IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 6, 2020, accepted July 12, 2020, date of publication July 17, 2020, date of current version August 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3010057

Evaluation of Bus Accessibility Based on Hotspot Detection and Matter-Element Analysis

CHAO SUN AND WEI QUAN
Department of Traffic Science and Engineering, Harbin Institute of Technology, Harbin 150090, China
Corresponding author: Wei Quan (weiquan@hit.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 51578197 and Grant 51578198.

ABSTRACT The paper aims to assess bus accessibility considering the matching between supply and demand for effectively optimizing the level and fairness of urban public transport service, which realizes the quantification of regional balance and accurate positioning the area with the worst balance. We firstly employ hotspot detection procedure based on taxi trajectory data and kernel density analysis to identify the travel sensitive areas, the heat values of which are deployed to represent travel demand spatial-temporally and evaluate weight factors for bus accessibility modeling. Matter-element theory is selected to establish multi-parameter evaluation model of bus accessibility, which has potential for solving incompatibility problems by systematically considering all factors. The correlations between accessibility indexes and heat value are deployed to evaluate weight factors rather than analytic hierarchy process or expert assessment method to improve subjectivity and dynamic updating. An index called the Level Ratio of Accessibility to Demand (*LRAD*) is addressed finally to quantify the balance between accessibility supply and travel demand of travel sensitive areas, which identifies the regional imbalance to assist the public transport system assignment. Xi'an, a large city, is selected as a case study for methodology verification. Bus accessibility degree of the whole city as well as its travel sensitive areas is evaluated by the matter-element model. It is found the bus transport accessibility of Xi'an is moderate level(M_3). The *LRAD* results identify the priority-processing area with the poor balance between accessibility supply and travel demand, which is cross referenced with local urban plans for verification.

INDEX TERMS Traffic engineering, transportation data analysis, bus accessibility, hotspot detection, matter-element analysis, balance between supply and demand.

I. INTRODUCTION

With the acceleration of urbanization, the urban spatial structure is in the condition of constant change, which is characterized as increasing area and polycentric layout. However, fringing into a new area will lead to the aggravation of job-housing imbalance and finally to results in larger travel impedance including the increasing distance and time. Accessibility is the primary capability to achieve mobility for people and goods, the constraint on which has a bad influence on the balance of development between different regions[1], even leading to deeper poverty[2]. Obviously, accessibility improvement by public transport is commonly regarded as the most effective way to deliver large numbers of passengers and even ease congestion in the city[3].

Public transport accessibility can be defined as a measure to access the possibility of travelers moving quickly and conveniently from one place to another[4]. It can be assessed in several ways[5], such as Time-based Transition Service Area Tool (TTSAT)[6], Land Use & Public Transport Accessibility index(LUPTA)[7], General Transit Feed Specification (GTFS)[8], Public Transport Accessibility Levels(PTAL)[9], etc., which are well published and applied. Moreover, many studies analyze the bus accessibility level to various infrastructures such as hospitals[10], schools[11] and entertainment facilities[12], whose results are generally used to test the rationality of site selection of common services. These works are mainly focused on evaluating the public transport accessibility in terms of specific travel demands, whereas only a few systematically consider the balance of regional supply and common travel demands.

Additionally, we have found that it has the potential to improve the current spatial region division method because

The associate editor coordinating the review of this manuscript and approving it for publication was Bohui Wang.

the emergence of new data sources enables researchers and practitioners to overcome the obstacle by better incorporation of more detailed and time-effective data. Generally, the spatial region division method is mainly based on its spatial attribution including the geographic administrative unit, road network condition, etc., and the element is zoned by the spatial position and land use without considering the actual travel characteristics[13]. Grid approach is widely adopted for the segmentation of the study area by mapping into square grids with a reasonable side length. Intense computations are required as the key indexes for accessibility assessment need a calculation for each permutation-combination of the Origin-Destination (OD) travel matrix, whose complexity depends on the grid dimension[14]. Additionally, a more efficient way is selecting typical regions for the city-level assessment. Hotspot detection based on the travel trajectory is a preferable way to identify the typical areas, which dynamically reflects the actual traffic operation[15].

Public transport is regarded as a welfare policy provided by the government, the connotation of which is to improve the public transport accessibility for everyone. One way to enhance the public transport system is to improve the accessibility based on population distribution. Because of the rapid urban expansion and the regional inequality of the economy, the imbalance of public transport supply and travel demand is exacerbating. Evaluation of the coupling degree between supply and demand regionally is beneficial for the assignment of the public transport system at the city level. The key required data are the regional travel demand based on the spatial element. One way is to combine land-use and population with new technology like mobile phone location signal data to estimate the demand[16]. However, as essentially a probability model, it cannot accurately confirm the quantitative relationship between the actual travel demand and land-use type. Utilizing trajectory data to extract the OD pairs is a relatively efficient and accurate way to achieve travel characteristics regionally[17]. Trajectory-based hotspot detection can easily describe the relative size of travel demand according to heat value.

In response to the above-mentioned challenges, the overarching theme of this paper is to explore the imbalance degree between bus accessibility supply and travel demand of travel sensitive areas. These areas are identified by hotspot analysis of taxi trajectory data. The main contributions of this work lie in three aspects: 1) Hotspot detection based on kernel density analysis is deployed to identify the travel sensitive area of OD, whose travel impedance and demand are easily achieved and updated. 2) Matter-element theory is firstly selected to establish the bus accessibility model, which systematically considers all factors to solve incompatibility problems and dynamically guarantee the implementation of the evaluation system. 3) The paper addresses an index to assess the coupling degree between accessibility supply and demand of travel sensitive areas, which assists in the identification of the priority-processing area for improving regional imbalance.

The remainder of the paper is structured as follows. Section II provides an overview of the methodology, as well as a detailed description of hotspot detection and travel sensitive area establishment. Additionally, Section II describes the selection of accessibility indexes and emphasizes the bus accessibility evaluation model based on matter-element theory. Section III takes Xi'an city as the study case to demonstrate the specific operation process of bus accessibility assessment, the coupling degree calculation of supply and demand and the priority area identification for transport system optimization. Section IV concludes and discusses this study.

II. METHODOLOGY

A. METHODOLOGY FRAMEWORK

The overarching theme of this paper is to explore the coupling degree between bus accessibility supply and travel demand of travel sensitive areas. The methodology framework mainly contains three components, which are hotspot analysis, accessibility modeling and analysis respectively, as shown in Fig.1.

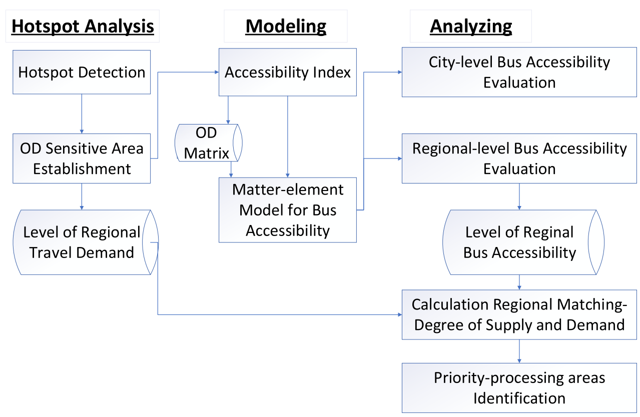


FIGURE 1. **Methodology Framework.**

B. HOTSPOT ANALYSIS

1) HOTSPOT DETECTION BASED ON KERNEL DENSITY ANALYSIS

A hotspot is generally defined as an area where some similar events take place and are geo-referenced on the map. With the development of the algorithm about spatial density analysis (including spatial relative analysis and clustering), it is relatively easy to precisely master the positions and effect degree of hotspots. For instance, the distribution characteristics of criminals, diseases and traffic collisions have been calculated via hotspot detection. The rapid development of navigation technology offers huge amounts of trajectory data from moving agents, which provides insights into both individual behaviors and group mobility. Based on the Big Data and Urban Computing, the study on travel characteristics is no longer limited to the traditional coarse-grained classification. The refined hotspot extraction model based on the density field has become an important tool for analyzing the characteristics of OD and tracking in recent years[18], [19].

Density analysis is spreading the data onto the earth's surface with spatial relationships to form density surface and demonstrate the distribution of spots. Kernal density estimation is universally employed in geography spatial analysis recently, which is very suitable for density estimation of large-scale spatial point data[20], [21]. Fig.2 illustrates the working principle of kernel density estimation. In the research area R, the kernel density estimation model regards an arbitrary spot as the center (called kernel k) and calculates the density value of the targeted spot in bandwidth r, which is determined by the number and distance between matter spots within the bandwidth range. Kernel density analysis calculates the density of points around each output raster cell. The density of each output raster cell is calculated by the sum of the values of all the kernel surfaces where they overlay the raster cell center.

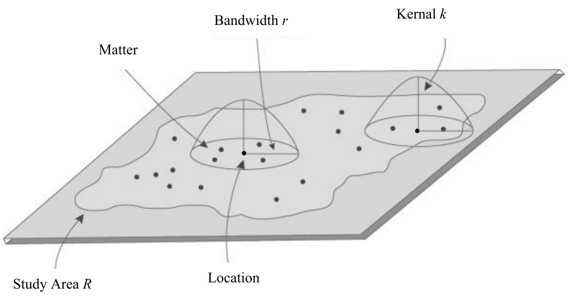


FIGURE 2. The model of Kernel Density estimation.

Fig.3 illustrates the flowchart of hotspot detection based on kernel density analysis. Data preprocessing is firstly performed, such as eliminating abnormal data, completing missing data and extracting OD from the taxi's trajectory data. Then, the OD data are exported into ArcGIS, and kernel density estimation based on spatial analyst tools is applied to output raster cells with kernel density as the raster value. The tools including *Kernal Density Analysis, Window analysis, Minus, Reclassification, Raster to Polygon* and *Turning Feature into Point* are used successively to obtain the travel 'O' and 'D' hotspots respectively and the corresponding raster

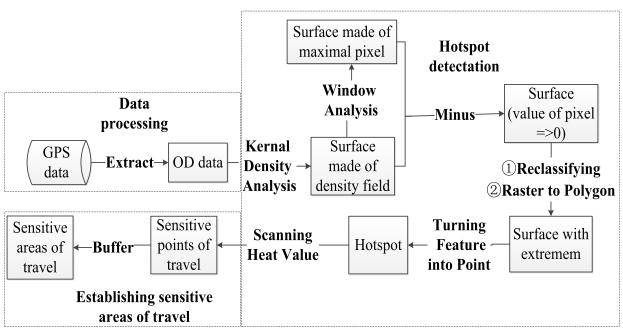


FIGURE 3. The flowchart of hotspot detection.

value. The working principle is demonstrated in the Fig.4. "O" hotspots are calculated by data of Origin (on the taxi) while "D" hotspots are calculated by data of Destination (off the taxi).

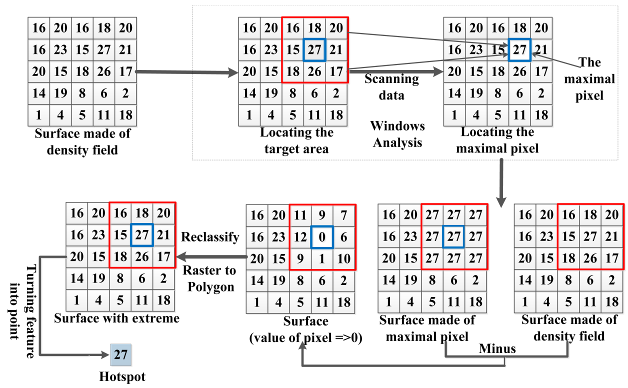


FIGURE 4. Demonstration of hotspot detection.

2) ESTABLISHMENT OF TRAVEL SENSITIVE AREA

The raster values of 'O' hotspots (getting into the taxi) indicate the levels of regional travel demands, and raster values of 'D' hotspots (off the taxi) delegate regional attractiveness, which are expressed as Heat Value (HV) for easy understanding as travel demands. The inflection point of HV distribution is set as the research scale for dividing the hotspots into two groups according to continuity[22]. Hotspots with larger HV are selected as travel sensitive points, which are imported into geographic database to establish the sensitive areas around hotspots by the Geographic Information System. According to evaluation system in the Transport Metropolis, the buffer radius r of travel sensitive region is set as 500 m in the paper, as shown the Fig.5.

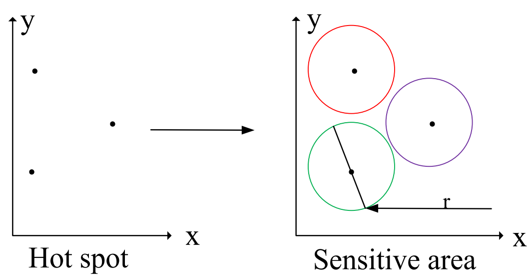


FIGURE 5. Establishment of sensitive area.

3) GRADING METHOD

In order to describe the relative level of HV between different sensitive areas, the raster value is divided into several grades by (1) [23]

$$Q_G = 1 + 3.322 \lg n \tag{1}$$

where the Q_G is quantity of the HV grade, and n is the total number of sensitive areas. Combined with the Q_G, the natural break tool in ArcGIS is deployed to grade the travel demand.

C. ACCESSIBILITY INDEX

In this paper, public transport accessibility is defined as a measure to access the possibility of travelers moving quickly and conveniently from one place to another based on the impedance measure. Bus accessibility can be viewed as the quantified ability to reach desired destinations by bus, owing to ease of travel. From the traveler's point of view, the decisive parameters of their travel accessibility are mainly quantity of opportunities, travel quality and travel cost[4]. Correspondingly, four key indicators, including bus station density, running time, service frequency and transfer number, are deployed to establish multi-parameter evaluation model of bus accessibility in this paper [24], [25].

1) BUS STATION DENSITY

Walking time from the start point to the nearest public transport stop and the time from the stops to the desired destination can be considered as the walking accessibility during a bus trip. Apparently, bus station density in a region reflects the walking accessibility, which can be attained by using (2) according to Floyd Shortest Path Algorithm. The number of bus stations in sensitive area can delegate the density because of the same size of sensitive area.

$$D_d = \sum_{i=1}^{s} I_i = \begin{cases} 1 & d[u] < r \\ 0 & d[u] > r \end{cases} \tag{2}$$

where D_b is bus stop density of the sensitive area with buffer radius r. $d[u]$ is distance between bus station u and central point of research area. s is the number of bus station of the whole research area.

2) RUNNING TIME

In-vehicle time is the main component of the whole bus trip, which indicates the ability to arrive at the destination on time. Many literatures require the bus running time as a key factor for evaluation accessibility[26]. In this study, the running time index of a region is calculated by (3).

$$T_r = \frac{\sum_{u=1}^{w} \sum_{d=1}^{z} t_u^d}{wz} \tag{3}$$

where T_r is the average running time from research area to all other sensitive areas, and t_u^d is average running time of all bus routes from the research area (on the bus) to the dth bus station of the uth 'D' sensitive area (off the bus), and w is the number of 'D' sensitive areas, and z is the number of bus stations of uth 'D' sensitive areas.

3) SERVICE FREQUENCY

Frequency of service refers to the total number of bus transit runs available to travelers per unit time within a certain area. The increasing of service frequency apparently decreases the waiting time which can be calculated by following equation.

$$F_s = \sum_{u=1}^{w} \sum_{d=1}^{z} Q_u^d \tag{4}$$

where F_s is the service frequency, and Q_u^d is the total bus transit runs available from the research area (on the bus) to the dth bus station of the uth 'D' sensitive area (off the bus).

4) AVERAGE TRANSFER NUMBER

Reducing the number of bus transfers is often an effective way to improve the efficiency of the public transport system. The transfers number is traditionally measured by the ratio between the total number of bus trips and the number of trips without transfer, which can reflect the ease of travel, cost, travel time, etc. This paper aims to evaluate regional bus accessibility. Consequently, the average number of transfers from one sensitive area to all other sensitive areas is calculated as an index by following equation:

$$N_t = \frac{\sum_{u=1}^{w} \sum_{d=1}^{z} c_u^d}{wz} \tag{5}$$

where N_t represents the number of transfer, and c_u^d is average number of transfer from the research area (on the bus) to the dth bus station of the uth 'D' sensitive area (off the bus) per unit time.

D. BUS ACCESSIBILITY EVALUATION MODEL BASED ON MATTER-ELEMENT THEORY

For reducing the influence of human factors and increasing the ability of dynamic updating, the matter-element analysis is introduced into this study. As a new interdiscipline, the Extenics has been introduced into many domains such as food industry, ecosystem safety, etc. The procedure of establishing bus accessibility model is listed as below.

1) MATTER-ELEMENT DEFINITION OF BUS ACCESSIBILITY

In matter-element analysis, the level of accessibility can be divided into several minimum units which are represented by M[27]. The matter-element of bus accessibility can be represented by the orderly ternary array R as shown in (6).

$$R = (M, c, v) = \begin{bmatrix} M & c_1 & v_1 \\ & c_2 & v_2 \\ & \dots & \dots \\ & c_i & v_i \\ & \dots & \dots \\ & c_n & v_n \end{bmatrix} \tag{6}$$

where M is class of bus accessibility, and c is the characteristic of M, and v is the value of c. If the bus accessibility is depicted by n characteristics, R would be expressed as a n-dimension array.

In this study, four indicators (including bus station density, running time, service frequency and transfer number) are employed to evaluate the bus accessibility. Therefore, the bus accessibility is a 4-dimensional matter-element, in which c_1 is D_d, and c_2 is T_r, c_3 is F_s and c_4 is N_t.

2) CLASSICAL DOMAIN AND SEGMENTED DOMAIN OF MATTER-ELEMENT

The bus accessibility is divided into several classes M_{cj} ($j = 1, \ldots, m$). For each M_{cj} subjected to class j, the value ranges of all c_i are called the classical domain, which is expressed as R_{cj}. In matter-element R_{cj}, the whole value range v_{ci} of characteristic c_i would be divided into m intervals within the range of a_{cji}-b_{cji}($j = 1, 2, \ldots, m$) [27], [28].The matrix of the classical domain R_{cj} can be achieved as following equation:

$$R_{cj} = (M_{cj}, c, v_{cj}) = \begin{bmatrix} M_{cj} & c_1 & v_{cj1} \\ & c_2 & v_{cj2} \\ & \cdots & \cdots \\ & c_i & v_{cji} \\ & \cdots & \cdots \\ & c_n & v_{cjn} \end{bmatrix} = \begin{bmatrix} M_{cj} & c_1 & [a_{cj1}, b_{cj1}] \\ & c_2 & [a_{cj2}, b_{cj2}] \\ & & \cdots \\ & c_i & [a_{cji}, b_{cji}] \\ & \cdots & \cdots \\ & c_n & [a_{cjn}, b_{cjn}] \end{bmatrix} \tag{7}$$

Segmented domain of matter-element indicates the whole value range of c_i. For c_i in (7), the segmented domain would be the interval a_{pi}-b_{pi}, as expressed in (8).

$$R_p = (M_p, c, v_p) = \begin{bmatrix} M_p & c_1 & v_{p1} \\ & c_2 & v_{p2} \\ & \cdots & \cdots \\ & c_i & v_{pi} \\ & \cdots & \cdots \\ & c_n & v_{pn} \end{bmatrix} = \begin{bmatrix} M_p & c_1 & [a_{p1}, b_{p1}] \\ & c_2 & [a_{p2}, b_{p2}] \\ & \cdots & \cdots \\ & c_i & [a_{pi}, b_{pi}] \\ & \cdots & \cdots \\ & c_n & [a_{pn}, b_{pn}] \end{bmatrix} \tag{8}$$

where R_p represents the matter-element of segmented domain. M_p and v_p indicate the same definition as same as M and v in (7) but apply to the segmented domain.

3) CORRELATION DEGREE CALCULATION

Based on the classical domain and the segmented domain, correlation degree between each characteristic c_i and each class j is calculated as the following equations:

$$K_j(v_i) = \begin{cases} \dfrac{\rho(v_i, v_{oji})}{\rho(v_i, v_{pi}) - \rho(v_i, v_{oji})} & v_i \notin v_{oji} \\ -\dfrac{\rho(v, v_{oji})}{|v_{oji}|} & v_i \in v_{oji} \end{cases} \tag{9}$$

$$\rho(v_i, v_{oji}) = |v_i - 0.5(a_{oji} + b_{oji})| - 0.5(b_{oji} - a_{oji}) \tag{10}$$

$$\rho(v_i, v_{pi}) = |v_i - 0.5(a_{pi} + b_{pi})| - 0.5(b_{pi} - a_{pi}) \tag{11}$$

where $K_j(v_i)$ is the correlation degree between ith character c_i and class j. $\rho(v_i, v_{oji})$ is the distance between the v_i and v. $\rho(v_i, v_{pi})$ is the distance between the v_i and v_{pi}.

4) WEIGHT CALCULATION AND CLASS EVALUATION

The integrated correlation degree between each matter-element object and evaluation class is essential to understanding the research object. The integrated correlation degree of each matter-element M_{cj} can be calculated by following equation:

$$K_j(M_{cj}) = \sum_{i=1}^{n} \omega_i K_j(v_{cji}) \tag{12}$$

where $K_j(M_{cj})$ indicates the integrated correlation degree between bus accessibility (M_c) and class j. ω_i is the weighted value of characteristic c_i. If $K_l = \max[K_j(M_{cj})]$ ($j = 1, \ldots, m$), M_{cj} belongs to class j.

Instead of analytic hierarchy process (AHP) or expert assessment method (EAM), this paper applies the coefficient of association between indicators and HV as weight to reduce the subjectivity, which is demonstrated minutely in the case studying.

III. CASE STUDYING

A. STUDYING AREA AND DATA DESCRIPYION

With a superior geographical position, Xi'an is the capital city of Shanxi province located in western China. Xi'an has a total area of more than 1 000 km^2 with 9.228 million registered populations. The studying area is an urban area with the most taxi and bus service, which includes 6 districts of Baqiao, Weiyang, Lianhu, Xincheng, Beilin, and Yanta, as Fig.6 shows.

FIGURE 6. The urban area of Xi'an.

The data for evaluation of bus accessibility in this study are drawn from several primary sources and basic processing. The basic data for hotspot detection and sensitive area establishment are the GPS data provided by the Xi'an taxi company, containing more than 40 million records. Data attributes include vehicle license plate number, longitude, latitude, time and status. After data preprocessing, such as coordinate conversion and data cleaning, OD points in the main urban area are extracted and displayed on the map in the coordinate system of WGS1984.

The basic data for the calculation the four accessibility indexes are the longitude and latitude data of all bus stations, bus route information, estimated travel time and transfer times between different stations, which can be obtained by Auto Navi Map. The number of stations in a sensitive area is used to calculate the density (D_b). According to the theoretical departure interval, the average hourly departure frequency of each station is obtained to calculate the service frequency (F_s). The average running time (T_r) and the average transfer times index (N_t) are calculated with the estimated travel time and transfer times to reach all stations in other sensitive areas.

B. HOTSPOT DETECTION AND TRAVEL SENSITIVE AREAS ESTABLISHMENT

Hotspot detection is deployed to mine the sensitive point of travel and establish the sensitive area. In this study, the pixel value is set as 20 and the bandwidth is set as 300 during the process of kernel density analysis. According to the methodology described in Section 2, the tools for hotspot detection in ArcGIS are used to calculate all hotspots, including 134 'O' hotspots and 159 'D' hotspots. By the scanning of the *HV* of all hotspots, an inflection point around 1000 is detected and is set as the research scale. Hotspots with *HV* over 1000 are selected and defined as travel sensitive spots, including 62 'O' sensitive areas and 64 'D' sensitive areas. "O" hotspots are calculated by data of Origin (on the taxi) while "D" hotspots are calculated by data of Destination (off the taxi).

Subsequently, with the travel sensitive spots as the center, travel sensitive areas are built as the buffer zone with a search radius of 500 *m* and exported into the geographical database. The *HV* of the hotspots indicates the travel demand degree of sensitive areas. Studying of areas with high *HV* is a precise and efficient way to assess and optimize the whole city. Fig.7 shows the distribution of all 'D' sensitive areas and 'O' sensitive areas.

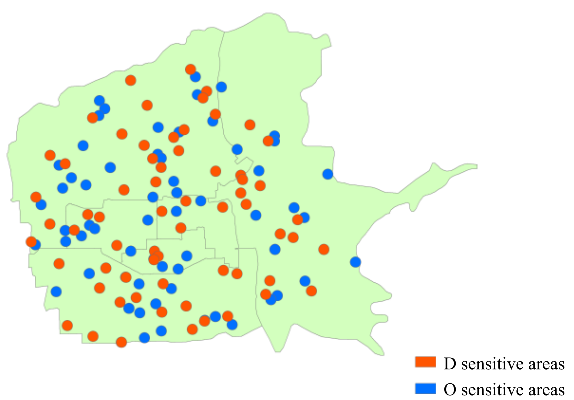


FIGURE 7. The distribution of "O" and "D" sensitive area.

C. BUS ACCESSIBILITY MODELING

1) STEP 1: BUILDING MATTER-ELEMENT OF BUS ACCESSIBILITY

Based on the natural breaks[29] and the results calculated by (1), the bus accessibility is divided into 5 levels, which are *Very Good*, *Good*, *Moderate*, *Bad* and *Very Bad* respectively, corresponding to M $= \{M_{c1}, M_{c2}, M_{c3}, M_{c4}, M_{c5}\}$ as shown in Table 1 [30].

TABLE 1. Gradation of accessibility.

	M_{c1}	M_{c2}	M_{c3}	M_{c4}	M_{c5}
D_d	>15	(12,15]	(8,12]	(5,8]	<5
T_r	<65	[65,70)	[70,75)	[75,80)	=>80
F_s	>80	(60,80]	(40-60]	(20,40]	<=20
N_t	<1.2	(1.2,1.4]	(1.4,1.6]	(1.6,1.8]	=>1.8

2) STEP 2: BUILDING THE CLASSICAL DOMAIN AND SEGMENT DOMAIN AFTER NORMALIZATIONS

Because all indicators have different scales, normalization using the efficacy coefficient method by using (13) is performed in each category. The results are listed in the Table 2.

$$c_i^* = f_{pi}(d_i) = \begin{cases} 1 & d_i \leq m_i \\ \dfrac{M_i - c_i}{M_i - m_i} & c_i \in p_i \\ 0 & c_i \geq M_i \end{cases} \tag{13}$$

where c_i^* isnormalized value of c_i, and p_i is the value range of c_i, and M_i is the maximum of c_i, and m_i and minimum of c_i.

TABLE 2. Normalization of gradation of accessibility.

	M_{c1}	M_{c2}	M_{c3}	M_{c4}	M_{c5}
D_d	(0,3/16)	[3/16,3/8)	[3/8,5/8)	[5/8,13/16]	(13/16,1)
T_r	(5/6,1]	(4/6,5/6]	(3/6,4/6]	(2/6,3/6]	[0,2/6]
F_s	(0,3/11)	[3/11,5/11)	[5/11,7/11)	[7/11,9/11)	[9/11,1]
N_t	(8/13,1]	(6/13,8/13]	(4/13,8/13]	(2/13,4/13]	[0,2/13]

After normalization, the classical domain for M_{c1} is shown as below.

$$R_{c1} = \begin{bmatrix} M_{c1} & c_1 & [0, 3/16) \\ & c_2 & (5/6, 1] \\ & c_3 & [0, 3/11) \\ & c_4 & (8/13, 1] \end{bmatrix} \tag{14}$$

where c_1 is D_d, and c_2 is T_r, and c_3 is F_s, and c_4 is N_t. R_{c2}, R_{c3}, R_{c4}, R_{c5} can be calculated similarly as R_{c1}.

Because of the normalization process, the segment domain is expressed as (15)

$$R_p = \begin{bmatrix} M & c_1 & [0, 1] \\ & c_2 & [0, 1] \\ & c_3 & [0, 1] \\ & c_4 & [0, 1] \end{bmatrix} \tag{15}$$

3) STEP 3: WEIGHT EVALUATION BASED ON HV

In existing theoretical studies, AHP and EAM are widely employed to determine the weight of index. AHP combines qualitative and quantitative analysis to make a decision[31], but the existing subjectivity may affect result objectivity. The paper deploys the associate coefficients between key indexes and *HV* of 'O' travel sensitive areas for weight calculation, the values of which embody the influence of specific indicators on travel demand. The Pearson Correlation Coefficients between average *HV* and the characteristics of bus accessibility are used for the bivariate correlation test by SPSS, as shown in Table 3. And the weight of each c_i can be calculated by (16):

$$\omega_i = \frac{|coa_i(HV, c_i)|}{\sum_{i=1}^{n} |coa_i(HV, c_i)|} \tag{16}$$

where the coa_i is the associate coefficient between c_i and *HV*, and ω_i is weighted value of c_i. Based on the Table 3, the result of weight evaluation is "$\omega = (0.481, 0.167, 0.125.0.227)$".

TABLE 3. The results of correlation test.

		HV	D_d	T_r	F_s	N_t
HV	Pearson correlation	1	0.151	0.436	0.113	-0.206
	Significant(two-Tailed Tests)		0.241	0.564	0.923	0.109
	N	62	62	62	62	62
D_d	Pearson correlation	0.151	1	-0.582	0.474	0.196
	Significant(two-Tailed Tests)	0.241				0.127
	N	62		62	62	62
T_r	Pearson correlation	0.436	-0.582	1	0.826	-0.521
	Significant(two-Tailed Tests)	0.564				
	N	62	62		62	62
F_s	Pearson correlation	0.113	-0.474	0.826	1	-0.361
	Significant(two-Tailed Tests)	0.923				0.004
	N	62	62	62		62
N_t	Pearson correlation	-0.206	0.196	-0.521	0.361	1
	Significant(two-Tailed Tests)	0.109	0.127		0.004	
	N	62	62	62	62	62

4) STEP 4: EVALUATION THE BUS ACCESSIBILITY LEVEL OF XI'AN CITY

The average values of four indexes(D_d, T_r, N_f, N_t) of all O sensitive areas in Xi'an city are substituted into (6) to build the matter-element of the evaluation object, which is expressed as (17):

$$R_x = [M, c, v] = [M_x, c_i, v_i] \tag{17}$$

By using (10)-(12), the correlation degree between each index and class was calculated and presented in Table 4.

TABLE 4. Correlation degree.

	M_{c1}	M_{c2}	M_{c3}	M_{c4}	M_{c5}
D_d	-0.656	-0.305	0.032	-0.381	-0.524
T_r	-0.176	0.278	-0.361	-0.574	-0.745
N_f	-0.478	-0.304	0.044	-0.220	-0.343
N_t	-0.795	-0.304	-0.233	-0.574	-0.652

The integrated correlation degree was calculated as $\boldsymbol{K_j}(M_i) = \boldsymbol{\omega}^*\boldsymbol{M_{rd}} = (-0.585, -0.207, -0.092, -0.437, -0.567)$.

Because the maximum integrated degree is $K_j(M_3)$, the level of bus accessibility in Xi'an city is M_3, which belongs to 'Moderately' level.

5) Step 5: EVALUATION AND GRADING THE REGIONAL BUS ACCESSIBILITY OF SENSITIVE AREA

With the similar way in step 3, the model is deployed to assess the accessibility of each "O" sensitive area. One of the main outputs is a classification map of bus accessibility as shown in Fig.8. The areas with *Very Bad* level marked red are mostly located at urban peripheries. This distribution is in line with the status that the development of urban public transport is slower than land use. The areas with *Very Good* level are mostly located centrally because of the long period of mature development. Fig.9 shows the proportion of different accessibility.

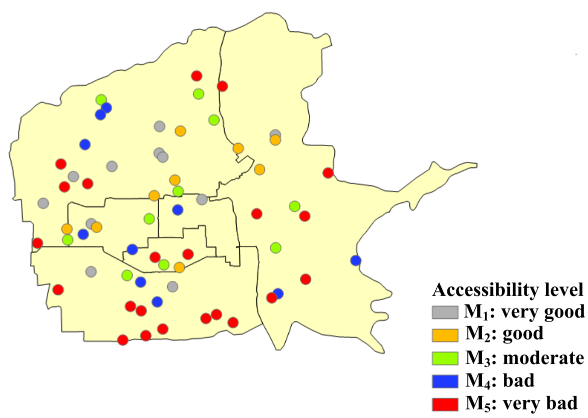


FIGURE 8. The accessibility of different sensitive areas.

D. EVALUATION OF REGIONAL COUPLING DEGREE OF ACCESSIBILITY SUPPLY AND DEMAND

To promote balanced allocation and improve the supply level of urban public transport resources can effectively reduce the travel differences between groups with different travel environments and improve the fairness of urban transport. In order to evaluate the regional balance between bus accessibility supply and travel demand, an index entitled Level Ratio of Accessibility to Demand (*LRAD*) is addressed in the

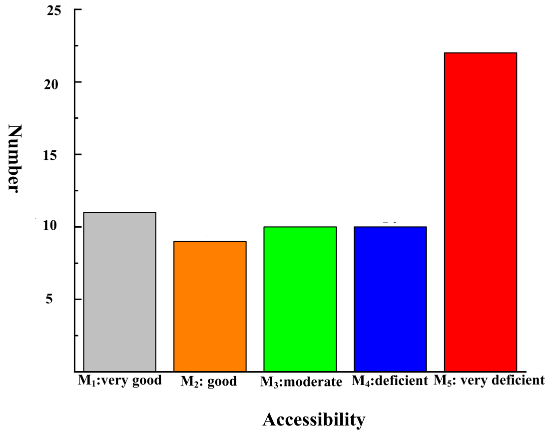


FIGURE 9. The number of different accessibility.

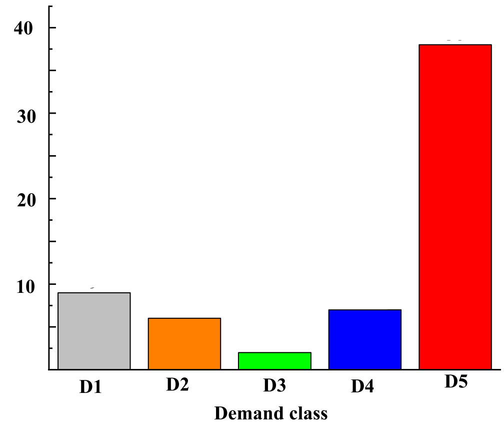


FIGURE 11. The number of different demand level.

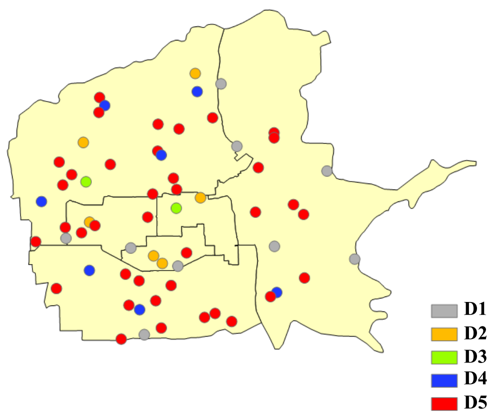


FIGURE 10. The demand of different sensitive areas.

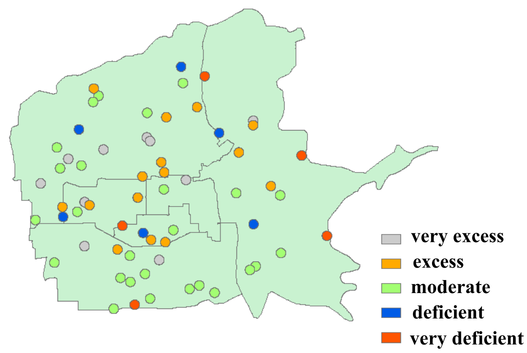


FIGURE 12. LRAD of O sensitive areas and distribution.

paper. Firstly, the grading process of the travel demands of 'O' sensitive areas is performed to divide the demands into five levels by the nature break method. The classification map of travel demand is shown in Fig.10, in which *D1* represents the highest demand and *D5* represents the lowest demand. Fig.11 shows the proportion of different travel demand levels. *LRAD* can be calculated by using (18).

$$LRAD_i = \frac{M_i}{D_i} \tag{18}$$

where $LRAD_i$ is the Level Ratio of Accessibility to Demand of the *ith* sensitive area, and M_i is the accessibility level of the *ith* sensitive area, and D_i is the travel demand level of the *ith* sensitive area.

The larger value of *LRAD* indicates the smaller value of the supply-demand ratio. If the value of *LRAD* is more approximated to 1, the coupling degree between bus accessibility supply and travel demand is preferable. Even though the regional accessibility level is not high, it is still able to meet the actual travel need from the perspective of balance. Fig.12 shows the *LRAD* distribution map of 'O' sensitive areas, which is graded into five levels as *Very Excess*, *Excess*, *Moderate*, *Deficient*, and *Very Deficient* in terms of supply compared to regional travel demand by the natural break method. Fig.13 shows the proportion of different travel demand level.

As demonstrated in Fig.12, the areas with *Very Excess*, *Excess* accessibility supply are mainly concentrated in the urban center, while the areas with *Deficient*, *Very Deficient* accessibility supply are mainly located at the urban periphery. Rapid urbanization and motorization lead to the urban form of low-density diffusion in a wider range, which makes the urban spatial layout more dispersed. It results in certain spatial differences in the density and service scope of the public transport network.

The areas with *Deficient*, *Very Deficient* accessibility supply compared to travel demand should be considered as the Priority-Processing Area (PPA) in order to improve the balance and fairness of urban transport efficiently. Such regionally targeted treatment measures can avoid resource waste in the optimization of urban public travel. Furthermore, because all sources of processing data can be easily obtained, repeating this method can dynamically evaluate and track the level of accessibility improvement temporally and spatially.

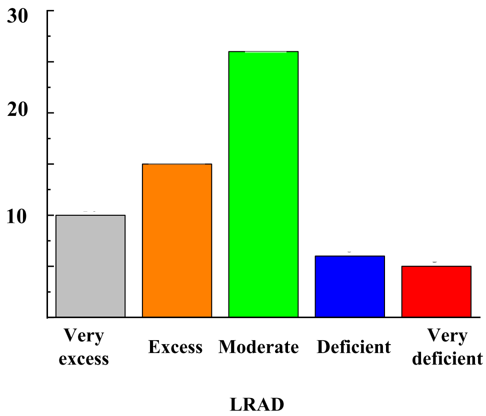


FIGURE 13. The number of different LRAD level.

IV. CONCLUSION

This paper explores a method to evaluate the bus accessibility of the urban area from the perspective of the balance between bus supply and travel demand, which realizes the quantification of regional balance and accurate positioning of the area with the worst balance for effectively optimizing the level and fairness of urban public transport service.

The paper proposes the hotspot detection method based on kernel density analysis and taxi trajectory data to explore and establish the urban travel sensitive areas with a relatively high heat value corresponding to travel demand, which can dynamically monitor and track the demand variation temporally and spatially. The accessibility analysis of the travel sensitive areas can mitigate the interference caused by the regions with insufficient data, effectively reducing the overall calculation amount of travel characteristics between regions.

Matter-element theory is selected to establish a multi-parameter evaluation model of bus accessibility because the model systematically considers all factors, which has the potential to solve incompatibility problems and dynamically guarantee the implementation of the evaluation system. Furthermore, the paper improves the weight evaluation method to increase the objectivity and dynamics in the criterion setting of the accessibility model, which is based on the mining of the correlation between travel demand and key indexes instead of AHP and EMP methodologies.

City-level and regional bus accessibility can be evaluated and graded by the proposed matter-element model. The *LRAD* index is finally used to quantify the regional balance of accessibility supply and travel demand. Based on the *LRAD* list, PPA can be identified and located, which can help to form a more targeted optimization scheme.

Xi'an, a large city, is selected as a case study for methodology verification. By hotspot detection, 62 'O' sensitive areas and 64 'D' sensitive areas are established. OD matrix and key indexes are calculated and substituted into the matter-element model to evaluate the bus accessibility level of the whole city and 'O' sensitive areas. Furthermore, the paper calculates the *LRAD* of each 'O' sensitive area and outputs the distribution map. The *LRAD* results identify 11 PPAs with a poor balance of low accessibility supply and high travel demand, which can be cross-referenced with local urban plans for verification. Based on the PPA results, targeted optimization measures can be formulated to effectively reduce the waste of public resources and quickly improve the overall level of bus service.

With the support of big data acquisition and analysis technology, all data sources for accessibility evaluation can be easily obtained. Repeating the proposed procedure can dynamically evaluate and track the level of accessibility improvement temporally and spatially. With the development of urban public transport, the methodology can be therefore used to promote balanced allocation and improve the supply level of urban public transport resources.

There are still some points needing further discussion: 1) The conclusion is only verified by the general characteristics of the rapid urbanization and motorization. Compared with accessibility model based on the bus trajectory data, the accuracy and efficiency of this method need to be further discussed. 2) We mainly focus on the accessibility of traditional bus but ignore the influences of other public transport means, such as subway and even shared bikes that have been developing fast in Xi'an. 3) Trajectory data for one workday are deployed to detect travel hotspots, which lacks consideration in influence of the other objective factors (non-working day, weather, etc.) on travel characteristics. Future researches should take the detailed consideration in correlations between different public transport means in order to offer a cross validation with matter of accessibility in this paper. Additionally, future researches should detect travel hotspots and calculate the bus accessibility in different conditions in order to establish a more refined matter-analysis model of accessibility.

DATA AVAILABILITY

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

SUPPLEMENTARY MATERIAL

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 there are no conflicts of interest regarding the publication of this paper.

REFERENCES

[1] R. Benevenuto and B. Caulfield, "Measuring access to urban centres in rural northeast brazil: A spatial accessibility poverty index," *J. Transp. Geography*, vol. 82, Jan. 2020, Art. no. 102553, doi: 10.1016/j.jtrangeo.2019.102553.

[2] G. Porter, "Transport services and their impact on poverty and growth in rural sub-saharan africa: A review of recent research and future research needs," *Transp. Rev.*, vol. 34, no. 1, pp. 25–45, Jan. 2014.

[3] Z. Song, M. Cao, T. Han, and R. Hickman, "Public transport accessibility and housing value uplift: Evidence from the docklands light railway in london," *Case Stud. Transp. Policy*, vol. 7, no. 3, pp. 607–616, Sep. 2019, doi: 10.1016/j.cstp.2019.07.001.

[4] N. Nassir, M. Hickman, A. Malekzadeh, and E. Irannezhad, "A utility-based travel impedance measure for public transit network accessibility," *Transp. Res. A, Policy Pract.*, vol. 88, pp. 26–39, Jun. 2016.

[5] B. Adhvaryu, A. Chopde, and L. Dashora, "Mapping public transport accessibility levels (PTAL) in india and its applications: A case study of surat," *Case Stud. Transp. Policy*, vol. 7, no. 2, pp. 293–300, Jun. 2019, doi: 10.1016/j.cstp.2019.03.004.

[6] C.-L. Cheng and A. Agrawal, "TTSAT: A new approach to mapping transit accessibility," *J. Public Transp.*, vol. 13, no. 1, pp. 55–72, Mar. 2010, doi: 10.5038/2375-0901.13.1.4.

[7] M. Pitot, T. Yigitcanlar, N. Sipe, and R. Evans, "Land use and public transport accessibility index (LUPTAI) tool—The development and pilot application of LUPTAI for the gold coast," *Proc. 29th Australas. Transp. Res. Forum*, vol. 6, 2006, pp. 1–20.

[8] C. Narboneta and K. Teknomo, "OpenTrip planner, openStreetMap, general transit feed specification: Tools for disaster relief and recovery," Tech. Rep., 2016.

[9] M. Hassan, "A multi-criteria approach of assessing public transport accessibility at a strategic level," *J. Transp. Geography*, vol. 57, pp. 19–34, Dec. 2016.

[10] L. Ma, N. Luo, T. Wan, C. Hu, and M. Peng, "An improved healthcare accessibility measure considering the temporal dimension and population demand of different ages," *Int. J. Environ. Res. Public Health*, vol. 15, no. 11, p. 2421, Oct. 2018, doi: 10.3390/ijerph15112421.

[11] H. C. Chin and K. W. Foong, "Influence of school accessibility on housing values," *J. Urban Planning Develop.*, vol. 132, no. 3, pp. 120–129, Sep. 2006.

[12] O. Järv, H. Tenkanen, M. Salonen, R. Ahas, and T. Toivonen, "Dynamic cities: Location-based accessibility modelling as a function of time," *Appl. Geography*, vol. 95, pp. 101–110, Jun. 2018, doi: 10.1016/j.apgeog.2018.04.009.

[13] H. Dong, M. Wu, X. Ding, L. Chu, L. Jia, Y. Qin, and X. Zhou, "Traffic zone division based on big data from mobile phone base stations," *Transp. Res. C, Emerg. Technol.*, vol. 58, pp. 278–291, Sep. 2015, doi: 10.1016/j.trc.2015.06.007.

[14] X. Albacete, D. Olaru, V. Paäl, and S. Biermann, "Measuring the accessibility of public transport: A critical comparison between methods in helsinki," *Appl. Spatial Anal. Policy*, vol. 10, no. 2, pp. 161–188, Jun. 2017, doi: 10.1007/s12061-015-9177-8.

[15] Y. Shen, L. Zhao, and J. Fan, "Analysis and visualization for hot spot based route recommendation using short-dated taxi GPS traces," *Information*, vol. 6, no. 2, pp. 134–151, Apr. 2015.

[16] J. Osorio-Arjona and J. C. García-Palomares, "Social media and urban mobility: Using Twitter to calculate home-work travel matrices," *Cities*, vol. 89, pp. 268–280, Jun. 2019, doi: 10.1016/j.cities.2019.03.006.

[17] G. D. Erhardt and L. Rizzo, "Evaluating the biases and sample size implications of multi-day GPS-enabled household travel surveys," *Transp. Res. Procedia*, vol. 32, pp. 279–290, 2018, doi: 10.1016/j.trpro.2018.10.051.

[18] W. Yu, "Discovering frequent movement paths from taxi trajectory data using spatially embedded networks and association rules," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 3, pp. 855–866, Mar. 2019.

[19] R. Bandyopadhyaya and S. Mitra, "Fuzzy cluster–based method of hotspot detection with limited information," *J. Transp. Saf. Secur.*, vol. 7, no. 4, pp. 307–323, Oct. 2015.

[20] A. Gramacki, *Nonparametric Kernel Density Estimation and Its Computational Aspects*, vol. 5. Berlin, Germany: Springer, 2018, pp. 85–118.

[21] J. Aitchison and I. J. Lauder, "Kernel density estimation for compositional data," *Appl. Stat.*, vol. 34, no. 2, pp. 129–137, 2018.

[22] K. Handley, T. Boerma, C. Victora, and T. G. Evans, "An inflection point for country health data," *Lancet Global Health*, vol. 3, no. 8, pp. e437–e438, Aug. 2015.

[23] J. Johnson, *Probability and Statistics*. Hoboken, NJ, USA: Wiley, 2007.

[24] D. A. Lessa, C. Lobo, and L. Cardoso, "Accessibility and urban mobility by bus in belo horizonte/minas Gerais—Brazil," *J. Transp. Geography*, vol. 77, pp. 1–10, May 2019, doi: 10.1016/j.jtrangeo.2019.04.004.

[25] J. C. Handley, L. Fu, and L. L. Tupper, "A case study in spatial-temporal accessibility for a transit system," *J. Transp. Geography*, vol. 75, pp. 25–36, Feb. 2019, doi: 10.1016/j.jtrangeo.2019.01.005.

[26] P. He, G. Jiang, S.-K. Lam, and D. Tang, "Travel-time prediction of bus journey with multiple bus trips," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 11, pp. 4192–4205, Nov. 2019.

[27] P.-F. Gu, W. Xi, W.-P. Ye, J. Shi, and J. Zhao, "Extenics matter-element analysis on dilemma problem in HMI design of nuclear power plant," *Nucl. Eng. Des.*, vol. 350, pp. 176–181, Aug. 2019.

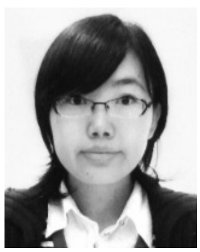
[28] Q. Xiao, R. He, and J. Yu, "Evaluation of taxi carpooling feasibility in different urban areas through the K-means matter–element analysis method," *Technol. Soc.*, vol. 53, pp. 135–143, May 2018.

[29] G. Alberti, "PlotJenks—R function for plotting univariate classification using Jenks' natural break method," Tech. Rep., 2017.

[30] A. Mohammadi, L. Amador-Jimenez, and F. Nasiri, "A multi-criteria assessment of the passengers' level of comfort in urban railway rolling stock," *Sustain. Cities Soc.*, vol. 53, Feb. 2020, Art. no. 101892, doi: 10.1016/j.scs.2019.101892.

[31] A. E. Wolnowska and W. Konicki, "Multi-criterial analysis of oversize cargo transport through the city, using the AHP method," *Transp. Res. Procedia*, vol. 39, pp. 614–623, Oct. 2019, doi: 10.1016/j.trpro.2019.06.063.

CHAO SUN was born in Linyi, China, in 1997. He received the B.S. degree in traffic engineering from Shandong Jianzhu University, in 2019. He is currently pursuing the master's degree with the Department of Traffic Science and Engineering, Harbin Institute of Technology, China. His research interest includes optimization for traffic network-based on big data.



WEI QUAN was born in Harbin, China, in 1980. She received the B.S. degree in measurement engineering from the Harbin Institute of Technology, China, in 2003, and the M.S. and Ph.D. degrees in electrical and electronic engineering from Saga University, Japan, in 2007. From 2007 to 2009, she was a Researcher working in biomedical engineering with Tohoku University, Japan. Since 2015, she has been an Associate Professor with the Department of Transportation Information and Control Engineering, Harbin Institute of Technology. Her research interests include advanced sensing and controlling technologies for ITS and transportation data analysis and management.