Journal of Cleaner Production 280 (2021) 124281

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro

Identifying the nonlinear relationship between free-floating bike sharing usage and built environment

Enhui Chen [a, b], Zhirui Ye [a, b, *]

[a] School of Transportation, Southeast University, 2 Sipailou, Nanjing, Jiangsu, 210096, China
[b] Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, 2 Southeast University Road, Nanjing, Jiangsu, 211189, China

ARTICLE INFO

Article history:
Received 19 February 2020
Received in revised form
9 September 2020
Accepted 15 September 2020
Available online 18 September 2020

Handling editor; Zhifu Mi

Keywords:
Free-floating bicycle
Bike sharing usage
Built environment
Threshold effect
Policy

ABSTRACT

A free-floating bike sharing system is an up-and-coming and marketable solution to promote transport flexibility and health benefits, which many people regard as a realistic way of generating more environmentally-friendly trips. Although many studies have investigated the associations between bike sharing usage and built environment, the existing literature has limited evidence about the relative importance of different built environment elements and their threshold impacts on cycling trips. This study contributes to the literature by proposing a modeling framework to explore the nonlinear impacts of built environment on bike sharing demand. A case study is conducted using Mobike bike sharing data in Chengdu. The analytical results indicate that population density and employment density are the two most significant factors that influence bike sharing usage. Total effects of land use variables rank the highest, followed by accessibility variables and transport facility variables. We then analyze the nonlinear impacts of different built environment elements on bike sharing usage to identify their effective ranges and threshold effects. These findings are important for planning departments to boost the share of non-motorized trips and embrace a cyclist-friendly design.

1. Introduction

Rising road congestion is an inescapable condition and has become a modern menace in large and growing cities throughout the world. Heavy traffic is closely related to air pollution, environmental degradation, safety risks, and a decrease in urban vitality. Apart from public transit (Chen et al., 2020), bike sharing is an innovative alternative to driving, and more and more people regard it as a realistic way of boosting sustainable growth. In addition to extending the reach and scope of public transit, bike sharing can reduce people's dependence on private cars and provide a convenient and healthy means of transport for locals and tourists (Fishman et al., 2013; Kaplan et al., 2015). With the increasing development of smartphone technology and high-speed internet, the emergence of free-floating bike sharing (FFBS) systems offers a much more flexible cycling solution than traditional station-based bike sharing. FFBS, seen as one of the hottest new transportation industries, has experienced an explosive growth since 2016 in China, Singapore, England, and many other countries. From 2016 to 2019, the number of Chinese registered FFBS users increased from 30 million to 300 million (Tan and Zhao, 2018). Supported almost entirely by venture capital funding, the FFBS industry is reviving the "kingdom of bicycles" by providing a low-cost and healthy lifestyle. By the middle of 2019, about 19.5 million shared bicycles "floated" on 360 cities' streets to transport users in China.

Many studies have attempted to investigate station-based bike sharing usage based on land use characteristics of station areas (e.g., population, employment, and land use design), transport service features (e.g., metro and bus services), and road network configuration (e.g., road length and intersection density) (Alcorn and Jiao, 2019; Faghih-Imani et al., 2017; Mateo-Babiano et al., 2016; Zhang et al., 2017). However, daily station-based bike sharing usage had declined from 66,040 trips in 2015—22,614 trips in downtown of Beijing in 2017, with the rate of decrease being over 65% (Cityif, 2018). Many other Chinese cities have also been experiencing this same situation. The dramatic growth of FFBS systems has overtaken the traditional bike sharing market (Gu et al., 2019). As station-based bike sharing has indeed suffered great losses in market share, one naturally wonders whether findings of station-

* Corresponding author. School of Transportation, Southeast University, 2 Sipailou, Nanjing, Jiangsu, 210096, China.
E-mail addresses: chenenhui@seu.edu.cn (E. Chen), yezhirui@seu.edu.cn (Z. Ye).

https://doi.org/10.1016/j.jclepro.2020.124281

based bike sharing research is applicable for guiding future bike sharing policy. Since FFBS is not confined to the dock, investigating FFBS requires different methods than traditional bike sharing. Meanwhile, although existing literature finds bike sharing travel demand is highly associated with urban built environment, questions remain regarding the extent to which different built environment elements contribute to bike sharing usage. The answers may underscore the relative importance of a built environment variable and help us implement targeted policies to effectively promote bike sharing ridership. In addition, traditional analytical models are based on the assumption of a linear relationship between the dependent variable and independent variables. In fact, the effects of built environment elements on urban mobility may not exactly conform to this assumption in the real world (Van Wee and Handy, 2016). On the one hand, the effect of a built environment element would not become obvious unless its amount reaches a certain level. It is important to identify the effective range of each built environment element that substantially influences the demand for bike sharing. On the other hand, when the amount of a built environment element is growing, it is not possible for its effect on bike sharing ridership to increase continually, because it may become saturated at the threshold. Detecting the threshold for each built environment element is helpful to efficiently allocate urban resources and achieve the targeted travel demand.

More intriguing questions should be asked about the underlying mechanisms that drive the relationship between bike sharing usage and built environment. For example, to what extent do different built environment elements contribute to increasing the demand for FFBS? Does the bike sharing usage have a nonlinear association with built environment? How do we identify the effective range and threshold effects of built environment, in order to more easily promote the market share and support the sustainable development? To seek answers for these questions, this study uses the large-scale bike sharing data collected in Chengdu, China. We propose a modeling framework to unravel the nonlinear relationship between bike sharing usage and built environment elements. These answers can offer us strong and coherent plans to actively promote urban cycling. Recognizable bike sharing systems not only enable us to embrace the concept of shared economy but also to form a strong and positive image of environmentally-friendly travel.

The remainder of the paper is organized as follows. The second section presents the literature review on the bike sharing usage and potentially contributing variables; the third section introduces the factors influencing FFBS usage and the modeling framework; the fourth section describes the study area and research data; the fifth section discusses the modeling results; and the final section concludes the paper, summarizes contributions, and recommends future studies.

2. Literature review

To reduce people's dependence on motor vehicles, bike sharing has gained a great popularity in transportation and environmental studies in the past decade (Si et al., 2019). Many studies focused on the potentially contributing factors influencing bike sharing trip demand, which can be divided into four main categories: weather conditions, socio-demographic effects, spatial and temporal characteristics, and built environment. It is widely accepted that weather conditions (e.g., temperature, precipitation, and wind speed) have a significant effect on bike sharing ridership (Lin et al., 2018). However, the weather condition is a short-term variable that influences the demand for trip generation (Eren and Uz, 2019). It changes continuously and cannot be controlled by human beings, which is not suitable for long-term policy planning. Socio-demographic factors (e.g., gender, age, education, income, and vehicle ownership) have been examined and have been found to strongly relate to the cycling trip demand in descriptive analysis (Feng and Li, 2016; Fishman et al., 2014). Previous studies mainly adopted revealed preference and stated preference research approaches to acquire the detailed socio-demographic information about participants. Although these researchers have identified some constructive conclusions, the descriptive data may incur some bias due to its small size and participants' subjective inclinations. Bike sharing usage varies at both spatial and temporal scales. Some approaches or models have been applied to investigate the dynamics of bike sharing usage at a finer granularity (Du et al., 2019; Yang et al., 2019). They found urban spatial attributes (e.g., geographic form and land use) and daily activities (e.g., commuting activities and non-commuting activities) significantly influence the spatial distribution of demand and temporal flow patterns. In fact, daily activities are highly dependent on spatial attributes that generate different trips with time-varying purposes throughout a day. Existing literature thus focuses on built environment elements in order to explore their impacts on bike usage (Mateo-Babiano et al., 2016; Wang and Chen, 2020; Zhang et al., 2017). The urban environment offers a suitable environment for cycling trips with a greater density of population, employment, and commercial services. These studies have found that bike sharing demand at stations near city centers, entertainment areas, and transit facilities is expected to increase. However, the relative importance of different built environment elements has not been identified, which merits further investigation to provide more specific information for promoting bike sharing usage.

Many studies investigated impacts of the station-area built environment on the bike sharing trip demand (Alcorn and Jiao, 2019; Liu and Lin, 2019; Zhang et al., 2017). Station-based bikes must be borrowed from and returned to a bike station, while free-floating bikes are not constrained to the location of stations. FFBS bikes are scattered throughout every location that is accessible to users (Ji et al., 2020). The method that regards a bike station as an analytical unit is thus not suitable for comprehensively analyzing FFBS trip demand. It is necessary to propose a new method for FFBS analysis. Researchers have attempted to develop an analytical framework to discover FFBS system dynamics, bike use characteristics influencing the demand, and the time-varying characteristics of cycling activities (Du et al., 2019; Xu et al., 2019). Xu et al. (2018) proposed the long short-term memory neural networks to forecast the cycling trip generation and attraction during different time intervals in a traffic analysis zone (TAZ). This method has achieved a better forecasting accuracy for FFBS trip demand compared to traditional statistical models during different time intervals. Yang et al. (2019) applied spatial statistics and graph-based approaches to explore changes in local travel flows before and after a new metro service, which enhances the understanding of last-mile trips. Some studies also analyzed the impacts of a FFBS system on station-based bike sharing usage and compared the usage regularity between these two types (Ji et al., 2020; Li et al., 2019). Hua et al. (2020) estimated the parking demand of FFBS bikes and its impacts on greenhouse gas emissions by using the clustering technique and identifying the critical locations for bike parking. As FFBS is a relatively new means of transport, the existing studies on FFBS have been mainly conducted in the past two years. More efforts are still needed to supplement and improve FFBS studies.

When modeling bike sharing usage and built environment, the existing literature applies the global model (e.g., multiple linear regression model) and local model (e.g., spatial autoregressive model) to detect their relationship. Multiple linear regression models were widely used to explore impacts of built environment on bike sharing usage at station level using cycling trip data (Alcorn and Jiao, 2019; Mateo-Babiano et al., 2016; Zhang et al., 2017). They have examined the significance of different built environment variables and their positive and negative effects on bike sharing usage. To investigate the impacts of built environment on trip demands and bicycle reallocation, some studies introduced generalized linear regression models to deal with count data with excess zeros at bike stations (Wang and Chen, 2020; Zhao et al., 2019). Considering the spatial dependence, researchers used a spatial autoregressive model to explore the impacts of contributing factors on the demand for bike sharing services (Shen et al., 2018). However, the linear assumption of these models has not taken into account the nonlinear relationship between bike sharing usage and built environment elements. It has been claimed that nonlinear effects result in variations in the role of built environment elements, which is critical to study sustainable urban mobility and spatial attribute (Ding et al., 2018; Van Wee and Handy, 2016). Empirical study for this claim on bike sharing remains lacking in the literature.

3. Methodology

3.1. Factors influencing FFBS usage

Trip records can be aggregated by check-in longitudes and latitudes at the TAZ scale as the dependent variable. To examine the relationship between FFBS usage and built environment, this study divides built environment elements into three categories, including land use, transport facility, and accessibility. Table 1 presents the descriptive statistics of variables. As to land use characteristics, population and employment are widely recognized as two significant factors influencing urban trips (Jun et al., 2015; Tu et al., 2018). The population density and employment density denote the number of people living in an area and the number of jobs in an area per km^2 respectively. Due to commuting activities, more or fewer cycling trips are expected based on the variation in population density and employment density across different TAZs. As each land use type can generate specific human activities and thus have quite different trip patterns, we apply the ratio of the area (RA) used for different purposes to quantify the residential, commercial, governmental, college, cultural, hospital, and industrial land uses. In addition, considering the intensity of land use (Chen et al., 2019), we introduce the density of restaurants, shopping services, leisure services, and hotels to measure their effects on bike sharing.

Transport facility is also an important category that significantly affects bike sharing. Bike sharing is often seen as a solution to the first and last mile issue of transit connectivity, which can promote the convenience of getting people to and from transit stations. In this study, to quantify the supply of transport facilities, we use the density of metro stations, bus stops, bus routes, and parking lots. Urban road network configuration including the length of road network and the intersection density is considered as a potential indicator of accessibility influencing the usage of bike sharing. In light of the special role of bike sharing in an urban transit system, three more factors are introduced to examine the intrinsic association between bike sharing.

And other transport modes. Considering the distance-decay effect, metro accessibility denoting the average impedance of road network nodes to metro stations can be calculated in a linear approximation (Qian and Ukkusuri, 2015; Tu et al., 2018). It characterizes an increase in usage as the cycling distance to metro stations decreases. The metro accessibility to bike sharing is thus calculated as:

$$\mathrm{MA}_i = \frac{1}{K}\sum_k \sum_j \frac{1}{d_{kj}^i} \tag{1}$$

where d_{kj}^i is the Manhattan distance from road node (intersection) k to metro station j in TAZ i. In view of the average length of adjacent metro stations, the upper bound of distance is set to be 2 km. A higher value of MA_i indicates a greater access to the metro station in TAZ i. The calculation of bus accessibility and parking lot

Table 1
Descriptive statistics of all variables.

Variable	Description	Mean	Std. Dev.	Min	Max
FFBS bike sharing usage	Average daily FFBS bike sharing usage per km^2 in each TAZ	1316	42	0	5993
Land use					
Population density	Population per km^2 in each TAZ	10,798	120	0	40,011
Employment density	Employment per km^2 in each TAZ	5952	112	0	36,606
Residential area	RA used for residential purposes in each TAZ	0.442	0.399	0	0.984
Commercial area	RA used for commercial purposes in each TAZ	0.101	0	0	0.911
Governmental area	RA used for government purposes in each TAZ	0.02	0.071	0	0.688
College area	RA used for colleges in each TAZ	0.026	0	0	0.857
Cultural area	RA used for cultural purposes in each TAZ	0.007	0	0	0.205
Hospital area	RA used for hospitals in each TAZ	0.007	0	0	0.154
Industrial area	RA used for industrial purposes in each TAZ	0.058	0.801	0	0.947
Restaurant	Number of restaurants per km^2 in each TAZ	136	6	0	548
Shopping service	Number of shopping services per km^2 in each TAZ	227	5	0	1427
Leisure service	Number of leisure services per km^2 in each TAZ	16	2	0	91
Hotel	Number of hotels per km^2 in each TAZ	43	2	0	727
Transport facility					
Metro station	Number of metro stations per km^2 in each TAZ	0	1	0	2
Bus stop	Number of bus stops per km^2 in each TAZ	6	3	0	20
Bus route	Number of bus routes per km^2 in each TAZ	25	6	0	137
Parking lot	Number of parking lots per km^2 in each TAZ	8	2	0	39
Accessibility					
Road density	Length of road per km^2 in each TAZ	6.792	2.505	0.349	102.807
Intersection density	Number of intersections per km^2 in each TAZ	19	5	0	88
Metro accessibility	Accessibility to metro stations in each TAZ	0.958	1.066	0	12.680
Bus accessibility	Accessibility to bus stops in each TAZ	2.282	1.558	0	21.872
Parking accessibility	Accessibility to parking lots in each TAZ	2.032	1.302	0	5.078

accessibility is similar to that of metro, except the upper bound of distance is set to be 1 km and metro stations are replaced with bus stops and parking lots.

3.2. The modeling framework

This study proposes a modeling framework to identify the nonlinear relationship between FFBS usage and built environment. The framework builds the gradient boosted regression trees (GBRT) and employs the relative importance index and the partial dependence (PD) plot to examine the individual contribution, effective range, and threshold effect of different variables. GBRT, different from the traditional boosting method, is known as multiple additive regression trees and has an advanced advantage in data mining. It combines two algorithms: regression trees are from a collection of regression models, and boosting is an adaptive method for integrating a series of models and improving predictive performance.

The tree-based regression method segments the predictor space (independent variables) into a group of distinct and non-overlapping regions based on a set of splitting rules. It then fits the mean response of the training observations falling in each region. For instance, we can use built environment elements as independent variables and the average usage of bike sharing as the response in each region to make predictions. The segmented regions are denoted terminal nodes or leaves. Predictor variables and splitting rules are determined by minimizing prediction errors.

Boosting is an ensemble method for promoting model accuracy. It integrates the output of many “weak learners” into a powerful “committee” (Hastie et al., 2009). Gradient boosting sequentially adds predictors to an ensemble and each one corrects its predecessor. It is a numerical optimization technique and applies the gradient descent to minimize the expectation of loss function. This sequential process focuses on the residuals and continues to iterate until the model fits the observations with minimized residuals.

The specific process of GBRT is described as follows. Assuming a response variable y and a set of independent variables $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$, the objective is to build a function $F^*(\mathbf{x})$ mapping $\mathbf{x}$ to y, which minimizes the expectation of loss function $\Psi(y, F(\mathbf{x}))$. We define the number of terminal nodes or leaves is J for each tree and divide the predictor space of the mth tree into J regions, such as $R_{1m}, R_{2m}, \ldots, R_{jm}$. The prediction for region R_{jm} of the mth tree is a constant value b_{jm}. The regression tree is expressed as

$$g_m(x) = \sum_{j=1}^{J} b_{jm} I, \ \ x \in R_{jm} \tag{2}$$

$$I = \begin{cases} 1, & x \in R_{jm} \\ 0, & \text{otherwise} \end{cases}$$

We then initialize a function $F_0(\mathrm{x})$ and set

$$F_0(\mathbf{x}) = \operatorname{argmin}_{\rho} \sum_{i=1}^{N} \Psi(y_i, \rho) \tag{3}$$

For the tree index m in 1, …, M, we iteratively build M regression trees. For the observation i in 1, …, N, we calculate the negative gradient of the loss function as the working response.

$$z_{im} = -\left[\frac{\partial \Psi(y_i, F(\mathbf{x}_i))}{\partial F(\mathbf{x}_i)}\right]_{F(\mathbf{x}) = F_{m-1}(\mathbf{x})} \tag{4}$$

Then, given the regression tree $g_m(\mathbf{x}_i)$, the optimal step length ρ_m of the gradient is defined as

$$\rho_m = \operatorname{argmin}_{\rho} \sum\nolimits_{i=1}^{N} \Psi(y_i, F_{m-1}(\mathbf{x}_i) + \rho g_m(\mathbf{x}_i)) \tag{5}$$

GBRT fits the training data in a stage-wise manner and updates the model by minimizing the expectation of loss function. Although adding regression trees contributes to achieving small training residuals, the model may have poor generalization ability with a continuous increase in the number of iterations. When the model conforms too closely to the training data, minor fluctuations are exaggerated, which reduces the predictive power of the model in the testing data. Thus it is important to determine the optimal number of trees and prevent the overfitting problem. For this reason, the shrinkage λ, also referred to as learning rate of the procedure, is introduced to scale the contribution of each tree. The updating function is described as

$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \lambda \rho_m g_m(\mathbf{x}_i) \tag{6}$$

The shrinkage parameter ($0 < \lambda \leq 1$) determines the contribution of each regression model to the $F^*(\mathbf{x})$. A smaller value of shrinkage parameter could decrease the training risk with the same number of iterations. It has been found that a small value of λ ($\lambda \leq 0.1$) leads to a much better generalization error (Friedman, 2002; Zhang and Haghani, 2015). Another parameter J, tree complexity, referring to the number of leaves, also affects the performance of GBRT. The optimal size of tree complexity can be estimated by improving the model performance. In addition, we also introduce a 10-fold cross validation method to avoid overfitting and promote the robustness of modeling results.

After achieving an accurate and stable model fitting, the modeling framework employs the relative importance index to interpret the individual contribution of different built environment variables. As the input variables are not relevant, it can help us identify the critical variables in predicting the outcome by quantifying the relative importance. We measure the relative importance by averaging the squared relevance over all of the additive trees. It is given as follows

$$IM_{x_i}^2(g_m) = \sum_{j=1}^{J-1} \hat{\iota}_j^2 I \tag{7}$$

$$IM_{x_i}^2 = \frac{1}{M} \sum_{m=1}^{M} IM_{x_i}^2(g_m) \tag{8}$$

where $\hat{\iota}_j^2$ is the maximal estimated improvement in squared error risk over that for a constant fit at the jth node of the mth tree for the input variable x_i. The sum of all relative importance of independent variables is 100%. Meanwhile, the modeling framework also provides a comprehensive summary of the dependence of the response on input variables through the PD plots. The PD plot has a causal interpretation and shows the marginal effect an input variable has on the predicted outcome (Molnar, 2019). It gives us a quantitative description of properties of a selected variable and helps us identify the nonlinear effects of built environment elements on FFBS usage. The PD function for regression is estimated by

$$F_s(x_s) = \mathrm{E}_{x_c}[F(x_s, x_c)] \tag{9}$$

$$\overline{F}_s(x_s) = \frac{1}{N} \sum_{i=1}^{N} F(x_s, x_{ic}) \tag{10}$$

where xs is the targeted input variable for which the PD function should be calculated and xc are the other variables used in GBRT. PD shows the causal relationship between the targeted input variable and the predicted outcome by marginalizing the GBRT output over the distribution of other variables. It represents the effect of the targeted input variable on the outcome after accounting for the average effects of other variables on the outcome.

4. Case study

4.1. Study area

The case study is conducted in Chengdu for further investigation. Chengdu, as the capital of Sichuan Province, is one of the largest and most populous cities in Western China. Currently, Chengdu houses 16,330,000 inhabitants with an urban population of 11,940,000. This city boasts an efficient public transit service and safe infrastructure for non-motorized transport such as cycling and walking. In order to promote the convenience of urban travel and accelerate the transition to sustainable growth, most big cities in China have launched the FFBS system since 2017. As of 2018, Chengdu has approximately 1,450,000 free-floating shared bikes, allocated in the downtown of this city with 240 TAZs. The downtown area is outlined by the fourth ring of the city with an area of 560 km^2, as shown in Fig. 1.

4.2. Data description

The FFBS dataset used in this study ranges from June 2 to 29, 2018, provided by the Mobike company, one of the largest FFBS service providers in China. Each record in the dataset includes user number, bike number, coordinates of origins and destinations, departure time, and arrival time. As suggested by Du et al. (2019) and Shen et al. (2018), we first remove records with abnormal information and trip distances less than 100 m or longer than 10 km. Over 15, 500, 000 valid trips are obtained after the data preprocessing stage, accounting for 91.7% of raw data records. Built environment data is partially obtained from Chengdu Municipal Bureau of Planning and Natural Resources, including the area of land use in downtown. Other data sources come from points of interest (POIs) in Gaode Map, one of the most popular digital maps in China. POIs record detailed information of specific services including their category, name, address, coordinate, and postcode. Massive POIs can fully reflect the functions of a region, and are an increasingly prevalent data source in transportation fields (Chen et al., 2019; Faghih-Imani et al., 2017; Lin et al., 2018; Shen et al., 2018; Zhao et al., 2019).

5. Results

5.1. Density variation of the FFBS usage

Kernel density estimation (KDE) is applied to investigate the local distribution patterns of FFBS behaviors at the spatial scale. It can identify hot spots and visualize the distribution and intensity of bike sharing usage by estimating the probability density function and creating a smooth surface. As shown in Fig. 2, main agglomeration areas are evident in the center of the city with high frequencies and densities. Hot spots are diversely distributed and cyclists' activity frequencies gradually decline when approaching the frontier of the downtown.

5.2. Model specifications

In this study, we use the variance inflation factors (VIFs) to examine the multi-collinearity among selected variables before developing the GBRT. Results show values of VIFs are all less than 10, indicating no multi-collinearity exists among independent variables in this study (Tu et al., 2018). We then build a grid of hyperparameter combinations to determine optimal values of shrinkage, trees, and tree complexity. Through iterating over every combination of hyperparameter values, it allows us to perform a grid search and evaluate which combination performs well. To enhance the robustness of the modeling results, we incorporate the 10-fold cross validation method into the estimation process by partitioning the dataset into 10 folds for training and testing. We then convert the cross validation error to the root mean square error (RMSE) to select the best combination of hyperparameter values. As shown in Fig. 3, RMSE substantially decreases for all values of shrinkage. The dashed line with the blue color has the smallest value of RMSE when the tree complexity increases to 8. We thus choose 0.01 and 8 as the value of shrinkage and tree complexity respectively, with the number of trees being 413.

5.3. Relative importance of built environment elements

Table 2 provides the relative importance of built environment elements on FFBS usage. The sum of the relative importance for all built environment elements is 100%. Among three categories of the potentially contributing variables, the collective contributions of land use variables account for 80.9%, followed by the accessibility variables (13.2%) and the transport facility variables (5.9%).

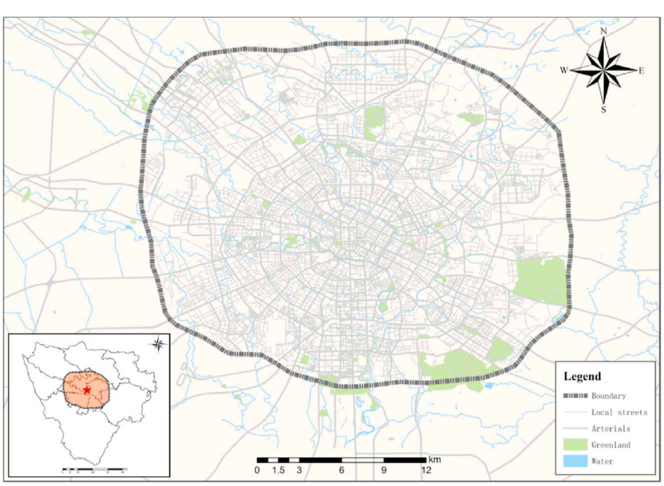

Fig. 1. Study area.

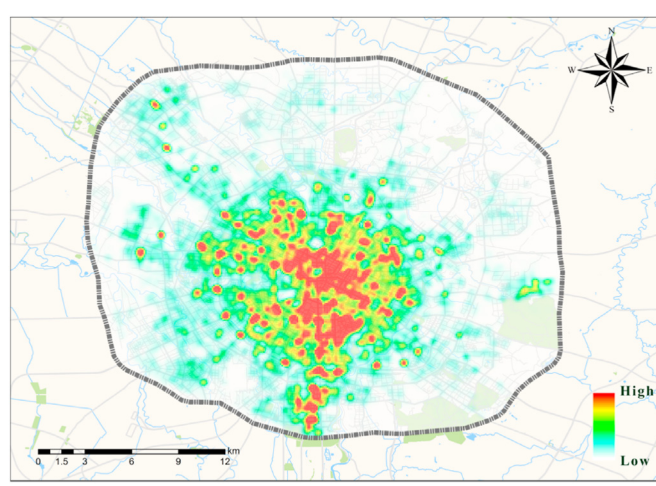

Fig. 2. Density estimation of the FFBS usage.

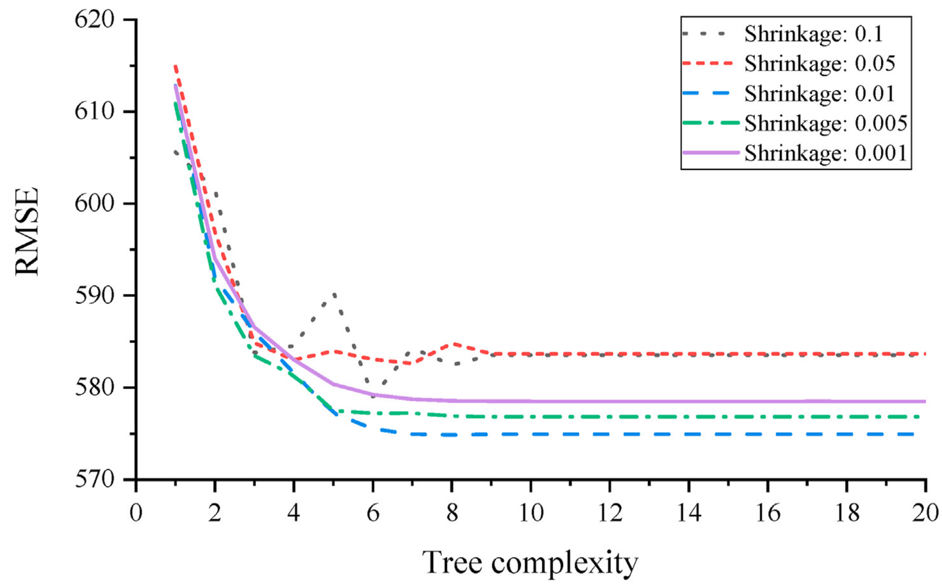

Fig. 3. Evaluation of a grid of hyperparameter combinations.

Table 2
Relative importance of built environment elements on FFBS usage.

Category	Variable	FFBS usage		
		Overall rank	Relative importance (%)	Total (%)
Land use	Population density	2	23.71	80.9
	Employment density	1	24.58	
	Residential area	10	2.05	
	Commercial area	9	2.98	
	Governmental area	20	0.09	
	College area	19	0.12	
	Cultural area	15	0.59	
	Hospital area	21	0.04	
	Industrial area	22	0.03	
	Restaurant	3	15.75	
	Shopping service	14	0.91	
	Leisure service	6	4.01	
	Hotel	5	6.03	
Transport facility	Metro station	12	1.03	5.9
	Bus stop	11	1.28	
	Bus route	8	3.31	
	Parking lot	17	0.27	
Accessibility	Road density	4	7.69	13.2
	Intersection density	7	3.76	
	Metro accessibility	13	1.00	
	Bus accessibility	16	0.54	
	Parking accessibility	18	0.21	

Population density and employment density are the two most significant variables in the land use category, and they collectively contribute to about 48% of the total effects on bike sharing usage. Bike sharing plays an important role in commuting activities. As non-commuting activities account for a larger proportion than ever before, restaurants, leisure services, and other non-commuting elements have a collective contribution of about 24% of total effects. Compared to metro and bus, bike sharing provides a more convenient mode for unfamiliar tourists, because it offers flexible schedules, routes, and online payment. Thus, bike sharing has gained a great popularity in areas with high density of hotels that have a relative importance of 6.03%. However, industrial, hospital, and governmental areas have a low influence on the bike sharing usage, probably because people tend to arrive by a car and shared bikes are not allowed to enter these areas for the sake of management. For example, people usually choose to take a car or a bus to the hospital, not a bike.

In terms of transport facility, metro and bus modes have more influence on transfers from/to bike sharing than the private car. It is worth noting that the number of bus routes (3.31%) in a TAZ has a larger influence on bike sharing usage than the number of bus stops (1.28%). Parking lots have a low influence on bike sharing demand. An explanation is that people who drive a car have more freedom than transit passengers. Walking may satisfy their needs of traveling from a parking lot to the destination. More efforts are needed to enhance the integration between bike sharing and transit. As to accessibility of the TAZ, road density and intersection density collectively contribute to 11.45% of total effects on bike sharing. The relative importance of road density is larger than that of intersection density, probably because road density directly reflects the length of road network for cycling in the TAZ. Neighborhoods with high road density have a large road network both allowing pedestrians to easily access a large number of bikes within a short walking distance and allowing cyclists to conveniently connect to their destinations in multiple directions within a short cycling distance. Accessibility to the metro station has the largest contribution in bike sharing usage among three intermodal accessibility indexes, followed by bus accessibility and parking accessibility, probably because of the large catchment area of a metro station.

5.4. Nonlinear effects of built environment on FFBS usage

Based on the modeling results of GBRT, we apply PD plots to identify the effective ranges and thresholds of built environment elements to promote bike sharing. PD plots illustrate the nonlinear effects of built environment on FFBS usage, which can provide useful guidance for repositioning in bike sharing systems and urban planning. Fig. 4 shows how FFBS usage responds to changes in a land use variable. Employment density and population density are positively associated with the bike sharing demand, which is consistent with another empirical study (Wang and Chen, 2020). Cycling trips double from 1000 to 2000 when employment density increases from 2000 jobs to 12,000 jobs per km^2. No additional effect of employment density on bike sharing usage exists while it exceeds 12,500 jobs/km^2. Population density first has a moderate effect on bike sharing usage with the population per km^2 increasing from 0 to 12,000. However, bike sharing ridership experiences a dramatic increase from 1200 to 1900 when population density ranges from 12,000 to 20,000. Cycling trips almost double in areas with high density of population and employment, compared to areas with low levels of density. As China is still undergoing a rapid urbanization, more and more people keep shifting from rural areas to urban areas (Ta et al., 2017). It is thus important for planners to direct future transportation planning in response to population and employment growth. In addition, cycling trips experience a remarkable increase when the number of restaurants increases from 130 to 200. Beyond this range, bike sharing has a slight increase and remains stable at 1700 trips. Different from other variables' moderate effects at the first stage, the presence of hotels substantially increases the number of trips by 320 with the effective range being from 0 to 70. After about 70 hotels/km^2, demand for bike sharing services has a slight increase from 1520 to 1570. Leisure services have two effective ranges from 12 to 20 and 30 to 37 respectively. Bike sharing usage then remains stable at 1590 at the threshold of effects of leisure services. Commercial areas are definitely associated with urban travel trips (Eren and Uz, 2019). When the proportion of commercial area is more than 0.3 in a TAZ, the demand for bike sharing dramatically increases. When the proportion exceeds 0.4, the rate of increase in demand for bike sharing becomes low. The proportion of commercial area then reaches the threshold of 0.52 and bike sharing usage remains stable at 1560.

Fig. 5 shows the nonlinear effects of transport facility variables on FFBS usage. A metro station can significantly promote daily bike sharing usage, which increases by 130 cycling trips in a day. When the number of bus stops per km^2 is less than 7, effects of bus stop

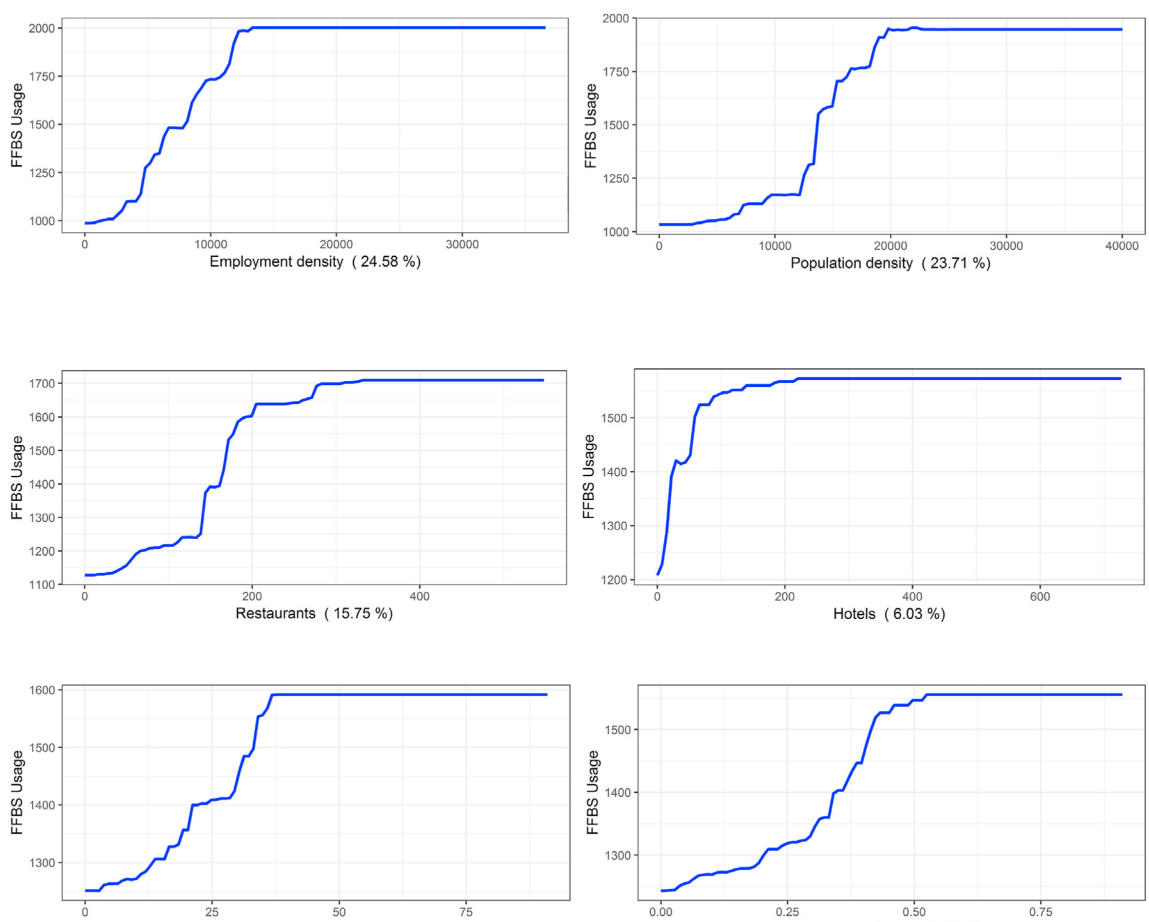

Fig. 4. Nonlinear effects of land use variables on FFBS usage.

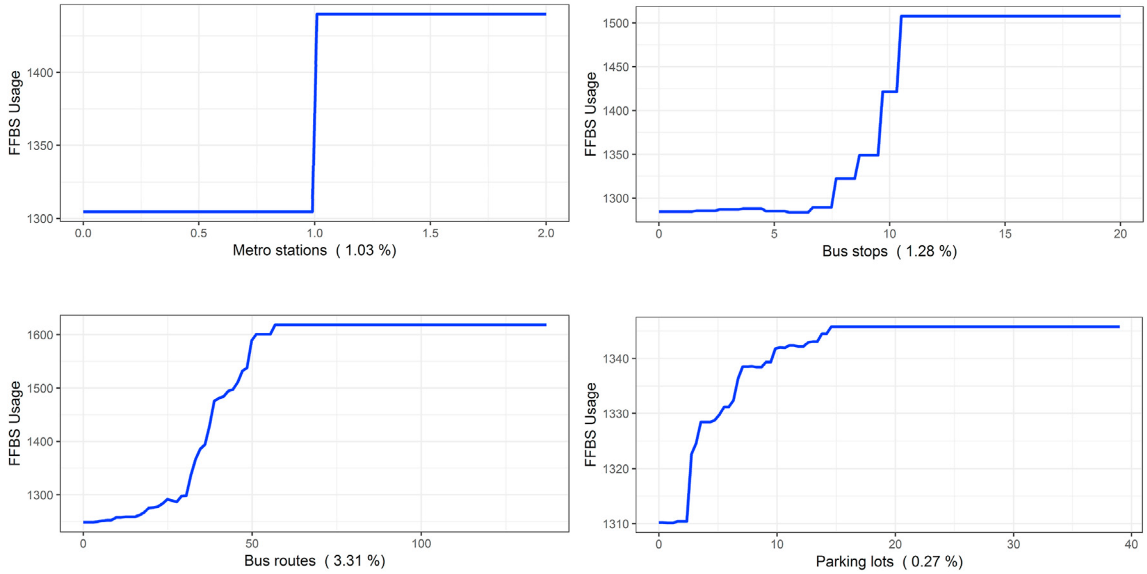

Fig. 5. Nonlinear effects of transport facility variables on FFBS usage.

density on the demand for bike sharing are almost nonexistent. Cycling trips increase rapidly after 7 stops/km^2 and the curve is saturated at 11 stops/km^2. It manifests that the connection between transit and bike sharing is strengthened when the density of bus stops is relatively high in a TAZ. Effects of bus routes on bike sharing usage are similar to that of bus stops. The demand for bike sharing is relatively stable from 0 to 30 routes/km^2 but dramatically increases after 30 routes/km^2. Bike sharing usage rapidly increases when the number of public parking lots ranges from 2 to 7. Then, it experiences a stable increase and remains unchanged after 15 parking lots.

As to the accessibility variables, their effects on bike sharing usage are shown in Fig. 6. Length of road per km^2 is positively associated with the demand for bike sharing. Within the range of 8–12, bike sharing usage linearly increases by about 360 trips. When the value of road density is more than 12, its effects on the increase of bike sharing usage remain stable. The effects of intersection density on cycling trips range from 20 to 35. Cycling trips then slightly increase and remain stable at 1570. Both road density and intersection density have positive influences on FFBS usage. This finding is consistent with the results of previous studies (Shen et al., 2018); (Xu et al., 2019). They capture the neighborhood connectivity and directness with which a traveler can traverse in a road network. High density of road and intersection indicates high accessibility, which provides convenient access to the neighborhood for travelers. In the interval of metro accessibility from 0.6 to 1.4, the demand for bike sharing rapidly increases, indicating free-floating bikes in the distance between 0.71 and 1.67 km to a metro station are inclined to be used to access the station. If the distance is beyond this range, taking a bus or walking may be a preferable choice for a traveler. As the distribution of bus stops is more denser than that of metro stations, people's tolerance of cycling distance to a bus stop is lower than that of a metro station. It is widely believed the catchment area size of a bus stop is less than that of a metro

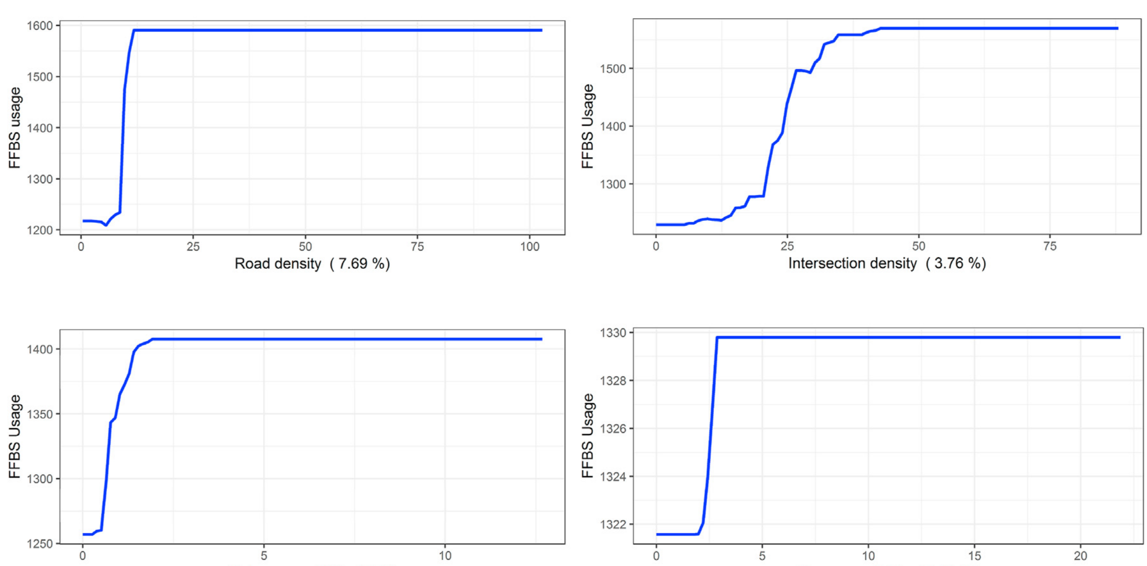

Fig. 6. Nonlinear effects of accessibility variables on FFBS usage.

station (Jiang et al., 2012). The effective range of bus accessibility is thus between 2 and 2.8. That is, free-floating bikes located in a distance between 0.36 and 0.5 km from a metro station can significantly influence their usage for travel to a bus stop.

6. Conclusions

This study provides a new perspective for investigating the impacts of built environment on FFBS usage. By building the proposed modeling framework, our work makes important contributions to the existing literature and puts forward policy suggestions for promoting bike sharing. We first aim to examine the relative importance of different built environment elements to the demand for bike sharing. It underscores the importance of independent variables influencing the predictive power. To identify the effective ranges and threshold effects of built environment elements, we then calculate and plot partial dependence to explore the nonlinear relationship between bike sharing usage and potentially contributing factors. As many Chinese cities are undergoing a transition to alleviate environmental degradation and traffic congestion, the free-floating bike sharing system, as an emerging mode of transport, has been widely adopted to encourage non-motorized trips and extend the reach and scope of the public transit. Solving these research problems can help operators efficiently allocate shared bikes for different areas and planners change the built environment elements to enhance the share of non-motorized trips.

The modeling results indicate that population and employment are two significant factors that collectively contribute to about 48% of the total effects on bike sharing usage. This underscores the critical role of commuting activities in influencing the demand for bike sharing. The relative influence of non-commuting activities on bike sharing usage accounts for 24%, including restaurants, leisure services, and other non-commuting elements. Hotels have a relative importance of 6.03% on bike sharing usage, indicating FFBS has gained a great popularity among tourists due to its flexible attributes. Industrial, hospital, and governmental areas have an insignificant influence on the cycling trips in a TAZ. Compared to land use factors, the collective contribution of transport facility factors to bike sharing is relatively low, indicating integration between bike sharing and other modes still needs to be further strengthened, especially for the transit modes (metro and bus). In addition, this study has found that road density and intersection density have stronger predictive power than intermodal accessibility indexes.

The modeling framework proposed by this study is more effective in unveiling the complex and nonlinear relationship between bike sharing usage and built environment elements than traditional linear models that have a limited capacity of identifying the varying patterns. When the employment density of a TAZ is over 2000 jobs per km^2, bike sharing usage experiences a rapid increase and remains stable up to 12,500 jobs per km^2. Population density has an effective influence on cycling trips when it exceeds 12,000 people per km^2, and its influence is saturated with 20,000 people. The impact of population density and employment density on bike sharing usage can help planners formulate the number of bikes launched in different areas of a city. Restaurants have the largest effects on cycling trips when their number reaches 200. Different from other variables, hotels significantly influence cycling trips at the early stage. Bike sharing has gained a great popularity among tourists, because it helps them enjoy meandering throughout a city without hassling with multiple transit transfers, high taxi fares and sore feet. When the number of leisure services reaches 37, it has the largest influence on bike sharing usage. The demand for bike sharing gradually increases until it reaches the threshold of commercial areas with the value being 0.52. These findings will help operators to allocate the appropriate number of bikes to meet the demand in areas with different land use types. In turn, the findings also offer guidance for communities implementing bicycle-oriented development to adjust land use types and invest in corresponding cyclist-friendly infrastructure.

Bike sharing is regarded as a solution to the first and last mile issue of transit connectivity. A metro station could increase the number of cycling trips by 130 in a day. Bus stops and bus routes would have the largest impacts on bike sharing usage when the number of bus stops reaches 11 and the number of bus routes reaches 60. The parking lot becomes effective when there are more than 2 and its effect is saturated when the number reaches 15. When road density reaches 12 km/km^2, it has the largest effect on bike sharing usage. The demand for bike sharing gradually increases until it reaches the threshold of intersection density with the value being 35. When bicycles are allocated in the distance between 0.71 and 1.67 km to a metro station, it can effectively influence the usage of bike sharing to access the metro. For a bus stop, the effective range is between 0.36 and 0.5 km. These findings provide concrete evidence for promoting sustainable transportation, which contributes to forming a powerful and positive image of shared resource economies and creating a vibrant, sustainable, and eco-friendly community. Whether residents cycle or not, they stand to benefit from healthier communities and cleaner air.

Although this study conducts a comprehensive analysis of the nonlinear relationship between bike sharing usage and built environment elements, it also has several limitations. First, as this paper empirically investigates the effective ranges and threshold effects of built environment elements in Chengdu (a monocentric city), findings of this study may not be directly transferable to other cities with different geographic and geometric properties. Therefore, researchers are encouraged to conduct more empirical studies using the modeling framework in other cities (e.g., a polycentric city) worldwide. Second, the demand for bike sharing varies during different periods in a day. Future studies may investigate divergent threshold effects of built environment on time-varying bike sharing usage. Third, this paper focuses on built environment factors influencing the demand for bike sharing. Other factors at the sociodemographic level are omitted, which may potentially affect the number of cycling trips.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study is sponsored by the National Key R&D Program of China (No. 2018YFB1601000) and the Scientific Research Foundation of Graduate School of Southeast University (No. YBPY1976).

References

Alcorn, L.G., Jiao, J., 2019. Bike-Sharing station usage and the surrounding built environments in major Texas cities. J. Plann. Educ. Res. 1–14 https://doi.org/10.1177/0739456X19862854.

Chen, E., Ye, Z., Wang, C., Zhang, W., 2019. Discovering the spatio-temporal impacts of built environment on metro ridership using smart card data. Cities 95, 102359. https://doi.org/10.1016/j.cities.2019.05.028.

Chen, E., Ye, Z., Wang, C., Xu, M., 2020. Subway passenger flow prediction for special events using smart card data. IEEE Trans. Intell. Transport. Syst. 21 (3), 1109–1120. https://doi.org/10.1109/TITS.2019.2902405.

Cityif, 2018. Do we still need station-based bikes with the rise of free-floating bike sharing (in Chinese). http://www.sohu.com/a/222173312_651721/. (Accessed 8 January 2020).

Ding, C., Cao, X.J., Næss, P., 2018. Applying gradient boosting decision trees to

examine non-linear effects of the built environment on driving distance in Oslo. Transport. Res. Part A. 110, 107−117. https://doi.org/10.1016/j.tra.2018.02.009.
Du, Y., Deng, F., Liao, F., 2019. A model framework for discovering the spatio-temporal usage patterns of public free-floating bike-sharing system. Transport. Res. Part C. 103, 39−55. https://doi.org/10.1016/j.trc.2019.04.006.
Eren, E., Uz, V.E., 2019. A review on bike-sharing: the factors affecting bike-sharing demand. Sust. Cities Soc. 54 https://doi.org/10.1016/j.scs.2019.101882, 101882.
Faghih-Imani, A., Hampshire, R., Marla, L., Eluru, N., 2017. An empirical analysis of bike sharing usage and rebalancing: evidence from Barcelona and Seville. Transport. Res. Part A. 97, 177−191. https://doi.org/10.1016/j.tra.2016.12.007.
Feng, P., Li, W., 2016. Willingness to use a public bicycle system: an example in Nanjing City. J. Publ. Transp. 19 (1), 6. https://doi.org/10.5038/2375-0901.19.1.6.
Fishman, E., Washington, S., Haworth, N., 2013. Bike share: a synthesis of the literature. Transport Rev. 33 (2), 148−165. https://doi.org/10.1080/01441647.2013.775612.
Fishman, E., Washington, S., Haworth, N., 2014. Bike share's impact on car use: evidence from the United States, Great Britain, and Australia. Transport. Res. Part D. 31, 13−20. https://doi.org/10.1016/j.trd.2014.05.013.
Friedman, J.H., 2002. Stochastic gradient boosting. Comput. Stat. Data Anal. 38 (4), 367−378. https://doi.org/10.1016/S0167-9473(01)00065-2.
Gu, T., Kim, I., Currie, G., 2019. To be or not to be dockless: empirical analysis of dockless bikeshare development in China. Transport. Res. Part A. 119, 122−147. https://doi.org/10.1016/j.tra.2018.11.007.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, second ed. Springer, New York.
Hua, M., Chen, X., Zheng, S., Cheng, L., Chen, J., 2020. Estimating the parking demand of free-floating bike sharing: a journey-data-based study of Nanjing, China. J. Clean. Prod. 244, 118764. https://doi.org/10.1016/j.jclepro.2019.118764.
Ji, Y., Ma, X., He, M., Jin, Y., Yuan, Y., 2020. Comparison of usage regularity and its determinants between docked and dockless bike-sharing systems: a case study in Nanjing, China. J. Clean. Prod. 255, 120110. https://doi.org/10.1016/j.jclepro.2020.120110.
Jiang, Y., Zegras, P.C., Mehndiratta, S., 2012. Walk the line: station context, corridor type and bus rapid transit walk access in Jinan, China. J. Transport Geogr. 20 (1), 1−14. https://doi.org/10.1016/j.jtrangeo.2011.09.007.
Jun, M.J., Choi, K., Jeong, J.E., Kwon, K.H., Kim, H.J., 2015. Land use characteristics of subway catchment areas and their influence on subway ridership in Seoul. J. Transport Geogr. 48, 30−40. https://doi.org/10.1016/j.jtrangeo.2015.08.002.
Kaplan, S., Manca, F., Nielsen, T.A.S., Prato, C.G., 2015. Intentions to use bike-sharing for holiday cycling: an application of the Theory of Planned Behavior. Tourism Manag. 47, 34−46. https://doi.org/10.1016/j.tourman.2014.08.017.
Li, H., Zhang, Y., Ding, H., Ren, G., 2019. Effects of dockless bike-sharing systems on the usage of the London Cycle Hire. Transport. Res. Part A. 130, 398−411. https://doi.org/10.1016/j.tra.2019.09.050.
Lin, P., Weng, J., Liang, Q., Alivanistos, D., Ma, S., 2018. Impact of weather conditions and built environment on public bikesharing trips in Beijing. Network. Spatial Econ. 1−17 https://doi.org/10.1007/s11067-019-09465-6.
Liu, H.C., Lin, J.J., 2019. Associations of built environments with spatiotemporal patterns of public bicycle use. J. Transport Geogr. 74, 299−312. https://doi.org/10.1016/j.jtrangeo.2018.12.010.
Mateo-Babiano, I., Bean, R., Corcoran, J., Pojani, D., 2016. How does our natural and built environment affect the use of bicycle sharing? Transport. Res. Part A. 94, 295−307. https://doi.org/10.1016/j.tra.2016.09.015.
Molnar, C., 2019. Interpretable Machine Learning. Lean Publishing.
Qian, X., Ukkusuri, S.V., 2015. Spatial variation of the urban taxi ridership using GPS data. Appl. Geogr. 59, 31−42. https://doi.org/10.1016/j.apgeog.2015.02.011.
Shen, Y., Zhang, X., Zhao, J., 2018. Understanding the usage of dockless bike sharing in Singapore. Int. J. Sustain. Transp. 12 (9), 686−700. https://doi.org/10.1080/15568318.2018.1429696.
Si, H., Shi, J.G., Wu, G., Chen, J., Zhao, X., 2019. Mapping the bike sharing research published from 2010 to 2018: a scientometric review. J. Clean. Prod. 213, 415−427. https://doi.org/10.1016/j.jclepro.2018.12.157.
Ta, N., Chai, Y., Zhang, Y., Sun, D., 2017. Understanding job-housing relationship and commuting pattern in Chinese cities: past, present and future. Transport. Res. Part D. 52, 562−573. https://doi.org/10.1016/j.trd.2016.11.011.
Tan, J., Zhao, Z., 2018. Bike-sharing growth hits the brakes. https://www.caixinglobal.com/2018-06-22/bike-sharing-growth-hits-the-brakes-101283480.html. (Accessed 8 January 2020).
Tu, W., Cao, R., Yue, Y., Zhou, B., Li, Q., Li, Q., 2018. Spatial variations in urban public ridership derived from GPS trajectories and smart card data. J. Transport Geogr. 69, 45−57. https://doi.org/10.1016/j.jtrangeo.2018.04.013.
Van Wee, B., Handy, S., 2016. Key research themes on urban space, scale, and sustainable urban mobility. Int. J. Sustain. Transp. 10 (1), 18−24. https://doi.org/10.1080/15568318.2013.820998.
Wang, K., Chen, Y.J., 2020. Joint analysis of the impacts of built environment on bikeshare station capacity and trip attractions. J. Transport Geogr. 82, 102603. https://doi.org/10.1016/j.jtrangeo.2019.102603.
Xu, C., Ji, J., Liu, P., 2018. The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets. Transport. Res. Part C. 95, 47−60. https://doi.org/10.1016/j.trc.2018.07.013.
Xu, Y., Chen, D., Zhang, X., Tu, W., Chen, Y., Shen, Y., Ratti, C., 2019. Unravel the landscape and pulses of cycling activities from a dockless bike-sharing system. Comput. Environ. Urban Syst. 75, 184−203. https://doi.org/10.1016/j.compenvurbsys.2019.02.002.
Yang, Y., Heppenstall, A., Turner, A., Comber, A., 2019. A spatiotemporal and graph-based analysis of dockless bike sharing patterns to understand urban flows over the last mile. Comput. Environ. Urban Syst. 77, 101361. https://doi.org/10.1016/j.compenvurbsys.2019.101361.
Zhang, Y., Haghani, A., 2015. A gradient boosting method to improve travel time prediction. Transport. Res. Part C. 58, 308−324. https://doi.org/10.1016/j.trc.2015.02.019.
Zhang, Y., Thomas, T., Brussel, M., Van Maarseveen, M., 2017. Exploring the impact of built environment factors on the use of public bikes at bike stations: case study in Zhongshan, China. J. Transport Geogr. 58, 59−70. https://doi.org/10.1016/j.jtrangeo.2016.11.014.
Zhao, D., Ong, G.P., Wang, W., Hu, X.J., 2019. Effect of built environment on shared bicycle reallocation: a case study on Nanjing, China. Transport. Res. Part A. 128, 73−88. https://doi.org/10.1016/j.tra.2019.07.018.