

Research Article

Modeling Spatial Riding Characteristics of Bike-Sharing Users Using Hotspot Areas-Based Association Rule Mining

Chao Sun  and Jian Lu 

School of Transportation, Southeast University, Nanjing 210000, China

Correspondence should be addressed to Jian Lu; lujian_1972@seu.edu.cn

Received 7 January 2022; Revised 1 March 2022; Accepted 10 March 2022; Published 1 April 2022

Academic Editor: Fei Hui

Copyright © 2022 Chao Sun and Jian Lu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study aims to investigate the spatial riding characteristics under different demand scenarios using association rule mining with hotspot detection, and to establish the subordinate rules between bike-sharing demand and land elements and between land elements. To reduce deviation from modifiable areal unit problem (MAUP) and improve objectivity and accuracy, we impose spatial constraints using the hotspot detection model instead of the square grid and traditional traffic zone. The bike-sharing trajectory-based kernel density algorithm is employed to explore the optimum analysis locations and the analysis areas with the relatively high demand. More importantly, the research featured here involves five demand scenarios for the differentiation of riding characteristics. The results show that the most significant influencers on bike-sharing demand include financial insurance facilities, dining facilities, and landscapes. As for characteristics of riding destination, the combinations between landscapes and financial insurance facilities, between landscapes and companies/enterprises, and between companies/enterprises and financial insurance facilities are more likely to be visited simultaneously. These findings make us understand urban spatial structure in response to traffic plan and provide evidence for bike-sharing dispatch optimization.

1. Introduction

Bike-sharing has improved the efficiency of the traffic system, but it has also faced many problems in its development [1]. For example, how to effectively explore the riding characteristics and the relationship between land use and bike-sharing demand is a fundamental problem to be solved [2]. The land use-based demand forecast is helpful to grasp the potential trend and to find out the connection and coordination control methods with other travel models, especially in cities that are just starting to develop bike-sharing. Analysis of the spatial riding characteristics also nudges the optimization of land use structure. Not surprisingly, a well-planned zone will naturally attract more bike-sharing users and encourage visitors to prefer bike-sharing travel.

Generally speaking, bike-sharing trips mainly consist of riding distance, riding time, riding purpose, riding volume, and other characteristics [3]. As one of the most important research instruments, spatial riding characteristic analysis

fulfills its role in wisely expanding the bike-sharing stations and codesigning a premium user experience with the management department plan. Like other travel characteristics, spatial riding characteristics can be divided into two categories, namely, origin characteristics and destination characteristics:

- (1) Trip-generation is the focus point of origin characteristics. For example, Amiri et al. [4] study the riding behaviors in freezing weather using intercept survey with cross-tabulation. Based on barrier models, Ahmadrza et al. [5] explore the spatial-temporal interaction of bike-sharing demand in New York City. Afterward, data mining technology makes researches more accurate and objective. As bike-sharing big data and traffic networks make solid ground, clustering [6, 7], regression analysis [8], and time series [9, 10] have been widely introduced to model the characteristics of bike-sharing generation. The usefulness of these methods goes much beyond

improving the bike-sharing service quality, especially as the traffic system keeps constantly changing. In recent years, more novel algorithms make breakthroughs for interpreting the multiscale interactions between land use and bike-sharing demand: fusion modules consisting of random forest, probability fitting, and time-domain analysis [11]; a spatial-temporal flow model (DestiFlow) [12]; and a gravity model-based Bayesian algorithm [13]. It is found that different land use and built-up environment impose different influences on bike-sharing trip-generation. Conversely, bike-sharing popularity can also change the direction of urban land plan. Not surprisingly, the interesting conclusions published in existing literature enable us to better understand the changing mechanism for bike-sharing demand and to improve service efficiency.

- (2) Well-rounded researches regarding the travel destinations especially for taxi users have ruled out the bike-sharing [14, 15]. Moreover, the existing travel destination inference model in bike-sharing essentially draws lessons from the ideas of other trip modes. For example, Zhang et al. [16] put forward a trip behavior-based regression method to infer trip destinations and predict tourism destinations [17]. Considering the individual heterogeneity of bike-sharing users, many destination inferencing models study the activeness of macro-land use by integrating multiple factors. Obviously, a pool of likely candidate destinations from aggregation narrows the research scope and improves efficiency [18, 19]. Additionally, in addition to machine learning, the most common destination choice model is logit and its various improved forms [20, 21], which can be used to evaluate the influencing factors of riding destinations. Relevant results show that destination choices are determined by multiple elements, such as bike lane [22] and weather [23]. Based on these models for inference and prediction, it is possible to further grasp the spatial variations of bike-sharing in real time and to promote regional economic development.

Nevertheless, how to determine optimum combination of facilities remains unknown. Most significantly, the existing square grid methods are rather no-brainer, missing a comprehensive investigation of the influence of data aggregation. For example, MAUP can lead the study into the error caused by scale or partition problem [24, 25]. Due to the complexity of traffic problem, most of the previous researches on characteristics and behaviors have neglected this potential problem. Generally, traffic zones play an important role in traffic engineering, and most early models pay extra attention to the division method [26, 27]. With deepening the research, researchers find that the zoning size has a tremendous negative impact on the subsequent practical applications [28]. In other words, errors may have preceded the investigation of the traffic characteristics based on square grids or other arbitrary zoning methods. In

response to the above-mentioned challenges, this paper will propose a more-oriented method for illustrating spatial riding characteristics in order to address the following questions:

- (1) Which facility combination can generate maximum travel demand for bike-sharing?
- (2) If there are the associations between visiting destinations of bike-sharing users, how different are they in demand scenarios?
- (3) How to establish spatial constraints for characteristics modeling to avoid MAUP?

To answer these questions, first, the study explores optimum analysis areas with relatively high heat values corresponding to bike-sharing travel demand. Compared with the square grid and traditional traffic zone, hotspot areas-based spatial constraints can identify the key problem and clear away the MAUP triggered troubles. Next, the study implements the Apriori algorithm with the riding demand and land elements as result markers to establish the origin- and destination-based subordinate rules. In addition, by dividing the analysis areas into five demand scenarios, the study differentiates the spatial riding characteristics under different demand levels.

2. Methodology

2.1. Establishing Analysis Areas by Detecting Hotspots. As expected, both the grid method and traffic zone method may take on the MAUP effect and lead to the accumulation of errors. Therefore, instead of dividing the entire research zone, we directly identify areas where the bike-sharing demand is the greatest. After all, taking the areas with greater bike-sharing demand as spatial constraints contributes to the analysis of spatial characteristics by weakening individual heterogeneity. Based on this, we can improve the regional attraction in the fastest way and avoid many unnecessary troubles from zoning scale.

“The hotspot (HS) detection model regulates good positive progress for the research of this paper. Exploring the HSs of various elements has always been the central of analyzing urban mobility and spatial-temporal patterns. HSs are usually defined as areas where features are the same as geo-references on the map [29]. The advances of spatial density analysis algorithms make it simpler to identify the exact location and extent of range effects. Density analysis distributes the data in a spatial relationship across the ground to calculate a density surface and to show the allocation of elements. The kernel density analysis (KDA) is the most popular in geospatial analysis and is very suitable for estimating the density of given large-scale spatial elements [30, 31]. Generally, there are three density analysis methods, namely, dot density, line density, and kernel density. Obviously, the bike-sharing data herein are dot elements, and therefore line density analysis is no more applicable. Despite the applicability of dot density analysis to the data type herein, application of such simple analysis results to the hotspot detection process fails. More

importantly, we may take a knock because dot density analysis requires an assignment of neighborhood zone to calculate the density around each output image element. By contrast, KDA employs a kernel function to calculate the amount of points per unit area based on the point elements, so as to fit each point to a smooth conical surface with a continuous digital field pattern. KDA allows the dispersion of all known points towards all directions, starting from the point location. In KDA, the quadratic formula used to generate the surface gives the highest value to the center of the surface (the point location) and reduces to zero within the search radius distance. For each output image element, intersection points that accumulate for each dispersed surface shall be calculated. Essentially, KDA deploys a similar Gaussian kernel function for interpolation, which makes the results more valid and reliable. Furthermore, the smooth density field surfaces formed by the KDA provide a stable basis for accurate hotspot detection. To sum up, KDA is selected as the main tool for analyzing spatial riding characteristics of bike-sharing users herein.

The KDA scans the measurement area, casts the mesh density according to (1), and produces a smooth surface. After converting the discrete target model to a continuous field model, we can intuitively visualize the density around the target.

$$\hat{f}(a, b) = \frac{3}{Nr^2\pi} \left[1 - \frac{(a - a_i)^2 + (b - b_i)^2}{r^2} \right]^2, \quad (1)$$

where N is the number of points, r is the bandwidth, a and b are the coordinates of the center point k , a_i and b_i are the coordinates of the sampling point, and $f(a, b)$ is the density of the center point $p(a, b)$ on the grid cell. Figure 1 shows the principle of KDA. In the study area R , the KDA model takes any point as the center (kernel K) and calculates the density value of target points in bandwidth R , determined by the number and distance of material points in bandwidth. The KDA calculates the point density around each output grid unit. The density of each output grid unit is calculated from the sum of all values that cover the central core area of the grid unit.

As shown in Figure 1, the bandwidth usually determines the fineness of the KDA results, so it is necessary to choose a reasonable bandwidth according to the requirements. The bandwidth selection model adopted in this paper is as follows: firstly, determine the average center of n element points, then take the median D_m of the distance from the average center to each event point, and calculate the standard distance S_D of event points, with the equation being as follows:

$$r = 0.9 \times \min \left(S_D, \sqrt{\frac{1}{\ln 2}} \times D_m \right) \times n^{-0.2}. \quad (2)$$

In addition, this paper introduces the K -adjacency distance method as an auxiliary method to determine the optimal bandwidth, as shown in the following equation:

$$r = \sum_{i=1}^n \sum_{j=1}^k \frac{d_{ij}}{kn}, \quad (3)$$

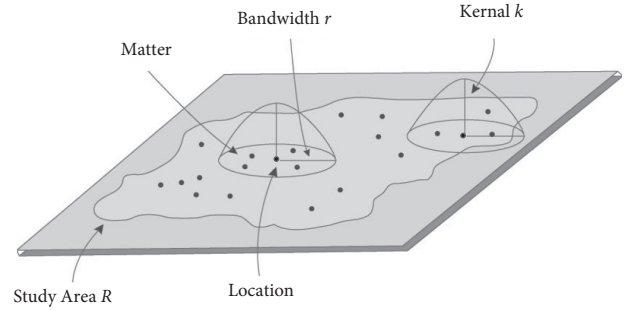


FIGURE 1: The demonstration of KDA principle.

where d_{ij} represents the nearest distance of order k , that is, the average distance from one event point to the k th element point. The k determines the smoothness of density surface. The larger the k is, the larger the bandwidth r and the smoother the generated density surface will be. The first method is the default method of calculating the optimal bandwidth in ArcGIS software, which only requires importing data to obtain the results without any other complex processing, while the second method needs to be implemented in ArcGIS with the help of Python programming. The results of the two calculations are compared, and the median value is chosen as the optimal bandwidth for the KDA in this paper.

The KDA algorithm can be employed to obtain a density surface with a continuous digital field pattern. As a result, HS detection model is constantly introduced to compensate for the accurate hot areas, which helps model the spatial riding characteristics. Figure 2 exhibits the algorithm flow of HS detection according to density field. Firstly, data pre-processing is carried out, such as eliminating abnormal data, filling up missing data, and extracting origin-destination data from the bike-sharing track. Then, export the origin-destination data to ArcGIS, and apply the KDA based on spatial analysis tools to output raster cells with kernel density as the raster value. Using KDA, Window Analysis, Minus, Reclassification, Raster to Polygon, and *Turning Feature into Point*, the travel HSs and the corresponding raster value of bike-sharing are obtained. The working principle is demonstrated in Figure 3.

The bike-sharing HSs are defined as analysis points, which are inputted into the geographic database to constitute the buffer areas by the GIS and to determine the analysis areas for spatial riding characteristics. According to the bus evaluation system in the transit metropolis, the buffer radius r of the analysis area for riding characteristics is set as 500 m herein, as shown in Figure 4. On the one hand, bike-sharing is a service open to the public, and it also has a parking station. On the other hand, in terms of the accessibility of conventional public transport stations, the maximum tolerance level for walking distance is mostly 500 m. That is to say, the majority of users prefer to walk for POI (point of interest) facility within 500 m radius. More importantly, the end riding point is usually at a little distance from the final destination as a result of the constraints of various factors such as bike-sharing

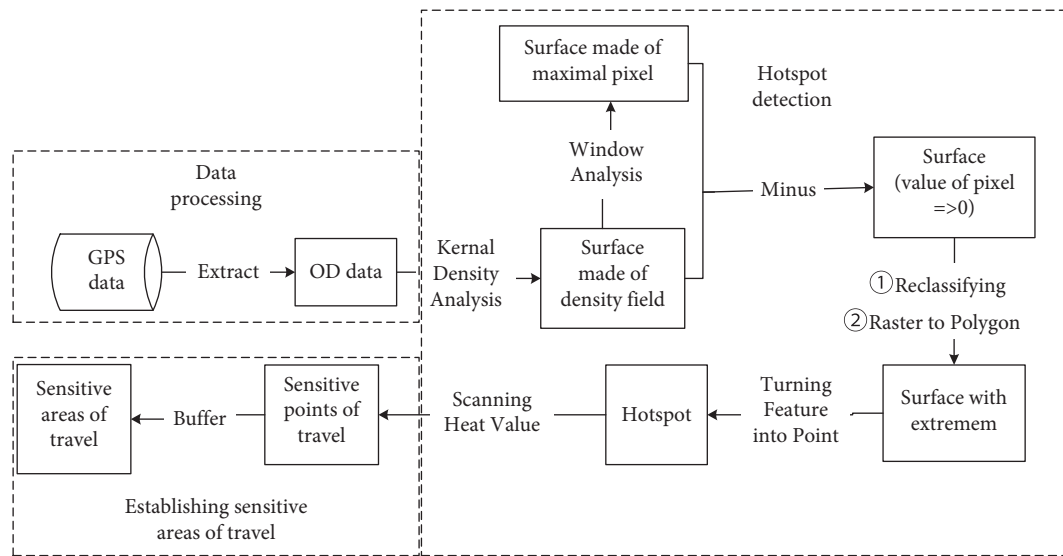


FIGURE 2: The flowchart of hotspot detection.

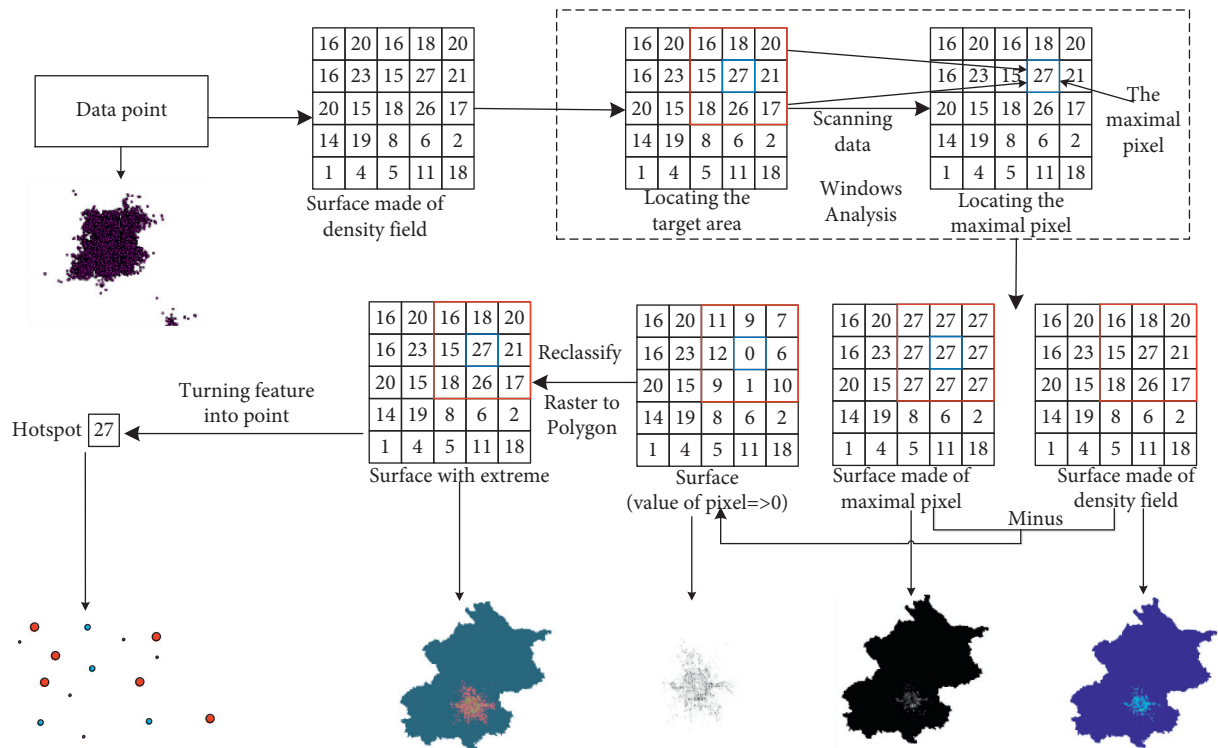


FIGURE 3: Principle demonstration of hotspot detection.

stations. Taken together, this distance often cannot be greater than 500 m; otherwise, it will exceed the user’s walking tolerance level. Therefore, it is relatively reasonable and realistic to set the buffer radius of the analysis area at 500 m.” (Sun C, Quan W. Evaluation of Bus Accessibility Based on Hotspot Detection and Matter-Element Analysis[J]. IEEE Access, 2020, PP(99):1–1). In fact, hotspot-based analysis areas are deployed for spatial constraints of association rule mining, which improve the efficiency of modeling.

2.2. Exploring Riding Characteristics Using Association Rules. As an essential algorithm in machine learning, association rule mining (ARM) is first proposed for market basket analysis (MBA). For example, the rule of “{onions, potatoes} ⇒ {burger}” in the market may indicate that if a customer buys onions and potatoes at the same time, then the customer is likely to buy hamburgers meat as well [32]. This information can be used to guide marketing activities, such as commodity pricing and commodity delivery. In this paper, the analysis areas are similar to different market

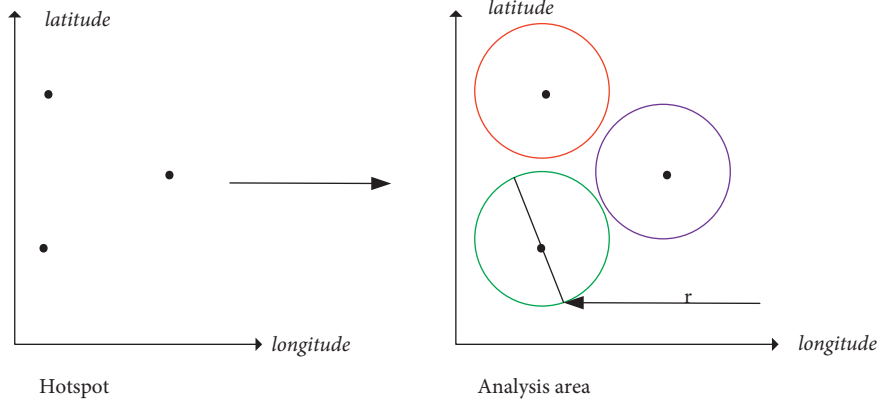


FIGURE 4: The representation for establishing spatial riding characteristic analysis area.

baskets, bike-sharing users are similar to customers, and POI facilities are similar to commodities, as shown in Figure 5. Therefore, we aim to establish the subordinate rules between bike-sharing demand and POI facilities and between POI facilities by modeling riding characteristics in hot areas.

2.2.1. Basic Definition. $I = \{i_1, i_2, \dots, i_n\}$ is defined as item-sets of spatial riding characteristics of bike-sharing users (SRCBU), i_i is defined as items of SRCBU, $D_b = \{t_1, t_2, \dots, t_n\}$ is defined as database for analyzing SRCBU, and t_k is regarded as transaction for presenting characteristics. A transaction is a collection of items; i.e., a transaction is a subset of I , $t_k \in I$ [33]. Each transaction is identified with a unique transaction ID. The SRCBU can be defined as follows:

$$X \Rightarrow Y, X, Y \subseteq I. \quad (4)$$

Each SRCBU consists of two different item-sets, where X is called premise or left-hand side (POI) and Y is called conclusion or right-hand side (POI or travel demand of bike-sharing).

2.2.2. Important Conceptions. In order to select interesting SRCBU from the set of all possible rules, various significance and interest constraints are employed, the best known of which are support and confidence of SRCBU [34].

(1) *Support.* Support is used to represent the occurrence frequency of SRCBU in the database. For item-set X in database D_b , its support is defined as the ratio of the number of transactions t containing item-set X to the number of all transactions T , as shown in the following equation:

$$\text{supp}(X) = \frac{|\{t \in T; X \subseteq t\}|}{|T|}. \quad (5)$$

(2) *Confidence Coefficient.* Confidence is introduced to measure the credibility of a SRCBU. For SRCBU $X \Rightarrow Y$, the confidence is defined as the ratio of the number of transactions in the database of SRCBU that contain both X and Y to the number of transactions that contain X . Therefore, confidence of a SRCBU can be regarded as conditional probability, as shown in the following equation:

$$\text{conf}(X \Rightarrow Y) = \frac{\text{supp}(X \cup Y)}{\text{supp}(X)}. \quad (6)$$

(3) *Lift.* A lift of SRCBU is defined as follows:

$$\text{lift}(X \Rightarrow Y) = \frac{\text{conf}(X \Rightarrow Y)}{\text{supp}(Y)} = \frac{\text{supp}(X \cup Y)}{\text{supp}(X) \times \text{supp}(Y)}. \quad (7)$$

(4) *Conviction.* The conviction of a SRCBU is as follows:

$$\text{conv}(X \Rightarrow Y) = \frac{1 - \text{supp}(Y)}{1 - \text{conf}(X \Rightarrow Y)}. \quad (8)$$

The conviction of SRCBU denotes the probability that X occurs but Y does not; i.e., the probability that the prediction of rule is wrong.

2.2.3. Association Rule Processing. An association rule between different POIs or between POIs and bike-sharing demand can only be considered interesting and key if it satisfies a minimum support threshold and a minimum confidence threshold. Association rule generation for SRCBU is split into two separate steps [35, 36].

- (i) Locate all frequent item-sets of SRCBU from the database using the minimum support threshold.
- (ii) Rules are generated from these frequent item-sets of SRCBU using minimum confidence thresholds. Although the phase of generating rules is straightforward, finding the frequent item-set of SRCBU requires more effort as it involves searching for the set of all possible items of SRCBU. The size of the item-set is a powerful set of I , which is $2^n - 1$ (excluding the meaningless empty set). Frequent item-sets have two very fundamental properties.

Property 1. All nonempty subsets of a frequent item-set of SRCBU are also frequent.

Property 2. All supersets of an infrequent set of SRCBU are infrequent.

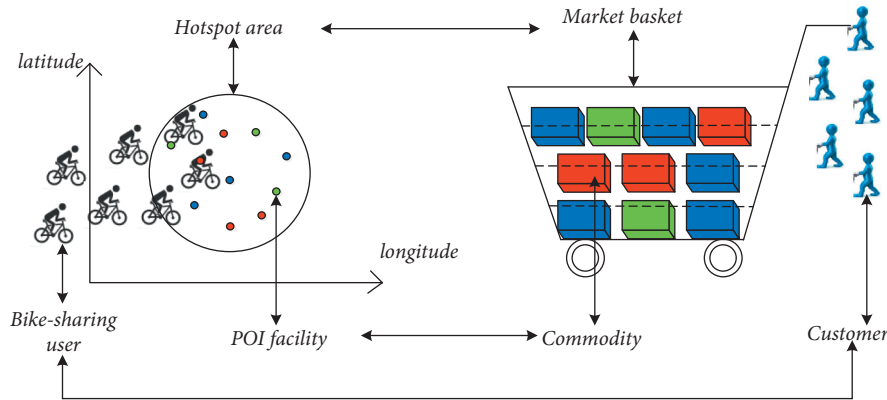


FIGURE 5: Model for riding characteristics similar to market basket analysis.

As shown in Figure 6, the color denotes the number of transactions containing SRCBU item-set and the SRCBU item-set at a lower level can contain at most the minimum number of SRCBU items of all its parents, e.g., {SRCBU item 1, 2} has at most $\min(\text{SRCBU item 1}, \text{SRCBU item 2})$ items. Based on this law, many efficient algorithms (e.g., Apriori, FP-Growth) make all the frequent item-sets of SRCBU available. The Apriori algorithm first generates frequent 1-item-set L_1 for SRCBU, and then combines two item-sets of SRCBU which only contain one different item in L_1 to generate frequent 2-item-set L_2 . The process is repeated until some value of r makes L_r null. The objective dominating Apriori is to get the largest frequent SRCBU item-set in a transactional dataset and use the same with a predetermined minimum confidence threshold to generate strong association rules between different POIs and between POIs and bike-sharing demand. Additionally, one fundamental feature of Apriori is that all nonempty subsets of a frequent item-set of SRCBU must also be frequent item-sets of SRCBU. Thus, the Apriori algorithm is processed as follows: ① finding all frequent item-sets of SRCBU (support must be greater than or equal to the given minimum support threshold for SRCBU herein); ② generating strong association rules between different POIs and between POIs and bike-sharing demand. From process ①, it is known that the items-sets of SRCBU that do not exceed a predetermined minimum support threshold have been removed, and if these remaining rules again satisfy a predetermined minimum confidence threshold for SRCBU, then a strong association rule between different POIs and between POIs and bike-sharing demand would be presented.

3. Results

3.1. Dataset. This paper selects Beijing as a case study, obtaining open-access bike-sharing travel records as the research data (<https://www.biendata.xyz/competition/mobike/data/>). The main fields on the dataset are orderid, bikeid, etc. (as shown in Table 1).

In view of some errors and inconsistent formats, we need to preprocess the data. The preprocessing procedure of data is as follows:

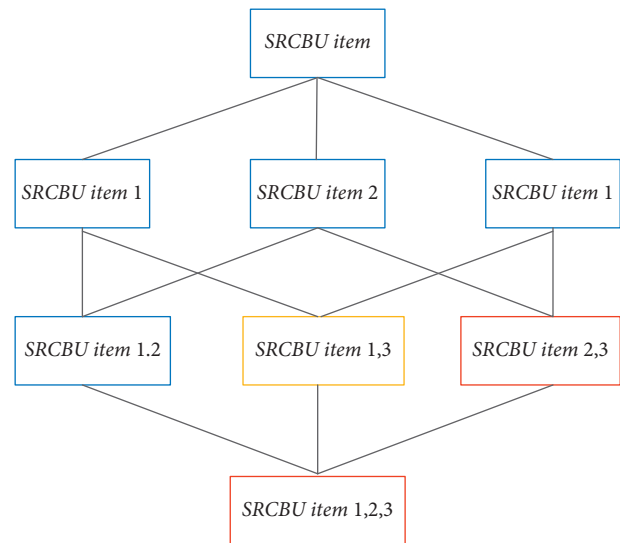


FIGURE 6: The property of frequent item-sets of SRCBU.

- (1) Coordinate transformation. Coordinate system transformation refers to the transformation of space points in different coordinate forms under the same Earth ellipsoid. The original location data of shared bikes is in Geohash format, which is converted into geodetic coordinate system (WGS84) according to research needs.
- (2) Data cleaning. In the original data of shared bikes, the equipment cannot send GPS data back, or the returned GPS data is in error due to GPS equipment failure, GPS signal shielding, and other factors. Therefore, these noise data should be cleared away before use. In the operation, totaling of types of errors was found, as shown in Table 2.
- (3) OD extraction. In fact, after the coordinate transformation, each piece of data has two latitude and longitude pieces of information, which are the corresponding origin and destination of each bike-sharing travel. We just need to distinguish the latitude and longitude of the origins from those of the destinations using the segmentation tool.

TABLE 1: The description of bike-sharing data.

Name of fields	The description of fields
orderid	The order reference
bikeid	The vehicle number
start_time	The initial time of the order
start_location_x	The longitude of the starting position of the order
start_location_y	The latitude of the starting position of the order
end_time	The ending time of the order
end_location_x	The longitude of the ending position of the order
end_location_y	The latitude of the ending position of the order

TABLE 2: Data cleaning.

The type of data to process	Form	Processing method
GPS records are not in Beijing	Location records are abnormal	Delete
GPS records are missing	Some data do not record the origin and destination	Complete the data according to the same bikes that have been used twice in a row, or delete them if not
Time records are missing	Some records have no corresponding time field	Delete
The time interval is too short	Users ride for a short time, only a few seconds	Delete
Some records are coded incorrectly	The encoding of some records is inconsistent with the encoding format of other data	Correct the code and re-read

Then, we employ a crawler tool to obtain all the POI data of Beijing with the help of Amap (<https://www.amap.com/>). The preprocessing for POI data is the same for bike-sharing data as described above. Further, all the POI data are introduced and divided into 13 categories to analyze the spatial riding characteristics including dining facilities and landscapes, as shown in Table 3.

The number of POIs within each analysis area is a continuous variable, but the Apriori association rule algorithm cannot deal with continuous numerical variables. Therefore, to adjust the format of data as required by modeling, nonhierarchical clustering algorithm (also called K-Means) is applied to discretize the data and cluster the attributes of each interest point into five categories. The principle of K-Means is to divide the data into predetermined class on a basis of a minimum error function and take distance as a similarity evaluation standard. This feature means that the closer the distance between two objects, the greater the similarity will be. For example, the discretization result of the dining facilities is shown in Table 4. A1 represents the minimum quantity, and A5 represents the maximum quantity, which is also valid for other POI facilities.

3.2. Analysis Location and Area. The KDA algorithm-based origin-destination data are employed to obtain the density field of bike-sharing travel, as shown in Figure 7(a). Through tools in ArcGIS, we recognize all hotspots with tools in ArcGIS and clustering method described in Section 2.1, as shown in Figure 7(b). Subsequently, with all the travel hotspots (analysis points) of bike-sharing as the center,

analysis areas are built as the buffer zone with a search radius of 500 m and exported to the geographical database, as shown in Figure 8. To discretize the raster value within each analysis area is also the premise of Apriori association rule algorithm. We use the method that processes the POI data to classify the raster values into five levels under which there are five demand scenarios. R1 is the highest level (Level 1), corresponding to the maximum demand of bike-sharing travel; on the contrary, R5 is the lowest level (Level 5), corresponding to the minimum demand of bike-sharing travel. In fact, R1–R5 simply indicate the demand scenarios from high to low; in other words, we define the group with the larger raster value in the clustering results as the high demand scenario R1, and so on. Essentially, R1–R5 are not different from A1–A5, B1–B5, etc. It is only that in this paper, for convenience of representation in GIS, R1 denotes the high values while the others, A1, B1, etc., denote the low values.

3.3. The Fitting Results of ARM. The model mainly consists of input, algorithm processing, and output. The input part includes the POI data, bike-sharing demand data, and modeling parameters. The processing part of algorithm is the Apriori, while the output part is the association rules between different POIs and between POIs and bike-sharing demand. We make the modeling a reality by first setting the minimum support and confidence of bike-sharing modeling parameters. Next, we input modeling data of POIs and then analyze them by the Apriori association rule algorithm conditional on the minimum support and confidence levels. In the application of association rules, there is absence of

TABLE 3: The description of POI data.

Name	Symbol	Description
Dining facilities	A	Restaurants, snack bars, canteens, etc.
Landscape	B	The Great Wall, Old Summer Palace, etc.
Public facilities	C	Public toilets, newsstands, emergency shelters, etc.
Company enterprise	D	Lear Group, 58 Tongcheng, GAC Toyota, etc.
Transport facilities	E	Railway stations, airports, etc.
Education facilities	F	Schools, universities, training institutions, etc.
Financial insurance facilities	G	Banks, insurance centers, investment agencies, etc.
Hotels	H	Dabao Apartment, Seven Day Hotel, etc.
Living facilities	I	Cemeteries, baths, health clinics, barber shops, etc.
Sports facilities	J	Soccer fields, basketball courts, video game studios, etc.
Medical facilities	K	Hospitals, health service stations, pharmacies, etc.
Government departments	L	People's Congress, mutual committees, police stations, etc.
Residences	M	Agricultural Community, Bright City, Rongxin Building, Silicon Valley, etc.

TABLE 4: The discrete number of dining facilities.

Range marker	Range	Number
A1	0-3	6060
A2	3-108.5980	1366
A3	108.5980-262.5235	490
A4	262.5235-641.7826	156
A5	>641.7826	17

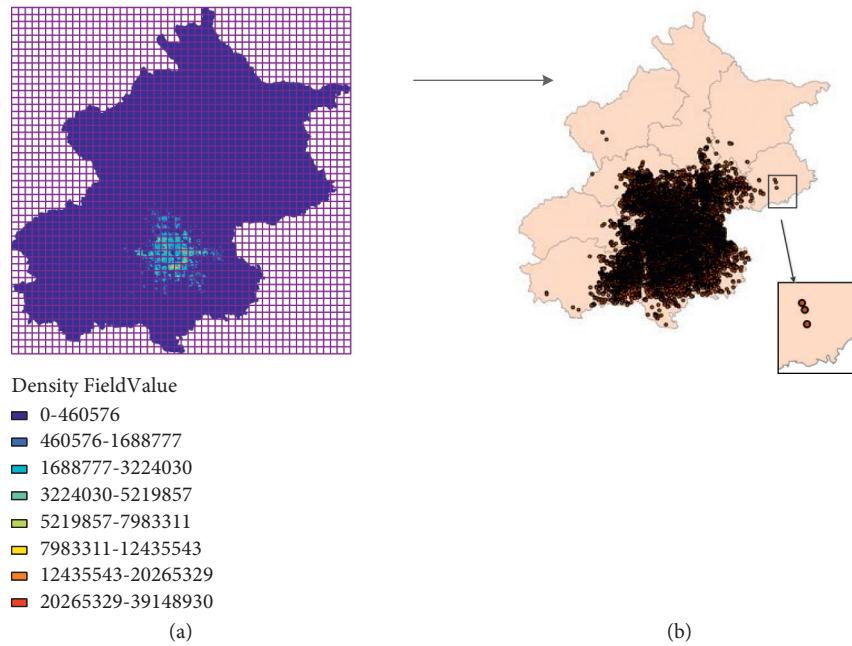


FIGURE 7: Demonstration density field and hotspots.

unified theory relative to the selection of relevant parameters, and the selection usually depends on different actual cases.

The model has been fitted twice. For the first time, with POI as X and bike-sharing demand level as Y , the model employs origin data to demonstrate spatial characteristics for bike-sharing users' origins, namely, Model 1. R1, R2, R3, R4, and R5 represent Level 1 to Level 5 travel demand. For

the second time, with POI as X and POI as Y , the model employs destination data to demonstrate spatial characteristics for bike-sharing users' destinations, namely, Model 2. By considering the initial sample size and the distribution characteristics of bike-sharing origin-destination data, we adjust the threshold value according to the parameter characteristics. After continuous manual debugging, the minimum support is selected to be 0.3 and 0.01, and the

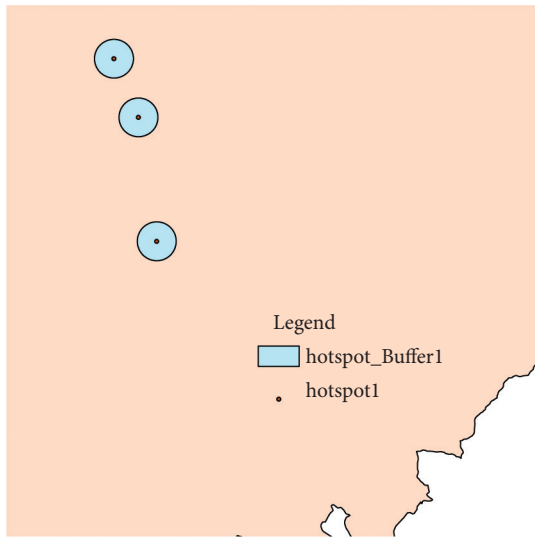


FIGURE 8: Examples of bike-sharing analysis areas (hotspot_buffer).

minimum confidence is selected to be 0.5 and 0.4, respectively, in the two fitting models. On the one hand, this is already the smallest threshold that can be adjusted and any smaller value would lose practical significance. On the other hand, it is only at this threshold that we can uncover the riding characteristics of bike-sharing users corresponding to the high-level analysis zones such as R4. The partial fitting results of the first model are shown in Table 5, while the partial fitting results of the second model are shown in Table 6.

The most important conclusions are drawn from the fitting results. For example, “A3, J3 -> R3” has a maximum support of 2.09% and a maximum confidence level of 51.52%. This result means that when dining facilities are at Level 3 and sports facilities are at Level 3, the probability of the demand for bike-sharing in the area being at Level 3 is 51.52%. This interpretation method is also suitable for the other cases.

Table 5 shows partial association rules between different POI facilities in the analysis area corresponding to Level 3 demand. For example, “B1, D2 -> G2” can reach maximum support of 34.99% and maximum confidence of 79.12%. Within the analysis area of Level 3 demand, when landscapes are at Level 1 and companies/enterprises are at Level 2, the probability of the financial insurance facilities being at Level 2 is 79.12%. The probability of this happening is as high as 34.99%. This result means that when POI facilities and bike-sharing demand meet the corresponding requirements, users will likely ride to financial facilities after visiting landscapes and companies/enterprises at the confidence level of 79.12%. The occurrence rate of this kind of associated visit is 34.99%. This interpretation method is also suitable for the other cases.

3.4. Further Interpretation Based on Statistic Index. Under the spatial constraints, there are significant differences in the association rules at different demand levels, no matter whether the bike-sharing demand or the POI itself is taken as

the result mark Y. This indicates the relevance of bike-sharing demand to different spatial riding characteristics. There is a significant difference between the POI association results corresponding to high and low demand, which may be related to spatial aggregation and dispersion effects. For example, when the bike-sharing demand is regarded as the outcome marker Y, the association rules between Level 1 demand and POI facilities cannot be mined, indicating that POI facilities cannot identify the analysis area with Level 1 demand. The only association rule obtained for R2 is “I2, J3 -> R2 (1.00%, 50%)”. The probability of Level 2 for demand is 50% in the area with Level 2 for living facilities and Level 3 for medical facilities. This incidence rate only accounts for 1.00%. In other words, when the two facilities satisfy the above requirements, the zone is more likely to be at the second-level bike-sharing demand. By contrast, 12 association rules belong to Level 2 bike-sharing demand, the most significant of which is “E4, I4 -> R3 (1.03%, 76.15%)”. The combination of transport facilities and living facilities has a great impact on the third-level bike-sharing demand, with a 76.15% probability of the analysis area falling into Level 3 demand when transport facilities and living facilities meet the requirements of Level 3 and Level 4, and an incidence rate of 1.0261%. There are more than a few dozens of association rules obtained from both Level 4 and Level 5 demand. Still, as for Level 4 analysis area, the support and confidence corresponding to the rules are significantly lower than those for Level 5 analysis area. For example, the highest confidence in Level 4 analysis area is “A2, G2 -> R4 (6.24%, 47.73%)”, while the highest confidence in Level 5 analysis area is over 96%, including “B1, H1 -> R5 (72.67%, 96.12%)” and “G1, K1 -> R5 (71.91%, 97.27%)”. Based on this, it is possible to obtain a probability of 47.73% that the bike-sharing demand in the analysis area falls into Level 4 when dining facilities and financial insurance facilities all fall into Level 2. However, when landscapes, hotels, and medical facilities belong to Level 1, the probability of bike-sharing demand belonging to Level 5 is higher than 71.91%. Further, the frequency of various POI facilities within each hotspot analysis area has been calculated, as shown in Table 7.

It can be seen from Table 7 that under different demand scenarios, dining facilities (8), landscapes (7), financial insurance facilities (10), hotels (5), and so on have the highest frequency. However, some POI facilities (public facilities and living facilities) only appear once or twice, which suggests that the impact of these facilities on bike-sharing demand is low for the corresponding support and confidence. When the POIs are deployed as the outcome marker Y, the relatively significant partial association rules and frequency statistics for POI combinations are shown in Table 8.

According to Table 8, landscapes and financial facilities frequently appear as the antecedents or consequences of the association rule simultaneously as high as 12 times. This result indicates that within all analysis areas, the frequency of simultaneously visiting financial facilities and landscapes hits high level, although there are differences at the different demand levels. This frequency is followed by a combination between landscapes and companies/enterprises, and

TABLE 5: Fitting results for the bike-sharing association rule Model 1 ($X \Rightarrow Y$).

Rule number	X		Result mark	Y	
	Range mark 1	Range mark 2		Support degree	Confidence coefficient
1	I2	J3	R2	1.0014%	50%
2	A3	J3	R3	2.0893%	51.5244%
3	D2	G2	R4	7.1084%	45.1689%
.....
n	B1	H1	R5	72.6666%	96.1243%

TABLE 6: Fitting results for the bike-sharing association rule Model 2 ($X \Rightarrow Y$).

Rule number	X		Result mark	Y	
	Range mark 1	Range mark 2		Support degree	Confidence coefficient
1	B1	D2	G2	34.9911%	79.1165%
2	H2	-	B1	45.4707%	78.0488%
.....
n	A3	-	J3	30.0178%	71.6102%

TABLE 7: The statistics of rules from Model 1.

Demand scenarios	The most significant rules	The most essential POI (frequency is in parentheses)
R1	-	-
R2	I2, J3 -> R2 (1.0014%, 50%) J3 -> R3 (3.301%, 46.354%); A3 -> R3 (2.918%, 48.163%) B1, K3 -> R3 (2.213%, 44.307%); A3, J3 -> R3 (2.089%, 51.524%)	I2 (1), J3 (1)
R3	E4, L3 -> R3 (1.001%, 49.390%); E4, H3 -> R3 (1.001%, 49.390%) C3, M3 -> R3 (1.00%, 40.5%); A3, G3, J3 -> R3 (1.013%, 55.405%) E4, I4 -> R3 (1.026%, 76.147%); A3, G3, J3 -> R3 (1.014%, 55.405%) G2 -> R4 (8.023%, 43.324%); A2 -> R4 (7.158%, 42.387%) D2, G2 -> R4 (7.108%, 45.169%); A2, D2 -> R4 (6.342%, 43.771%)	A3 (3), B1 (1), C3 (1), E4 (3), F2 (1), G3 (2), H3 (1), I4 (1), J3 (4), K3 (1), L3 (1), M3 (2)
R4	B1, G2 -> R4 (6.293%, 42.488%); K2 -> R4 (6.268%, 41.728%) A2, G2 -> R4 (6.243%, 47.732%); D2, H2 -> R4 (6.1565%, 41.3965%) D2, F2 -> R4 (6.144%, 40.605%) G2, H2 -> R4 (6.144%, 45.764%) B1 -> R5 (77.661%, 82.680%) G1 -> R5 (73.9028%, 96.3261%) B1, G1 -> R5 (73.198%, 96.433%); H1 -> R5 (73.186%, 96.057%)	A2 (3), B1 (1), D2 (4), F2 (1), G2 (5), H2 (2)
R5	K1 -> R5 (73.1611%, 96.8893%); A1 -> R5 (72.704%, 97.046%) B1, H1 -> R5 (72.667%, 96.124%); B1, K1 -> R5 (72.555%, 96.928%) A1, B1 -> R5 (72.222%, 97.075%); G1, K1 -> R5 (71.913%, 97.274%)	A1 (2), B1 (5), G1 (3), H1 (2), K1 (3)

TABLE 8: The statistics of rules from Model 2.

Demand scenarios	The most significant rules	The most obvious POI combination (frequency is in parentheses)
R1	B1 -> G2 (56.6667%, 70.8333%); G2 -> B1 (56.6667%, 85%)	
	B1 -> D2 (53.3333%, 66.6667%); D2 -> B1 (53.3333%, 84.2105%)	
	B1 -> H2 (53.3333%, 66.6667%); H2 -> B1 (53.3333%, 84.2105%)	
	D2 -> G2 (50%, 78.9474%); G2 -> D2 (50%, 75%)	
	G2 -> H2 (50%, 75%); H2 -> G2 (50%, 78.9474%)	
R2	B1, G2 -> D2 (43.3333%, 76.471%); D2, G2 -> B1 (43.3333%, 86.667%)	B and G (12)
	B1 -> I2 (62.8415%, 85.1852%); I2 -> B1 (62.8415%, 70.5521%)	B and D (11)
	H2 -> I2 (49.7268%, 91.9192%); I2 -> H2 (49.7268%, 55.8282%)	D and G (11)
	G2 -> I2 (46.4481%, 90.4255%); I2 -> G2 (46.4481%, 52.1472%)	B and H (7)
	J3 -> I2 (44.2623%, 98.7805%); D2 -> I2 (43.7158%, 93.0233%)	G and K (4)
R3	K3 -> I2 (43.1694%, 100%); C3 -> I2 (42.0765%, 100%)	H and I (4)
	B1, H2 -> I2 (37.1585%, 89.4737%); B1, I2 -> H2 (37.1585%, 59.130%)	B and I (3)
	D2 -> G2 (46.3588%, 79.0909%); G2 -> D2 (46.3588%, 81.3084%)	D and H (2)
	B1 -> H2 (45.4707%, 60.0939%); H2 -> B1 (45.4707%, 78.0488%)	A and D (2)
	B1 -> D2 (44.2274%, 58.4507%); D2 -> B1 (44.2274%, 75.4545%)	A and G (2)
R4	B1 -> G2 (43.5169%, 57.5117%); G2 -> B1 (43.5169%, 76.324%)	A and B (2)
	D2 -> H2 (39.7869%, 67.8788%); H2 -> D2 (39.7869%, 68.2927%)	G and H (2)
	B1, D2 -> G2 (34.991%, 79.117%); B1, G2 -> D2 (34.991%, 80.408%)	B and K (2)
	D2 -> G2 (62.7045%, 85.1852%); G2 -> D2 (62.7045%, 88.5978%)	G and I (2)
	B1 -> D2 (57.7972%, 72.5034%); D2 -> B1 (57.7972%, 78.5185%)	D and I (1)
R5	A2 -> D2 (55.9433%, 88.601%); D2 -> A2 (55.9433%, 76%)	C and I (1)
	B1 -> G2 (55.5071%, 69.6306%); G2 -> B1 (55.5071%, 78.4284%)	J and I (1)
	A2 -> G2 (55.0709%, 87.2193%); G2 -> A2 (55.0709%, 77.812%)	K and I (1)
	B1, D2 -> G2 (49.5093%, 85.660%); B1, G2 -> D2 (49.509%, 89.195%)	
	B1 -> G1 (92.5735%, 94.2534%); G1 -> B1 (92.5735%, 99.0465%)	
R5	B1 -> H1 (91.9012%, 93.5689%); H1 -> B1 (91.9012%, 99.2905%)	
	B1 -> K1 (91.7605%, 93.4257%); K1 -> B1 (91.7605%, 99.172%)	
	A1 -> B1 (91.3383%, 99.3368%); B1 -> A1 (91.3383%, 92.9959%)	
	G1 -> K1 (90.9475%, 97.3068%); K1 -> G1 (90.9475%, 98.2933%)	
	B1, G1 -> K1 (90.338%, 97.5850%); B1, K1 -> G1 (90.338%, 98.450%)	

between companies/enterprises and financial facilities. A frequency value exceeding 11 provides a high probability of users visiting these POI combinations simultaneously. By contrast, the frequency of combinations between companies/enterprises and living facilities and between public facilities and living facilities is only 1, which means that the probability of bike-sharing users simultaneously visiting these POI facilities is low. More obviously, some combinations will never appear at all, with a frequency of only 0. It is impossible for users to access these facilities of POI at the same time.

4. Conclusion and Discussion

This paper implemented modeling spatial riding characteristics of bike-sharing users under five demand scenarios based on the hotspot detection and the association rule mining, which has established the subordinate rules between bike-sharing demand and POIs and between POIs. As far as origin characteristics of riding are concerned, it is most important to investigate the bike-sharing demand level from different POI types. The analysis area with Level 1 demand is more complex and cannot be directly investigated from the type combinations of POI facilities. This situation is reflected in the fact that no corresponding association rules can be

found out when Level 1 demand is deployed as a result marker. However, Level 2 analysis area has a certain degree of differentiation. For example, when living facilities are at Level 2 and sports facilities are at Level 3, the probability that the bike-sharing demand is at Level 2 is higher. Therefore, when these facilities in an area meet the requirement as mentioned above at the same time, it is necessary to increase the dispatching number and to approach the two facilities as close as possible. By contrast, as for the analysis area with Level 3 demand, more factors are associated with the bike-sharing demand. More importantly, most range markers of POI facilities corresponding to Level 3 analysis area are at Level 3, with only the transport facilities being at Level 4. This means that if most POIs are at Level 3 and the transport facilities are at Level 4, then the bike-sharing demand is likely to be at Level 3, which is a medium demand level. However, Level 4 demand for bike-sharing travel is closely related to dining facilities, landscapes, companies/enterprises, educational facilities, financial insurance facilities, and hotels. In particular, the financial insurance facilities are most closely related to companies/enterprises. When these POI facilities in an area are at Level 2, especially when financial insurance facilities and companies/enterprises are present simultaneously, it is more likely that the bike-sharing demand is at Level 4, which is a lower demand level.

The POIs closely related to Level 5 demand for bike-sharing travel are dining facilities, landscapes, financial insurance facilities, hotels, and medical facilities. When these POI facilities are at Level 5 in an area, especially when landscapes, financial insurance facilities, and hotels tend to zero simultaneously, it is more likely that the bike-sharing demand is at Level 5, with almost no demand. Overall, the POI facilities that have the greatest impact on the bike-sharing trip-generation are financial insurance facilities, which play a large role in determining the demand level for bike-sharing. We infer that white-collar workers from financial insurance facilities seem to prefer bike-sharing for their off-duty commute, which is an important finding as it helps to better serve these “environmentalists.” Secondly, the influence of dining facilities and landscape is also greater, in terms of multiple rules of bike-sharing travel. The rules of dining facilities are easy to understand. After all, the bike-sharing can be employed to help burn calories, and therefore it becomes the first choice for users after consuming abundant food. For landscapes, rules may be related to mood or traffic jam nearby. After enjoying the landscapes, one may not care about the size of the return journey time. More importantly, Beijing is a fast-paced city, where most working people only have weekends to see the landscapes and to have fun. People gathering around scenic spots on weekends can lead to increased traffic congestion, which is the place where bike-sharing travel has obvious advantage. However, there is not enough evidence that public facilities and government departments have a strong effect on bike-sharing demand, given certain support and confidence requirements. This shows that travelers leaving public facilities (such as public toilets) and government officials at work may have little interest in riding shared bicycles. The former mainly do not go further and have lower travel demand, while the latter pays more attention to time.

As for characteristics of riding destination, multiple POI facilities are strongly correlated, which helps to better deploy bike-sharing parking and optimize land use structure in a region. Most notably, landscapes are closely related to financial insurance facilities. Within Level 1 analysis area, when the landscapes belong to Level 1 and the financial insurance facilities belong to Level 2, there is the 70.83% probability of cyclists visiting the landscapes and then visiting the financial insurance facilities, or 85% probability of visiting the financial insurance facilities and then visiting the landscapes. Even within Level 3 analysis area, the support for both rules comes to 43.52%, still with a confidence level greater than 40%. This is an interesting discovery because two seemingly contradictory individuals are connected. Therefore, it is reasonable to believe that the above-mentioned facilities in an area are more attractive to bike-sharing users when they meet the corresponding requirements. Similarly, the combinations between landscapes and companies/enterprises and between companies/enterprises and financial insurance facilities are found 11 times more frequently as either the antecedents or consequences of the association rule simultaneously. For example, within the first-level R1 analysis area, there is “B1, G2 -> D2 (43.33%, 76.47%) or D2, G2 -> B1 (43.33%, 86.67%)”. Within an area

with Level 1 landscapes, Level 2 companies/enterprises, and Level 2 financial insurance facilities, the probability of cyclists visiting companies/enterprises after visiting landscapes and financial insurance facilities is 43.33%. The probability of this occurrence is 76.471%. Instead, the probability of cyclists visiting landscapes after visiting companies/enterprises and financial insurance facilities is 43.33%. The probability of this occurrence is over 80%, with a value of 86.67%. The next strongest connection is between landscapes and hotels, which exists seven times and ranks fourth. Within Level 1 analysis area, when landscapes belong to Level 1 and hotels belong to Level 2, the probability of cyclists visiting hotels after visiting landscapes is 66.67%. The probability of this situation occurring is 53.33%. On the contrary, the probability of cyclists visiting landscapes after visiting hotels is 84.21%. Therefore, the coexistence of these two facilities is conducive to attracting bike-sharing users and improving usage frequencies even for short distances. The frequency of coexistence of financial insurance facilities and medical facilities or of hotels and living facilities is four times, demonstrating a strong correlation between them. However, the frequency of coexistence of living facilities and sports facilities or of living facilities and medical facilities is only 1, which shows that these POI facilities have little correlation with other facilities. More importantly, many POI facilities have nothing to do with other facilities. For example, the frequency of coexistence of landscapes and public facilities is 0. There is no evidence that their combinations have contributed to the attraction of bike-sharing users.

This study implements an analysis of spatial riding characteristics from the perspective of demand differences based on hotspot detection and association rule mining, which demonstrate the subordinate rules between bike-sharing demand and POI as well as between POIs. In general, the superiority of this study, compared to other riding characteristics models, and its importance in bike-sharing dispatch or urban infrastructure plan are as follows.

- ① Being more reliable and practicable. Most of the existing researches only present some strict and definite characteristics and seldom demonstrate the confidence level at which the POI facilities can be accessed simultaneously as the travel demand changes. Confidence level is the degree to which traffic managers rely on the effectiveness of bike-sharing stations and scheduling quantities, and it is essentially a rationality test of urban structure from the perspective of users. Based on this, we can adjust the layout of POI facilities according to the travel demand that the system need to meet, so as to improve the attractiveness of bike-sharing travel and reduce environmental pollution.
- ② Being more directed. Most studies have been concentrated on the cross-spatio-temporal riding characteristics of bike-sharing users, but analysis for single small-scale POI structure is relatively little. The hot areas enable us to establish spatial constraints according to demand of bike-sharing travel. Compared with other methods, the characteristic method applied to areas with large trip-generation and trip-attraction can enable us to understand the most important intrinsic connection between different facilities. Similar to the basket analysis in

Figure 5, we just need to grasp the relationships between the most important items in customers' shopping baskets, but not all of them. More significantly, the spatial constraints of hot area greatly reduce the data amount and improve the fitting efficiency of algorithm. ③ Low difficulty. We just require origin-destination and POI data within each analysis area and never have to spend a great deal of money on complex spatial-temporal information, which would be accomplished by simple programs and geographical software. On the one hand, we need no micro-data manipulation like map matching, just regular data quality improvement. Compared to other complex algorithms such as deep learning interactive model and so on, the Apriori employed never calls for a specially designed solving process and has a low threshold for refitting in other zones with different environments. As mentioned above, the solution speed herein is very fast, so efficiency is also an advantage of this study. However, there are still some points needing further discussion: ① The riding characteristics of bike-sharing users are only discussed within each hotspot analysis area. Compared with spatial geographic models and regression methods, this study lacks discussion on the correlation between different analysis areas. ② We mainly focus on the riding characteristics of bike-sharing users but ignore the relations with other modes of transportation, such as subway and bus that have been developed fast in many cities. ③ The bike-sharing origin-destination data for one week are deployed to model association rules and to analyze riding characteristics, which lacks consideration of other objective elements, such as weather and sudden events.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 51478110.

References

- [1] A. Soltani, T. Mátrai, R. Camporeale, and A. Allan, "Exploring shared-bike travel patterns using big data: evidence in Chicago and Budapest," *Lecture Notes in Geoinformation and Cartography*, pp. 53–68, Springer, Cham, Switzerland, 2019.
- [2] Y. Zhang, T. Thomas, J. G. Brussel, and M. Maarseveen, "The characteristics of bike-sharing usage: case study in Zhongshan, China," *Int. J. Transp. Dev. Integr.*, vol. 1, no. 2, pp. 245–255, 2018.
- [3] E. Fishman, S. Washington, and N. Haworth, "Bike share: a synthesis of the literature," *Transport Reviews*, vol. 33, no. 2, pp. 148–165, 2013.
- [4] M. Amiri and F. Sadeghpour, *Cycling Characteristics in Cities with Cold Weather*, Sustain. Cities Soc., 2015.
- [5] A. A. F. and E. B. N. A., "Incorporating the impact of spatio-temporal interactions on bicycle sharing system demand: a case study of New York CitiBike system," *Journal of Transport Geography*, vol. 54, pp. 218–227, 2016.
- [6] L. Mingyuan and W. Guan, "A study of parking space of shared bikes near rail Transit stations, Chengdu," *Planners*, vol. 35, no. 18, pp. 54–61, 2019.
- [7] Y. Xing, K. Wang, and J. J. Lu, "Exploring travel patterns and trip purposes of dockless bike-sharing by analyzing massive bike-sharing data in Shanghai, China," *Journal of Transport Geography*, vol. 87, Article ID 102787, 2020.
- [8] T. Ma, C. Liu, and S. Erdogan, "Bicycle sharing and public Transit: does capital bikeshare affect metrorail ridership in Washington, D.C.," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2534, 2015.
- [9] Z. Gao, S. Wei, L. Wang, and S. Fan, "Exploring the spatial-temporal characteristics of traditional public bicycle use in Yancheng, China: a perspective of time series cluster of stations," *Sustainability*, vol. 12, no. 16, pp. 1–17, 2020.
- [10] A. Kaltenbrunner, R. Meza, J. Grivolla, J. Codina, and R. Banchs, "Urban cycles and mobility patterns: exploring and predicting trends in a bicycle-based public transport system," *Pervasive and Mobile Computing*, vol. 6, no. 4, pp. 455–466, 2010.
- [11] Y. Du, F. Deng, and F. Liao, "A model framework for discovering the spatio-temporal usage patterns of public free-floating bike-sharing system," *Transportation Research Part C: Emerging Technologies*, vol. 103, no. 6, pp. 39–55, 2019.
- [12] J. Dong, B. Chen, L. He et al., "A spatio-temporal flow model of urban dockless shared bikes based on points of interest clustering," *ISPRS International Journal of Geo-Information*, vol. 8, no. 8, p. 345, 2019.
- [13] S. Li, C. Zhuang, Z. Tan, F. Gao, Z. Lai, and Z. Wu, "Inferring the trip purposes and uncovering spatio-temporal activity patterns from dockless shared bike dataset in Shenzhen, China," *Journal of Transport Geography*, vol. 91, no. 1, Article ID 102974, 2021.
- [14] F. Zong, Y. Tian, Y. He, J. Tang, and J. Lv, "Trip destination prediction based on multi-day GPS data," *Physica A: Statistical Mechanics and Its Applications*, vol. 515, pp. 258–269, 2019.
- [15] Y. Endo, K. Nishida, H. Toda, H. Sawada, and D. M. Sawada, "Predicting destinations from partial trajectories using recurrent neural network," *Advances in Knowledge Discovery and Data Mining*, vol. 10234, pp. 160–172, Springer, Cham, Switzerland, 2017.
- [16] J. Zhang, X. Pan, M. Li, and P. S. Yu, "Bicycle-sharing system Analysis and trip prediction," *2016 17th IEEE International Conference on Mobile Data Management (MDM)*, vol. 1, pp. 174–179, 2016.
- [17] A. Emanuela, H. Norbert, N. I. Costel, and E. Tudor, "Measuring the intention to return to a foreign tourism destination in the cases of two age layers of Generation Y - a logistic regression-centred approach. Evidence from Romania and South Africa," *Eur. J. Interdiscip. Stud.*, vol. 1, 2019.
- [18] J. Jiang, F. Lin, J. Fan, H. Lv, and J. Wu, "A destination prediction network based on spatiotemporal data for bike-sharing," *Complexity*, vol. 2019, pp. 1–14, 2019.
- [19] Y. Li and B. Shuai, "Origin and destination forecasting on dockless shared bicycle in a hybrid deep-learning algorithms," *Multimedia Tools and Applications*, vol. 79, no. 7–8, pp. 5269–5280, 2020.

- [20] A. Faghih-Imani and N. Eluru, "Analysing bicycle-sharing system user destination choice preferences: c," *Journal of Transport Geography*, vol. 44, no. apr, pp. 53–64, 2015.
- [21] A. Faghih-Imani and N. Eluru, "A finite mixture modeling approach to examine New York City bicycle sharing system (CitiBike) users' destination preferences," *Transportation*, vol. 47, no. 2, pp. 529–553, 2020.
- [22] F. González, C. Melo-Riquelme, and L. De Grange, "A combined destination and route choice model for a bicycle sharing system," *Transportation*, vol. 43, no. 3, pp. 407–423, 2016.
- [23] V. Lucas and A. R. Andrade, "Predicting hourly origin-destination demand in bike sharing systems using hurdle models: l," *Case Studies on Transport Policy*, vol. 9, no. 4, pp. 1836–1848, 2021.
- [24] A. Pani, P. K. Sahu, A. Chandra, and A. K. Sarkar, "Assessing the extent of modifiable areal unit problem in modelling freight (trip) generation: relationship between zone design and model estimation results," *Journal of Transport Geography*, vol. 80, no. September, Article ID 102524, 2019.
- [25] A. Cartone and P. Postiglione, "Principal component analysis for geographical data: the role of spatial effects in the definition of composite indicators," *Spatial Economic Analysis*, vol. 16, no. 2, pp. 126–147, 2020.
- [26] D. Honghui, "Traffic zone division based on big data from mobile phone base stations - ScienceDirect," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 278–291, 2015.
- [27] L. Wang, J. Tang, X. Fei, and M. Gong, "A mixed integer programming formulation and solution for traffic analysis zone delineation considering zone amount decision," *Information Scientist*, vol. 280, pp. 322–337, 2014.
- [28] M. Garreton, A. Basauri, and L. Valenzuela, "Exploring the correlation between city size and residential segregation: comparing Chilean cities with spatially unbiased indexes," *Environment and Urbanization*, vol. 32, no. 2, pp. 569–588, 2020.
- [29] S. E. Wilson, A. Bunko, S. Johnson, J. Murray, and Y. Li, "The geographic distribution of un-immunized children in Ontario, Canada: hotspot detection using Bayesian spatial analysis," *Vaccine*, vol. 39, no. 1, 2021.
- [30] A. Gramacki and Artur, "FFT-based algorithms for kernel density estimation and bandwidth selection," *Studies in Big Data*, vol. 10, no. 5, pp. 85–118, 2018.
- [31] J. Aitchison and I. J. Lauder, "Kernel density estimation for compositional data," *Appl. Stat.*, vol. 34, no. 2, pp. 129–137, 2018.
- [32] Y.-L. Chen, K. Tang, R.-J. Shen, and Y.-H. Hu, "Market basket analysis in a multiple store environment," *Decision Support Systems*, vol. 40, no. 2, pp. 339–354, 2005.
- [33] L. Bing, W. Hsu, and Y. Ma, "Integrating classification and association rule mining," *Proc Kdd*, vol. 1711, no. 8, pp. 27–31, 1998.
- [34] C. Q. Zhang and S. C. Zhang, *Association Rule Mining: modeling and Algorithms*, Springer, Sydney, 2002.
- [35] P. Smart, K. K. Thanammal, and S. S. Sujatha, "A novel linear assorted classification method based association rule mining with spatial data," *Sādhanā*, vol. 46, no. 1, pp. 1–12, 2021.
- [36] P. Matapurkar and S. Shrivastava, "Comparative analysis for mining fuzzified dataset using association rule mining approach," in *Proceedings of the 8th International Conference on Reliability*, pp. 383–387, Noida, India, June 2020.