Mapping connection and substitution behaviors between shared E-bicycles and public transportation

Original
Mapping connection and substitution behaviors between shared E-bicycles and public transportation / Ye, J., Bai, J., Diana, M.. - In: TRANSPORTATION. - ISSN 0049-4488. - STAMPA. - 53:(2026), pp. 1911-1944. [10.1007/s11116-024-10550-0]

Availability:
This version is available at: 11583/2997024 since: 2026-07-30T09:15:46Z

Publisher:
Springer

Published
DOI:10.1007/s11116-024-10550-0

(Article begins on next page)

Mapping connection and substitution behaviors between Shared E-bicycles and Public Transportation

Jianhong Ye, Jiahao Bai, Marco Diana

9th September 2024

Preferred citation: this document may be cited directly referring to the above mentioned final published version:

Ye, J., Bai, J., & Diana, M. (2026). Mapping connection and substitution behaviors between shared E-bicycles and public transportation. *Transportation*, 53, 1911–1944, doi: 10.1007/s11116-024-10550-0.

Mapping connection and substitution behaviors between Shared E-bicycles and Public Transportation

Jianhong Ye[1,2] · Jiahao Bai[1,2*] · Marco Diana[3]

Abstract The emergence and development of shared E-bicycles (SEB) has resulted in complex connection and substitution relationships with public transportation (PT) in urban areas, impacting the social benefits of integrated transportation systems and the return on investment of PT subsidies. Accurately identifying how SEB complement and/or compete with PT is still a crucial challenge. Therefore, this research proposes a new discrimination method based on multi-source data fusion to address this issue. The method's discrimination accuracy is evaluated based on both telephone interviews data and transaction data from SEB users in Shenyang City, China. The results indicate that SEB are primarily used for single-mode trips (75%). When SEB are used to combine with other modes, they mainly serve as connecting modes to rail transit (22.1%). On the other hand, shared bicycles (SB) and buses are the two modes most replaced by SEB, followed by private cars and walking. 63% of SEB rides can be accurately classified by the proposed method in terms of their connection and substitution relationship with PT. Incorporating anonymized socioeconomic attribute information from big data holds promise for further enhancing the method's performance. This research provides key insights for assessing the societal benefits of SEBs on PT, offering valuable theoretical and methodological support for transportation planning.

Keywords: shared e-bicycles; transit; complementarity and substitution; multimodality; multinomial logit; random forest.

1 Introduction

The emergence and development of shared mobility services, including shared electric bicycles (SEBs), ride-hailing, and car-sharing, have significantly influenced people's travel behaviors and the mode share of traditional travel modes in many urban areas around the world. Shared bikes (SBs) emerging prior to SEBs have demonstrated a notable substitution behavior with public transportation (PT), which means that trips that were previously completed by PT, particularly by buses, are diverted to SBs when the latter service appears (Wang, 2017, Tang et al., 2011, Yang et al., 2016, Ye et al., 2024). Additionally, SBs also form connection behavior with PT (Ji et al., 2016, Lin et al., 2018, Molinillo et al., 2019, Ye et al., 2024), since SB are used as access or egress mode between transit stops and trips origins or destinations. More recently, SEBs have demonstrated a stronger substitution behavior with PT than SB due to their advantages in speed (Campbell et al., 2016, Lazarus et al., 2020).

Beyond substitution effects, previous research has also investigated the role of SB and SEB connecting to PT, in order to complement fixed line services. The riding distance of SEB is significantly longer than that of SB (Lazarus et al., 2020), which means SEBs may promote a greater distance range ridings to connect to PT than SBs. Jointly considering both connection and substitution behaviors, SEBs can both promote and inhibit PT ridership, therefore the net impact on PT ridership is vague. Developing PT has always been a key strategy to address urban transportation challenges such as traffic congestion and environmental pollution, and it has also become one of the main parts of fiscal subsidies. The intricate connection and substitution behavior between SEB and PT directly impacts the social benefits derived from PT and the effectiveness of fiscal subsidies. Therefore, accurately discriminating the connection and substitution behaviors between the two modes of transportation is an essential work for the assessment of the impact of SEB on PT ridership, the social benefits derived from PT and the effectiveness of public fiscal subsidies for PT.

Jianhong Ye
E-mail: 12107@tongji.edu.cn
Jiahao Bai (corresponding author)
E-mail: bjh0321@tongji.edu.cn
Postal address: Tongji University, 4800 Cao'an Road, Shanghai, 201804, China
Marco Diana
E-mail: marco.diana@polito.it

[1] Urban Mobility Institute, Tongji University, 4800 Cao'an Road, Shanghai, 201804, China
[2] The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, 4800 Cao'an Road, Shanghai, 201804, China
[3] Department of Environment, Land and Infrastructure Engineering, Politecnico di Torino, Turin, Italy

The aforementioned connection and substitution behaviors can be jointly referred to as interactive behaviors. More in details, the interactive behaviors between SEB and PT consists of four different behaviors: connection and substitution behavior with PT, connection and non-substitution behavior with PT, non-connection and substitution behavior with PT and non-connection and non-substitution behavior with PT. It is noteworthy that connection and substitution are not necessarily mutually exclusive, since for example the use of SEB could shorten the travel distance covered by PT to complete a given trip, without completely substituting PT itself. We finally note that interactive behaviors are not completely explaining the demand for SEB, since part of it is induced and therefore related trips would not have been made at all if the shared service had not existed. However, the focus of this paper is on interactive behaviors rather than on the complete characterization of SEB demand, therefore induced trips will be not considered in the following.

Three main methods have been used in the literature to discriminate and analyze interactive behaviors between PT and different shared micro-mobility modes (SMMs), including SBs (Kong et al., 2020, Wu et al., 2022), shared E-scooters (Vinagre Díaz et al., 2023, Luo et al., 2021) and SEBs (Suchanek et al., 2021, Fukushige et al., 2021). The first method is based on administering a survey (Wang, 2017, Tang et al., 2011, Yang et al., 2016), where micro-mobility services users are asked about their modal choices if the service were not available and whether they rode micro-mobility to connect to other travel modes (refer to the following subsection 2.1.1 for details). This type of method can accurately identify the interactive behaviors, but only for those specific trips that are considered in the survey. Therefore, the characterization of all trips involving the use of SMMs and the actual share of each kind of interactive behaviors cannot be directly observed, only statistically inferred at best.

The second method is based on the analysis of transaction data from the service operator (Kong et al., 2020, Wu et al., 2022), which infers interactive behaviors based on the spatial-temporal characteristics of recorded trips (refer to the following subsection 2.1.2 for details). Currently, this type of method simply relies on the relationship between riding distance, distances between the riding starting or ending points and their respective nearest PT stops, and a threshold, which is based on empirical judgment, ignoring the complex and varying modal interactions that might occur in practice in different situation. Furthermore, due to the lack of observed behaviors, this method simplistically categorizes interactive behaviors into connection behavior, substitution behavior, and supplementary behavior whenever the former two behaviors are not observed. Therefore, we can observe that this method treats substitution and connection behaviors as mutually excluding, overlooking situations where both substitution and connection behaviors occur simultaneously, as discussed above. The third method is based on utility theory (refer to the following subsection 2.1.3 for details) and to date and to the best of our knowledge it was only used to model substitution behaviors between SMMs and other modes of transportation, without considering the connection behaviors.

Therefore, the first contribution of this study is to propose a more comprehensive method framework, according to which interactive behaviors can be categorized into the above mentioned four different cases, thus overcoming the simplistic assumption of mutually exclusive connection and substitution behaviors. To achieve this result, transaction data will be jointly analyzed with survey data, thus merging two of the above-mentioned methods and showing that this approach can outperform the sole consideration of transaction-based spatio-temporal relationships. Besides that, the second contribution is that we analyze the interactive behaviors adopting a multimodality perspective, thus considering impacts on all modes used to complete a trip, rather than only considering how the main mode (usually referred to as the longest-distance mode in a trip) is affected by SEB.

2 Literature review

2.1 Interactive behaviors and their discrimination methods between SMMs and PT

As mentioned above, three methods are used in the existing literature to research the interactive behaviors between SMMs and PT. Each method and related applications will be introduced below.

2.1.1 Questionnaire-based methods

The questionnaire-based method directly observes connection and substitution behaviors by asking participants specific questions. Respondents are typically asked to recall the most recent trip involving the use of a SEB (or another kind of SMMs) and to indicate what travel modes would have been used for the trip, if no SEB were available (Fukushige et al., 2021, Teixeira et al., 2020, Fishman, 2015, Wang et al., 2022). Other researchers used a different approach, namely retrospective general questions (LDA Consulting, 2017), thus asking what travel mode has the SMM most often replaced (or how has the use of other travel modes changed) since the respondent started using the SMM. Additionally, connection behaviors can be obtained by asking to detail all travel modes beyond micro-mobility that have been used to complete a trip, or that are typically used jointly with micro-mobility services. It is however important to note that only approximate figures on substitution and connection behaviors can be obtained for all trips with such trip-specific or general questions.

Coming to findings of works using this approach, SBs, as one of the most early developing and researched SMMs, were found acting a crucial role in serving the "last mile" of a PT journey. A survey conducted in Malaga province, Spain, showed that SB most frequently combined with PT (14.1%), while single mode SB trips were 79.6% and combination with other modes were 6.3% (Molinillo et al., 2019). In an earlier work in Dublin, 39% of SB trips involved the use of other modes, of which 22% are combined with rail and 14% are combined with bus (Murphy and Usher, 2014). In contrast, Asian cities show larger percentages of multimodal SB trips, with 50.2% being combined with PT while 48.8% being single-mode trips (Wang, 2017). Concerning shared e-scooters, a survey conducted in France indicated that 23% of respondents combined shared e-scooter with other modes during their last trip, with 66% of these 23% combining shared e-scooter with PT (6t-Bureau De Recherche, 2019). Similar results were found in Oslo, where 57% of the last shared scooter trip was multimodal, in most cases in combination with PT and especially with rail transit (Fearnley et al., 2020, Aarhaug et al., 2023). In comparison, there is relatively limited research on SEBs. A survey conducted in three Poland cities, Gdansk, Gdynia and Sopot, showed 39% of respondents have ever used shared E-bicycles to connect to PT (Bieliński et al., 2021).

In terms of substitution, it is proved that most trips substituted by SB derive from sustainable travel modes (especially from PT and walking), with only a small part shifting from car (Teixeira et al., 2020). In a meta-analysis of several researches in this field, the median percentage of respondents replacing car trips with SB was 10%, whereas the median of replaced trips from PT or walking was 70% (Teixeira et al., 2020). Even though PT and walking are the most substituted modes in absolute terms, it is in general the mostly used mode in a given city that shows stronger substitution patterns (Fishman et al., 2014). For example, some North American cities (Fishman et al., 2014, Shaheen et al., 2014) have higher car substitution but lower PT substitution, while European (Midgley, 2011, Transport for London, 2010) and Asian cities (Wang, 2017, Tang et al., 2011) showed much higher PT substitution but lower car substitution. Concerning shared e-scooters, walking and PT are the most substituted modes in European cities (Christoforou et al., 2021, Fearnley et al., 2020, Laa and Leth, 2020). Conversely, taxis/TNCs (25%) and private cars (14%), only behind walking (41%), outnumbered PT (9%) in terms of median of substitution rates in 23 U.S. cities, according to the review by Wang et al. (2022). Finally, there is limited research on the substitution rate of SEBs for other travel modes. Bieliński et al. (2021) examined an electric bike-sharing system in Tricity, Poland, and found that SEBs were predominantly used as substitutes for PT (48.5%) or for walking (21.4%). A similar result was found by Campbell et al. (2016) from a SP experiment in Beijing, China, to estimate the likelihood of travelers switching to conventional SBs or to SEBs: PT travelers mostly preferred to shift to SEBs, followed by walking travelers. Another survey conducted in Sacramento, California, found that a higher percentage of PT trips were replaced by SEB for commuting purposes than for non-commuting purposes (Fukushige et al., 2021).

The questionnaire-based method obtains interactive behaviors by directly asking the respondents, ensuring the accuracy of the results for trips under analysis. However, due to both maximum sample size and respondent burden issues such trips are only a tiny minority of all trips. Statistical results are then inferred on the universe of trips; however, they may have estimation errors due to sampling error, and possibly sampling bias depending on the survey design. This bias is then affecting the substitution rate derived from the pool of observed trips and projected to all trips carried out in the study area e.g. to estimate environmental effects (Zhu and Lu, 2023, Felipe-Falgas et al., 2022). Moreover, in case of multimodal trips, current research using this method only roughly considers whether the mode used to cover the longest distance is substituted by shared micro-mobility. For example, when SEB substitute a combination of SEBs and PT, current surveys typically only capture the substitution of PT but they cannot analyze the actual situation in case both shared electric bicycles and PT are simultaneously substituted.

2.1.2 Methods based on transaction data analysis

The discrimination method based on transaction data analysis firstly derives specific variables from analyzing transaction data of shared micro-mobility operators, along with the location of transportation facilities (shared means pickup and drop-off points, train stations, bus stops, ...) and operational data of PT services. In reviewing applications of this method, we first introduce a common framework based on the definition of the following quantities:

- The first is the radius, r_c as shown in Figure 1, of the potential circular area, A, around a PT station within which the drop-off or pickup of SMM is considered as a connection behavior with a PT service in that station (Wu et al., 2022, Kong et al., 2020). Trips 1-2 in Figure 1 would have been categorized into connection behaviors. Other cases are discussed below.

- The second is the radius, r_w, of the potential circular area, B, around a PT station to detect substitution behaviors. If a SEB trip, like trips 5-7 in Figure 1, have origin and destination falling in two different areas B, and also there is a PT service between these two areas, then the SEB trip is potentially substituting a PT trip (Wu et al., 2022, Kong et al., 2020). r_w is normally larger than r_c and it depends on the maximum traveler's walking distance.
- The third is the maximum riding distance, d_c, by SMM suitable for connecting to PT (Wu et al., 2022). As an example, trip 3 in Figure 1 should have been categorized as connection behavior. However, it is classified as supplementary behavior because connection trips are usually shorter than other trips. Current research suggests that long-distance SMM trips, with a distance exceeding d_c, are not in connection with PT, even if one of their endpoints is within area A as shown in Figure 1.
- The fourth is t_c, i.e. the maximum riding duration by SMM suitable for connecting to PT (Kong et al., 2020), which is sometimes considered instead of d_c, using it to classify supplementary behaviors as shown again by trip 3 in Figure 1.
- The above four quantities, reflecting spatial, technological and economic characteristics of SMMs trips are commonly used in current studies. Moreover, some other studies introduce additional quantities, reflecting the attributes of PT and the differences between PT and SMM, as introduced below: w is the maximum time interval between the SMM pickup or drop-off event and the PT service time (i.e. arriving or leaving at a given station) that is considered suitable to consider a connection between the two services (Liu et al., 2022, Luo et al., 2021, Qiu and Chang, 2021).
- f is the maximum number of transfers of a PT service between two stations to consider such service as potentially substituted by SMM (Kong et al., 2020). A SMM trip like trips 5-7 in Figure 1 is not classified as substitution, but rather has complementary behavior, if the f value of the potentially substituted PT trip is above the threshold.
- h is the maximum ratio of the in-vehicle distance of the PT trip over the riding distance of SMM for the same origin and destination (Luo et al., 2021). Like for f, it is considered to distinguish potential substitution behavior from potential supplementary behavior in cases like trips 5-7 in Figure 1.

In many of the below reviewed studies, these values are assumed based on conventional wisdom or practical experience. Based on these parameters, the following sets and variables can then be defined:

- S_s is the set of PT stations within a circular radius of r_w from the starting point of a SMM trip. Similarly, S_e is the set of PT stations within a circular radius of r_w from the ending point of a SMM trip (Wu et al., 2022, Kong et al., 2020).
- SS_s is the set of PT stations within a circular radius of r_c from the starting point of a SMM trip, while SS_e is the set of PT stations within a circular radius of r_c from the ending point of a SMM trip (Vinagre Díaz et al., 2023, Wu et al., 2022).
- Y_p is a binary variable indicating whether at least one PT service is available from any station in S_s to any station in S_e (Wu et al., 2022, Kong et al., 2020). In case there is more than one PT service, the most convenient one is considered along with the corresponding starting station ss, $ss \in S_s$, and ending station se, $se \in S_e$. The most convenient PT route is the route with the shortest total travel time, denoted as rt_{min}, among the all-possible routes from any station in S_s to any station in S_e (Luo et al., 2021).
- F is the number of transfers required when opting for PT to travel between ss and se (Kong et al., 2020).
- T_{pt}, is the travel time by PT from ss to se, including walking times (Luo et al., 2021).
- Y_{pt} is a binary variable, indicating whether the starting time of a SMM trip falls within the PT operation time range (Wu et al., 2022).
- D_c is the riding distance of a SMM trip (Wu et al., 2022).
- T_c, is the riding duration of a SMM trip (Kong et al., 2020).
- W_e is the minimum positive time interval between the ending time of a SMM trip and the departure time of a PT service from the PT station ss to consider the two trips as connected. Similarly, W_s is the minimum positive time interval between the arrival time of a PT service in the PT station se and the start time of a SMM trip (Luo et al., 2021).
- D_{pt} is the travelled distance from ss to se when using PT (Luo et al., 2021).

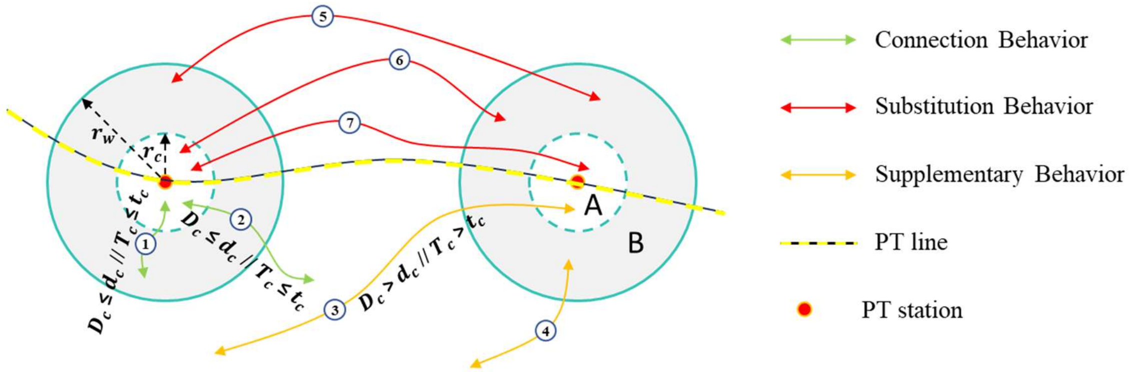

Fig. 1 The spatial characteristics of potential interactive behaviors.

Several algorithms can be envisaged to classify SMM trips from transactions. They typically first distinguish between connection and complementary behaviors, and then between substitution and complementary behaviors. An example of classification algorithm based on the above introduced notation, which is a useful reference to review the literature in this sector, is reported in Figure 2.

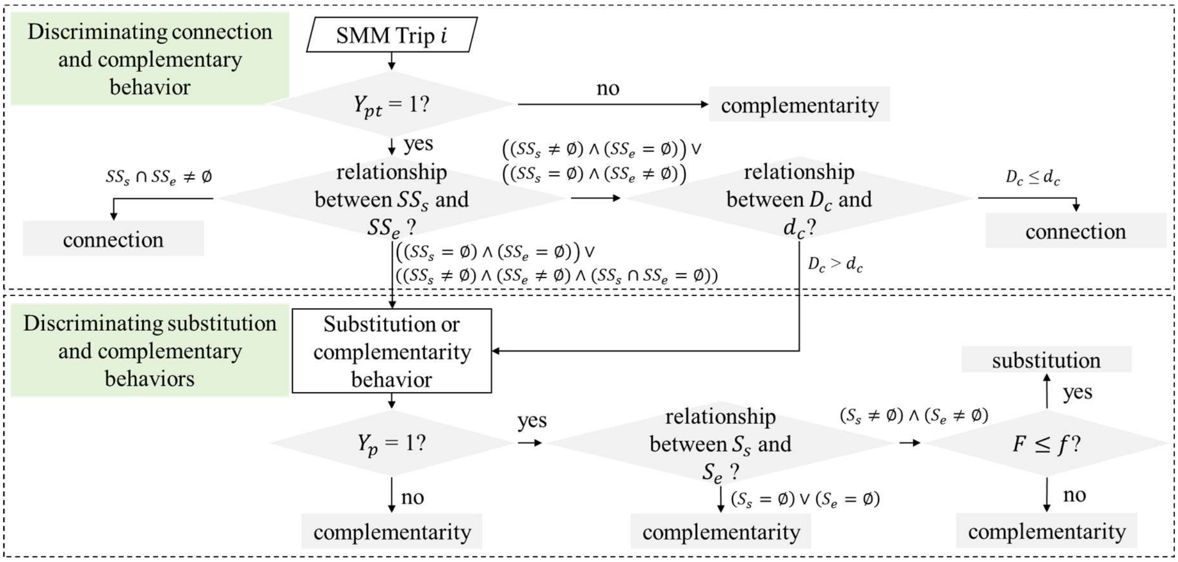

Fig. 2 Discrimination logic for potential interactive behaviors.

Some studies explored the spatiotemporal characteristics of interactive behaviors between SMM and PT across different city contexts using the aforementioned method. Wu et al. (2022) examined the interactive behaviors between the dockless SB system and metro system in Shanghai based onY_{pt}, S_s, S_e, and D_c and using a simpler discrimination algorithm than the one shown in Figure 2. Their findings revealed that dockless SB primarily serves to connect and complement the metro system. Kong et al. (2020) investigated the interactive behaviors between dock-based SB and PT in four American cities utilizing all the variables from Figure 2, except considering T_c rather than D_c. They found that on weekends, substitution behaviors are most prominent, followed by connection behaviors. On weekdays, connection behaviors exceed substitution behaviors. In terms of shared e-scooter, Luo et al. (2021) also considered all the variables from Figure 2, along with three above introduced additional variables, namely W_e, W_s, and D_{pt}. They discovered that, in Indianapolis, most of the shared e-scooter trips were more likely to substitute the existing bus service. In that city, the shared e-scooter system mainly connects or supplements the bus system during the time when bus operation is limited or in areas with low bus coverage. Conversely, in Roma, the majority of the e-scooter trips in Rome are connected with subway or railway stations (Vinagre Díaz et al., 2023).

This type of method can effectively discriminate interactive behaviors for all recorded SMM trips by examining the value of some of the above defined quantities, based on empirical rules such as those reported in the flowchart of Figure 2. These values are typically derived from experience (Kong et al., 2020), sensitivity analysis (Wu et al., 2022), or clustering algorithms (Vinagre Díaz et al., 2023). However, the resulting interactive behaviors are potential rather than actual, as the discriminating values of the parameters are determined based on aggregate-level considerations. Yin et al. (2024) verified that the actual connection behavior between shared e-scooter and rail transit significantly differ from the potential connection behavior among quantitative, spatial and temporal dimensions. Moreover, this type of method can only categorize potential interactive behaviors into three types: potential connection, potential substitution, and potential complementarity, without considering the possibility of simultaneous connection and substitution behavior, that could happen, for example, when a SMM replaces a less convenient bus connection to rail transit, the latter still being used.

2.1.3 Method*s based on utility theory*

Only few studies elicit interactive behaviors from the discrimination method based on utility theory. To the best of our knowledge, two studies utilized the mixed logit model to derive substitution behaviors between SMMs and other modes, including PT (Li et al., 2021, Reck et al., 2022). Despite the fact that both studies use the mixed logit model, they differ in their ideas.

Li et al. (2021) utilized a mode choice behavior survey conducted before the introduction of SBs in Shanghai in 2017 and SB transaction data recorded from August to September 2018. They calibrated a mixed logit model with the choice set containing walking, private bike, PT and car, and independent variables included in-vehicle time, walking time, access/egress time and transfer number. The study assumes that the introduction of SBs is expected to replace the existing mode among the previously listed one having the maximum utility when used to complete the same recorded SBs trips Consequently, independent variables extracted from each SB trip was input into the mixed logit model, which predicts those modes that are substituted by SBs.

On the other hand, Reck et al. (2022) utilized observed mode choice data tracked by GPS, collected during the operation of SMM, to calibrate the mixed logit model. The choice set included PT, car, private bike, private E-scooter, private E-bike, SEB and shared E-scooter. Subsequently, in this study, all observations related to shared E-scooter trips were selected, and their characteristics were input into the mixed logit model. The alternative mode having the maximum utility, excluding shared E-scooter, was then considered as the substituted mode.

Although this type of method has its strengths, it was used specifically to study substitution behaviors between SMM and other modes, without considering the connection behavior. Additionally, and in common with questionnaire-based methods, the focus in only on the mode that covers the longest distance in case of multimodal trips. Moreover, this is an indirect method for which the travel mode with maximum utility, excluding SMM, is consider as the substituted mode. However, travelers might replace the mode with second-largest or third-largest utility with SMM due to travelers' bounded rationality or model mis-specification. Directly asking the travel mode substituted by SMMs and then including it in the choice set might thus be preferable.

2.2 Summary and contribution

In conclusion, a substantial body of research has been dedicated to discriminating interactive behaviors between shared SMM and PT, resulting in three types of primary methods: the questionnaire-based discrimination method, the method based on transaction data analysis and the method based on utility theory. However, each of these methods has some limitations as summarized in Table 1.

Table 1 Existing methods for discriminating interactive behaviors and their drawbacks.

Method	Discrimination Principle	Drawbacks
Questionnaire-based Method	Directly asking the respondents.	1. It can only investigate interactive behaviors in small samples. 2. Estimation errors on the universe of trips due to possible sampling bias, limited sample size and respondent burden. 3. For multimodal trips, it can only assess whether the longest distance mode is substituted by SMM.
Transaction Data Analysis	Examining the spatiotemporal relationship between SMM trips and public transportation.	1. It can discriminate interactive behaviors for all recorded SMM trips, but the results are hypothetical rather than actual. 2. Problems in jointly considering connection and substitution behaviors.
Utility Theory	The travel mode with maximum utility, excluding SMM, is considered as the substituted mode.	1. Typically used to study substitution behavior between SMM and other modes, without considering connection behavior. 2. For multimodal trips, it usually only considers whether the longest distance mode is substituted by SMM. 3. It is an indirect method that may cause inaccurate predictions due to the behavioral assumptions of econometric models.

These limitations could be overcome by developing discrimination methods that can combine the advantages of survey data and transaction data to achieve more accurate discrimination of interactive behaviors at the trip level. Additionally, as mentioned in the introduction section, SMMs and PT can give rise to four types of interactive behaviors. Up to now, none of the above-mentioned studies is effectively discriminating these four behaviors.

Thus, this study introduces a novel discrimination method, named the interactive behaviors discrimination method based on multi-source data fusion and utility (IBMMU). This method integrates survey and transaction data, within a utility maximization framework. Unlike existing methods, IBMMU achieves discrimination of interactive behaviors between SMMs and PT for all trips in a transactional dataset also based on empirical evidence. Additionally, it can simultaneously consider both connection and substitution behaviors in all legs of multimodal trips. Furthermore, the method can work with both econometric models and machine learning approaches.

3 Data and modelling framework

3.1 Study area

The study area is located in Shenyang, a city situated in the northeast of China, comprising 10 districts, 1 county-level city, and 2 counties, covering a total area of 12,860 km² with a population of 7.56 million and a population density of 588 people per km². The downtown area of Shenyang includes Heping District, Shenhe District, Dadong District, Huanggu District, and Tiexi District, spanning an area of 571 km² with a population of 3.93 million and a population density of 6,889 people per km². In 2020, Shenyang's GDP reached 657.16 billion yuan, ranking third in China, with a per capita GDP of 72,800 yuan. The average number of private cars per hundred people in the city is 19.8. By the end of 2020 when field activities took place, Shenyang had four operational subway lines, covering a total length of 214.59 km, ranking 13th in China. Shared bicycles were introduced in Shenyang in 2017, followed by shared electric bicycles in 2018. As of the end of 2020, three major shared mobility companies, namely Hello, Meituan, and Didi, dominated the shared bicycle and shared electric bicycle market in Shenyang, with over 4,000 virtual shared E-bicycle stations[1] (as depicted in Figure 3) and more than 100,000 shared electric bicycles deployed. In the east part of the SEB operation area in Figure 3, there is a swath close to the border without virtual stations. This is mainly because at that time, the SEB virtual stations in this area had not yet been approved by the government, so they had not been put into operation.

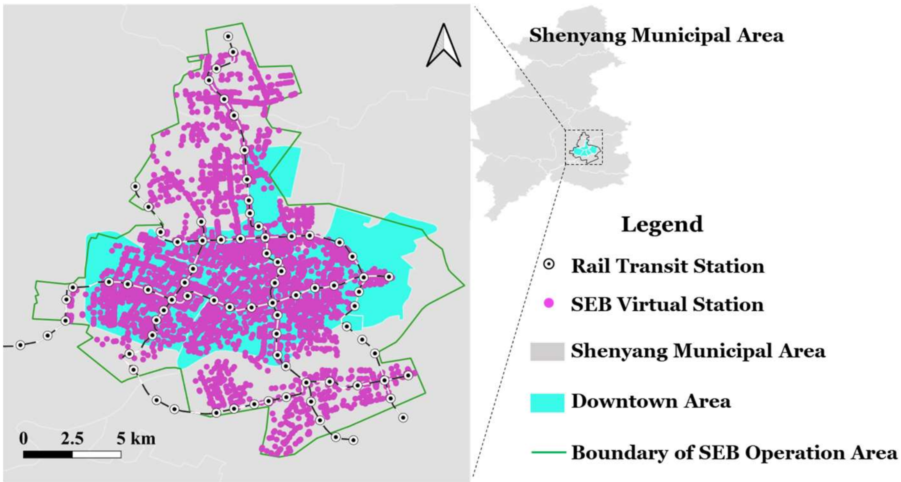

Fig. 3 Map of shared e-bicycle virtual stations in Shenyang City

3.2 Data sources and data processing methods

3.2.1 Telephone survey data of SEB users

In order to understand the interactive behaviors between SEBs and PT, as well as develop and test performance of the IBMMU, a telephone survey was conducted with SEB users in Shenyang City.

We posted messages on the operator’s app to recruit volunteers for telephone interviews through one of the three shared mobility companies mentioned above, namely Hello. The Shenyang municipal government allocates equally SEB quotas to the three above mentioned companies, Hello, Meituan and Didi, which are all sharing all the same virtual stations. That means the three companies are allowed to operate up to a maximum number of bikes within the same operational area, although the actual fleet size of each company is unknown. Also, although each company is

[1] Shared E-bike parking area constrained by electronic fences, with no physical parking facilities.

proposing different flat rates or packages, the unit price is the same: 3 yuan for the first 30 minutes, with an additional 1 yuan for every extra 10 minutes, and any time less than 10 minutes is counted as 10 minutes. Besides, the style of the E-bicycles and its range are similar for the three companies. Since we could only obtain data from Hello company, and all three companies have the same operational permits in terms of time and space, we assume that the patterns of use of Hello service are not radically different from those of the other two, even if the market shares are not known. Data from Hello can therefore be considered as representative of BES usage in the city of Shenyang.

Eligibility for participation was limited to users with network IP addresses located in Shenyang City. To incentivize participation, we offered each volunteer who completed the interview a 3-day free riding reward with no more than twice ridings on one day, as mentioned in the recruitment information. Due to business restrictions issues, Hello authorized no more five days to advertise the survey. Thus, in order to balance the number of samples on weekdays and weekends, the recruitment information was posted from 12:00 AM on October 15, 2020, to 12:00 AM on October 20, 2020. Interested volunteers could leave their phone numbers on the platform. In total, we collected 2,081 volunteers, out of which we randomly selected 500 volunteers, given the budget – including the survey duration, number of interviewers, labor costs, and respondent incentive costs - of this survey, for structured telephone interviews. Hello did not provide us with user and transaction data before the telephone survey, so we couldn't apply stratified sampling based on temporal and spatial distribution or riding frequency or sociodemographic characteristics. Instead, we used random sampling to minimize bias. Interviews were conducted each evening from October 26th (Monday) to November 1st (Sunday). To ensure the quality of interviews and avoid disrupting volunteers' work and daily life, we sent text messages to the respondents in the daytime to inform them of the interview time. Each night, 10 interviewers conducted interviews simultaneously from 7:30 PM to 9:30 PM. On average, each interviewer conducted interviews with 5 individuals per night, and the average interview duration was approximately 17 minutes. Because we need to obtain details of respondents' multimodal trips at the trip leg level, including transfer locations and times within these trips, which are complex and difficult to gather through online questionnaires, we conducted surveys via telephone interviews to limit misunderstandings and underreporting of more complex multimodal trips through interaction with an interviewer. Furthermore, some elderly people are less accustomed in using internet technology. Finally, it is important to note that during the whole experiment, including the completion of structured telephone interviews, the study area was not affected by the COVID-19 pandemic restrictions.

The interview comprises two sets of questions. The first part was focused on the most recent trip by SEB that could be recalled by the participant. We first inquired whether he/she used the shared E-bicycle from the origin directly to the destination of the whole trip or only to complete part of it. Secondly, if the participant used shared E-bike for the whole trip, we would inquire the names/street address numbers of the origin and the destination, which were then converted into geographic coordinates. If more than one mode were used to complete the trip, we asked to report all of them from origin to destination in chronological order, along with the names/street address numbers of all connection points between two consecutively used modes. For the second part, participants were asked which travel modes would have they used to complete that whole trip, if no shared E-bicycle were available at that time. We asked participants to give all the travel modes they would have used from origin to destination of that trip in chronological order, along with the names/street address numbers of all connection points between two modes. If the participants said they wouldn't have made this trip if no shared E-bicycle were available, the interview ended. It is noted that although we tried to collect the starting/ending time information about the trip, quite a lot of respondents could not remember the exact starting/ending time of the trip, nor could they specify the exact day, especially for the non-commuting respondents. When we tried to ask for the time information, most of the participants gave uncertain responses, so we did not obtain the validated time information. During the interview process, some volunteers occasionally forgot certain details, leading to incomplete records for a few samples. Consequently, we excluded these samples with missing information. After filtering, we obtained 429 valid complete samples for this study.

An example of telephone survey data collected is shown in Table 2 in which all information on multimodality is reported in chronological order. Additional information was also available that is not displayed here for the sake of briefness, including phone numbers, timestamps and geographic coordinates of the above points. The mode chains, referring to sequences of travel modes used during a whole trip, reported by the respondents involve up to two transfers, as exemplified by sample 4 in Table 2.

For later use, we manually analyzed the travel mode chains of each sample to extract the interactive behaviors between SEB and PT. In order to ensure the independence of samples, with much of the travel mode chains only including one SEB trip, we only extracted the interactive behaviors of the last SEB trip contained in each sample, along with transferring the starting and ending point location names to coordinates (e.g. in sample 4 from Table 2). Following the above introduced conceptual framework, the interactive behavior was coded in the categorical variable shown in the last column of Table 2 according to the following legend: 1 is a connection and substitution behavior with PT, 2 is a connection and non-substitution behavior with PT, 3 is non-connection and substitution behavior with PT and 4 is a non-connection and non-substitution behavior with PT.

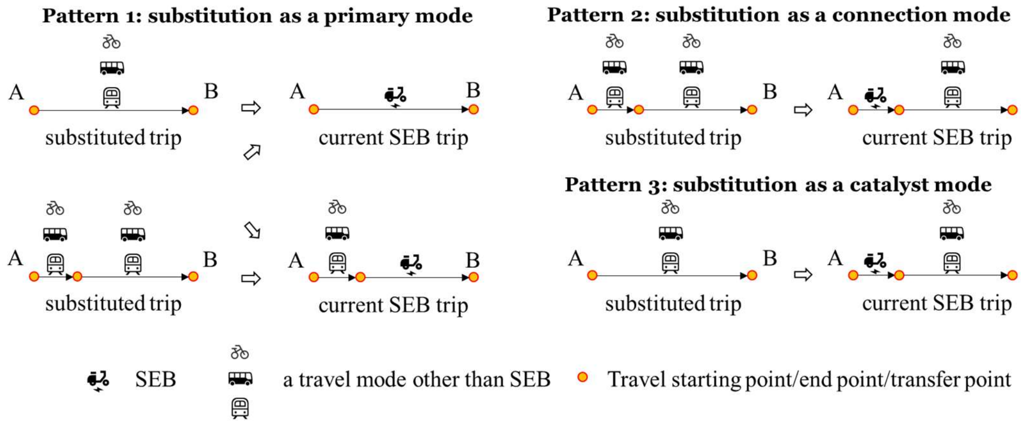

Fig. 4 The three substitution patterns between SEBs and other travel modes.

The specific method to assign each sample to one of the four interactive behaviors is as follows. If a sample involves either a single-mode trip with SEB or a multimodal trip with SEB and non-PT only, then the sample is considered a non-connection behavior with PT; otherwise, it is a connection behavior with PT, such as a multimodal trip involving SEB connecting to the metro or a multimodal trip involving SEB separately connecting to the metro and train at both endpoints. Then, we determine whether the SEB of the sample is substituting either PT or non-PT. This is however a relatively complex task. To understand how it is performed, we introduce three definitions as follows:

- Substitution as a primary mode: SEB serves as the primary mode for the current trip, where "primary mode" means that it is covering the longest distance of the trip, and it replaces a former primary mode in the substituted trip, as shown in Pattern 1 in the left half of Figure 4.
- Substitution as a connection mode: SEB serves as the connection mode for the current trip, therefore not covering the longest distance of the trip, and it is substituting another connection mode within the substituted trip, as shown in Pattern 2 in Figure 4.
- Substitution as a catalyst mode: SEB serves as a connector another travel mode for the current trip, and it is only partly substituting another primary travel mode within the substituted trip, as shown in Pattern 3 in Figure 4.

Table 2 Telephone survey data and derived interactive behavior (last column).

Sample number	Origin name	Mode one	Transfer point one	Mode two	Transfer point two	Mode three	Destination name	Substituted mode one	Substituted transfer point one	Substituted mode two	Substituted transfer point two	Substituted mode three	IB
1	Bei guan road 42	SEB	NULL	NULL	NULL	NULL	Sunshine Community	Metro	NULL	NULL	NULL	NULL	3
2	Shenyang North Railway Station	Metro	Qingnian Street Metro Station	SEB	NULL	NULL	Nanshi Department Store	Metro	NULL	NULL	NULL	NULL	1
3	Xiang Feng Hua Yuan	Metro	Dawn Square Metro Station	SEB	NULL	NULL	Dongling Road 19-6	Car	NULL	NULL	NULL	NULL	2
4	Shenyang Hetai Xincheng	SEB	Yingbin Road Metro Station	Metro	Dongzhong Street Metro Station	SEB	Shenyang Dadong Vehicle Management Office	SB	Yingbin Road Metro Station	Metro	Dongzhong Street Metro Station	SB	1
5	Xinyu Garden	SEB	NULL	NULL	NULL	NULL	Tawan Street 38	Private E-bicycle	NULL	NULL	NULL	NULL	4
……	……	……	……	……	……	……	……	……	……	……	……	……	……

Note: 1, 2, 3, 4 representing connection and substitution behavior with PT, connection and non-substitution behavior with PT, non-connection and substitution behavior with PT, and non-connection and non-substitution behavior with PT, respectively.

If the relationship between SEB and PT conforms to any of these three patterns, it is mapped as substituting for PT; otherwise, it is mapped as substituting for non-PT. Finally, the intersection of the above two classifications (namely, connection versus non-connection behavior with PT and substitution behavior for PT versus for non-PT) forms the above four categories of interactive behaviors. Descriptive statistics on the substitution and connection behaviors between SEB and other travel modes (including PT) are provided in Section 4.1.

3.2.2 Transaction data of the SEB service

As we mentioned, survey respondents could not always precisely recall the details of their trip involving the use of SEB. Additionally, recorded trips only represent a tiny fraction of the overall travel demand served by the service. In order to obtain the precise trip information for extracting IBs discrimination metrics, we collected riding transaction data from these same 429 interviewed users on the shared mobility platform during the period from October 10 to November 30, resulting in a total of 8,323 records. The main fields included in transaction data are shown in Table 3.

Table 3 Transaction data main fields.

Sample number	Transaction ID	Phone number	Starting time	Ending time	Starting coordinate	Ending coordinate	Riding distance
1	TP20201015…114	188…7416	2020-10-15 08:03:47	2020-10-15 08:51:07	123.4727, 41.8193	123.3699, 41.8124	9,631 m
2	TP20201016…138	189…8232	2020-10-17 13:21:31	2020-10-17 13:29:09	123.4428, 41.822	123.4387, 41.7986	1219 m
3	TP20201030…497	155…0893	2020-10-30 09:05:24	2020-10-30 09:12:10	123.5099, 41.8144	123.5338, 41.8149	2342 m
4	TP20201023…118	136…5768	2020-10-23 10:01:58	2020-10-23 10:13:11	123.3850, 41.8441	123.3158, 41.8069	1744 m
5	TP20201029…917	132…8900	2020-10-29 15:49:39	2020-10-29 16:06:38	123.3610, 41.8264	123.3674, 41.8327	3189 m
……	……	……	……	……	……	……	

3.2.3 Data from the Baidu map Application Programming Interface (API)

We complemented the above information by considering PT station data, including the names of bus stops and metro stations and their corresponding geographic coordinates, using the Baidu Maps API. Additionally, the same tool supplied the main attributes (travel duration, cost, distance, walking connection distance, and walking connection duration) of different modal alternatives (PT, private cars, bicycles, and walking) related to the same trip by SEB that was reported by the interviewee in the survey. The Baidu Map API could output the attributes of different modal alternatives after we input the starting point coordinates, ending point coordinates and starting time, which were the same with the SEB trip. Since the SEB trip had completed and our analysis work began after the completion of the survey and transaction data collection, we were unable to obtain attributes information to a timepoint in the past through the API. Therefore, we obtained the data generated from the week as close as possible to the actual SEB riding date (with sunny weather) to approximately represent the actual attributes information. The time frame for acquiring the attribute information is from December 7, 2020, to December 13, 2020, where the attribute acquisition time for each sample coincides with the actual start time of the SEB trip within a 24-hour period and corresponds to the same day of the week.

3.3 Framework of IBMMU

Figure 5 illustrates the framework of IBMMU, consisting of 4 steps, for discriminating interactive behaviors between SEBs and PT. The first step is preparing data for training and validating the interactive behaviors discrimination model. The second step is calculating the values of independent variables considered to discriminate interactive behaviors (IBs) and deriving the dependent variable representing IBs (as shown in the last column of Table 2). In step 3, an IBs discrimination model is constructed, calibrated and validated based on the variables calculated in step 2. In step 4, the validated IB discrimination model is used to determine the IBs for the large-scale SEBs transaction trips. The subsequent sections will elaborate on the details of each of these 4 steps.

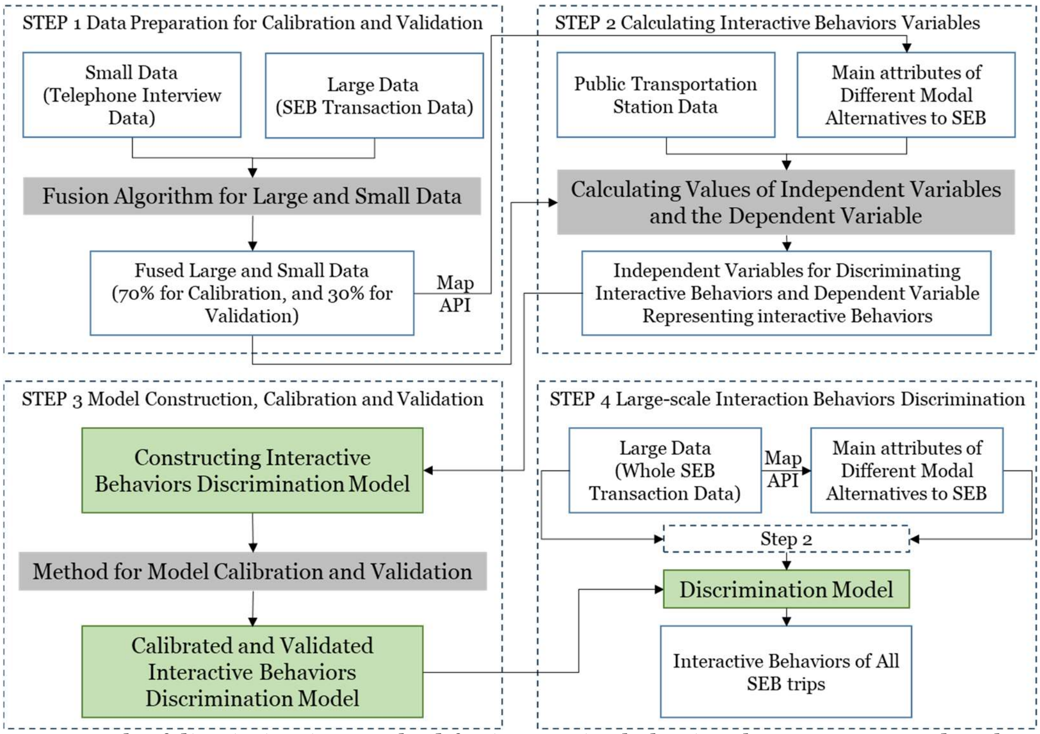

Fig. 5 Framework of discrimination method for interactive behaviors between SEB and PT based on both surveys and transaction data.

3.3.1 Data preparation for calibration and validation

Step one of the IBMMU is to achieve the merger of Table 2 and Table 3 mentioned above through a fusion algorithm represented in Figure 6. The final purpose of the IBMMU is to discriminate interactive behaviors only based on SEB transaction data. However, in order to calibrate and validate the method it is needed to consider observed interactive behaviors that are reported in the last column of Table 2. Therefore, there is a need to merge the “IBs” field from the telephone survey dataset with the transaction data.

More in details, for each respondent i we have one interview observation in Table 2 and several transaction samples in Table 3 generated from October 10th to November 30th. In order to extract the transaction sample corresponding to the interview sample, we firstly extracted all respondent i’s transaction samples based on the phone number, thereby forming the set U_i. Secondly, we calculate the Euclidean distance between each origin point in set U_i and the origin point of the interview sample, all the resulting values of distances forming the set O_{si}. Through the same method we can get the D_{si} set, consisting of distances from all destination points in set U_i to the destination point of the interview sample. Thirdly, we select all transactions in U_i for which both the corresponding elements in sets O_{si} and D_{si} have a value of less than 200 m (threshold that will be discussed in the next paragraph), and we consider this the subset of potential matching candidates with the interview sample i. If there is no transaction satisfying both conditions and the subset is empty, then the interview sample is invalid and the corresponding observation is discarded. Otherwise, the transaction sample in the latter subset with the shortest date interval from the interview date is merged with the interview sample, therefore information on the observed interactive behavior is matched with the corresponding transaction. This process will be looped until all interview samples are processed. The final merged data is allocated with 70% for model calibration and 30% for model performance evaluation (validation) in Step 3.

In theory, the geographic coordinates of starting/ending points of the transaction are matching those of the interview samples. However, since the respondents can only report the names or street numbers around the starting/ending points, there could be a lack of precision of the coordinates from survey data. We assume that the actual starting/ending points occur within the SEB virtual station next to the reported names or street numbers of places. Otherwise, it likely means that the two trips do not match. Thereby, we calculated the number of virtual stations within 100 m, 200 m, and 300 m range of both starting and ending points of each trip in the interview sample and found there is at least one virtual station within the range of 200 m radius for each interview sample. Therefore, we selected 200 m as the threshold to match potential transactions with interview samples.

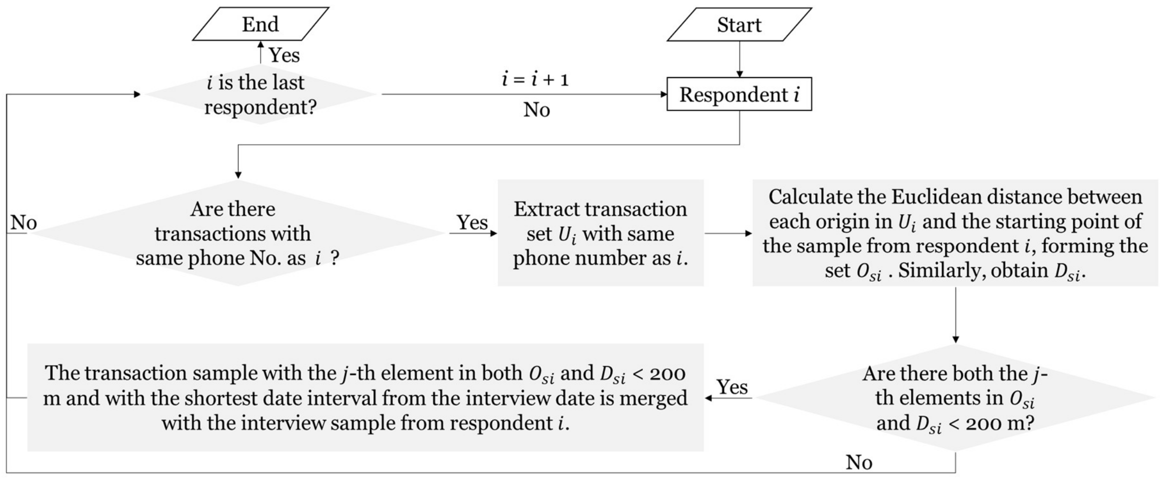

Fig. 6 Algorithm for merging survey and transaction datasets.

3.3.2 Interactive behaviors discrimination variables

In step two of IBMMU, three data sources are considered, namely merged data output from step one, the localization of PT stations and the main attributes of different modal alternatives to SEB, which were introduced in section 3.2.3. The goal is to identify interactive behaviors, thus defining the categorical dependent variable named INTR_BEHAVIOR shown in Table 4, which should be matched with the observed IB variable.

Table 4 Coopetition behavior discrimination variables.

Group	Variable	Definition	Expected impact
Dependent Variable	INTR_BEHAVIOR	Nominal variable with categories labelled as 1, 2, 3 and 4, respectively representing connection and substitution behaviour with PT, connection and non-substitution behaviour with PT, non-connection and substitution behaviour with PT, and non-connection and non-substitution behaviour with PT.	NA
Spatiotemporal relationships between SEB and PT	OP_YN	Whether the SEB riding occurs during PT operating hours.	C: + S: +
	PT_YN	Whether both the starting and ending points of a SEB trip are located within a 150-meter walking distance of PT stations (either rail transit station or bus stop).	C: / S: +
	OD_200	Whether only one (not both) end of a SEB trip's starting and ending points is within a 200-meter walking distance of a rail station.	C: + S: /
	EBIKE_DIS	Riding distance of a SEB trip.	C: - S: +
Relative performances of SEB vs. PT	PT_EBIKE_DUR	The ratio of the shortest PT duration including waiting time to SEB riding duration from the same origin to the same destination.	C: / S: -
	PT_CON	Whether there are congested segments on the shortest bus route, in terms of time, with the same origin and destination as a SEB option.	C: / S: +
	AC_WPEBIKE	The ratio of walking connection distance for the most convenient PT option to the riding distance of a SEB option, with the same origin and destination for both options.	C: / S: +
	TRANS	The number of transfers in PT for the same origin and destination when using a SEB.	C: / S: +
Relative performances of additional competing means	CAR_PT_ DUR	The ratio of car (including taxi, ride-sharing, and private car) duration to the shortest PT duration to from the same origin to the same destination.	C: / S: -
	BIKE_PT_ DUR	The ratio of bicycle (including shared bicycle and private bicycle) duration to the shortest PT duration from the same origin to the same destination.	C: / S: -
	WALK_PT_ DUR	The ratio of walking duration to the shortest PT duration from the same origin to the same destination.	C: / S: -
	CAR_PT_ DIS	The ratio of car (including taxi, ride-hailing, and private car) distance to the total distance of the shortest-duration PT option from the same origin to the same destination.	C: / S: -
	BIKE_PT_ DIS	The ratio of bicycle (including shared bicycle and private bicycle) distance to the total distance of the shortest-duration PT option from the same origin to the same destination.	C: / S: -
	WALK_PT_ DIS	The ratio of walking distance to total distance of the shortest-duration PT option from the same origin the same destination.	C: / S: -

CAR_PT_ PR	The ratio of private car cost to the cost of the shortest-duration PT option from the same origin to the same destination.	C: / S: -
TAXI_PT_ PR	The ratio of taxi (including ride-hailing) cost to the cost of the shortest-duration PT option from the same origin to the same destination.	C: / S: -
BIKE_PT_ PR	The ratio of bicycle (shared bicycle) cost to the cost of the shortest-duration PT option from the same origin to the same destination.	C: / S: -

Note: "C: +" indicates the variable potentially positively correlated with the connection behavior between SEB and PT; "S: +" indicates the variable potentially positively correlated with the substitution behavior between SEB and PT; "C: -" indicates the variable potentially negatively correlated with the connection behavior between SEB and PT; "S: -" indicates the variable potentially negatively correlated with the substitution behavior between SEB and PT

Independent variables in later models belong to three groups. The first group captures the spatiotemporal relationships between SEB trips and PT trips (as shown from the second line of Table 4). First and foremost, SEB connecting or substituting PT can only occur during PT operating hours. Therefore, we introduce variable OP_YN.

All else being equal, when both the starting and ending points of a SEB ride fall within the walking range of PT stations (considering both rail transit stations and bus stops), there is a higher probability of SEB substituting PT. Clearly, we anticipate that the performances of PT services connecting the two stations play also a role, that will however be considered in the next group of variables. On the other hand, due to the fact that SEBs' starting and ending points are restricted by the location of virtual stations of these services, the distance of the SEBs' starting and ending points to public transit stations might not be good predictors of the probability of substitution behavior. Therefore, we did not use such distance as discrimination variables. Instead, we defined the binary variable PT_YN based on a walking distance threshold. As shown in the left plot of figure 7, when both the distance from starting point to the nearest PT station and from ending point to the nearest PT station are lower than 150 *m* in our dataset, the observed percentage of substitution behavior (thus, the value of IB is equal to 1 or 3) is higher than that of non-substitution behavior (IB = 2 or IB = 4). Therefore, the distance threshold of PT_YN was set to 150 *m*.

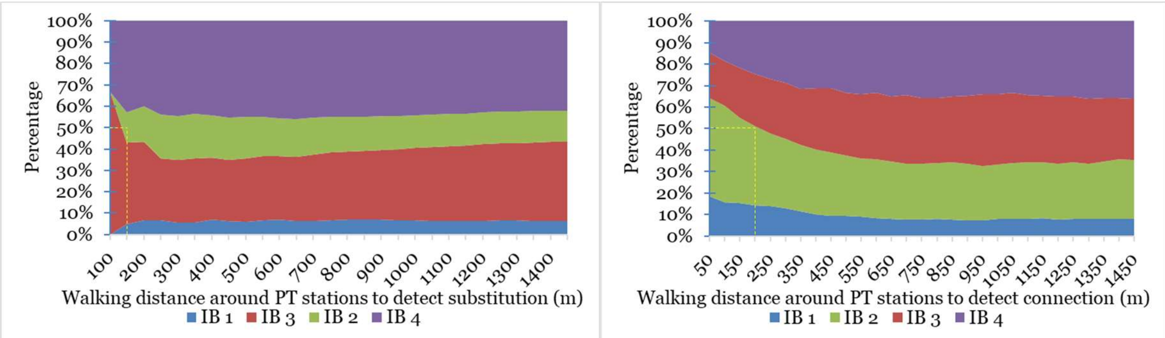

Fig. 7 Observed proportions of the four IBs as a function of walking range of PT stations (left plot) and of walking distance suitable for SEB connecting to PT (right plot).

It is noteworthy that the left plot of Figure 7 illustrates a reversal in the downward trend of interactive behaviors 3 as the walking distance increases above 500 m, which seems counterintuitive. To interpret this finding, we suggest that there may be two distinct categories of travelers with different decision-making behaviors. One group relies primarily on PT and when these travelers arrive at a PT station and find that the bus has not arrived or the electronic sign indicates a delayed arrival, they switch SEBs. The other group prioritizes SEBs over PT due to inconvenience in using PT (long walking transfer distances). They preferentially use PT only when no available SEBs are found within the SEB stations.

Concerning connection behaviors, when only one (not both) of the starting and ending points of the SEB riding falls within the walking distance suitable for connecting to PT between such point and a rail transit station, the probability of observing a connection behavior is higher. Please note that, in this case, only rail services are considered rather than bus stops, as previously done when analyzing the substitution behavior. This simplifying assumption is taken based on both the above literature review outcomes and empirical results in our dataset, that will be presented in Section 4.1. Namely, out of the 23.3% of SEB trips that are in combination with PT, 22.1% are connecting to rail, and only 1.2% are connecting to bus (see the right part of Figure 8). Hence, the binary variable OD_200 was defined. As shown in the right plot of Figure 7, when the distance is lower than 200 m, the observed percentage of connection behavior (IB = 1 or IB = 2) is higher than that of non-connection behavior (IB = 3 or IB = 4). Therefore, the distance threshold of OD_200 was set to 200 m. Finally, connecting ridings usually have shorter distance than non-connecting ridings. Therefore, we consider the distance covered by the SEB riding (EBIKE_DIS) among the predictors.

The second group of variables (listed in the middle of Table 4) is designed to capture the competitive advantage between SEB and PT. When riding a SEB offers significant advantages over using PT for the same origin and destination, travelers are more likely to use SEB as a substitute for PT. In particular, if the duration of riding a SEB is shorter than PT when traveling from the same origin to the same destination, there is a high probability that the SEB substitute PT, and thus the variable PT_EBIKE_DUR is introduced. Besides, a larger number of PT transfers between the origin and destination increases the probability of SEB substituting PT. Therefore, this factor is considered through the variable TRANS.

Bus is influenced by traffic congestion, which will decrease the relative advantage of PT over SEB. Therefore, the binary variable PT_CON is introduced with value equal to zero when the PT is rail or there is no congestion segment on the bus trip. Furthermore, when the walking distance to access PT is relatively long compared to the riding distance for the same trip by SEBs, there is a greater likelihood of substitution behavior between SEB and PT. Therefore, the variable AC_WPEBIKE is included.

The third group of variables (shown in the lower part of Table 4) is introduced to enlarge the perspective, observing that SEB might divert trips from other modes beyond PT. In general, SEBs should be compared with the best alternative travel option to complete a given trip, rather than just with TP, to have a clearer picture of the potential substitution effects. For instance, if the travel time by car is significantly shorter than that by PT, then the car is more likely to be the travel mode in use before the introduction of SEB. Therefore, the variable CAR_PT_DUR is introduced. We similarly compare costs between PT and car, through the variable CAR_PT_PR. In this research we consider five travel modes potential being substituted by BES, namely PT, car, taxi, bike and walking, so we introduce variables comparing travel duration, travel distance and travel cost of PT versus the latter four. It is noted that travel times and distances for taxis and cars are similar while walking has no cost, so the final list of variables in this group includes CAR_PT_DUR, BIKE_PTW_DUR, WALK_PT_DUR, CAR_PT_DIS, BIKE_PT_DIS, WALK_PT_DIS, CAR_PT_PR, TAXI_BI_PR, and BIKE_PT_PR were designed. All these variables are listed in Table 4.

It is worth noting that the IBMMU framework we proposed ultimately aims to discriminate large-scale IBs between SEBs and PT using only SEB transaction data, as will be described in section 3.3.4. Therefore, variables that cannot be extracted from transaction data, such as socio-economic attributes, were not considered.

3.3.3 Model training and validation

We use two different techniques to discriminate interactive behaviors, namely a Multinomial Logit (MNL) model from econometrics and a Random Forest (RF) model from machine learning. The main reasons for adopting these two models are as follows. The MNL model is a classic model in the field of traffic travel behavior choice, which can be used for behavior prediction and it has good interpretability. On the other hand, the RF model automatically constructs the relationship between independent and dependent variables by learning from the data, often demonstrating strong predictive accuracy. Moreover, RF is effective in addressing the issue of sample imbalance, given by the sample distribution of the 4 types of interactive behaviors in this study, as already apparent in Figure 7 and better described in section 4.1.

The core concept of the MNL model is that the choice yielding the highest utility reflects the actual choice of individual travelers. The utility function of MNL is showed by Equation 1

$$U_{nj} = \alpha_j + \beta_j X_{nj} + \epsilon_{nj} \tag{1}$$

where U_{nj} is a linear combination of predictor variables that determines the utility of the interactive behavior j (j = 1, ..., 4) for individual n. α_j represents the intercept of the utility function for the interactive behavior j. β_j is the parameter vector (coefficients). X_{nj} is the observed feature vector (independent variables), which are the above introduced interactive behaviors discrimination variables. Finally, ϵ_{nj} represents the random error. In this study, we used the maximum likelihood estimation method and Newton algorithm provided by the “statsmodels” package in Python to estimating model parameters. A stepwise regression approach and Akaike Information Criterion (AIC) were used to select significant variables with strong predictive performance and explanatory power. After getting the calibrated model, we validated the discrimination performance based on the validation data as described in section 3.3.1, by comparing predicted INTR_BEHAVIOR values with observed IB variables.

As shown in Equation 2, we define Acc to evaluate the discrimination performance of the MNL model:

$$Acc = \frac{\sum_{i=1}^{4} P_i}{\sum_{i=1}^{4} N_i} \tag{2}$$

where N_i represents the number of observations having shown an interactive behavior i, i = 1, ..., 4, and P_i represents the number of samples correctly predicted as having interactive behavior i. However, the sample sizes of the 4 types of interactive behaviors are imbalanced, as shown in Figure 7. Therefore, it seems wise to separately evaluate the discrimination accuracy for each type of interactive behaviors, beyond considering the model performance from an overall perspective, through the following quantities Acc_i:

$$Acc_i = \frac{P_i}{N_i}, \qquad i = 1, \ldots, 4\,. \tag{3}$$

On the other hand, the RF model is a machine learning method that belongs to the ensemble learning family. It combines a large number of uncorrelated decision classification trees, and the final prediction is made by averaging the results of each tree's independent voting. The concept of this model is similar to bagging (bootstrapped aggregation of weak learners), but the key difference lies in how Random Forest breaks the correlation between trees through random selection of input variables. At each split node, the Random Forest grows trees by randomly selecting $m \leq p$ input variables, where p represents the total number of independent variables, and m is the number of input variables included in each sub-decision tree. It has been proven that Random Forest can reduce variance compared to bagging, making it a powerful statistical tool for classification problems (Hastie et al., 2009). Not only is it easy to train and adjust, but it also performs exceptionally well on nonlinear data.

To obtain a calibrated RF model for discriminating IBs, two crucial steps are involved: hyperparameter optimization and model training. The hyperparameters subjected to optimization are presented in Table 5 below, together with their definition and candidate value set (Pedregosa et al., 2011). These candidate value sets for the hyperparameters ensures that each optimal hyperparameter falls within the interval, excluding the values at both ends, thus avoiding local optima. Notably, the candidate value set #5 in Table 5 includes the three methods provided by Python's "sklearn" package for selecting the maximum number of features (Pedregosa et al., 2011). Similarly, candidate value set #7 in Table 5 encompasses two commonly used methods for assessing node purity. Bayesian optimization is employed in this study for hyperparameter optimization, which combines the prior distribution of a function with sample information to derive the posterior distribution of the function. Utilizing the posterior information, the algorithm identifies the location of the maximum value of the function based on a specific criterion, resulting in a faster convergence speed compared to other optimization methods. Thus, Bayesian optimization from Python's "sklearn" package is utilized to determine the optimal hyperparameters for the Random Forest model in this research. After obtaining the optimal hyperparameters, the model training process utilizes the "RandomForestClassifier" tool again from Python's "sklearn" package.

Table 5 Hyperparameters of RF and candidate value set.

#	Hyperparameter	Definition	Candidate value set
1	n_estimators	Number of base classifiers	[50, 100, 150, 200, 250, 300, 400, 500]
2	max_depth	Maximum depth of decision trees	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15]
3	min_samples_split	Minimum number of samples required to split a node	[2,3,4,5,6,7,8,9,10]
4	min_samples_leaf	Minimum number of samples required to form a leaf node	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15]
5	max_features	Maximum number of features considered during the search for the best split	[auto, sqrt, log2]
6	ccp_alpha	Cost complexity parameter	[0.001, 0.002, 0.003, 0.004, 0.005, 0.006, 0.007, 0.008, 0.009, 0.01, 0.05, 0.1]
7	max_leaf_nodes	Maximum number of leaf nodes	[15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40]
8	min_impurity_decrease	Minimum impurity decrease required for a split	[0, 0.00001, 0.0001, 0.001, 0.01, 0.1]
9	criterion	Criterion used to assess node purity	[gini, entropy]

3.3.4 Large-scale interactive behaviors discrimination

Once we have a validated discrimination model from the previous step, we can employ it to discriminate the interactive behaviors of all SEB trips contained in the above introduced transaction dataset. As illustrated in step four of Figure 5, the SEB transaction data, along with the main attributes extracted from the Map API for various modal alternatives to SEB, is transformed into values for the discrimination variables used as input for the discrimination model. Subsequently, the discrimination model produces the interactive behavior for each SEB trip.

This last step is the application phase of the IBMMU. The focus of this study is on proposing and validating a framework of IBMMU, namely step one to step three. Therefore, in the following section we omit the presentation of results related to the forecast of interactive behaviors for the whole dataset to save on space, also considering that relevant findings can already be obtained when considering the training and validation dataset.

4 Results and discussion

4.1 Descriptive statistics on interactive behaviors between SEB and other modes

Data from telephone surveys provide a detailed understanding of the interaction behaviors between SEB and PT, as depicted in Figure 8. Among all SEB rides, a majority of SEB rides, specifically 74.6%, are single mode trips, indicating exclusive reliance on SEB for a whole trip. In contrast, 25.4% of trips are multimodal, integrating SEB with other travel modes. This represents a non-negligible proportion in the dataset. Among these multimodal trips, 23.3% involve SEB used in conjunction with PT. When breaking down these PT connections, rail transit emerges as the predominant companion to SEB, representing 22.1% of all trips, while buses account for a smaller fraction at 1.2%. Additionally, a minimal 2.1% of all SEB trips link with non-PT travel modes. Overall, SEB is primarily utilized for single mode trips or in combination with rail transit, those two behaviors constituting 74.6+22.1 = 96.7% of its usage. This indicates a distinct difference in the connecting targets between SBs and SEBs. The former is mainly used for connecting to both buses and rail transit (Murphy and Usher, 2014), while the latter is mainly suitable for connecting to rail transit. This difference may be attributed to the faster riding speed, effortless riding pattern, and the absence of detours of SEBs. Sharing similar advantages with SEBs, previous studies have suggested that shared e-scooters also primarily connect to rail transit in multimodal scenarios (Fearnley et al., 2020).

The lower part of Figure 8 shows for both unimodal and multimodal SEB trips whether they are substituting PT (rail or bus) or non-PT modes. Different cases can be jointly considered, as shown by the colored arrows to distinguish among the four above introduced interactive behaviors between SEB and PT. As illustrated in Figure 8, connection and substitution with PT accounts for 6.1%, connection and non-substitution with PT accounts for 17.2%, non-connection and substitution with PT accounts for 35.2%, and non-connection and non-substitution with PT accounts for 41.5%. This implies a close relationship between SEBs and PT, as about 60% of SEB rides are either substitutes for or connected to PT. Some studies suggest that the emergence of SMMs has diminished buses ridership while enhancing travel efficiency for certain travelers (Godavarthy et al., 2022, Luo et al., 2021, Shi et al., 2021, Ma et al., 2019, Campbell and Brakewood, 2017). However, the underlying phenomenon implies a trade-off between the losses in government subsidies for public transit and utility gains for travelers. How to strike a balance between these aspects and foster the collaborative development of public transit systems and shared micro-mobility systems is a topic worthy of research.

The above four interactive behaviors' riding distance as shown in Table 6, in which the distance of connection behavior with PT is much shorter than that of non-connection behavior with PT, while the distance of substitution behavior with PT is much longer than that of non-substitution behavior with PT.

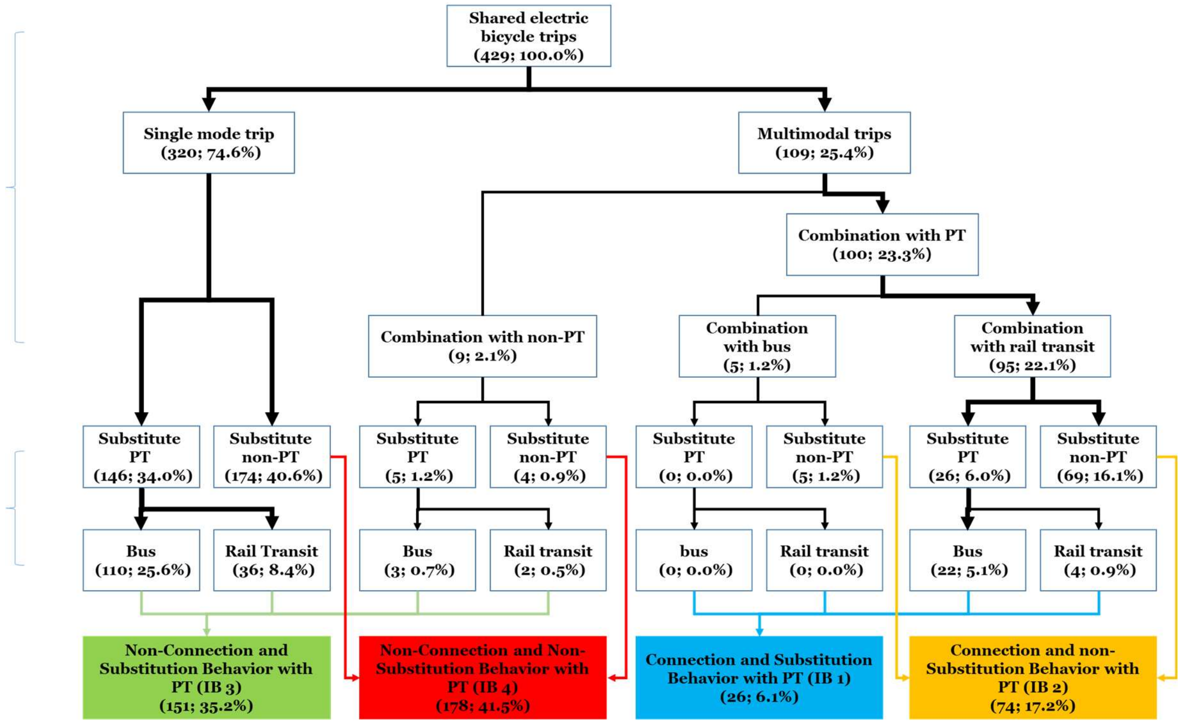

Fig. 8 The distribution of substitution and connection behaviors between SEBs and PT, as well as other non-PT modes.

Table 6 Average riding distance for four interactive behaviors (unit: km).

	Substitution with PT	Non-substitution with PT	Total
Connection with PT	2.37	1.76	1.82
Non-connection with PT	5.40	3.13	4.08
Total	4.92	2.64	3.58

As illustrated in the lower half of Figure 8, the four most commonplace situations in our dataset concerning substitution patterns are single mode trips substituting PT (146 cases, 34.0%) or non-PT modes (174 cases, 40.6%), followed by multimodal trips involving the use of both SEB and rail that are substituting either PT (26 cases, 6.0%) or non-PT modes (69 cases, 16.1%). Figure 9 shows the breakdown of each of these four groups into the three substitution patterns introduced in section 3.2.1, namely (1) substitution as a primary mode, (2) substitution as a connection mode and (3) substitution as a catalyst mode. The three patterns are represented by blue columns, while labels to the right of blue columns indicate the corresponding primary mode being substituted within the pattern and labels to the right of green columns indicate all the specific travel modes substituted by SEB, irrespective of the pattern. Figure 9A reveals that all 34% monomodal SEB trips that replace PT correspond to pattern one. This includes 25.6% where SEB substitutes for buses (21.9% for sole bus and 3.7% for a combination of SB and bus) and 8.4% where SEB substitutes for rail transit, encompassing three specific scenarios. Figure 9B indicates that also all monomodal SEB trips replacing non-PT modes align with pattern one. Within these trips, the percentage of SEB substituting SB, private cars, taxis (inclusive of ride-hailing services), walking, and private e-bicycles is 14.5%, 12.8%, 6.5%, 3.7%, and 3.0%, respectively. It is noteworthy that when monomodal SEB trips substitute trips involving the use of non-PT modes, the latter are always monomodal trips (see Figure 9B). On the contrary, when monomodal SEB trips substitute trips in which PT is used, the latter are also multimodal trips, thus exhibiting more complex substitution patterns that were not investigated by previous research just focusing on the main substituted mode (see Figure 9A).

Concerning the lower half of Figure 8, within the 22.1% of multimodal trips where SEB connects to rail transit, 6% of SEB serves as a substitute for PT, and 16.1% as a substitute for non-PT modes. Specifically, as depicted in Figure 9C, of the 6% of combined SEB and rail transit trips replacing PT, 4.2% adhere to pattern two, functioning as connection modes where SEB linked to rail transit takes the place of bus connections to the rail transit. Additionally, 0.9% of SEB trips linked to rail transit substitute for transfers between different rail lines, while another 0.9% fit into pattern three as a catalyst mode, where SEB connected to rail transit replaces singular bus trips. Furthermore, Figure 9D shows that in the 16.1% of SEB trips substituting non-PT, 13.9% are categorized under pattern two as a connector, and 2.2% under pattern three as a catalyst. Finally, given the minor share of multimodal SEB trips connected to non-PT and buses (respectively 2.1% and 1.2% as shown in Figure 8), the specific substitution patterns and details are not elaborated further.

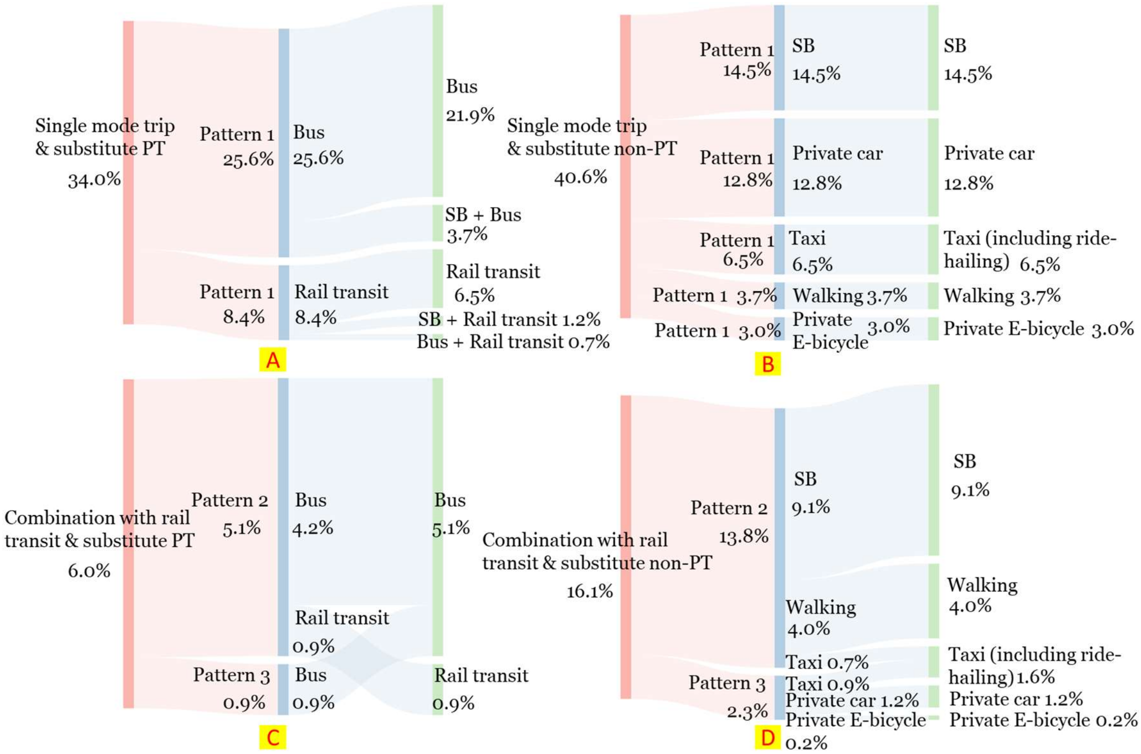

Fig 9 SEB substitution patterns. Red columns indicate four connection and substitution behaviors selected from Figure 8, blue columns indicate SEB substitution patterns and green columns indicate the specific travel modes replaced by SEB.

From the above analysis, actual substitution patterns are highly complex when viewed from the perspective of multimodal trips, so previous studies that only analyzed the main substitution mode (usually referred to as the longest-distance mode in a trip) of SMM are short-sighted (Fukushige et al., 2021, Christoforou et al., 2021, Fearnley et al., 2020, Laa and Leth, 2020). Under three substitution patterns, the overall substitution rate of SEBs for PT is 41.3%, consistent with previous research indicating that sustainable travel modes (PT and walk) are the main substitutes for SEBs (Fukushige et al., 2021, Bieliński et al., 2021). Among these, the substitution of buses accounts for 31.5%, while the substitution of rail transit accounts for 9.8%. Similar to other SMMs, when substituting for PT, SEBs primarily substitute buses (Campbell et al., 2016, Wang, 2017, Murphy and Usher, 2014, Fearnley et al., 2020). This suggests that rail transit holds a strong advantage over active travel modes. However, the issue of various SMMs strongly substituting buses and other active travel modes potentially conflicts with the expected environmental benefits, presenting a worthwhile topic for research.

4.2 Modelling results

Following the methodological steps on section 3, we have organized and obtained the input data for the third step of the IBMMU framework. The resulting dataset encompasses information about interactive behaviors as the dependent variable, along with data on factors that differentiate these behaviors as independent variables, totaling 429 samples. As already mentioned, this study assumes that SEB only combines with rail transit when it is used to connect to the other travel modes since SEB trips involving connections to buses and non-PT modes are only 3.3% of the total. Additionally, whenever SEB acts as a catalyst the real substitution is happening between other modes, which are also not considered in the modeling process. Omitting such cases, the final sample size is therefore 401. This refined dataset is divided into 280 samples (70%) for model calibration and 121 samples (30%) for validation. In order to control bias, we adopt stratified sampling, ensuring that the calibration samples and validation samples contain the same proportion of each of the four types of interactive behaviors.

For the third step in the IBMMU framework, we independently applied the Multinomial Logistic (MNL) and Random Forest (RF) models. To mitigate collinearity concerns, ten variables exhibiting a Variance Inflation Factor (VIF) above 10 - namely OP_YN, PT_EBIKE_DUR, PT_CON, CAR_PT_DUR, WALK_PT_DUR, CAR_PT_DIS, BIKE_PT_DIS, WALK_PT_DIS, CAR_PT_PR, TAXI_PT_PR, and BIKE_PT_PR - were excluded.

4.2.1 MNL model results

For calibrating the MNL model, we employed the stepwise regression approach and sequentially added interactive discrimination variables to calculate coefficients of the systematic utilities and their corresponding p-values. Variables with p-values > 0.1 were subsequently excluded, resulting in three significant variables as presented in Table 7.

The coefficient of OD_200 is significantly negative for "non-connection and substitution behavior" and "non-connection and non-substitution behavior". This indicates that the probability of connection is higher when either the starting or ending point of the SEB ride is within a 200m walking distance from a PT station.

Additionally, the coefficient of EBIKE_DIS for the "connection and non-substitution behavior" is not statistically significant, suggesting that the riding distance of SEBs does not affect their substitution for PT when used to connect to rail transit. In contrast, for the other two choices, the coefficients of the EBIKE_DIS are significantly positive, which means the longer distance from riding SEB, the lower likely of PT substituted by SEBs. Specifically, the coefficient for "non-connection and substitution behavior" greater than that for the "non-connect and non-substitution behavior" implies that as the riding distance increases, the probability of SEBs substituted PT increases under the non-connection situations.

Table 7 The MNL model parameters.

Variables	Coefficients		
	connection and non-substitution behaviour	Non-connection and substitution behaviour	non-connection and non-substitution behaviour
CONSTANT	2.8554***	0.5165	3.1719***
OD_200	-1.1204	-3.7080***	-3.5208***
EBIKE_DIS	0.0007	0.0015***	0.0012***
BIKE_PT_DUR	-5.0158**	-2.2317	-4.0903**
The reference choice	connection and substitution behaviour with PT		
N	280		
Pseudo R-square.	0.2676		
Log-Likelihood	-239.51		
LLR p-value	<0.001		

Note: *** indicates significance at the level of P≤0.01, ** indicates significance at the level of 0.01<P≤0.05, and * indicates significance at the level of 0.1<P≤0.05.

Finally, the BIKE_PT_DUR variable exhibits significant negative coefficients in both the "connection and non-substitution behavior" and "non-connection and non-substitution behavior" choices. This suggests that as the ratio of the duration of BIKE to the duration of PT decreases, the likelihood of SEBs not substituting for PT increases. This finding is logical, as a greater speed advantage of bikes over PT, including shared bicycles, would lead people to prefer bicycles instead of PT when SEBs are not available.

Interestingly, the variables relative performances of SEB vs. PT are all insignificant. It's possible, although it's not quite what we expected. Specifically, a traveler's mode choice is the result of a comprehensive comparison of various travel options. While people might switch from PT to SEBs due to the speed advantage of SEBs over PT, it is also possible that in some travel scenarios, even without SEBs, PT might not be the most efficient option.

4.2.2 RF model results

The RF model was implemented using the hyperparameter tuning and model training approach detailed in the Methodology section (Section 3.3.3). Initially, we employed Bayesian optimization to determine the optimal hyperparameters. Then, we conducted 100 pre-trainings on the training set (samples used for model calibration) with 100 different random seeds to mitigate randomness, resulting in 100 sets of variable importance scores. Thirdly, the average importance score for each interactive behavior discrimination variable over 100 pre-trainings was calculated, and we sorted the average scores in descending order. Subsequently, we added variables into the RF model one at a time, in descending order of average importance. At the same time, after each variable was added, a 5-fold cross-validation was performed to assess the relationship between the number of variables and the average predictive accuracy. As depicted in left picture of Figure 10, we observed that after adding 3 variables to the model, the accuracy tends to stable. After adding additional variables, the accuracy began to fluctuate steadily. This means that when the number of important variables exceeds 3, the improvement in prediction performance is minimal, and it is highly likely to lead to overfitting issues in the model. Finally, the top three important variables were utilized to retrain the model, and the importance score of each variable is shown in the right picture of Figure 10.

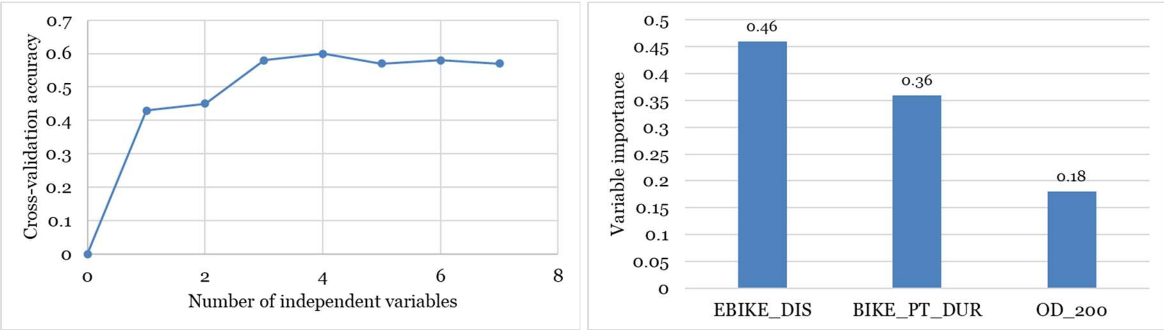

Fig. 10 The relationship between the number of importance variables and the cross-validation prediction accuracy in the RF model (left) and the ranking of variable importance scores (right).

The right picture of Figure 10 illustrates the importance score ranking of variables in the optimal RF model. Interestingly, the three variables — EBIKE_DIS, BIKE_PT_DUR and OD_200 —are the same as those considered in the MNL model. This indicates that these three variables exhibit high importance in both types of models and can be considered reliable variables suitable for discriminating interactive behaviors.

4.3 Comparing model accuracies

As shown in Table 8, the MNL model and RF model have similar overall discrimination accuracy for the four types of interactive behaviors, with a $Acc_{overall}$ value around 63%. However, the MNL model shows better balance in discrimination accuracy for each type of behavior, with a standard variance of 25.5%, while the RF model has a standard variance of 32.8%. Moreover, the RF model even failed to accurately identify any samples of the connection and substitution behavior with PT due to the low size of test samples, which may be due to the RF model struggling to demonstrate its predictive performance advantages and handle sample imbalance with a small sample size. In the context of the small-sample data used in this study, RF model as a representative of machine learning did not demonstrate a clear performance superiority over the MNL. This suggests that econometric models cannot be replaced by machine learning in problems with small samples.

Table 8 Discrimination accuracy of the MNL Model and RF Model.

Model	All Interactive Behaviours			Connection behaviour	Substitution behaviour
	Interactive behaviours	Acc_i	$Acc_{overall}$	$Acc_{overall-con}$	$Acc_{overall-sub}$
MNL	1	14.3%	62.8%	90.1%	69.4%
	2	77.8%			
	3	50.0%			
	4	75.0%			
RF	1	0.0%	63.6%	90.9%	68.6%
	2	83.3%			
	3	50.0%			
	4	76.9%			

Note: interactive behaviours 1-4 representing connection and substitution behaviour with PT, connection and non-substitution behaviour with PT, non-connection and substitution behaviour with PT, and non-connection and non-substitution behaviour with PT, respectively. $Acc_{overall-co}$ indicates the proportion of samples of connection behaviour identified among the total samples of connection behaviour, including both connection and substitution behaviours with PT and connection and non-substitution behaviour with PT. $Acc_{overall-sub}$ shares similar principles with $Acc_{overall-con}$.

When considering only the discrimination accuracy of connection behavior—meaning, for all samples belonging to behavior 1 and behavior 2, the proportion of samples accurately identified as either behavior (including cases where behavior 1 is identified as behavior 2) to the total number of behavior 1 and behavior 2 samples—the MNL model and RF model exhibit similar high accuracy, at 90.1% and 90.9%, respectively. Likewise, when focusing on the accuracy of substitution behavior, both models show comparable accuracy, with the MNL model at 69.4% and the RF model at 68.6%. The lack of personal socio-economic attribute data from transaction data is one of the reasons that limit the discrimination accuracy of our proposed method, especially when discriminating the substitution behavior. For instance, factors such as owning private cars, household income levels, and educational background can potentially influence people's choice of travel modes, thereby affecting the substitution relationship between SMMs and PT. Future research may benefit from incorporating personal socio-economic attribute data through smartphone apps, offering potential improvements in discrimination performance.

5 Conclusions

Accurately discriminating the interactive behaviors between SMMs and PT is of utmost importance for evaluating their social benefits and assessing the effectiveness of financial subsidies when integrating SMMs into the PT system. However, existing discrimination methods have not achieved precise discrimination of interactive behaviors across the entire spectrum of SMMs trips. Taking into account the strengths and limitations of commonly employed methods for discriminating interactive behaviors between SMMs and PT, such as questionnaire-based discrimination method, method based on transaction data analysis and method based on utility theory, this study proposes the IBMMU framework, which integrates the advantages of the above three methods. The IBMMU framework is not only applicable to the discrimination of interactive behaviors between SEB and PT but it can also be extended to study interactions between other SMMs and public transportation. Additionally, since the performance of different types of prediction models varies across different datasets, the IBMMU framework is suitable for embedding any supervised prediction model to discriminate interactive behaviors on diverse datasets. On theoretical grounds, the IBMMU framework can discriminate four types of actual interactive behaviors between SMMs and PT at the trip level across all transaction data, addressing the limitations of current methods that were summarized in Table 1. Additionally, three factors influencing these interactive behaviors are identified.

In practical applications, the IBMMU framework holds substantial value in transportation planning and management. It provides insights into the connection and substitution relationships between SMMs and public transportation, facilitating the assessment of their competitive and cooperative dynamics across regions. This, in turn, it supports the strategic planning and optimization of their integrated development therefore promoting transportation equity. Additionally, the framework can quantify the impact of SMMs on public transportation ridership and, when combined with energy and emission frameworks based on life cycle assessment (LCA), it allows for better assessment of their true environmental and energy impacts, including rebound effects on different modes.

To validate the practical performance of the IBMMU framework, we conducted an evaluation using survey data and transaction data from SEB users in Shenyang, China. The research findings indicate that the IBMMU framework, embedded with MNL or RF models as the discrimination mode, effectively achieves the discrimination of interactive behaviors between SEBs and PT, resulting in an overall discrimination accuracy of 63%. Moreover, the discrimination accuracy for single connection and substitution behaviors reaches about 90% and about 70%, respectively.

Moreover, this study, for the first time, explores the interactive behaviors between SEB and PT from the multimodal trip perspective rather than only considering the main mode (usually referred to as the longest-distance mode in a trip) substituted by SEB. The research findings indicate that differing from SBs, which are mainly used for connecting with both buses and rail transit, SEBs are only mainly used for connecting with rail transit. The substitution behaviors between SEBs and PT can be categorized into three patterns, namely substitution as a primary mode, substitution as a connection mode, and substitution as a catalyst mode. Consistent with previous research, PT is one of the primary substitutes for SEBs, accounting for a total proportion of 41.3%. Among them, the substitution rate for buses is higher than that for rail transit (31.5% vs. 9.8%). When simultaneously considering the connection and substitution behaviors, connection and substitution behavior with PT accounts for 6.1%, connection and non-substitution behavior with PT accounts for 17.2%, non-connection and substitution behavior with PT accounts for 35.2%, and non-connection and non-substitution behavior with PT accounts for 41.5%.

Based on the above findings, several recommendations can be made for planners and policymakers. The vast majority of multimodal SEB trips are integrated with rail transit, and the substitution effect of SEBs on buses is much higher than on rail transit. Therefore, it is advisable to prioritize the placement of SEB stations and fleets around rail transit stations, while reducing the integration of SEB facilities with bus stops, except in the case of high-demand, overcrowded bus routes when the bus service cannot be strengthened for any contingent reason. This study finds that the average riding distance for SEBs connecting to rail transit is 1.8 km, and the likelihood of connecting to rail transit is significantly higher when the endpoint of the shared e-bike trip is within 200 meters of the rail transit station. Thus, it is recommended to establish SEB stations within a 2 km radius of rail transit stations.

This study also has some limitations, which offers potentials for future research. Firstly, the proportion of "connection and substitution behavior with PT" is relatively low, and the limited number of calibration and validation samples may impact the robustness of the discrimination accuracy of our proposed method for this specific category of interactive behavior.

Secondly, our phone interview has shown that SEBs not only connect with rail transit but also exhibit connections with buses (1.2%) and with non-PT (2.1%). Due to the limited sample size in this study, we did not consider the two small probability behaviors in the model. Future research with larger sample sizes is required to address and distinguish these cases effectively.

Lastly, the samples of SEBs acting as a catalyst to substitute other travel modes (which accounted for 3.66% in this study), demonstrating significant differences in spatiotemporal characteristics from the modes being replaced, are excluded from the modeling process. It presents challenges in discriminating such behaviors based on spatiotemporal relationships in transaction data. Therefore, further research is needed to develop appropriate discrimination methods specifically for this substitution behavior as catalyst.

Author Contributions Jianhong Ye: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Project administration, Writing - review & editing. Jiahao Bai: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software; Validation; Visualization, Writing – original draft. Marco Diana: Formal analysis, Supervision, Validation, Writing – review & editing.

Acknowledgments We sincerely thank the HelloBike company for the cooperation in collecting the data.

Funding This work was supported by National Natural Science Foundation of China under Grant number 52172320, and by the Fundamental Research Funds for the Central Universities under Grant number 2023-4-ZD-01. Financial support was also provided by the China Scholarship Council (CSC) under Grant number 202306260133.

Declarations

Conflicts of interest The authors declare that they have no conflict of interest.

References

6T-BUREAU DE RECHERCHE 2019. Uses and Users of Free-floating Electric Scooters in France. 6t-Bureau De Recherche.

AARHAUG, J., FEARNLEY, N. & JOHNSSON, E. 2023. E-scooters and public transport – Complement or competition? *Research in Transportation Economics,* 98.

BIELIŃSKI, T., KWAPISZ, A. & WAŻNA, A. 2021. Electric bike-sharing services mode substitution for driving, public transit, and cycling. *Transportation Research Part D: Transport and Environment,* 96.

CAMPBELL, A. A., CHERRY, C. R., RYERSON, M. S. & YANG, X. 2016. Factors influencing the choice of shared bicycles and shared electric bikes in Beijing. *Transportation Research Part C: Emerging Technologies,* 67**,** 399-414.

CAMPBELL, K. B. & BRAKEWOOD, C. 2017. Sharing riders: How bikesharing impacts bus ridership in New York City. *Transportation Research Part A: Policy and Practice,* 100**,** 264-282.

CHRISTOFOROU, Z., DE BORTOLI, A., GIOLDASIS, C. & SEIDOWSKY, R. 2021. Who is using e-scooters and how? Evidence from Paris. *Transportation Research Part D: Transport and Environment,* 92.

FEARNLEY, N., JOHNSSON, E. & BERGE, S. H. 2020. Patterns of e-scooter use in combination with public transport. *Findings*.

FELIPE-FALGAS, P., MADRID-LOPEZ, C. & MARQUET, O. 2022. Assessing Environmental Performance of Micromobility Using LCA and Self-Reported Modal Change: The Case of Shared E-Bikes, E-Scooters, and E-Mopeds in Barcelona. *Sustainability,* 14.

FISHMAN, E. 2015. Bikeshare: A Review of Recent Literature. *Transport Reviews,* 36**,** 92-113.

FISHMAN, E., WASHINGTON, S. & HAWORTH, N. 2014. Bike share's impact on car use: Evidence from the United States, Great Britain, and Australia. *Transportation Research Part D: Transport and Environment,* 31**,** 13-20.
FUKUSHIGE, T., FITCH, D. T. & HANDY, S. 2021. Factors influencing dock-less E-bike-share mode substitution: Evidence from Sacramento, California. *Transportation Research Part D: Transport and Environment,* 99**,** 102990.
GODAVARTHY, R., MATTSON, J. & HOUGH, J. 2022. Impact of bike share on transit ridership in a smaller city with a university-oriented bike share program. *Journal of Public Transportation,* 24.
HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J. H. & FRIEDMAN, J. H. 2009. *The elements of statistical learning: data mining, inference, and prediction*, Springer.
JI, Y., FAN, Y., ERMAGUN, A., CAO, X., WANG, W. & DAS, K. 2016. Public bicycle as a feeder mode to rail transit in China: The role of gender, age, income, trip purpose, and bicycle theft experience. *International Journal of Sustainable Transportation,* 11**,** 308-317.
KONG, H., JIN, S. T. & SUI, D. Z. 2020. Deciphering the relationship between bikesharing and public transit: Modal substitution, integration, and complementation. *Transportation Research Part D: Transport and Environment,* 85.
LAA, B. & LETH, U. 2020. Survey of E-scooter users in Vienna: Who they are and how they ride. *Journal of Transport Geography,* 89.
LAZARUS, J., POURQUIER, J. C., FENG, F., HAMMEL, H. & SHAHEEN, S. 2020. Micromobility evolution and expansion: Understanding how docked and dockless bikesharing models complement and compete – A case study of San Francisco. *Journal of Transport Geography,* 84.
LDA CONSULTING, I. 2017. 2016 Capital Bikeshare Member Survey Report. LDA Consulting, Inc.
LI, A., GAO, K., ZHAO, P., QU, X. & AXHAUSEN, K. W. 2021. High-resolution assessment of environmental benefits of dockless bike-sharing systems based on transaction data. *Journal of Cleaner Production,* 296.
LIN, J.-J., ZHAO, P., TAKADA, K., LI, S., YAI, T. & CHEN, C.-H. 2018. Built environment and public bike usage for metro access: A comparison of neighborhoods in Beijing, Taipei, and Tokyo. *Transportation Research Part D: Transport and Environment,* 63**,** 209-221.
LIU, Y., JI, Y., FENG, T. & SHI, Z. 2022. A route analysis of metro-bikeshare users using smart card data. *Travel Behaviour and Society,* 26**,** 108-120.
LUO, H., ZHANG, Z., GKRITZA, K. & CAI, H. 2021. Are shared electric scooters competing with buses? a case study in Indianapolis. *Transportation Research Part D: Transport and Environment,* 97.
MA, X., ZHANG, X., LI, X., WANG, X. & ZHAO, X. 2019. Impacts of free-floating bikesharing system on public transit ridership. *Transportation Research Part D: Transport and Environment,* 76**,** 100-110.
MIDGLEY, P. 2011. Bicycle-sharing schemes: enhancing sustainable mobility in urban areas. *United Nations, Department of Economic and Social Affairs,* 8**,** 1-12.
MOLINILLO, S., RUIZ-MONTAñEZ, M. & LIéBANA-CABANILLAS, F. 2019. User characteristics influencing use of a bicycle-sharing system integrated into an intermodal transport network in Spain. *International Journal of Sustainable Transportation,* 14**,** 513-524.
MURPHY, E. & USHER, J. 2014. The Role of Bicycle-sharing in the City: Analysis of the Irish Experience. *International Journal of Sustainable Transportation,* 9**,** 116-125.
PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., VANDERPLAS, J., PASSOS, A., COURNAPEAU, D., BRUCHER, M., PERROT, M. & DUCHESNAY, E. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research,* 12**,** 2825-2830.
QIU, W. & CHANG, H. 2021. The interplay between dockless bikeshare and bus for small-size cities in the US: A case study of Ithaca. *Journal of Transport Geography,* 96.
RECK, D. J., MARTIN, H. & AXHAUSEN, K. W. 2022. Mode choice, substitution patterns and environmental impacts of shared and personal micro-mobility. *Transportation Research Part D: Transport and Environment,* 102.
SHAHEEN, S. A., MARTIN, E. W., CHAN, N. D., COHEN, A. M. & POGODZINSKI, M. 2014. Public bikesharing in North America during a period of rapid expansion: understanding business models, industry trends & user impacts, MTI Report 12-29.
SHI, X., LI, Z. & XIA, E. 2021. The impact of ride-hailing and shared bikes on public transit: Moderating effect of the legitimacy. *Research in Transportation Economics,* 85.
SUCHANEK, M., JAGIEŁŁO, A. & SUCHANEK, J. 2021. Substitutability and Complementarity of Municipal Electric Bike Sharing Systems against Other Forms of Urban Transport. *Applied Sciences,* 11**,** 6702.

TANG, Y., PAN, H. & SHEN, Q. 2011. Bike-Renting Systems in Beijing, Shanghai and Hangzhou and Their Impact on Travel Behavior. *International Association for China Planning, Shanghai, China*.

TEIXEIRA, J. F., SILVA, C. & MOURA E Sá, F. 2020. Empirical evidence on the impacts of bikesharing: a literature review. *Transport Reviews,* 41**,** 329-351.

TRANSPORT FOR LONDON 2010. Barclays Cycle Hire customer satisfaction and behaviour Autumn 2010 (Wave 1 - cycle hire members only) 10102. Transport for London.

VINAGRE DíAZ, J. J., FERNáNDEZ POZO, R., RODRíGUEZ GONZáLEZ, A. B., WILBY, M. R. & ANVARI, B. 2023. Blind classification of e-scooter trips according to their relationship with public transport. *Transportation***,** 1-22.

WANG, K., QIAN, X., FITCH, D. T., LEE, Y., MALIK, J. & CIRCELLA, G. 2022. What travel modes do shared e-scooters displace? A review of recent research findings. *Transport Reviews,* 43**,** 5-31.

WANG, R. 2017. Analysis of User Characteristics of Shared Bicycle in Shanghai and its Relationship with Public Transport Development. *Transport and Shipping,* 4**,** 46-52.

WU, Y., LI, W., YU, Q., LI, J. & LI, W. 2022. Analysis of the Relationship between Dockless Bicycle-Sharing and the Metro: Connection, Competition, and Complementation. *Journal of Advanced Transportation,* 2022**,** 1-16.

YANG, M., LIU, X., WANG, W., LI, Z. & ZHAO, J. 2016. Empirical Analysis of a Mode Shift to Using Public Bicycles to Access the Suburban Metro: Survey of Nanjing, China. *Journal of Urban Planning and Development,* 142**,** 05015011.

YE, J., BAI, J. & HAO, W. 2024. A Systematic Review of the Coopetition Relationship between Bike-Sharing and Public Transit. *Journal of Advanced Transportation,* 2024**,** 6681895.

YIN, Z., RYBARCZYK, G., ZHENG, A., SU, L., SUN, B. & YAN, X. 2024. Shared micromobility as a first- and last-mile transit solution? Spatiotemporal insights from a novel dataset. *Journal of Transport Geography,* 114**,** 103778.

ZHU, Z. & LU, C. 2023. Life cycle assessment of shared electric bicycle on greenhouse gas emissions in China. *Science of The Total Environment,* 860.