ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)

https://doi.org/10.17559/TV-20230426000571

Original scientific paper

Clustering of Passenger Flow and Land-Use of Beijing Urban Rail Transit Stations Based on Multi-Source Data

Xuanxuan XIA, Hongchang LI, Kexin LIN, Kun LING*

Abstract: With the deepening of transit-oriented development (TOD), the construction of high-density, multi-functional urban spatial structures around public transportation stations has become a significant focus in urban development planning. This study is aimed to explore the spatial and temporal cluster patterns and the relationship between taxi passenger flow and land use types around urban rail transit stations. Taking Beijing as an example, this study extracts the time series of pick-up and drop-off points from the taxi GPS track dataset and employs the affinity propagation (AP) method to conduct the spatial and temporal cluster analysis on these taxi pick-up and drop-off points. Then, based on the urban POI dataset, this study classifies these stations into four functional types by adopting the K-means method. Finally, the partial least squares (PLS) method has been used to explore the relationship between the taxi passenger flow and the land use types under different time intervals. The result indicates that: first, there is a regional aggregation effect for the taxi passenger source in the rail transit station area, and the cluster areas are mainly located in Chaoyang and Haidian districts, as well as along the circular metro line. Besides, there is a significant rush hour in the morning and evening for this passenger flow. Second, the commercial, residential, and comprehensive transit stations are mainly located in the central urban districts of Beijing with balanced spatial distribution, while the industrial transit stations are dispersed to the outer suburbs of the city, which conform to Alonso's spatial distribution of land. Third, there is a significant positive correlation between taxi passenger sources in the rail transit station area and land use types containing transportation service, public facility and daily life service. This study analyses the population mobility and land type cluster form around the urban rail transit from the micro-perspective, which verifies and supplements Alonso's transportation location theory. Therefore, this study contributes to the rational planning of urban spatial resource allocation and the construction of the urban rail transit network.

Keywords: clustering algorithm; land-use; passenger flow; rail transit station

1 INTRODUCTION

During the process of urban development, there is an imbalance in the spatial distribution of economic activities of the population, as well as the allocation of land resources. According to the census data, the population density of the core area of Beijing is 100 times that of the outer suburbs, and about 79.5% of the population is concentrated in the urban function expansion region and the urban development new area. The spatial dislocation between urban residence and employment leads to an increasingly serious Job-housing imbalance. The difference in population mobility and the spatial recourse allocation between the urban core area and the suburban area can be attributed to the agglomeration of different economic factors [1]. The transportation efficiency of products and the labor force is affected by the economic connection and agglomeration effect between urban regions [1] and affects the spatial resource allocation of different locations within the city. As we all know, the contradiction and imbalance between transportation supply and travel demand has always been an enduring issue [2]. While large-scale transportation supply effectively caters to centralized travel demands, it faces efficiency issues in meeting dispersed travel demands [3]. Public transportation generally integrates consumer travel demands within a certain range, extracting commonalities in terms of time, direction, origins and destinations, and speed, while providing a large-scale transportation supply. Urban rail transit plays a crucial role in urban economic growth and people's economic activities [4]. Urban rail transit is an important component of public transportation in large cities, and its stations, the nodes in the rail transit network [5], play the part of connectors between various transportation modes for catering to travel demands in different times and spaces [4]. It is widely recognized that there are substantial disparities in passenger flow among different rail transit stations and lines. By examining operational urban rail transit systems in China, there is a significant imbalance in the distribution of passenger flow across different rail lines. For instance, the daily passenger flow on Line 10 in Beijing exceeds one million, whereas that of the Fangshan Line is less than 200000. Urban rail transit serves as a typical example of a large-scale transportation supply, and the distribution of taxi pick-up and drop-off points reflects the dispersed travel demands of residents. Therefore, studying the spatiotemporal agglomeration patterns of taxi pick-up and drop-off points around urban rail transit stations can effectively understand the supply-demand imbalance in the transportation market [6]. In addition, in the process of urban spatial sprawl and land use density decreases, which will lead to urban decay, the break of urban internal links, as well as a series of Urban energy and environmental issues [7]. In 1992, Peter Calthorpe introduced the concept of TOD. In "The next American metropolis: Ecology, community, and the American dream", the development model of replacing suburban sprawl with TOD was proposed [8]. The main feature of the TOD development mode is to improve the mixed degree of land function in the public transport station area. The result is a compact urban form that promotes the use of public transport [9]. In the TOD model of rail transit, emphasis is not only placed on constructing centralized, high-density, and multi-functional urban spatial structures but also on the efficient integration of different transportation modes. Mega-cities also emphasize the development of large-scale station complexes and the creation of integrated demonstration projects that integrate high-density station areas and street environments [10]. Not only does unbanning rail transit commuting efficiency for residents [11, 12], but it plays a crucial role in urban economic growth and people's economic activities [4]. It has a significant effect on population agglomeration and spatial structure of cities [13] because of the distinct agglomeration patterns of land use types around different rail transit stations [4]. Therefore, analyzing the distribution of passenger flow and the agglomeration status of land types within rail transit station areas is of great importance for public transportation management and land planning in cities [5]. From the concept of traffic location, traffic location is the gathering place of traffic behavior and traffic resources [14]. Traffic resource is the condition of traffic activity, which is conducive to the aggregation of traffic behavior. Conversely, the increase in traffic activities will also promote the utilization of transportation resources and further the development of agglomeration. From a spatial perspective, the spatial form of traffic location is represented by the location of traffic nodes (traffic hubs), the location of traffic lines, and the location of traffic networks [15]. The different aggregation states of traffic resources lead to the characteristics of non-homogeneity in space, therefore the traffic behavior also produces a certain agglomeration in space. The gathering of traffic location mainly includes the traffic behavior and traffic resources generated by people and goods in space. Transportation resources can be divided into two categories: movable transportation resources and fixed transportation resources. In this paper, we take the taxi passenger flow around the urban rail transit station as the traffic behavior in the traffic location station. Meanwhile, the POI around the rail transit station is regarded as the traffic resource in the traffic location station. Based on the TOD theory and the traffic location theory, we study the spatial and temporal agglomeration of population and the agglomeration form of land resources in the location of urban rail transit stations. Then we analyze the correlation between them. This study aims to improve the efficiency of urban rail transit stations and lines and optimize the spatial layout of the urban rail transit network. It is important to study the heterogeneity of taxi passenger flow patterns and land agglomeration types around urban rail transit stations in different locations, as well as the interactions between them [16, 5]. This study focuses on the following questions:

1. The AP clustering algorithm based on time series is used to analyze the spatial and temporal clustering state of taxi passenger flow in the location of urban rail transit stations. By using the K-means clustering algorithm, we explore the spatial agglomeration state of POI in the location of urban rail transit stations.
2. From the perspective of rail transit lines and administrative districts, we analyze the spatial heterogeneity of taxi passenger flow and land use types within the location of urban rail transit stations.
3. The relationship between taxi passenger flow and land use type in urban rail transit stations is analyzed at different times of the day.

This paper includes the following sections: Section 2 provides a literature review on urban rail transit. Section 3 details the study area, experimental data, and data processing procedures. Section 4 introduces the AP clustering algorithm, K-means clustering algorithm, and relevant clustering validation models. Section 5 describes the clustering results and related research analysis. Finally, section 6 presents the conclusions.

2 LITERATURE REVIEW

Urban rail transit plays a significant role in urban development, facilitating transportation, driving economic growth in station areas, and promoting the improvement of land resource utilization types [17, 18]. There is a mutual influence between passenger flow and land use in urban rail transit station areas. Different land use types directly impact the volume of passenger sources, while the concentration and mobility of population activities drive the distribution of functional areas such as retail and life services around rail stations [19]. Therefore, understanding the distribution of passenger sources and land use types in urban rail transit station areas is crucial for the rational planning of rail transit routes, station locations, optimization of urban spatial layout, alleviation of urban traffic congestion, and promotion of urban sustainability. In recent years, scholars have focused on two main aspects in the study of urban rail transit using various data sources: the prediction and analysis of passenger flow, and the analysis of the relationship between urban rail transit and land use, including the characterization of surrounding land use patterns and the correlation between urban rail transit and land use. First, predicting and analyzing passenger flow in urban rail transit. With the rapid development of computer processing technology, a large amount of transportation data, spatial information data, and resident travel data have been explored, leading to an in depth analysis of passenger flow in urban rail transit by scholars. Firstly, scholars have made predictions on the passenger flow of urban rail transit by constructing different models and methods. Dong (2023) proposed the Temporal-Spatial Network Long Short-Term Memory (TNS-LSTM) model for short-term prediction of subway passenger flow at entry and exit stations, and designed a spatiotemporal network matrix using K-means clustering to forecast short-term passenger flow at multiple stations in the subway network [20]. To estimate the total passenger flow in urban rail transit, Shang (2012) constructed a linear regression model with multiple variables, predicting passenger flow in multiple cities and comparing it with the actual flow, demonstrating the accuracy of the model [21]. Wang (2023), based on the Transit-Oriented Development (TOD) model, classified urban rail transit stations using the hierarchical clustering algorithm and proposed a method for predicting passenger flow at entry and exit stations [22]. Next, scholars have also conducted research and analysis on the distribution of passenger sources and travel patterns in urban rail transit. Cao (2021) conducted a clustering analysis on the spatiotemporal distribution of subway passengers based on fare card data, thereby exploring passenger travel characteristics and patterns [17]. Chen (2021) used the K-means clustering method to establish Ordinary Least Squares (OLS) and Partial Least Squares (PLS) regression models to analyze the factors influencing the passenger flow of station-based bike sharing (SBBS) and free-floating bike sharing (FFBS) based on rail transit stations. The results revealed a substitution effect between the two bike-sharing modes in rail transit travel, with a higher frequency of bike-sharing usage in commuting trips [23]. In addition, as the development of urban rail transit puts pressure on land and housing prices, Delmelle (2020) examined the out-migration status of low-income residents in rail transit station neighborhoods and found that low income residents did not disproportionately leave areas with new transit investments [24]. Furthermore, scholars have employed various quantitative algorithms to conduct cluster analysis on the passenger flow of urban rail transit. Zhang (2023) developed a novel community detection algorithm based on nonnegative matrix tri-factorization for a higher granularity analysis of subway station passenger flow, enabling more flexible clustering analysis based on different passenger travel patterns [5]. Pang (2023) utilized smart card data to extract the dynamic characteristics of subway passenger flow and identified different patterns of subway station flow through hierarchical clustering and k-means clustering methods [25]. Ma (2013) combined the k-means++ clustering algorithm with a rough set theory to cluster and classify passenger travel patterns and regularities [26]. Jiao (2023) employed the autocorrelation function method to measure the temporal similarity of passenger flow time series between two independent stations, determining the spatiotemporal similarity between stations [27]. Second, the analysis of the relationship between urban rail transit and land use. Scholars used different data sources to infer urban land use types and explored the relationship between urban rail transit and land use types. Urban rail transit has a significant impact on real estate development [28]. The construction of rail transit stations and lines has different effects on different types of surrounding land use and drives the urban spatial layout [29]. Firstly, scholars analyzed the types of urban land use. Liu (2021) used the potential multi-perspective sub-spatial clustering method to infer the type of urban land use by using multi-source traffic data such as taxi data and bus smart card data [30, 31]. Gao (2020) calculated the utilization rate of land planning along urban rail transit based on GIS spatial clustering mining technology and association rules clustering algorithm [32]. Xia (2021) classified POI around urban rail transit stations based on the K-Means clustering algorithm [33]. Yuan (2012) proposed a framework (named DRoF) to explore the distribution of different functional areas in a city based on personnel mobility and points of interest (pol) [34]. Secondly, scholars have researched the correlation between urban rail transit and land use. Urban land use and transportation are two basic activities that form urban spatial structures, and they promote and influence each other. The larger the volume of urban rail transit, the stronger the urban cohesion, which can promote the development of land use along the line. Hou (2020) clustered the land use and population density around the rail transit station and explored the coordination relationship between rail transit and land use through the evaluation method of data envelopment analysis (DEA) [35]. Wang (2022) proposed an analytical framework combining spatial point pattern recognition technology and an OLS regression model to explore the dynamic relationship between population activities and functional facilities at different types of sites [36]. Based on the dynamic Time warping (DTW) affinity propagation (AP) algorithm based on distance, Liu (2020) found that urban functional landscape pattern has significant differences in passenger flow characteristics of different accessible station areas [31]. Pang (2023) used multinomial logistic regression analysis to investigate the relationship between passenger travel patterns of urban rail transit and environmental factors [25]. Deng (2015) studied the relationship between land development around different types of stations and the time distribution of passenger flow in and out of subway stations [4]. Previous studies on urban rail transit mainly focused on predicting and analyzing the passenger flow in and out of subway stations and passenger travel rules, as well as discussed the correlation between urban rail transit and land use. Despite the remarkable results of previous studies, relatively few studies have combined urban rail transit with other modes of transportation. Moreover, the spatial and temporal characteristics of residents' decentralized needs around urban rail transit stations and their correlation with land use types are not fully considered. The AP clustering algorithm of time series is used to cluster the taxi pick-up and drop off points in the location of urban rail transit stations in Beijing. And analyze the distribution of taxi passenger flow in different periods and geographical locations, which can reveal the change law of taxi demand around urban rail transit stations. Secondly, the K-means clustering algorithm is used to cluster the land use types around urban rail transit stations, and the stations with different functions are obtained. Finally, we explored the correlation between taxi passenger flows around urban rail transit stations and land use types and obtained the correlation between the two at different periods. This has a certain guiding significance for government departments to accurately plan urban rail transit lines and rationally plan urban spatial structure.

3 RESEARCH SETTINGS

3.1 Research Area

Beijing is the first city in mainland China to operate the subway, with a high density of population and rail transit lines. For example, by the end of 2021., the density of permanent residents in Beijing was 1323 people per square kilometer, and the length of urban rail transit reached 783 kilometers. To further combine rail transit with urban life, Beijing constantly promotes the process of "four-network integration" of rail transit and builds "urban life on rail". The research area of this paper mainly includes 12 administrative districts of Beijing (namely Dongcheng, Xicheng, Haidian, Chaoyang, Fengtai, Shijingshan, Changping, Fangshan, Tongzhou, Shunyi, Daxing, and Mentougou), with an area of 7664.5 km², accounting for 46.70% of the total area of Beijing. It is also the main area for population travel, as shown in Fig. 1.

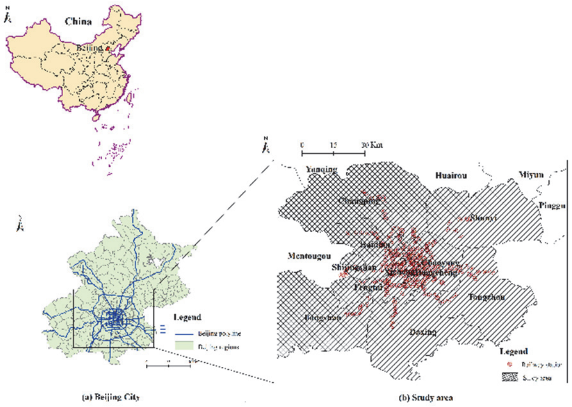


Figure 1 Beijing city and research area

3.2 Dataset Description

3.2.1 Urban POI Data

Point of Information (POI) data describes the basic information of functional areas of urban space and generally refers to point data in an internet electronic map, which can be used for urban spatial layout planning and urban functional land inference. The basic information of POI data includes four attributes: name, coordinate, category, and address. The detailed description is shown in Tab. 1. Through APIs (Application Programming Interface) provided by the Gaode platform, a total of 1152891 POIs open source data of Beijing are obtained by data mining technology. Gaode classifies POI data into three levels. There are 23 first-level categories, mainly covering catering service, shopping service, life service, leisure service, and so on.

Table 1 POI data field and description

POI	Longitude	Latitude	Functional zone	Administrative region
The authentic beef noodle from Anhui province	116.417259	40.088123	Catering service	Changping
(Zebra) Beijing Qingyun Hubang Network Technology Co., LTD	116.434090	39.959138	Companies	Chaoyang
Kyushu University Beijing Office	116.324959	39.964705	educational services	Haidian

3.2.2 Taxi GPS Trajectory Data

To a certain extent, taxi GPS track data can effectively reveal the travel behavior and traffic demand of residents. To better reflect the travel demand of residents, we obtained taxi track data of Beijing from the Didi platform from December 22 to 24, 2016, including taxi data (91986 orders in total) and ride-hailing data (473596 orders in total), and finally obtained a total of 506940 orders with valid data after data cleaning and processing. Each order is composed of a series of trajectory datasets, that is, the coordinates of longitude, latitude, and the instantaneous speed of the taxi on the corresponding timestamp in every 6 seconds. Specific data fields and descriptions are shown in Tab. 2.

Table 2 Taxi trajectory data field and description

Data field	Description
ID	3433d3171dc1df2365ae4280830ef17
Time	2016/12/22 18:27:23
Longitude	116.36145
Latitude	39.95524
Speed	8.4

3.2.3 Urban Rail Transit Data

During the formation of the urban rail transit network, the RTS, as the basic node of the rail transit network, directly affects the overall operating efficiency of the rail transit network and the degree of agglomeration of social and economic elements. According to the data from the Beijing metro operation company, Beijing had 24 subway lines by the end of 2021. 348 urban rail transit station data of Beijing are obtained through data mining technology. The rail transit data fields mainly include subway line, station name, station longitude, and latitude, as shown in Tab. 3.

Table 3 Rail transit data field and description

Rail transit line	Station name	Longitude	Latitude
Line 1	Babaoshan	116.242277	39.913185
Line 2	Xizhimen	116.362125	39.946021
Line 3	Anheqiaobei	116.276432	40.018657

3.3 Data Processing

Data processing mainly includes data preprocessing and the research framework. The data preprocessing contains the cleaning of taxi GPS trajectory data, the division of land-use type, and the clustering analysis of urban rail stations. Fig. 2 shows the framework of the research.

Step 1. Clean the taxi GPS trajectory data.

Due to some unreasonable and abnormal data in the initial dataset, the taxi GPS trajectory data needs to be cleaned. Therefore, we delete data outside the study area (taxi data not belonging to Beijing), remove data with abnormal taxi running statuses (non-0 and non-1) and eliminate data with an instantaneous speed greater than 1000 km/h.

Step 2. Divide urban POI data into land-use types.

POI data includes location name, specific address, longitude, latitude, etc. Each POI is categorized into different land-use types. Based on the classification of POI data by Gaode, we divide POI data into specific functional zones. We utilize the Gaode POI classification standard and keywords in the Beijing POI data to categorize the POI data into 23 functional land categories. Specific content can be found on the website: https://lbs.amap.com/api/webservice/download.

Step 3. Cluster analysis of urban RTS.

Scholars have considered a variety of connection methods and divided the attraction range of urban RTS into reasonable and irrational attraction ranges [37]. Regarding the attraction range of sites, scholars believed that the range should be between 400 m and 800 m according to the actual situation, and 800 m is widely used at present [33, 38, 39]. Firstly, the number of passenger and land types around urban rail stations is counted. We take the area within 800 m around the RTS as the space buffer of the RTS. Through cluster analysis, we analyze the number of passengers at taxi pick-up and drop-off points and the number of each land-use type within the 800 m coverage range of RTS, and thus obtain RTS with different attributes. Secondly, we count the number of RTS in different administrative districts and rail transit lines. We conduct the reverse address analysis by using the open platform of Gaode location-based services to obtain the detailed geographical location of each rail station, and make statistics on the administrative regions and rail transit lines where the RTS is located.

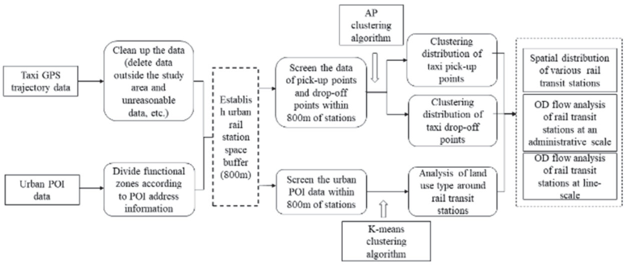

Taxi GPS trajectory data
Urban POI data
Clean up the data (delete data outside the study area and unreasonable data, etc.)
Divide functional zones according to POI address information
Establish urban rail station space buffer (800m)
Screen the data of pick-up points and drop-off points within 800m of stations
Screen the urban POI data within 800m of stations
AP clustering algorithm
K-means clustering algorithm
Clustering distribution of taxi pick-up points
Clustering distribution of taxi drop-off points
Analysis of land use type around rail transit stations
Spatial distribution of various rail transit stations
OD flow analysis of rail transit stations at an administrative scale
OD flow analysis of rail transit stations at line-scale

Figure 2 Framework of the research process

4 METHODOLOGY

4.1 AP Clustering Algorithm

Affinity Propagation (AP), also known as neighbor propagation or affinity propagation, was proposed by Brendan J. Frey and Delbert Dueck. The AP clustering algorithm is based on the "information transfer" between data points, and the similarity between data points is input. In this section, we want to categorise the passenger flow to different stations type based on the Affinity Propagation (AP). We regard all data points as the underlying centers of clustering (named exemplars) and construct a network (similarity matrix) by the connected data points. Then, we calculate the clustering centers for each sample based on the responsibility and availability messaging of each edge in the network. The distance function of static multi-attribute sequence data regards each attribute as an independent value, and the attribute values at different moments in the time series are correlated, so the similarity calculation method of time series needs to be defined. Calculating the similarity between time series can be complex due to their dynamic characteristics. Taxi GPS track data is a data set that consists of time series. We take the center of the urban rail transit station as the impact zone within the 800m area of the station. The AP cluster analysis of time series on the data points of taxi passengers can make us clearly understand the changing rules of passenger flow around the station in different periods. In this study, we employ the improved CORT method, which combines distance measurement between time series and the correlation of the time series themselves [40, 41]. The CORT method not only considers the shape similarity between time series but also measures the similarity between their wave features by using the first-order time correlation coefficient.

Step 1. Calculate the similarity between passenger flow. Assuming that the passenger flow of station i and station j can be regarded as t two-time series, T_i and T_j, respectively, and the $T_{i,t}$ refers to the passenger flow volume of station i in time t. Then, the similarity of passenger flow between station i and station j can be measured based on the Eq. (1) [40]:

$$d_{CORT}\left(T_i,T_j\right)=\frac{2\times d\left(T_i,T_j\right)}{1+\exp p\times\left(\frac{\sum_{t=1}^{24}\left(T_{i,t+1}-T_{j,t+1}\right)\times\left(T_{i,t+1}-T_{j,t}\right)}{\sqrt{\sum_{t=1}^{24}\left(T_{i,t+1}-T_{i,t}\right)^2}\times\sqrt{\sum_{t=1}^{24}\left(T_{j,t+1}-T_{j,t}\right)^2}}\right)} \tag{1}$$

where the parameter p is set to 2, and the Euclidean distance $d\left(T_i, T_j\right)$ represents the dynamic time-warping distance DTW between the passenger flow of station i and station j.

Step 2. Calculate AP clustering results. Here, our data is represented as $G=\left\{q_1, q_2, ..., q_n\right\}$, where q_n refers to the passenger flow volume of station N containing 24 hours. The similarity of N stations is calculated in pairs, and these similarities between station i and station j is represented as $S\left(i,j\right)$, which is shown in Eq. (2):

$$S\left(i,j\right)=\begin{cases}d_{CORT}\left(T_i,\ T_j\right),\ i\neq j\\ p,\ i=j\end{cases}\quad i,j\in\left[1,\ 2,\ ...,\ n\right] \tag{2}$$

Here, $S\left(i,j\right)$ is the similarity between g_i and g_j station i and station j. Negative CORT distance is generally adopted. Therefore, the greater $S\left(i,j\right)$ is, the closer the two stations are, the higher the similarity will be. p means the "preference", usually set to the minimum or median value of $S\left(i,j\right)$ similarity. A greater "preference" represents a stronger ability for stations to turn into clustering centers, that is, a larger number of clustering centers. For getting the optimal solution, we employ the iterative algorithm from the minimum value to the lower quartile to select p Firstly, measure Responsibility. Responsibility refers to the level at which a station q_k is fit for the clustering center for a data point q_i, denoted by $r\left(i,\ k\right)$. As exhibited in Fig. 3a, the red arrow represents that station i sends information to station j, that is, a course that point q_i selects station k. Exemplar refers to the clustering center. The iteration formula of attractiveness is as follows:

$$R_{t+1}\left(i,\ k\right)=\left(1-\lambda\right)\times R_{t+1}\left(i,\ k\right)+\lambda\times R_t\left(i,\ k\right) \tag{3}$$

Among,

$$R_{t+1}\left(i,\ k\right)=\begin{cases}S\left(i,\ k\right)-\max\limits_{j\neq k}\left\{A_t\left(i,j\right)+R_t\left(i,j\right)\right\},\ i\neq k\\ p\left(k\right)-\max\limits_{j\neq k}\left\{S\left(i,j\right)\right\},\ i=k\end{cases} \tag{4}$$

Here, $R_{t+1}\left(i,\ k\right)$ is the updated $R\left(i,\ k\right), R_t\left(i,\ k\right)$, is the previous one $R\left(i,k\right)$. λ refers to the damping coefficient, and its value range is [0.5, 1], which is used for the convergence of the algorithm. Secondly, measure Availability. Availability refers to the suitability that a station i selects the station k as the clustering center of it, denoted as $a\left(i,\ k\right)$. As exhibited in Fig. 3b, the red arrow shows that station k is sending information to the station i, that is, an approach where the station k selects the station i. The availability iteration formula is shown in Eq. (5) and Eq. (6).

$$A_{t+1}\left(i,\ k\right)=\left(1-\lambda\right)\times A_{t+1}\left(i,\ k\right)+\lambda\times A_t\left(i,\ k\right) \tag{5}$$

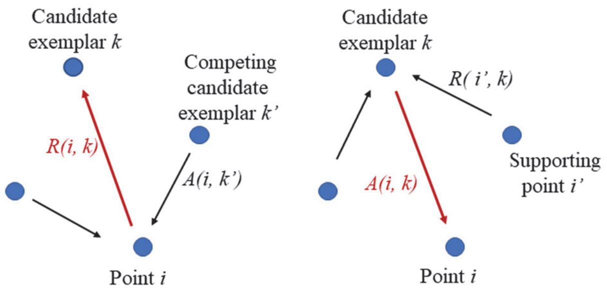
Here,

$$A_{t+1}(i,k)=\begin{cases}\min\left\{0, R_{t+1}(k,k)+\sum\limits_{j\in\{i,k\}}\max\{0, R_{t+1}(j,k)\}\right\}, i\neq k\\ \sum\limits_{j\neq k}\max\{0, R_{t+1}(j,k)\}, i=k\end{cases} \tag{6}$$

where, $A_{t+1}(i,k)$ is the updated one $A(i,k)$, and $A_t(i,k)$ is the previous one $A(i,k)$. λ refers to the damping coefficient, with a value of [0.5, 1), for the convergence of the algorithm. The larger the sum value of $R(i,k)$ and $A(i,k)$ is, the more likely station k is to be the cluster center and the more likely station i is to belong to the cluster with the point q_k as the cluster center. Thirdly, generate high-quality exemplars. Minimizing clustering energy $E(C)$ by iterating until high-quality clustering energy $C=\{o_1, o_2, ..., o_c\}$ is produced. Here C is the number of clusters and the expression of $E(C)$ is as follows:

$$E(C)=-\sum_{i=1}^{N}S(i,o_j) \tag{7}$$

where,o_j refers to the element and of the cluster C_j, $i\in c_j$, and $j\in[1, 2, ..., C]$.



(a) Responsibility (b) Availability

Figure 3 Responsibility and availability process

4.2 K-Means Clustering Algorithm

The K-means algorithm uses the method of finding the extreme value of the function to obtain the iterative operation, mainly by continuously taking the nearest mean value from the seed point. Given a set of data points and an initial set of cluster centers, the algorithm classifies each data point to the nearest class cluster firstly by calculating the distance between each data point and cluster center. Then, the center of the cluster is updated. Second, the allocation of each data point and the updating of the cluster center point is iterated until a condition that the small variation of the cluster center point or the stipulated number of iterations is reached. Our data includes n stations $Z=\{Z_1, Z_2, ..., Z_n\}$, each of which has attributes in m dimensions, that is, the POI counts under different land use types. The purpose of the K-means algorithm is to assign n stations into h class clusters in terms of the distance between objects. Note that each object only is assigned to one class cluster whose distance from the class cluster center to the object is the smallest. According to K-means, h clustering centers $\{P_1, P_2, ..., P_h\}$ need to be initialized first, $1<h\leq n$, and the Euclidean distance from each cluster center to each station is calculated, shown in Eq. (8).

$$dis(Z_i, P_j)=\sqrt{\sum_{t=1}^{m}(Z_{it}-P_{jt})^2} \tag{8}$$

Where Z_i refers to the object $i, 1\leq i\leq n, P_j$ denotes the cluster center $j, 1\leq j\leq h, Z_{it}$ refers to the property t of the object $i, 1\leq t\leq n, P_{jt}$ refers to the attribute t of cluster centers j. The distance between each object and each cluster center is compared by turns, and the objects are allocated to the cluster closest to the cluster center. Then h clusters $\{Q_1, Q_2, ..., Q_h\}$ are obtained. The prototype of the class cluster is defined based on the k-means algorithm, which quantizes the mean value of all stations in the class cluster for each dimension., and its calculation formula is as follows:

$$P_l=\frac{\sum\limits_{Z_i\in Q_l} Z_i}{|Q_l|} \tag{9}$$

where, P_l refers to the center of the cluster $l, 1\leq l\leq h, |Q_l|$ refers to the number of stations in the class cluster l, and Z refers to the object i in the class cluster l.

4.3 Evaluation of Clustering

Silhouette Coefficient, an evaluation method of clustering effect, combines two factors, namely cohesion and separation. Based on the same original data, the coefficient is adopted to assess the impact of different clustering algorithms or their operating modes on clustering effect results. The calculation formula of the coefficient $s(i)$ is as follows:

$$s(i)=\frac{b(i)-a(i)}{\max\{a(i), b(i)\}} \tag{10}$$

The value of the coefficient $s(i)$ ranges between [−1, 1], with a value closer to 1 indicating that there is a superior cohesion and separation level relatively. Averaging the contour coefficients of all points are the total contour coefficients of the clustering results. $a(i)$ refers to the average distance between the sample i and other samples in the cluster (red lines), and $b(i)$ refers to the minimum distance between the sample i and all samples of other clusters (blue lines).

Fig. 4 shows the contour coefficient of the K-means clustering algorithm. We find that the silhouette coefficient is highest when the number of clustering is 4. Therefore, the clustering effect will be optimal when the number of the clustering is equal to 4. Therefore, in the following clustering results, we divide the land use types around urban RTS into four types of functional RTS clusters.

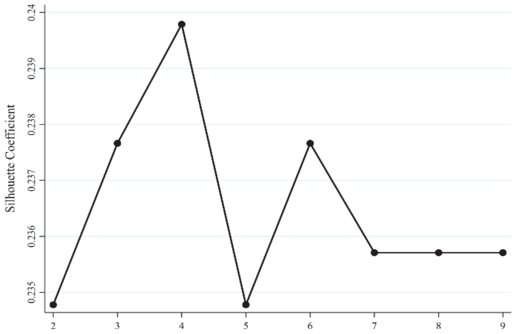


Figure 4 The silhouette coefficient

5 Results and Discussions

5.1 Overview

The hotspot analysis is conducted on taxi passenger flow within 800 m around urban RTS, and ArcGIS software is used for data visualization, as shown in Fig. 5. Since the status of taxi passenger flow can directly reflect the travel needs of residents, analyzing the distribution of taxi pick-up and drop-off points around urban rail stations, we can further understand the imbalance between large scale transportation supply and decentralized travel demand. This has a certain guiding significance for the spatial layout of urban public transport stations, alleviating the passenger source pressure of urban rail stations and sharing the urban traffic pressure. On the whole, the passenger flow around the RTS in the central districts of Beijing (including Dongcheng district, Xicheng district, Haidian district, Chaoyang district, Shijingshan district, and Fengtai district) presents an agglomeration pattern, while the density of the outer suburbs is lower. This indicates that there is obvious spatial heterogeneity in the decentralized travel demand of urban residents. Firstly, there are significantly more RTS in Beijing's downtown area than in its outer suburbs. Secondly, the permanent population of the sixth district is about 1.01 times that of the outskirts. Chaoyang district has a large number of CBD districts, while Haidian district is dominated by educational institutions and research units. Most residents go out to work, so their demand for public transportation is high.

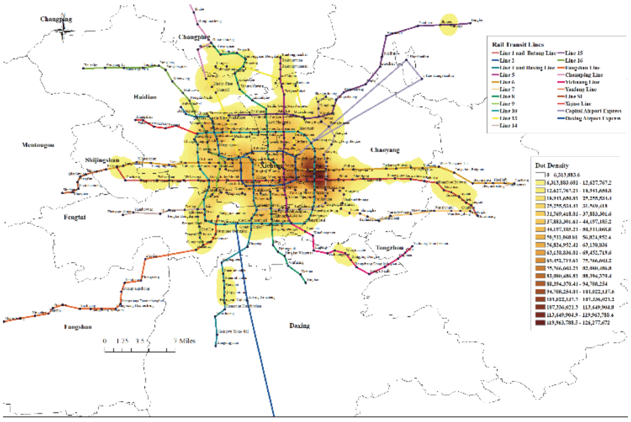


Figure 5 The hot spot distribution of passengers around RTS

From the perspective of the distribution of taxi passengers around RTS, the hotspots of taxi passengers are mainly distributed at the junction of urban rail transit lines, transportation hubs, commercial areas, tourist areas, and so on. For example, Hujialou station is the intersection of subway line 6 and line 10, and the majority of hotspots for taxi pickups and drop-offs are concentrated in this area, indicating that the Hujialou station carries a large passenger flow. Because the urban rail transit line handover is an important transfer point in passenger travel routes, passengers have a large demand for transfer stations, so the passenger flow in this area is relatively large.

5.2 Spatio-Temporal Distribution of Passengers Around RTS

According to the AP clustering algorithm, the clustering results of taxi pick-up and drop-off passengers around RTS are obtained, as shown in Fig. 6 to Fig. 9. In this way, we can clearly understand the spatiotemporal distribution of passengers around different RTS. The pick-up and drop-off passenger volume of taxis around RTS reflect the passenger flow in and out of RTS to some extent. This has a certain guiding significance for optimizing urban rail lines and urban public transport. We use O-Cluster and D-Cluster respectively to represent the clustering results of taxi pick-up passengers and drop-off passengers around RTS, in which O-Cluster contains 5 kinds of clustering and D-Cluster contains 6 kinds of clustering.

5.2.1 Clustering Characteristics of Taxi Pick-Up Points Around RTS

Fig. 6 depicts the spatial distribution of clustering results of taxi pick-up passengers. It can be found that various RTS are mainly concentrated in Chaoyang district, Haidian district, and Fengtai district, while Shijingshan district, Shunyi district, and Mentougou district have fewer RTS. The RTS belonging to O-Cluster 2 and O-Cluster 3 have a relatively wide distribution range, while the RTS in other clusters have a relatively small distribution, indicating that the passenger characteristics around most of the RTS are consistent with O-Cluster2 and O-Cluster3. The spatial distribution of O-Cluster3 is more centralized in the central urban area, suggesting that the geographical location of the RTS affects the passenger flow.

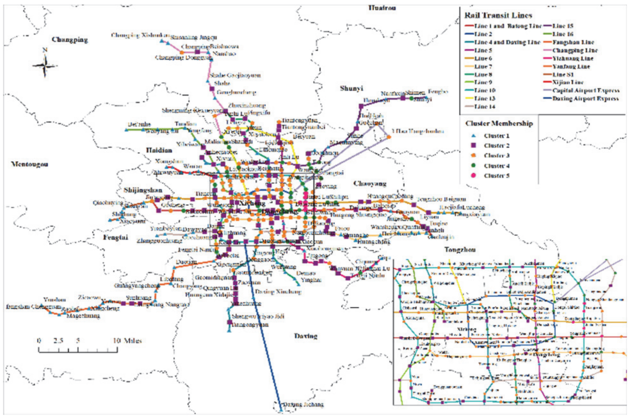


Figure 6 The spatial clustering distribution of taxi pick-up passengers

From the trend of passengers over time, the different clustering results have the same change trend in different periods.The passenger flow is relatively low from 00:00 to 6:00. From 7:00 to 10:00, the passenger flow grows faster and reaches the morning peak. Then, the passenger flow around the RTS is in a stable growth state, and passenger flow shows a downward trend during 21:00 - 0:00. Several factors contribute to these trends. Firstly, residents are mostly in a resting state during the period of 00:00 - 6:00 and 21:00 - 0:00, and urban rail transit is in a state of suspended operation, so the passenger flow around RTS is relatively low. Secondly, from 7:00 to 10:00 and 10:00 to 21:00, residents are actively engaged in work and travel, leading to a higher demand for urban rail transit. Consequently, the number of passengers around the RTS is relatively high during these periods. Thirdly, according to a report by the Beijing Intelligent Transportation Association in 2022, the daily passenger volume of the urban rail transit network has been increasing, and the proportion of public transport trips has increased from 32.3 percent to 57.4 percent. It has become the first choice for commuters due to its convenience, punctuality and speed.

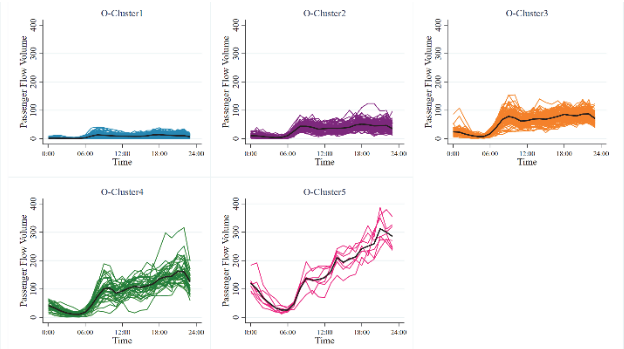


Figure 7 The statistical clustering characteristics of taxi pick-up passengers

In addition, the passenger volume of different clustering results is different in quantity and period. Specifically, the passenger volume of O-Cluster 1 and O-Cluster 2 is significantly lower than that of O-Cluster 3, O-Cluster 4, and O-Cluster 5. However, RTS belonging to O-Cluster 4 and O-Cluster 5 are at a relatively low level, among which RTS belonging to O-Cluster 5 are the lowest, accounting for only 2.01% of all stations, and only distributed on three rail transit lines, namely line 6, line 10 and line 14.This indicates that the RTS belonging to O-Cluster 5 is surrounded by dense passenger sources, with high travel intensity of residents and high pressure of passenger transportation. Similarly, the passenger flow around the RTS in O-Cluster 1 and O-Cluster 2 is relatively low. However, the rail transit lines and stations belonging to O-Cluster 1 and O-Cluster 2 are relatively large, so the passenger pressure of these RTS is relatively small.

5.2.2 Clustering Characteristics of Taxi Drop-Off Points Around RTS

We obtain the clustering result of passenger flow at taxi drop-off points around RTS, as shown in Fig. 8. The passenger of taxi drop-off points around RTS reflects residents' demand for rail transit to a certain extent, which is conducive to the location setting of the RTS and the adjustment of subway departure frequency.

On the whole, RTS with various clustering results is mainly distributed in Chaoyang and Haidian districts, accounting for 23.56% and 18.39% of all RTS. However, Shijingshan district, Shunyi district, and Mentougou district respectively have the fewest RTS, accounting for less than 10% of the total. D-cluster 1 has the largest number of RTS, which are mainly distributed in the outer suburbs. The number of RTS belonging to D-Cluster 2, 3 and 4 is almost uniform and relatively evenly distributed in the downtown area. The number of D-Cluster 5 and 6 is the least, mainly distributed in the Chaoyang district. This indicates that different attributes of RTS have obvious spatial heterogeneity. From the clustering results of taxi drop-off points around RTS, the variation trend of passenger flow is consistent in different periods, but D-Cluster 1, 2, and 3 show significantly lower passenger flows compared to D-Cluster 4, 5, and 6. In particular, the passenger volume is at a relatively low level during 0:00 - 6:00, reaches the morning peak from 7:00 to 10:00, and then remains at a relatively high level. At 17:00 - 19:00, the evening peak appears, and the passenger flow volume around the RTS decreases after 19:00. This is because the RTS belonging to D-Cluster 1, 2, and 3 are mainly distributed on the rail transit lines leading to the outer suburbs. For example, RTS belonging to D-Cluster 1 are mainly distributed on Fangshan Line (accounting for 10.64%). The economic level of the outer suburbs is not developed, the population density is small, and residents' transport demand is relatively small.

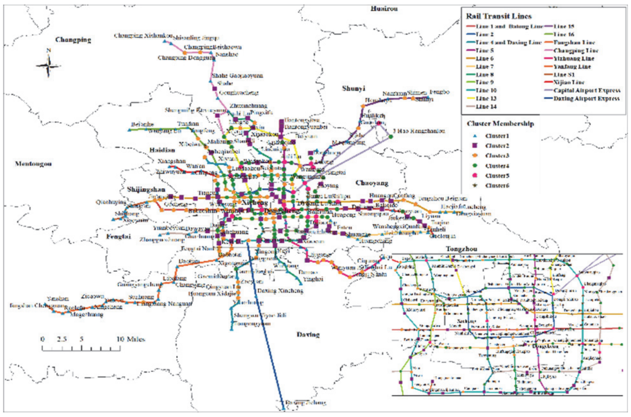


Figure 8 The spatial clustering distribution of taxi drop-off passengers

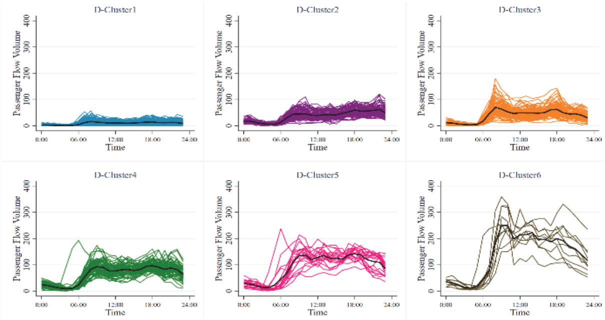


Figure 9 The statistical clustering characteristics of taxi drop-off passengers

5.3 Analysis of Land-Use Type Around RTS

POI data can effectively reflect urban land use type and land utilization rate, and the land type around RTS affects the number and distribution of passengers. By clustering POI data around the RTS, we can obtain the following four categories: commercial transit stations, residential transit stations, comprehensive transit stations, and industrial transit stations. The specific classification criteria of different functional stations are shown in Tab. 4, including the land types of different functional RTS, the representative RTS, and the number of RTS. Appendix Fig. A.1 is the description of the different functional RTS clustering. According to the land utilization situation around the RTS,we divide the station, explore the land development around the RTS, and put forward the optimization strategy.

According to the experimental results in Tab. 4, we find that the number of industrial transit stations is the largest, accounting for 39.37% of all stations, while the number of commercial transit stations is the least at 10.34%. Commercial transit stations, including 26 RTS, such as Hujialou station, are mainly surrounded by functional areas, such as shopping and dining services. Residential transit stations include 70 stations such as Suzhou Street station, which are mainly distributed in residential areas, specific place names, and other functional areas of residential types. Comprehensive and industrial transit stations contain the most RTS, which mainly include Anzhenmen station and Bagou station, accounting for 69.54% in total. The distribution of functional areas around these RTS is relatively balanced, and the land use type is relatively perfect, including education, recreation, and transportation facilities. In terms of the spatial distribution of functional RTS, commercial, residential, and comprehensive transit stations are mainly distributed in Dongcheng district, Xicheng district, Chaoyang district, Haidian district, and Fengtai district. The distribution of different types of RTS is relatively balanced. However, industrial transit stations are located in the outer suburbs, such as Tongzhou, Fangshan, and Changping districts. This can be attributed to two main factors. Firstly, the infrastructure in the central city of Beijing is relatively perfect, the distribution of the RTS is balanced and there are a large number of them, and the land utilization is full with a large number of residential and commercial zones. Therefore, the distribution of all kinds of cluster stations is relatively uniform, indicating that the land use types around RTS in the six districts are balanced to some extent. Secondly, some manufacturing and transportation hub facilities are mainly distributed in the outer suburbs due to the limited land availability in the downtown area of Beijing. Therefore, the land use types around the RTS in the outer suburbs are mainly traffic facilities, automobile sales, and other functional areas.

Table 4 Characteristics and description of clustering of urban RTS

Category	Particulars	Representative station	Number of stations
Commercial transit stations	Mostly for shopping services, life services and catering services and so on	Hujialou, Dazhong Temple, Jiulongshan, Shunyi, Dongdan, Chongwenmen, Qingnan Road, Dajiao Pavilion, Dahongmen South, Changping	36
Residential transit stations	Mostly for place names and address information, companies and enterprises (dormitories and apartments)	Suzhou Street, Dongzhimen, Pingleyuan, Guanzhuang, Suzhou Bridge, Fuxingmen, Yonghegong, Xuanwumen, Majiapu, Heping West Bridge, Shilibao, Wanzi, Anhua Bridge, Keyi Road, Wanyuan Street, Jiukeshu	70
Comprehensive transit stations	Education, recreation and other functional areas are evenly distributed	Anzhen Gate, Guangximen Gate, Puhuangyu Gate, Shimen Gate, Wanshou Temple, Sihui East, Gulou Dajie, Qingyuan Road, Tiantongyuan, Caofang, Hufang Bridge, Shichahai, Liuli Bridge East, Old Palace, Shuangqiao, East Guantou South, Life Science Park, Fangshan Chengguan	105
Industrial transit stations	The number of functional areas such as automobile sales and traffic facilities is small, but the distribution is more balanced	Bagou, Shangezhuang, South Faxin, North Anhe, Tian'anmen East, Tiangong Courtyard, East Gate of the Temple of Heaven, Lucheng, Shuanghe, Lincuiqiao, Wufutang, Liyuan Village, Secondary Canal, Baliqiao, Changyang, Beishaowa, Magezhuang, Tea Tent, Daxing Airport, Terminal 3	137

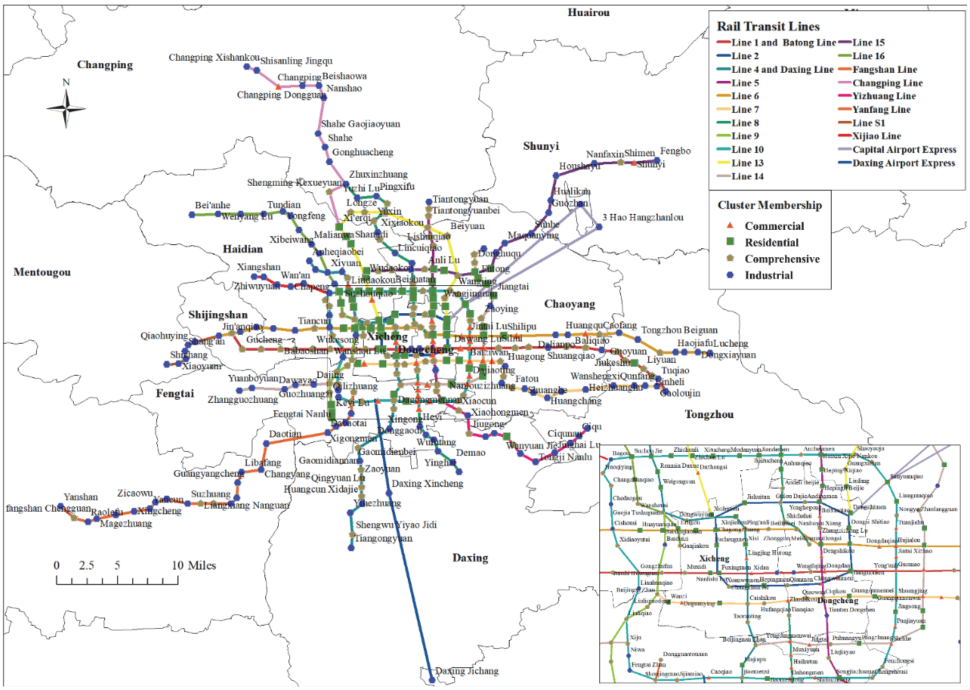
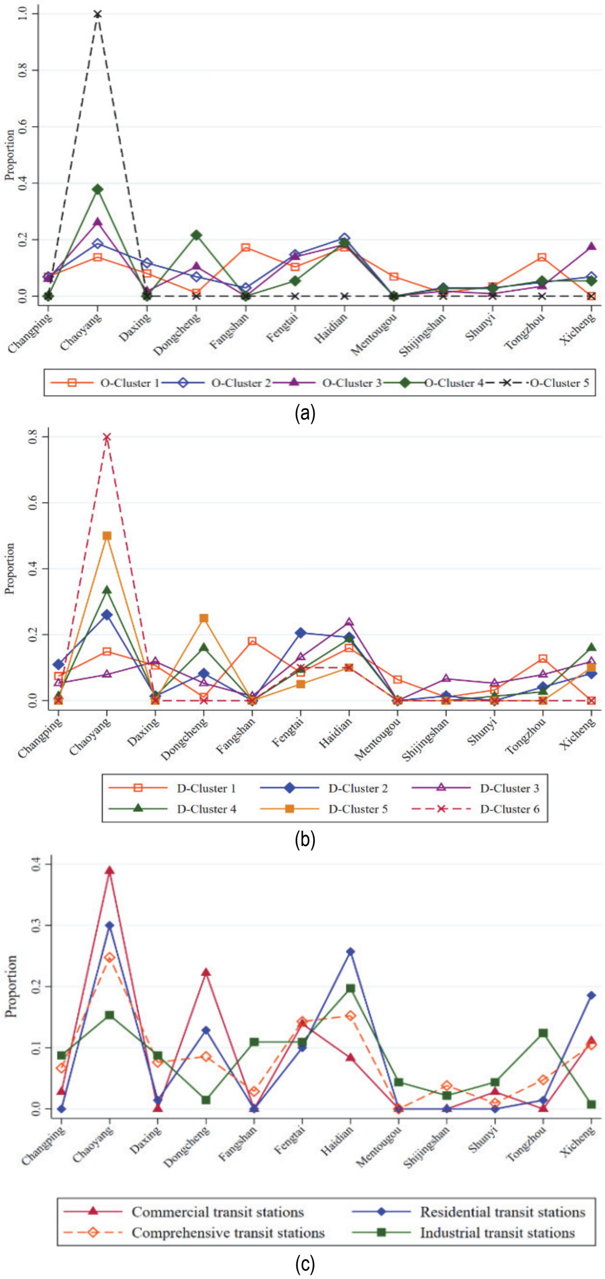


Figure 10 The spatial clustering distribution of the functional land-use

5.4 Correlation Analysis of Land Use Type and Passengers Around RTS

The taxi passenger volume around the RTS reflects the residents' commuting demand for rail transit to some extent, while the land use type around the RTS provides the refinement and a true reflection of human social and economic activities. This study discusses the distribution of land use characteristics and passenger volume for different types of stations in different administrative regions and rail lines. It aims to provide recommendations for optimizing land use around rail stations and managing passenger evacuation volume. In addition, to further explore the correlation between passenger and land use types around RTS, this study uses partial least squares regression [42] to calculate the correlation between passenger and land use types at different periods, and the calculation results are shown in Fig. A.2 and Fig. A.3. From the figures, we can find that there is a positive correlation between passenger flow around urban rail transit stations and land use types such as transportation service, public facility and daily life service. However, there was a negative correlation among the types of land use such as tourist attraction, science/culture education service and shopping. This indicates that functional land use types around urban rail transit stations are closely related to residents' travel needs, and further illustrates the importance of TOD development model, which has an important impact on urban development planning.

Step 1: Calculate the proportion of RTS with different attributes (passenger and land-use type) in each administrative district. Assuming that the proportion of the number of RTS with different passenger flow clusters in different administrative regions to the total number of corresponding cluster stations is denoted as P_i^O and P_i^D, respectively (where P_i^O represents the cluster proportion of taxi pick-up points, and P_i^D represents the cluster proportion of taxi drop-off points). The proportion of the number of functional cluster stations in different administrative regions to the total number of corresponding cluster stations is P_i^Z. Here, $P_i^O=\frac{A_i}{M_i}, P_i^D=\frac{B_i}{N_i}; P_i^Z=\frac{c_i}{Qi}\left(i=1, 2, ..., 12\right)$. A_i and B_i respectively represent the number of different clustering stations of taxi pick-up and drop-off points in each administrative region, and C_i represents the number of different functional stations in each administrative region. M_i, N_i, Q_i respectively represent the total number of RTS with different clustering attributes in each administrative region. Fig. 11 describes the distribution of RTS with different clustering attributes in different administrative districts. On the whole, the passenger flow volume around the RTS is decreasing from the central urban area to the outer suburban area. The RTS belonging to the passenger flow clusters 1, 2 and 3 stay a relatively low passenger flow. However, the O-Cluster 4 and 5 and the D-Cluster 4, 5 and 6, represent the stations with high passenger flow, and these RST are mainly distributed in Chaoyang, Dongcheng and Haidian. At the same time, we can see from Fig. 11c that Chaoyang, Dongcheng and Haidian have more commercial and residential transit stations. According to the calculation, the correlation between the RTS belonging to D-Cluster 4, 5 and 6 and the commercial and residential transit stations is above 0.6, suggesting that the commercial and residential land distribution around the stations will increase the passenger flow to some extent. Therefore, the construction of RTS should be appropriately increased in areas with more commercial and residential land.



(a)

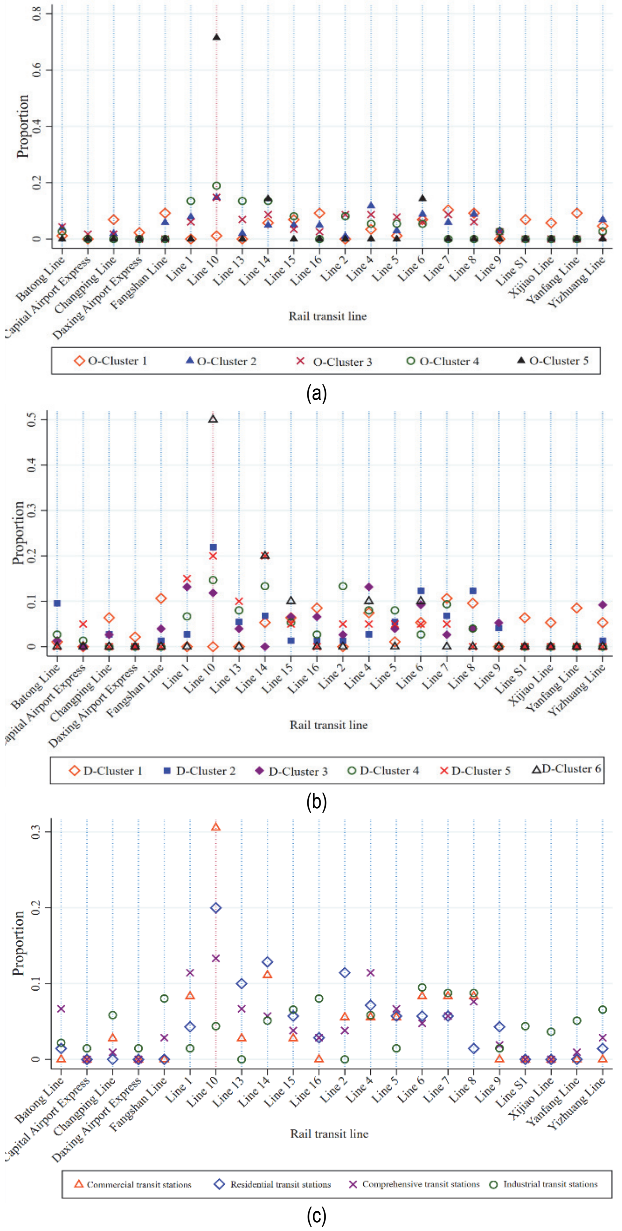
(b)

(c)

Figure 11 The distribution characteristic of passengers and land-use in different administrative districts

Step 2: Calculate the proportion of RTS with different attributes (passenger and land-use) on different rail transit lines. Similarly, the proportion of the number of RTS with different passenger flow clusters on different rail lines to the total number of corresponding cluster stations is P_j^O and P_j^D respectively (where P_j^O represents the cluster proportion of taxi pick-up points, and P_j^D represents the cluster proportion of taxi drop-off points).The proportion of the number of different functional clustering RTS on different rail lines to the total number of corresponding clustering stations is P_j^Z .

Here, $P_j^O = \frac{A_j}{M_j}, P_j^D = \frac{B_j}{N_j}, P_j^Z = \frac{c_j}{Q_j} = (j = 1, 2, ..., 24)$. A_j and B_j respectively represent the number of RTS in different clusters of the taxi pick-up and drop-off points of each rail line; C_j represents the number of different functional RTS on each rail line; M_j, N_j, Q_j respectively represent the total number of RTS in different clusters on each rail line. Fig. 12 shows the proportion of passengers and land use on different rail transit lines.



(a)

(b)

(c)

Figure 12 The distribution of passenger and land use in different rail transit lines

Firstly, the RTS belonging to O-Cluster 1 are mainly in line 7 and line 8, accounting for 19.54% (see Fig. 12a). These stations are mainly distributed outside the sixth ring road of Beijing. The RTS of O-Cluster 2 is mainly distributed on line 10 and line 4 which are far away from the downtown area, accounting for 26.47% and 23.48%, respectively. Therefore, taxi passenger flow around RTS belonging to O-Cluster 1 and O-Cluster 2 is relatively low. RTS belonging to O-Cluster 3, 4, and 5 mainly distribute the same lines, all of which are line 10, and the passenger flow is relatively high. Secondly, the RTS belonging to D-Cluster 4, 5, and 6 are mainly distributed in metro line 10 and line 14 (see Fig. 12b). Because metro line 10 is a circle line, and Line 14 is an anti-L-shaped backbone line, both of which have more interchange stations and relatively large passenger flow around the stations. To some extent, it indicates that areas with more passengers tend to have fewer RTS, while areas with more RTS have relatively fewer passengers. In addition, D-Cluster 1 has the largest number of RTS, accounting for 27.01% of all RTS, but it has the least passenger flow. Therefore, increasing the construction of loop subway lines and parallel lines can help alleviate traffic congestion caused by high passenger flow in the city center. Finally, in terms of the distribution of functional clustering on rail transit lines, industrial transit stations are the most widely distributed in each line, while commercial transit stations are relatively less distributed in rail transit lines. From Fig. 12c, we can find that commercial, residential, and comprehensive transit stations are mainly distributed in line 10. Among them, commercial transit stations account for the highest proportion of 30.56%, and they are least distributed in line 13, line 15, and Changping Line, accounting for only 2.78%. Commercial rail stations bear the higher passenger pressure, indicating that it needs to increase the number of RTS and parallel lines in commercial areas. This shows that the RTS in commercial areas has greater pressure on passenger transport, which can guide commercial station planning. In addition, the distribution of industrial transit stations on each line is relatively uniform, accounting for about 4.55%, which, due to the land-use type around industrial transit stations, is relatively balanced.

6 CONCLUSIONS

RTS plays a crucial role as nodes connecting different urban areas [4]. The spatial layout and optimization of RTS have a certain impact on promoting urban economic development and alleviating traffic congestion. As a transportation hub, the external effect of RTS cannot be ignored. The location of the RTS has a direct impact on human social activities and the diversity of land use types, thus affecting the compactness and sustainability of the city. By analyzing the distribution of passengers and land-use types around RTS in time and space, a series of conclusions are obtained as follows: Firstly, the passenger flows present different agglomeration states, with obvious spatial heterogeneity. From the spatial distribution, there are obvious differences between administrative districts and urban rail lines where different clustering results are located. It is mainly distributed in Chaoyang district, Haidian district, and Fengtai district, with the largest number of RTS belonging to Cluster 3, accounting for 33.05% of all stations. From the time distribution, passenger flow with different clustering results has the same change trend in different periods, but the drop-off point has an obvious evening peak compared with the pick-up point. Secondly, the spatial distribution of land-use types around RTS indicates that commercial, residential, and comprehensive transit stations are primarily concentrated in the Dongcheng district, Xicheng district, Chaoyang district, Haidian district, and Fengtai district, with relatively balanced distribution. From the clustering distribution of RTS, there are obvious differences in land-use types around different station types, and industrial transit stations are the most widely distributed on each line. However, the distribution of commercial transit stations on rail lines is relatively small. Among these, industrial transit stations account for the largest number, accounting for 39.37% of all stations, while commercial stations only account for 10.34% at least. Thirdly, the RTS with higher passenger flow volumes are mainly distributed in ring lines and downtown districts, such as metro line 10, Chaoyang district, Dongcheng district and Haidian district. In addition, these RTS are mainly surrounded by commercial and residential land-use types, while industrial RTS are mainly distributed in the outer suburbs. Based on multi-source data, AP and K-means clustering algorithms are used to build a spatial buffer around RTS (800 m as an example), and different attributes of passenger flow and land-use types around RTS are classified. By comparative analysis, we explore the spatial and temporal distribution of passengers and land-use types around RTS in different rail lines and administrative districts, revealing the correlation between passenger flow and land-use. These conclusions have certain guiding significance for the passenger flow evacuation of urban rail transit stations and the construction of urban functional facilities. In terms of passenger flow evacuation, more circular rail lines should be built in central cities and connected to outer suburbs. At the same time, in the life service, medical service, public facilities, and other functional areas increase the construction of fast and slow rail lines to achieve the purpose of easing traffic congestion. In terms of urban spatial planning, it is necessary to improve the construction of commercial functional areas in urban suburbs and promote the overall development of urban spatial layout. In the future, research can be carried out from the following perspectives: multi-dimensional data passenger flow analysis, clustering algorithm, clustering algorithm under different circumstances, and time granularity. Firstly, traffic data of different types and periods can be used to analyze passenger and land types around RTS by combining qualitative and quantitative methods. Secondly, the corresponding clustering model can be built according to the characteristics of the data, which can ensure the accuracy of data processing and avoid the simplification of the model. Finally, the clustering method and time granularity of passenger flow on urban rail lines are reasonably selected under different conditions such as holidays and emergencies.

Acknowledgements

This research was supported by the National Social Science Foundation research project "the Realization Path of Supporting the Comprehensive Construction of a Modern Country with Strong Transportation Network" (grant numbers 21AZD019).

7 REFERENCES

[1] Chen, X. (2021). Analysis on Agglomeration Effect of Transportation Infrastructure Based on Spatial Economics. *Journal of Hebei University of Economics and Trade*, *42*(4), 9.

[2] Zhu, S., Jia, S., Sun, Q., & Meng, Q. (2023). An empirical study of China–Singapore International Land-Sea Trade Corridor: Analysis from supply and demand sides. *Transport Policy, 135*, 1-10. https://doi.org/10.1016/j.tranpol.2023.03.001

[3] Rong, C., Li, X., Wang, X., & Yan, F. (2020).On the Transformation Trend of Distributed Transport Supply.*Journal of Beijing Jiaotong University (Social Sciences Edition)*, *19*(03), 18-31.

[4] Deng, J. & Xu, M. (2015). Characteristics of subway station ridership with surrounding land use: A case study in Beijing. *2015 International Conference on Transportation Information and Safety (ICTIS). IEEE.* https://doi.org/10.1109/ICTIS.2015.7232208

[5] Zhang, C., Zheng, B., & Tsung, F. (2023). Multi-view metro station clustering based on passenger flows: a functional data-edged network community detection approach. *Data Mining and Knowledge Discovery*, *37*(3), 1154-1208. https://doi.org/10.1007/s10618-023-00916-w

[6] Wu, C., Pei, Y., & Gao, J. (2015). Model for estimation urban transportation supply-demand ratio. *Mathematical Problems in Engineering*, *2015*(Pt.22), 1-12. https://doi.org/10.1155/2015/502739

[7] Long, F., Liu, J., & Zheng, L. (2022). The effects of public environmental concern on urban-rural environmental inequality: Evidence from Chinese industrial enterprises. *Sustainable Cities and Society*, *80,* 103787. https://doi.org/10.1016/j.scs.2022.103787

[8] Calthorpe, P. (1993). The next American metropolis: Ecology, community, and the American dream. *Princeton architectural press.*

[9] Pan, H., Li, J., Shen, Q., & Shi, C. (2017). What determines rail transit passenger volume? Implications for transit oriented development planning. *Transportation Research Part D: Transport and Environment*, *57*, 52-63. https://doi.org/10.1016/j.trd.2017.09.016

[10] Rong, C., Zhu, D., Liu, L., & Wang, D.(2023). Promoting Rail Transit TOD Development in Urban Regeneration with Floor Area Ratio(FAR) Bonus and Transfer. *Urban Development Studies*, *30*(04), 25-30.

[11] Schuetz, J. (2014). Do Rail Transit Stations Encourage Neighborhood Retail Activity? *Urban Studies*, *52*(14), 2699-2723. https://doi.org/10.1177/0042098014549128

[12] A, D. R. B. & B, K. R. I. (2001). Identifying the impacts of rail transit stations on residential property values. *Journal of Urban Economics*, *50*(1), 1-25. https://doi.org/10.1006/juec.2001.2214

[13] Chu, D. & Wei, S. (2017). From polysemous affect to city integration: the definition thinking and frontier method to radiation realm of city rail transit station. *Springer International Publishing.* https://doi.org/10.1007/978-3-319-48296-5_3

[14] Huang, Z. & Rong, C. (2011). Transport Location Performance and the Integration Effect of the Large-scale Railway Passenger Station on Transport Resources. *Journal of the China Railway Society*, *33*(06), 8-13.

[15] Zhang, S. (2017) Study on the Coupling Development and Optimization of Beijing Urban Rail Transit Network and Urban Spatial Structure. *Beijing Jiaotong University.*

[16] Zemp, S., Stauffacher, M., Lang, D., & Scholz, R. (2011). Classifying railway stations for strategic transport and land

use planning: context matters! *Journal of Transport Geography*. https://doi.org/10.1016/j.jtrangeo.2010.08.008

[17] Cao, J., Xu, Y., Sun, L., Zhao S., & Wang, Y. (2021). Passenger Flow Characteristics and Analysis of Urban Functional Structure Based on Rail Transit Data. *Urban Rapid Rail Transit*, *34*(2), 71-78+85.

[18] Bollinger, C. R. & Ihlanfeldt, K. R. (1997). The impact of rapid rail transit on economic development: the case of atlanta's marta. *Journal of Urban Economics*, *42*(2), 179-204. https://doi.org/10.1006/juec.1996.2020

[19] Huang, H. (1996). The Land-Use Impacts of Urban Rail Transit Systems. *Journal of Planning Literature*, *11*(1), 17-30. https://doi.org/10.1177/088541229601100103

[20] Dong, N., Li, T., Liu, T., Tu, R., Lin, F., Liu, H., & Bo Y. (2023). A method for short-term passenger flow prediction in urban rail transit based on deep learning. *Multimed Tools Appl*. https://doi.org/10.1007/s11042-023-14388-z

[21] Shang, B. & Zhang, X. N. (2012). Passengers flow forecasting model of urban rail transit based on the macro-factors. *Advanced Engineering Forum*, *6*, 688-693. https://doi.org/10.4028/www.scientific.net/AEF.6-7.688

[22] Wang, Q., Liu, Y., Zheng, S., Chen, B., & Li, Y. (2023). New stations of urban rail based on TOD model classification inbound and outbound passenger traffic forecasts: the case of Chengdu. *Sixth International Conference on Traffic Engineering and Transportation System (ICTETS 2022)*, *12591*, 163-172. https://doi.org/10.1117/12.2668791

[23] Chen, W., Chen, X., Chen, J., & Cheng, L. (2022). What factors influence ridership of station-based bike sharing and free-floating bike sharing at rail transit stations? *International Journal of Sustainable Transportation*, *16*(4), 357-373. https://doi.org/10.1080/15568318.2021.1872121

[24] Delmelle, E. & Nilsson, I. (2020). New rail transit stations and the out-migration of low-income residents. *Urban Studies*, *57*. https://doi.org/10.1177/0042098019836631

[25] Pang, L., Jiang, Y., Wang, J., Qiu, N., Xu, X., Ren, L., & Han, X. (2023). Research of Metro Stations with Varying Patterns of Ridership and Their Relationship with Built Environment, on the Example of Tianjin, China. *Sustainability*, *15*(12), 9533. https://doi.org/10.3390/su15129533

[26] Ma, X., Wu, Y. J., Wang, Y., Chen, F., & Liu, J. (2013). Mining smart card data for transit riders' travel patterns. *Transportation Research Part C: Emerging Technologies*, *36*, 1-12. https://doi.org/10.1016/j.trc.2013.07.010

[27] Jiao, H., Huang, S., & Zhou, Y. (2023). Understanding the land use function of station areas based on spatiotemporal similarity in rail transit ridership: A case study in Shanghai, China. *Journal of Transport Geography*, *109*, 103568. https://doi.org/10.1016/j.jtrangeo.2023.103568

[28] Wei, C., Fu, M., Wang, L., Yang, H., Tang, F., & Xiong, Y. (2022). The research development of hedonic price model-based real estate appraisal in the era of big data. *Land*, *11*(3), 334. https://doi.org/10.3390/land11030334

[29] Yang, J. & Zhang, D. (2008). Study on reasonable attraction range of urban rail transit station. *China Railway*, *3*, 72-75. https://doi.org/10.19549/j.issn.1001-683x.2008.03.018

[30] Liu, Q., Huan, W., Deng, M., Zheng, X., & Yuan, H. (2021). Inferring Urban Land Use from Multi-Source Urban Mobility Data Using Latent Multi-View Subspace Clustering. ISPRS *International Journal of Geo-Information, 10*(5), 274. https://doi.org/10.3390/ijgi10050274

[31] Liu, Y., Tang, M., Wu, Z., Tu, Z., An, Z., Wang, N., & Li, Y. (2020). Analysis of passenger flow characteristics and their relationship with surrounding urban functional landscape pattern. *Transactions in GIS*, *24*(6), 1602-1629. https://doi.org/10.1111/tgis.12665

[32] Gao, Y., Zhang, Y., & Alsulaiman, H. (2021). Spatial structure system of land use along urban rail transit based on GIS spatial clustering. *European Journal of Remote Sensing*, *54*(sup2), 438-445. https://doi.org/10.1080/22797254.2020.1801356

[33] Xia, X. & Gai, J. (2021). Classification of urban rail transit stations and points and analysis of passenger flow characteristics based on K-Means clustering algorithm. *Modern Urban Transit*, *4*, 112-118.

[34] Yuan, J., Zheng, Y., & Xie, X. (2012). Discovering regions of different functions in a city using human mobility and POIs. *In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 186-194. https://doi.org/10.1145/2339530.2339561

[35] Hou, Q., Xing, Y., Wang, D., Liu, J., Fan, X., & Duan, Y. (2020). Study on coupling degree of rail transit capacity and land use based on multivariate data from cloud platform. *Journal of Cloud Computing*, *9*, 1-12. https://doi.org/10.1186/s13677-020-0151-x

[36] Wang, D., Dewancker, B., Duan, Y., & Zhao, M. (2022). Exploring Spatial Features of Population Activities and Functional Facilities in Rail Transit Station Realm Based on Real-Time Positioning Data: A Case of Xi'an Metro Line 2. *ISPRS International Journal of Geo-Information*, *11*(9), 485. https://doi.org/10.3390/ijgi11090485

[37] Yang, J. & Zhang, D. (2008). Study on reasonable attraction range of urban rail transit station. *China Railway*, *3*, 72-75. https://doi.org/10.19549/j.issn.1001-683x.2008.03.018

[38] Hsiao, S., Lu, J., Sterling, J., & Weatherford, M. (1997). Use of geographic information system for analysis of transit pedestrian access. *Transportation Research Record*, *1604*(1), 50-59. https://doi.org/10.3141/1604-07

[39] Guo, G. (2020). Cluster Analysis and Application of Beijing Urban Rail Transit Stations Based on Passenger Flow Characteristics Data of POI. *Beijing Jiaotong University*. https://doi.org/10.26944/d.cnki.gbfju.2020.001760

[40] Chouakria, A. D. & Nagabhushan, P. N. (2007). Adaptive dissimilarity index for measuring time series proximity. *Advances in Data Analysis and Classification*, *1*(1), 5-21. https://doi.org/10.1007/s11634-006-0004-6

[41] Cheng, J., Liu, J. J., & Gao, Y. (2016). Analyzing the spatio-temporal characteristics of Beijing's OD trip volume based on time series clustering method. *Journal of Geo-information Science*, *18*(9), 1227-1239.

[42] Shen, W., Xiao, W., & Wang, X. (2016). Passenger satisfaction evaluation model for Urban rail transit: A structural equation modeling based on partial least squares. *Transport Policy*, *46*, 20-31. https://doi.org/10.1016/j.tranpol.2015.10.006

Contact information:

Xuanxuan XIA
School of Economics and Management, Beijing Jiaotong University,
Beijing 100044, China
E-mail: 20113017@bjtu.edu.cn

Hongchang LI
School of Economics and Management, Beijing Jiaotong University,
Beijing 100044, China
E-mail: hchli@bjtu.edu.cn

Kexin LIN
School of Business, Shandong Normal University,
Jinan, 250358, China
E-mail: 2021021045@stu.sdnu.edu.cn

Kun LING
(Corresponding author)
School of Economics and Management, Beijing Jiaotong University,
Beijing 100044, China
E-mail: 11862638@qq.com

Appendix A

See Fig. A.1, Fig. A.2 and Fig. A.3.

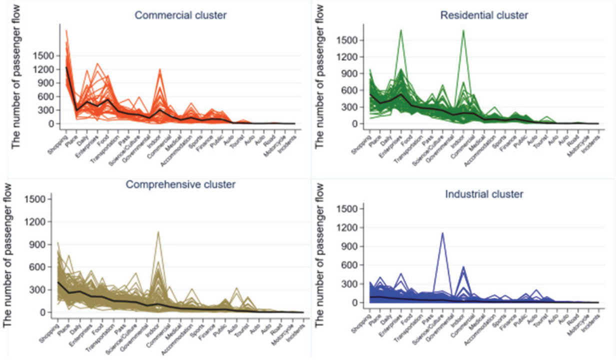


Figure A.1 The number of POI for different functional transit stations

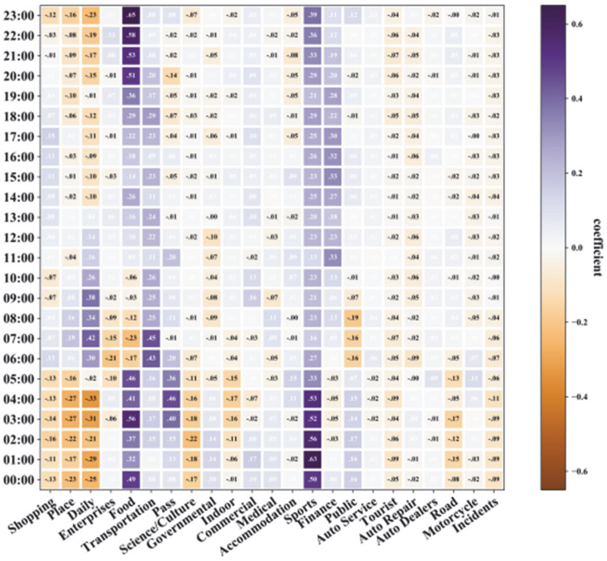


Figure A.2 The correlation between passenger flow and POI at taxi pick-up points

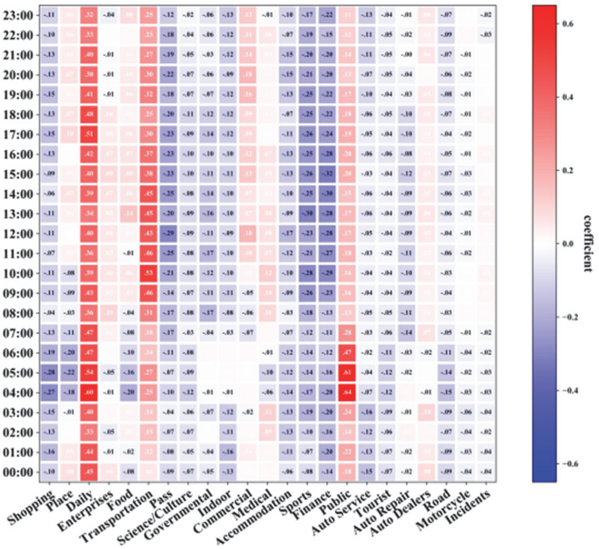


Figure A.3 The correlation between passenger flow and POI at taxi drop-off points