lots within 150 meters of 3A-level scenic spots and social security agencies in Beijing is about 22%. It can be seen from Table 1 that in the eastern position of the 3A-level scenic spot, the association features are mainly social security institutions, parking lots and medical facilities, while in the eastern position, the social security institutions are relatively complete. In the west, the related

features are mainly medical facilities, parking lots, Chinese restaurants, and car services. The appearance of car rental and maintenance services indicates to some extent that car rental and maintenance services will be provided for self-driving tourists at the west gate of the 3A-level scenic spots. In the south direction, the association features are mainly Chinese restaurants, parking lots and medical facilities. Different from the other locations, the northern location attracts the characteristics associated with the pick-up and drop-off points of taxis, and the northern location is more likely to attract tourists to take taxis in the four locations of the 3A-level scenic spots.

**Table 2.** The top five rules of 4A-level scenic spots are in descending order of support.

Orientation	Association rules	Support	Confidence	Lift
East	{public toilets} => {parking lot}	0.25	0.69	1.42
	{social security agencies} => {parking lot}	0.25	1	2.05
	{Medical facilities} => {parking lot}	0.25	1	2.05
	{public toilets, 4A-level scenic spots} => {parking lot}	0.25	0.69	1.42
	{social security agencies, 4A-level scenic spots} => {parking lot}	0.25	1	2.05
West	{hostel} => {Chinese restaurants}	0.41	1	2

	{Chinese restaurants} => {hostel}	0.41	0.81	2
	{Medical facilities} => {parking lot}	0.41	1	2
	{parking lot} => {Medical facilities}	0.41	0.81	2
	{public toilets} => {Chinese restaurants}	0.41	0.9	1.8
South	{bus stations} => {parking lot}	0.36	0.9	1.83
	{parking lot} => {bus stations}	0.36	0.75	1.83
	{bus stations,4A-level scenic spots} => {parking lot}	0.36	0.9	1.83
	{parking lot,4A-level scenic spots} => {bus stations}	0.36	0.75	1.83
	{public toilets} => {parking lot}	0.34	0.85	1.73
North	{drop-off points} => {pick-up points}	0.73	1	1.36
	{pick-up points} => {drop-off points}	0.73	1	1.36
	{4A-level scenic spots, drop-off points} => {pick-up points}	0.73	1	1.36
	{4A-level scenic spots, pick-up points} => {drop-off points}	0.73	1	1.36
	{parking lot} => {drop-off points}	0.65	1	1.3

As shown in Table 2, in the four directions at a distance of 150 meters, the northern position has the highest overall support, followed by the south, the west, and the east. According to the analysis of sub-direction, under the eastern position of the 4A level scenic spot, the associated features are mainly public toilets, social security organs, parking lots, scenic spots and other related features, while under the eastern position, the scenic spot itself attracts

more facilities. In the western bit, {hostel}=> {Chinese restaurant} is the highest association rule, which shows strong association characteristics. In the south, bus stations, parking lots, public toilets and scenic spots are the main associated features, which also show that the scenic spots attract other facilities. Similar to the 3A spots, the 4A spots are also more attractive to taxi pickup points in the north.

**Table 3.** The top five rules of 5A-level scenic spots are in descending order of support.

Orientation	Association rules	Support	Confidence	Lift
East	{Chinese restaurants} => {bus stations}	0.42	1	2.33
	{bus stations} => {Chinese restaurants}	0.42	1	2.33
	{Chinese restaurants} => {parking lot}	0.42	1	2.33
	{parking lot} => {Chinese restaurants}	0.42	1	2.33
	{Chinese restaurants} => {the convenience store}	0.42	1	1.75
West	{Chinese restaurants} => {bus stations}	0.6	1	1.6
	{bus stations} => {Chinese restaurants}	0.6	1	1.67
	{Chinese restaurants} => {parking lot}	0.6	1	1.25
	{parking lot} => {Chinese restaurants}	0.6	0.75	1.25
	{bus stations} => {parking lot}	0.6	1	1.25
South	{Chinese restaurants} => {parking lot}	0.76	1	1.18
	{parking lot} => {Chinese restaurants}	0.76	0.91	1.18
	{Chinese restaurants,5A-level scenic spots} => {parking lot}	0.76	1	1.18
	{parking lot,5A-level scenic spots} => {Chinese restaurants}	0.76	0.91	1.18

	{bus stations} => {parking lot}	0.69	1	1.18
North	{Chinese restaurants} => {public toilets}	0.57	1	1.75
	{public toilets} => {Chinese restaurants}	0.57	1	1.75
	{Chinese restaurants} => {social security agencies}	0.57	1	1.75
	{social security agencies} => {Chinese restaurants}	0.57	1	1.75
	{Chinese restaurants} => {Ordinary hotel}	0.57	1	1.75

As shown in Table 3, the support degree of association rules generated in the four directions is at a high level and the gap is small. The support degree of the south direction is the highest, followed by the west and north and south directions, and finally the east direction. The 5A scenic spots show strong association characteristics of Chinese restaurants, bus stations and parking lots in the east, west and south directions. Different from other level scenic spots, the characteristics of the highest association rules are less attracted by 5A level scenic spots and more attracted by each other's public facilities, resulting in the

phenomenon of aggregation.

The 3A-level scenic spots will be taken as an example to illustrate the difference of spatial association characteristics in different directions. This paper used arulesViz visualization package drawing association rules as a result, the horizontal axis LHS (left) the rules of the first two facilities said in the group of the most frequent two types, the vertical axis RHS (right) and the left side of the associated facilities. The size of the circle indicates the degree of support, and the shade of color indicates the degree of improvement.

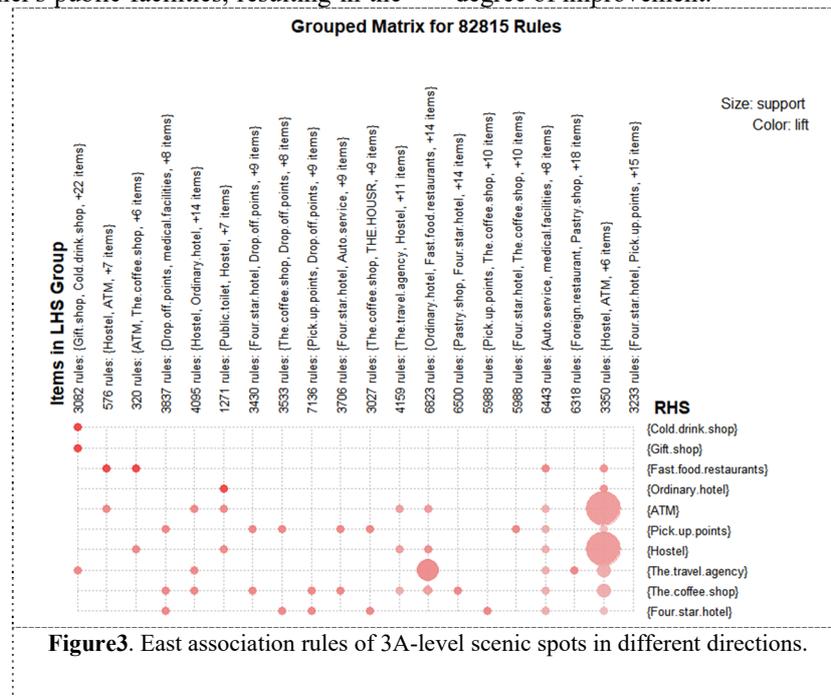


Figure3. East association rules of 3A-level scenic spots in different directions.

As can be seen from Figure3, the association characteristics of {gift shop, cold drink shop, etc.} and {3A-level scenic spots, hostel, ATM, etc.} are obvious. In the east, scenic spots will attract hostel, fast food

restaurants, ATM and other facilities. In addition, the 3A-level scenic spots show the aggregation phenomenon between different catering types and different accommodation types in the east.

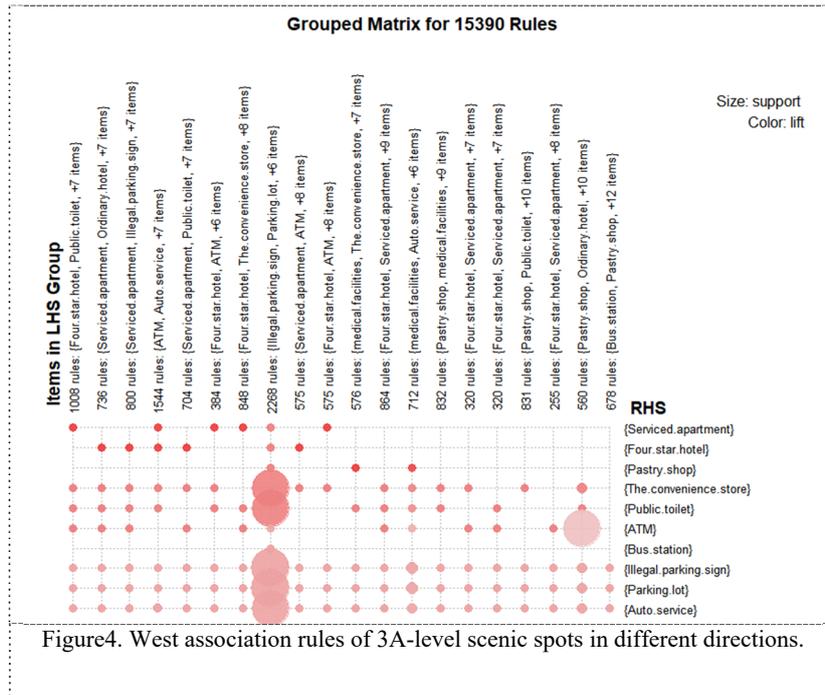


Figure4. West association rules of 3A-level scenic spots in different directions.

Under the western level, Figure4, {four-star hotel, ordinary hotel, serviced apartment, etc.}, {illegal parking signs, parking lots, public toilets, etc.} have obvious characteristics, and there is an obvious clustering phenomenon of accommodation types under the western

level, which indicates that the accommodation industry layout is compact under the western 3A-level scenic spot 150 meters away. Practitioners tend to locate with competitors.

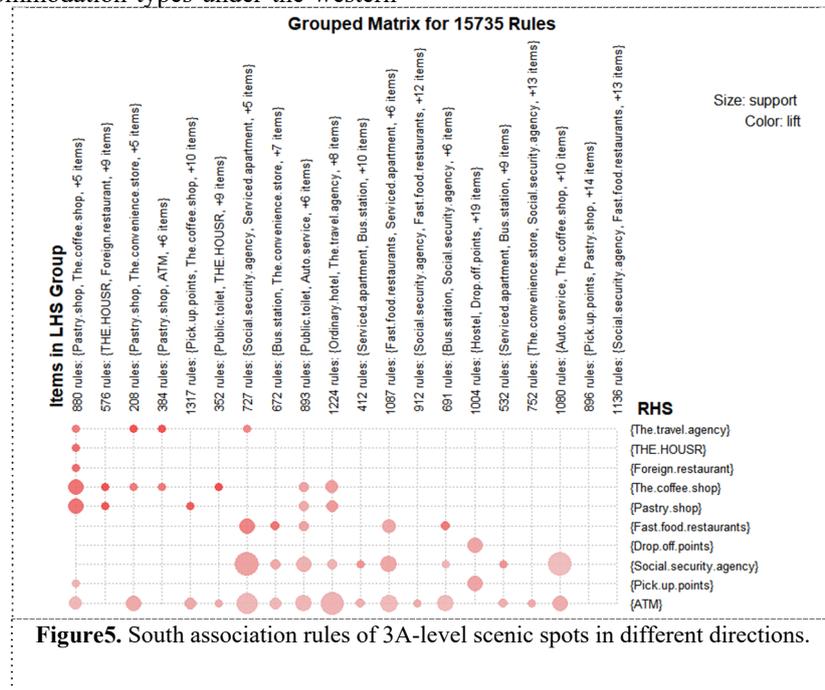


Figure5. South association rules of 3A-level scenic spots in different directions.

In the south direction, Figure5, {café, bakery, tea house, etc.}, {public toilet, tea house, fast food restaurant, etc.} and other types of leisure catering facilities have

obvious association characteristics. In this direction, a variety of catering types of facilities show spatial aggregation characteristics.

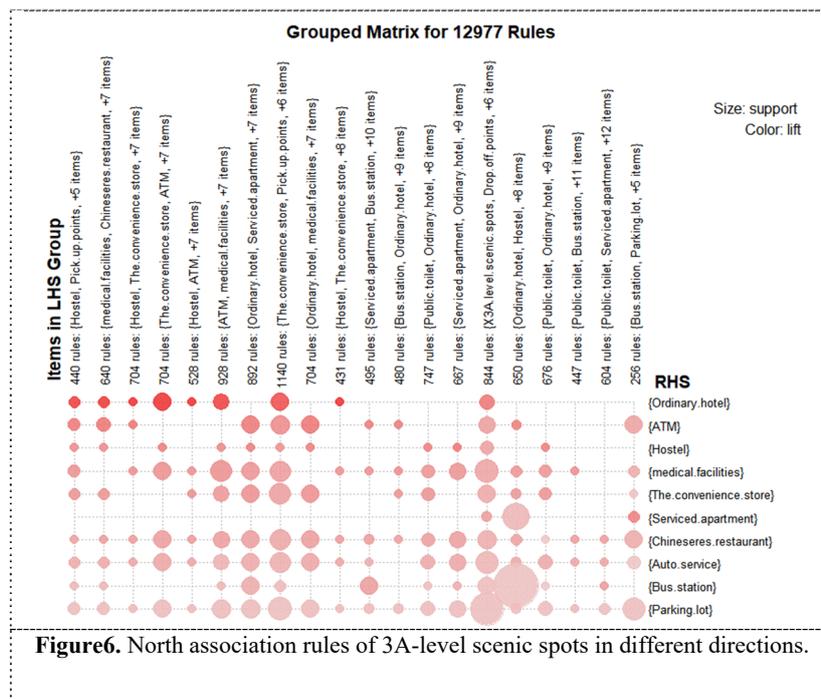


Figure6. North association rules of 3A-level scenic spots in different directions.

In the north, Figure6, {convenience stores, ATM machines, ordinary hotels, etc.}, {convenience stores, pick-up points, ordinary hotels, etc.}, {3A-level scenic spots, car services, ordinary hotels, etc.} have obvious features, and there are many types of facilities associated with the scenic spots. In the north, the scenic spots also attract other types of facilities. Different from the other three locations, the northern location does not have strong correlation characteristics with the catering types such as pastry shop, cold drink shop, fast food restaurant and tea house, indicating that the catering types in the northern direction are relatively single.

#### 4. Conclusions

In the face of a large number of data based on spatial location produced in the era of big data, mining of spatial association rules is an important part of urban research. This paper uses POI data and taxi data to calculate the association characteristics among facilities and study the spatial distribution among industries. In each city, facilities entities generally have symbiosis, connection and other topological relations. Apriori algorithm, as a classic spatial association rule mining method, can well mine the spatial association characteristics of urban facilities, indicating the simultaneous occurrence of facilities and the frequent simultaneous occurrence of facilities. The practical significance of the results in this paper is as follows :(1) For tourists, the association features of taxi trips in the results can provide some reference for tourists when taking a taxi trip, and the features of illegal parking signs can make self-driving tourists carefully consider when choosing appropriate places to park their cars. (2) For scenic spots, the experimental results also reflect the problem of unbalanced distribution of public service resources in scenic spots. From downtown to suburban areas, the

abundant allocation of public service resources in scenic spots gradually decreases. The experimental results provide reference for urban management departments to reasonably arrange scenic resources and meet the needs of tourists.

#### Acknowledgments

This research was funded by the National Nature Science Foundation of China (Grant No. 41771477).

#### References

1. Hu S, He Z, Wu L, et al. A framework for extracting urban functional regions based on multiprototype word embeddings using points-of-interest data[J]. Computers, Environment and Urban Systems, 2020,80:101442.
2. Luan X, Cheng L, Song Y, et al. Better understanding the choice of travel mode by urban residents: New insights from the catchment areas of rail transit stations[J]. Sustainable Cities and Society, 2020,53:101968.
3. Yue W, Chen Y, Thy P T M, et al. Identifying urban vitality in metropolitan areas of developing countries from a comparative perspective: Ho Chi Minh City versus Shanghai[J]. Sustainable Cities and Society, 2021,65:102609.
4. Koperski K, Han J. Discovery of Spatial Association Rules in Geographic Information Databases[J]. 1995.
5. Bo F. Hybrid spatial data mining methods for site selection of emergency response centers[J]. Natural Hazards, 2014,70(1):643-656.
6. Ma W, Xue C, Zhou J. Mining time-series association rules from Western Pacific spatial-temporal data[J]. IOP Conference Series: Earth and Environmental

- Science, 2014,17(1):12224-12226.
7. Chi J, Jiao L, Dong T, et al. Quantitative identification and visualization of urban functional area based on POI data[J]. Journal of Geomatics.
  8. Sheikh A S, Patidar M R. A NOVEL ALGORITHM DUFPP for MINING ASSOCIATION RULES in DYNAMIC DATABASES[J].
  9. Gang F, Wei Z, Qian Y. An Algorithm of Constrained Spatial Association Rules Based on Binary: International Symposium on Neural Networks: Advances in Neural Networks, 2008[C].
  10. Li X. Mining spatial association rules in spatially heterogeneous environment[J]. International Society for Optics and Photonics, 2008.
  11. Agrawal R. Mining Association Rules Between Sets of Items in Large Databases, SIGMOD Conference[J]. proc sigmod, 1993.
  12. Merdun H, Ozturk Y. arules - A Computational Environment for Mining Association Rules and Frequent Item Sets[J]. Journal of Statistical Software, 2005,014.