Hindawi
Journal of Advanced Transportation
Volume 2018, Article ID 4047682, 11 pages
https://doi.org/10.1155/2018/4047682

Research Article

An Association Rule Based Method to Integrate Metro-Public Bicycle Smart Card Data for Trip Chain Analysis

De Zhao,[1,2] **Wei Wang**,[3] **Ghim Ping Ong**,[4] **and Yanjie Ji**[3]

[1] *Jiangsu Key Laboratory of Urban ITS, Southeast University, Si Pai Lou No. 2, Nanjing 210096, China*
[2] *Department of Civil and Environmental Engineering, National University of Singapore, Engineering Drive 2, E1A 08-20, Singapore 117576*
[3] *School of Transportation, Southeast University, Si Pai Lou No. 2, Nanjing 210096, China*
[4] *Department of Civil and Environmental Engineering, National University of Singapore, Engineering Drive 2, E1A 07-03, Singapore 117576*

Correspondence should be addressed to De Zhao; zhaode_0@aliyun.com

Received 26 January 2018; Accepted 17 April 2018; Published 28 May 2018

Academic Editor: Lele Zhang

Smart card data provide valuable insights and massive samples for enhancing the understanding of transfer behavior between metro and public bicycle. However, smart cards for metro and public bicycle are often issued and managed by independent companies and this results in the same commuter having different identity tags in the metro and public bicycle smart card systems. The primary objective of this study is to develop a data fusion methodology for matching metro and public bicycle smart cards for the same commuter using historical smart card data. A novel method with association rules to match the data derived from the two systems is proposed and validation was performed. The results showed that our proposed method successfully matched 573 pairs of smart cards with an accuracy of 100%. We also validated the association rules method through visualization of individual metro and public bicycle trips. Based on the matched cards, interesting findings of metro-bicycle transfer have been derived, including the spatial pattern of the public bicycle as first/last mile solution as well as the duration of a metro trip chain.

1. Introduction

Public bicycle usage for metro access provides new opportunities for sustainable transportation, helping to address the "first-mile" and "last-mile" problems [1]. To understand the effect of integrating public bike and metro systems, transportation planners and researchers have been striving to evaluate the transfer efficiency and behavior through personal travel profiling [2], social-demographic information [1, 3], or public bicycle historical trips [4]. Previous attempts to understand metro and public bicycle transfer are limited, partly due to the difficulty in data collection. Conventional household travel surveys or diaries are time-consuming and laborious to carry out while convenient access to large travel datasets and integration across different data platforms are yet to be found.

In China, both metro and public bicycle transactions are made via Automatic Fare Collection (AFC) system, also known as smart card (SC). SC data, with a massive sample size, can provide valuable insights into the understanding of metro-public bicycle transfer behavior [5, 6]. Compared with conventional surveys, SC data collection, being a by-product of revenue collection, is a convenient method for retrieving travel patterns of commuters. Therefore, a great deal of research studies have emerged in terms of SC data mining [6–12].

However, the smart cards for metro and public bicycle systems are issued and managed by independent companies and in most cases, commuters need to hold two different smart cards to complete a public bicycle-metro trip chain. As such, the metro SC dataset and public bicycle SC dataset are saved independently without a common and unique identifier (ID) for a single commuter. This makes it difficult for researchers and transit agencies to leverage big SC data to investigate metro-bike transfer behavior effectively and efficiently, unless there exists a method to match the unique

card IDs within each system. It is not possible to directly and accurately match the smart cards from different datasets, as accessing personal information may offend the privacy. Nevertheless, the detailed individual travel pattern hidden in the SC data makes it possible to match card IDs of the same commuter.

Therefore, the primary objective of this study is to develop a data fusion methodology for matching metro and public bicycle smart cards of commuter identity in an integrated metro-public bicycle network. To achieve the aim, this study provides a novel approach using association rules (AR), a concept in the machine learning domain. The smart card data from Nanjing metro and public bicycle in China is used to demonstrate and validate our developed method. The remainder of the paper is structured as follows. In the literature review section, previous studies on smart card data for trip chain, multisource data fusion, and association rules are reviewed. The methodology and data source section formulates the association rules to match smart card travel and identifier data and presents the data source preprocessing procedures and validation approach. The results and validation section applies the proposed data fusion method to empirical analysis, calibrates the key parameters, and validates the proposed method. The section also presents possible application after matching metro and public bicycle card IDs. The conclusions and recommendations section concludes the paper and gives the research limitations as well as recommendations for future study.

2. Literature Review

Transit smart card data records transit riders' detailed trip log, which can be used to analyze the transit riders' trip chain. A huge body of literature has grown with regard to SC data analysis [7, 11, 14–17]. Means of mining SC data are various, e.g., data fusion and machine learning. Three streams of research are relevant to this study: (1) smart card data for trip chain; (2) multisource data fusion; (3) association rules.

2.1. Smart Card Data for Trip Chain. Past research studies in the literature have analyzed historical SC data to estimate transit origin location [18], destination location [6, 19], and total daily or monthly transit trip chain pattern [8, 10]. Furthermore, long-term year-to-year changes in transit users' trip habits could also be tracked and analyzed [12]. As for public bicycle, the research just started in recent years [20]. Notably, most of the research used bicycle trip data [21] rather than true SC data, since the trip data is easier to obtain. Trip data is usually open to the public in cities of United States or Europe, where the bicycle rental is accomplished via credit card or cell phone app. In general, compared with SC data, bicycle trip data lacks card ID and thus cannot be used to model users' travel behavior. By far, public bicycle SC data has been used to investigate public bicycle users' travel patterns [22] as well as bicycle trip chains for men and women [14, 23]. Public bicycle SC data could also help to classify different types of behaviors and compare the trip disparity [24].

2.2. Multisource Data Fusion. SC data provides much detailed information about each trip, but not the information about trip purpose, user assessment, and ultimate destination. When integrated with other data sources, SC data can play a greater role in mining transit riders' behavior and validating previous research approaches. By integrating both SC and Global Positioning System (GPS) data, Munizaga and Palma [9] estimated the OD of multimodal transit systems and validated the results against metro OD surveys in Santiago, Chile [16]. Ma and Wang [25] built a data-driven platform by integrating SC and GPS data to monitor transit performance in Beijing. Researchers can also examine the spatial-temporal dynamics of bus passengers and estimate the trip purposes when matching SC data, respectively, with General Transit Feed Specification (GTFS) data [15] and person trip survey data [17]. Yet, very little research attempted to integrate metro SC data with public bicycle SC data for investigating metro-bicycle transfer.

2.3. Association Rules. AR was first introduced by Agrawal et al. [26], and they applied this model to the supermarket transaction data to find out what items people would buy together. They also proposed algorithms for finding the AR. Shortly after that, the method was applied to other fields as a popular machine learning technique, including transportation. AR was firstly used in the transportation area by Keuleers et al. [27] to learn the travel patterns of multiday activity diaries. Soon after that, Keuleers et al. [28] tried to recognize temporal effects that may exist in the same data. AR showed high efficiency and convenience in rules mining. Later, Kusumastuti et al. [29] explored individuals' thoughts about leisure-shopping travel decisions by means of AR. Diana [30] used AR analysis to explore travel patterns of different modes based on 2009 US National Household Travel Survey and found the substitution effect between private modes and public transit. In particular, Chu and Chapleau [31] used AR to mine behavior rules of SC users and found some potential regularities with a high level of confidence.

Among all the relevant studies presented in this section, there still remains lack of a data fusion methodology to match smart cards from different sources. As AR is capable of identifying potential relationships between items, this paper attempts to develop AR-based algorithm to match metro and public bicycle SC data of the same person within an integrated bus-public bicycle network. We convert metro SC data and public bicycle SC data into transaction datasets and follow the method of Agrawal et al. [26] to match the card IDs. In our paper, we also propose an approach to validate the developed method.

3. Methodology and Data Source

3.1. Association Rules. Let $I = \{i_1, i_2, \ldots, i_k\}$ be a set of items. A transaction d_i is defined as a group of items, namely, a subset of I. $D = \{d_1, d_2, \ldots, d_n\}$ is a set of all transactions called the transaction database. Each transaction d_i in D has a unique transaction number. An association rule is used to describe potential relations of several items in the transaction

(1) L_1 = {large 1 − itemsets};
(2) for ($k = 2; L_{k-1} \neq \phi; k++$) do begin
(3) C_k = apriori-gen(L_{k-1}); //New candidates
(4) for all transactions $t \in D$ do begin
(5) C_t = subset(C_k, t); //Candidates contained in t
(6) for all candidates $c \in C_t$ do
(7) $c.count++$;
(8) end
(9) $L_k = \{c \in C_k | c.count \geq$ min support$\}$
(10) end
(11) All frequent sets = $\bigcup_k L_k$

Algorithm 1: Apriori algorithm [13].

database D and is expressed as $X=>Y$, where $X, Y \subseteq I$ and $X \cap Y = \emptyset$. X is called antecedent or left-hand side (LHS), and Y is called consequent or right-hand side (RHS). For example, in supermarket sales data mining, the rule {butter, bread}=>{milk} means if a customer buys both "butter" and "bread", he is also likely to buy "milk". In this research, we set all metro card IDs and public bicycle card IDs as items I. An association rule would indicate that there is potential association between card IDs in LHS and RHS. To better match two smart cards by AR, we should try our best to cluster two cards of the same person into one transaction d_i.

In association rules, there are three key parameters: support, confidence, and lift. The corresponding definitions are listed below. The support value of X, represented as supp (X), means the probability that the item-set X appears in the database D, defined as the proportion of transactions that includes the item-set X in D, as in (1). Accordingly, supp($X \Rightarrow Y$) can be expressed by (2). The generalized expression $X \cup Y$ in association rules means the union of the items in X and Y rather than either X or Y:

$$\text{supp}(X) = P(X) = \frac{\text{num}(X)}{\text{num}(D)}, \quad (1)$$

$$\text{supp}(X \Longrightarrow Y) = P(X \cup Y) = \frac{\text{num}(X \cup Y)}{\text{num}(D)}. \quad (2)$$

The confidence value of a rule, expressed as conf($X=>Y$), indicates the proportion of the transactions containing both X and Y in those that contain X:

$$\text{conf}(X \Longrightarrow Y) = P(Y \mid X) = \frac{\text{supp}(X \cup Y)}{\text{supp}(X)}. \quad (3)$$

The lift value is used to describe the effectiveness of the association rule, as defined in (4). A lift of 1.0 implies that the occurrence of X has nothing to do with that of Y. That is to say, no association rule can be found between X and Y when the lift = 1.0. When the lift is more than 1.0, $X=>Y$ is an effective association rule and greater value of lift indicates stronger association rule:

$$\text{lift}(X \Longrightarrow Y) = \frac{P(Y \mid X)}{P(Y)} = \frac{\text{supp}(X \cup Y)}{\text{supp}(X) \times \text{supp}(Y)}. \quad (4)$$

The Minimum Support (MS) value and the Minimum Confidence (MC) value are set as a constraint on measure of significance to ensure that the rules under consideration are sufficiently significant. MS is used to search the most frequent item-sets and MC is used to form rules based on these frequent item-sets. The former process is computationally intensive. To accomplish the former process, the commonly used Apriori algorithm is applied, as shown in Algorithm 1. As the Apriori algorithm can be found in previous research [13, 27, 28], we do not expatiate in this study.

3.2. Data Source and Preprocessing. Two major datasets were used in this study: metro SC data and public bicycle SC data, as shown in Figures 1(a) and 1(b). The datasets were recorded from November 1, 2015, to November 24, 2015, obtained from Nanjing Smart Card Company and Public Bicycle Company, respectively. The metro SC data contains the card ID, departure station, tap-in time, arrival station, and tap-out time. The original public bicycle data contains card ID, rent station, rent time, return station, and return time. Based on the location of the bicycle rent/return station, we added its nearby metro station in the 300 m buffer to the data frame. A buffer of 300 m radius was used as the walkable distance for public bicycle trips [32], because the planning standard promulgated by Nanjing government has suggested the walkable distance for community public facility is 300 m (5-min walk) [33].

The 24-day metro SC data contains over 34 million rows of trips and the 24-day public bicycle SC data includes nearly 1.2 million rows of trips. To mine AR between two databases, we need to merge and convert the SC data into transaction data. As mentioned above, we need to cluster two cards of the same person into one transaction to the greatest extent possible. The most likely transaction is metro-bicycle transfer or bicycle-metro transfer.

Firstly, we divide metro SC data and public bicycle SC data into subsets of regular time slot (TS) based on tap-in time and return time, respectively. Then, we merged the metro SC subset and the public bicycle SC subset with the same time slot and metro station (departure station and return nearby metro station, respectively) as one transaction. For bicycle-metro transfer, two cards of the same person may probably appear in the same transaction. Similarly, we created the metro-bicycle transfer transactions by cutting and merging

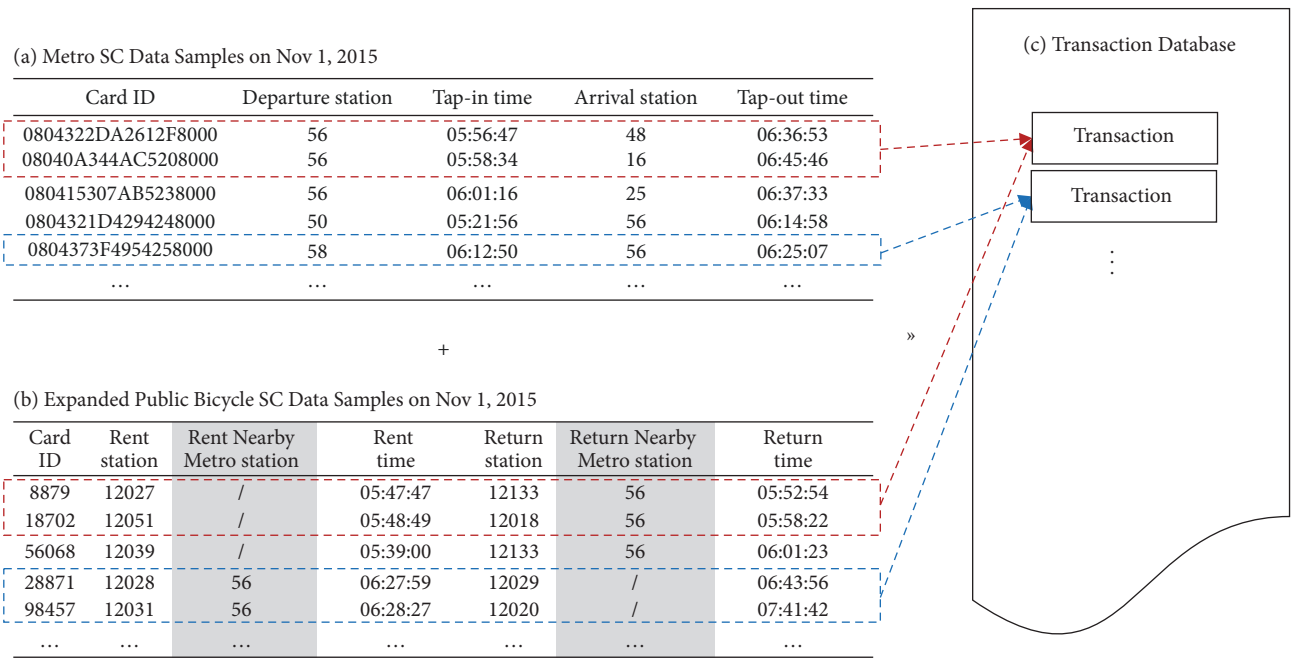


Figure 1: Smart card data and transaction dataset generation.

the datasets based on tap-out time, rent time, arrival station, and rent nearby metro station. All the transactions added up to create the transaction database D, as shown in Figure 1(c). In order to reduce the computational burden, we removed transactions with only metro card IDs or only public bicycle card IDs, because such transactions only contribute to finding internal relationship among metro cards or among public bicycle cards.

3.3. Validation Approach. Unfortunately, we cannot directly know whether two cards belong to the same person, because the only way to identify one person between different databases is to obtain nonanonymous personal information, which may violate individual privacy. Instead, we put forward a surrogated approach to validate the results: using the data in the first 20 days (train data) to train the association rules and the data in the last 4 days (test data) to validate the results. We assume the matched two card IDs do belong to the same person if they meet all the following three conditions.

(1) A transfer behavior of the two card IDs is also observed in the test data. A transfer behavior is defined as renting a public bicycle within 10 minutes after exiting the metro station or entering the metro station within 10 minutes after returning a public bicycle. We use 10 minutes as the maximum value of metro-bicycle (or bicycle-metro) transfer time. As mentioned above, the walkable distance between metro stations and public bicycle stations is 300 m, which is also equivalent to 5-min walk for an average person [33]. We set the maximum value of the transfer time as twice of 5 minutes.

(2) No time-overlap was observed between the matched two card IDs in 24-day datasets. In other words, one cannot take the metro while renting the public bike or rent a public bike during the metro ride.

(3) One metro SC ID only matches with one public bicycle SC ID. We assume each person usually owns one metro SC or public bicycle SC. Therefore, it makes sense only if one item (card ID) associated with only one other item.

Meeting all above three conditions by chance is quite a small probability event. Because there are over 1.8 million unique card IDs in the metro SC database and over 0.1 million unique card IDs in the public bicycle SC database. Based on our data, we randomly choose one card ID from the metro SC database and one card ID from the public bicycle SC database to check if they can meet the three conditions at the same time. We repeated 10,000 times of the selections; there are only 16 pairs of card IDs meeting all the validation conditions. Hence, the probability of meeting all above three conditions by chance is 0.16%.

4. Results and Validation

After cutting the SC datasets into thousands of transactions by the proposed method and learning the data with AR, associations between SC IDs are retrieved. The Apriori algorithm is used to identify the most frequent item-sets (metro SC and public bicycle SC). The associated rules learning in this research was performed using R 3.4.0. Given the three model parameters and the database of Nanjing (totally 1,149,335 items and 38,014 transactions), the average calculation time of associated rules is 18 seconds with the help of Apriori algorithm on a PC with Intel i7-6700 3.4 GHz and 16 GB DDR4 RAM.

To capture the meaningful results, we only selected rules with one metro SC ID as LHS and one public bicycle SC

Table 1: Extracted association rules.

No.	Association rules (LHS ⇒ RHS)	Support	Confidence	Lift
1	{0804322DA2612F8000} ⇒ {00008879}	5.78×10^{-4}	0.79	853.37
2	{08046A39EA02218000} ⇒ {00008429}	7.36×10^{-4}	0.97	1223.44
3	{0804142E2ABA268000} ⇒ {00196771}	5.52×10^{-4}	0.72	1019.53
4	{080E4711130A4D3C00} ⇒ {00168715}	5.52×10^{-4}	0.70	917.58
5	{0804293EE2382C8000} ⇒ {00151192}	7.36×10^{-4}	0.93	1182.66
6	{08042E156A85368000} ⇒ {00163152}	8.15×10^{-4}	0.91	936.75
...	...	...	...	...

ID as RHS. Because many-to-one rules, metro-metro rules, and bicycle-bicycle rules are all invalid, metro-metro rules probably mean that two or more cardholders often take the metro together. It is also true for bicycle-bicycle rules. These rules are not concerned in this research and thus removed from the association rules list. The extracted association rules are shown in Table 1. The metro card ID is represented as a string of 18 hex digits, while the public bicycle card ID is a string of 8 decimal digits. One AR indicates one metro SC ID matched with one public bicycle SC ID. All the extracted rules have a very high "lift" value, indicating the associations between cards are significantly strong. The support value scattered between 0 and 1.2×10^{-3}, indicating the parameter range for MS calibration.

There are three parameters MS, MC, and TS in our proposed model. They jointly determined the number of ARs (the number of matched IDs) and the accuracy of results. We need to obtain the optimal combination of three parameters in order to derive more ARs as well as better accuracy. Accuracy is defined as the ratio between the number of ARs meeting all the three validation conditions and the total number of ARs.

Figure 2 shows how the three parameters influenced the results. We set TS, respectively, as 2 min, 5 min, 10 min, and 20 min, as shown in Figures 2(a), 2(b), 2(c), and 2(d). In general, the accuracies of ARs under various parameter combinations are all very high, with most of them being over 90%. However, with increase of the number of ARs, the accuracy of results decreased. In other words, we cannot achieve the optimal levels of both number and accuracy simultaneously. To reach an accuracy of 100%, our proposed approach could at most identify 573 ARs (matched 573 pairs of SC IDs), with TS = 10 min, MS = 0.00055, and MC = 0.4 as marked in Figure 2(c). A too small a value of TS (e.g., 2 min) will reduce the maximum number of ARs because the metro card and public bicycle card of the same transfer could be divided into two transactions with high probability. On the other hand, too big TS (e.g., 20 min) could greatly improve the maximum number of ARs but may introduce many invalid ARs, thus decreasing the accuracy.

As mentioned above, there are totally 573 metro cards matched with bicycle cards under the accuracy of 100%. All the 573 paired cards have satisfied the 3 validation conditions. To ensure that the matched cards do belong to one person, we derived each individuals' trip log and visualized six of them as Figure 3 shows. The blue solid line segment indicates metro trip, while the red dashed line segment indicates public bicycle trip. All the metro trips and public bicycle trips of 24 days are displayed in this plot. The metro-bicycle transfer behavior of an individual can be easily identified by adjacent metro and public bicycle trip.

Apparently, Individuals 1, 2, 5, and 6 have a regular trip pattern. They used public bicycle as a daily first-/last-mile connection to metro. The majority of their metro trips are connected with public bicycle trips, showing specific symmetries between morning trip chain and afternoon/evening one. Individuals 3 and 4 are probably not traditional office workers with routine commutes, but they still take public bicycle as a good way to address the first-/last-mile problem.

After matching two cards of the same person, interesting findings of metro-bicycle transfer could be obtained. Around 2/3 of the transfers between metro and public bicycle occurred in the peak hours. Figure 4 shows the spatial analysis of public bicycle trips that connected with metro trips by peak hours of the day. The plot provides us with a visual impression of public bicycle trips as a first-/last-mile connection.

We can find that there are usually several public bicycle stops within the vicinity of metro stations and the trip demand is not shared equally among them. Users prefer to rent bicycles from only one or two of these stations, which further exacerbates the burden of rebalancing. In zones far away from the metro station, public bikes even take on longer-distance connections, for example, in the northern part of the research area. The maximum straight-line distance is over 4 km.

Another interesting finding from Figure 4 is that the first-mile trips during morning peak (7:00~9:00) have the same spatial pattern with last-mile trips during evening peak (17:00~19:00). It is also true for last-mile trips during the morning peak and first-mile trips during the evening peak. This makes sense since people who ride a public bicycle from home to metro station (first mile) in the morning tend to ride one from metro station (last mile) in the evening, and people who ride a public bicycle from metro station to workplace (last mile) in the morning tend to ride one back to metro station (first mile) in the evening.

The results of duration of the metro-bicycle trip chain are presented in Figure 5, based on the 573 matched cards. Outliers have been removed since some overlong trips will overestimate the average trip duration. The outlier is defined as any data point that is over 1.5 interquartile ranges below the first quartile or above the third quartile.

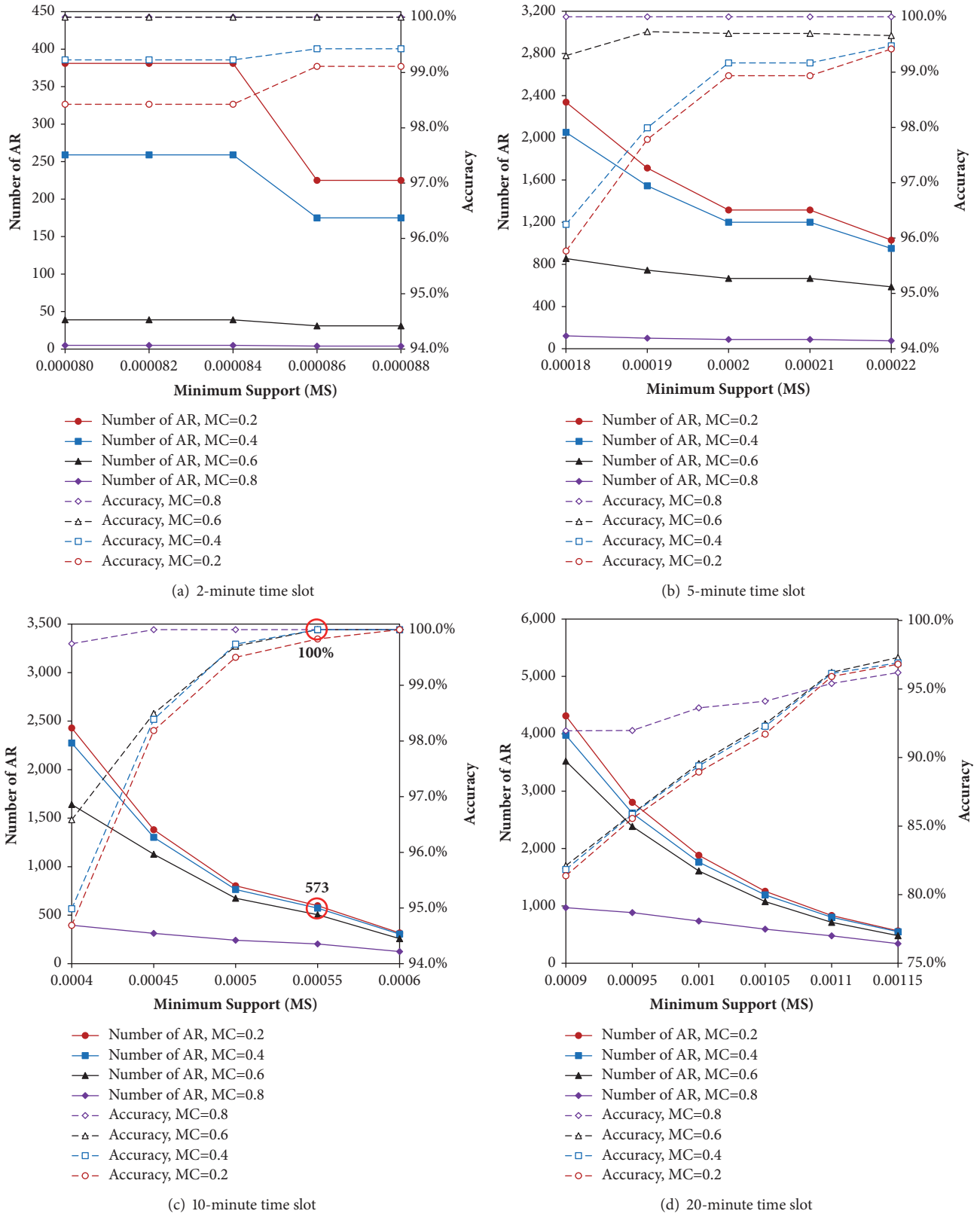


Figure 2: Parameter calibration.

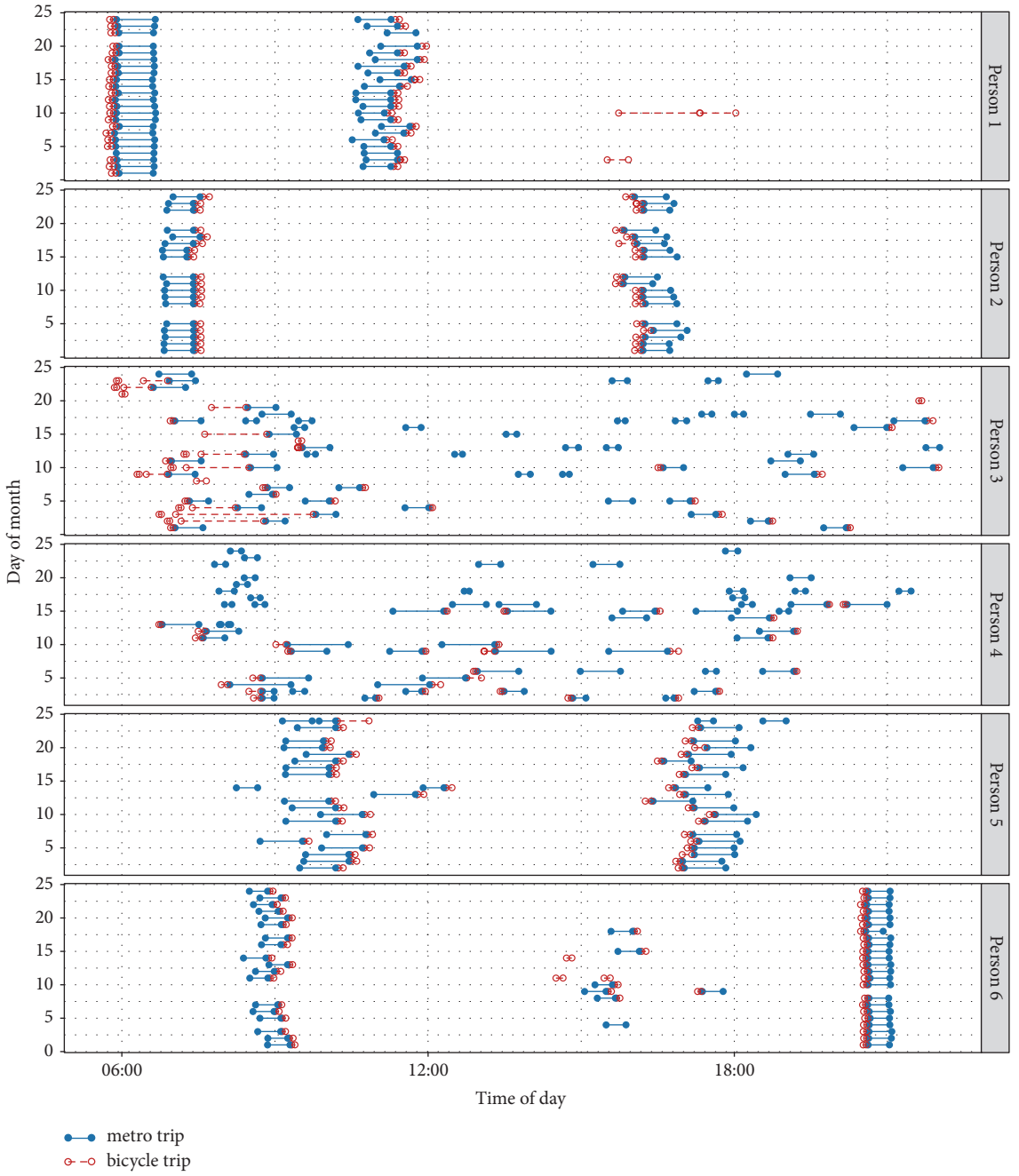


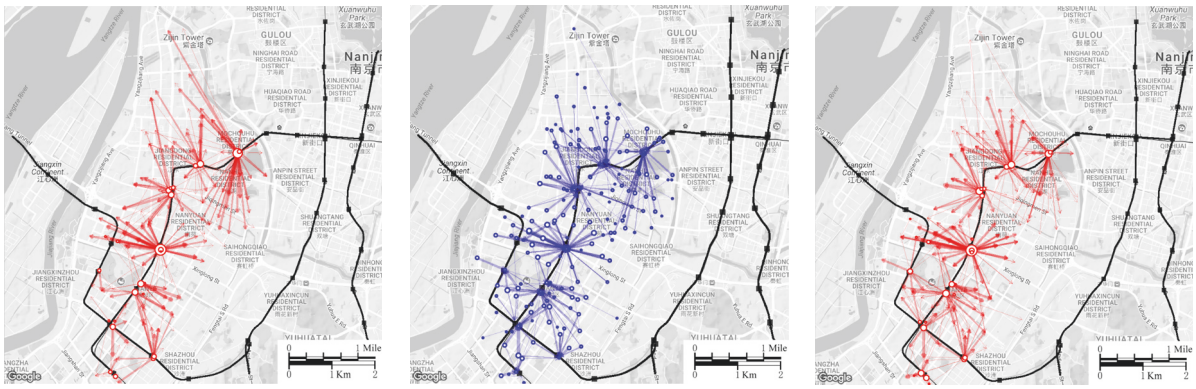
Figure 3: Individual trip log based on matched cards.

We divided metro-bicycle trip chain into metro trip, transfer, and public bicycle trip. For metro trip with last mile, the average metro trip time, transfer time and public bicycle time are respectively 23.59 min, 1.71 min, and 7.01 min. For metro trip with first mile, public bicycle time, transfer time and the average metro trip time are respectively 6.80 min, 1.75 min, and 23.50 min. The transfer time is very short, indicating that the walking distance between the metro station and bicycle stop is within a reasonable range. The connection time (public bicycle time and transfer time) takes around 27% of the total travel time.

When public bicycle was used as a first-mile mode, three parts of the trip chain in the evening peak are all longer than morning peak or even nonpeak hours. This makes sense since people have to hurry to work in the morning, but they can take their time back home in the evening. Counterintuitively, for both first-mile and last-mile mode, the public bicycle trip in nonpeak hours is shorter than morning peak. This is probably because, in rush hours, people are more likely to spend much time on finding an unoccupied slot to return the bicycle.

Travel patterns comparison between matched public bicycle SC and unmatched public bicycle SC are shown in Figure 6. Outliers have also been removed to prevent overestimating the average trip duration or trip distance. The results showed that the average trip duration for matched

(a) Morning peak, first mile (b) Morning peak, last mile (c) Evening peak, first mile



(d) Evening peak, last mile (e) Nonpeak, first mile (f) Nonpeak, last mile

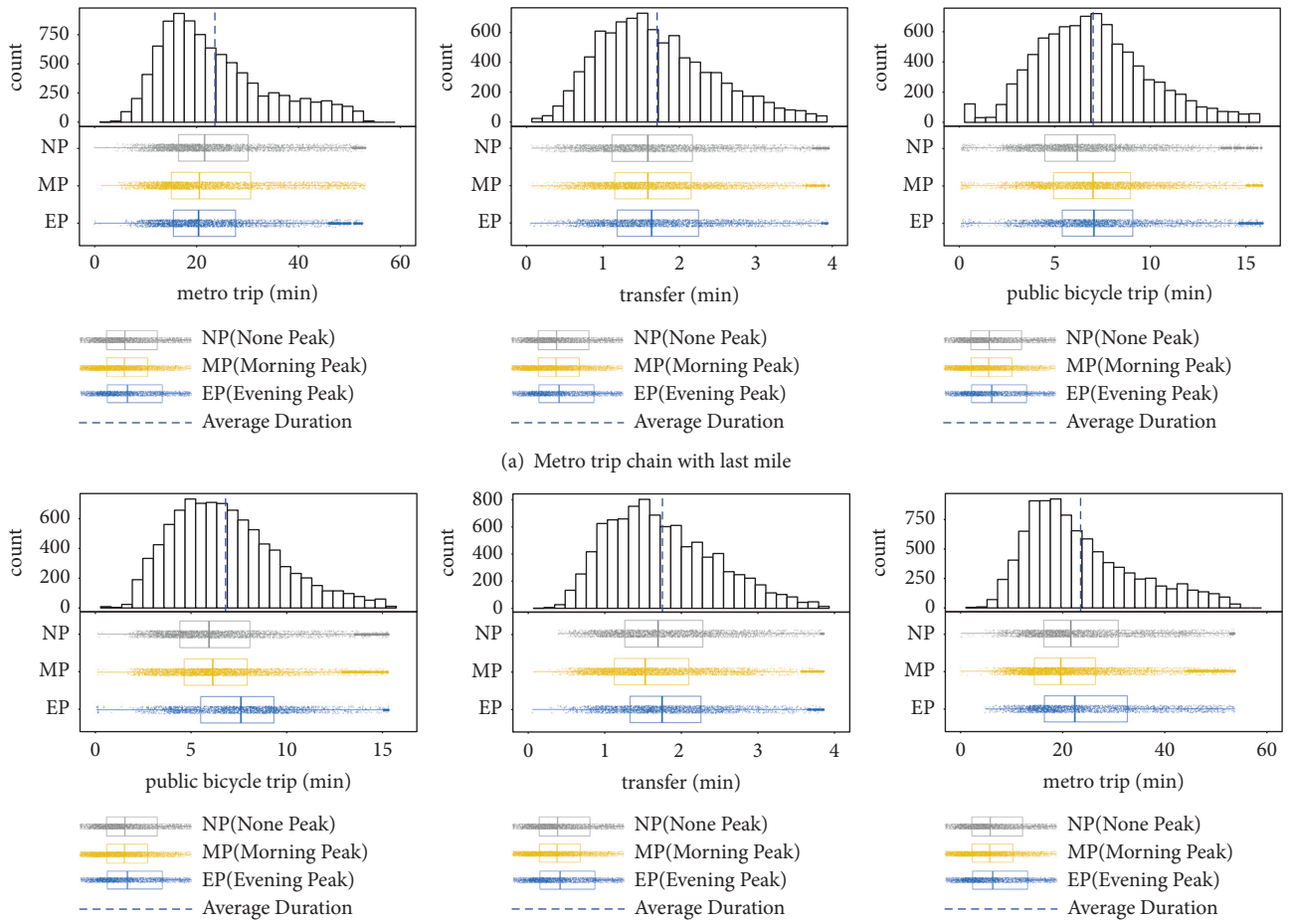
FIGURE 4: Spatial pattern of public bicycle as first-/last-mile solution.

public bicycle SC data is 6.80 min, while the average trip duration for unmatched public bicycle SC data is 10.02 min with 95% of the trip duration being less than 30 min, which is consistent with previous research studies by Zhao et al. [14]. The average trip distances for matched and unmatched public bicycle SC data are 0.95 km and 1.03 km, respectively, indicating the distance of first-mile or last-mile public bicycle trips is shorter than that of the other public bicycle trips. In contrast, the average trip distance is 0.99 km in Santander, Spain [24]. Notably, unmatched public bicycle SC data also contains first-mile or last-mile trips to connect metro.

5. Conclusions and Recommendations

This research has put forward a novel data fusion method using association rules mining to match metro and public bicycle smart cards of the same commuter. We attempt to match SC IDs from different sources and propose a validation approach. We calibrated the three key parameters MS, MC, and TS by demonstrating how they influenced the number of ARs and the accuracy of results. The validation process showed that, with increase of the number of ARs, the accuracy of results decreased. The individual trip log has also been derived to validate the association rules method by visualizing individual metro and public bicycle trip of each day. Based on the matched cards of the same person, interesting findings of metro-bicycle transfer have been found, including spatial pattern of public bicycle as first-/last-mile solution as well as duration of metro trip chain.

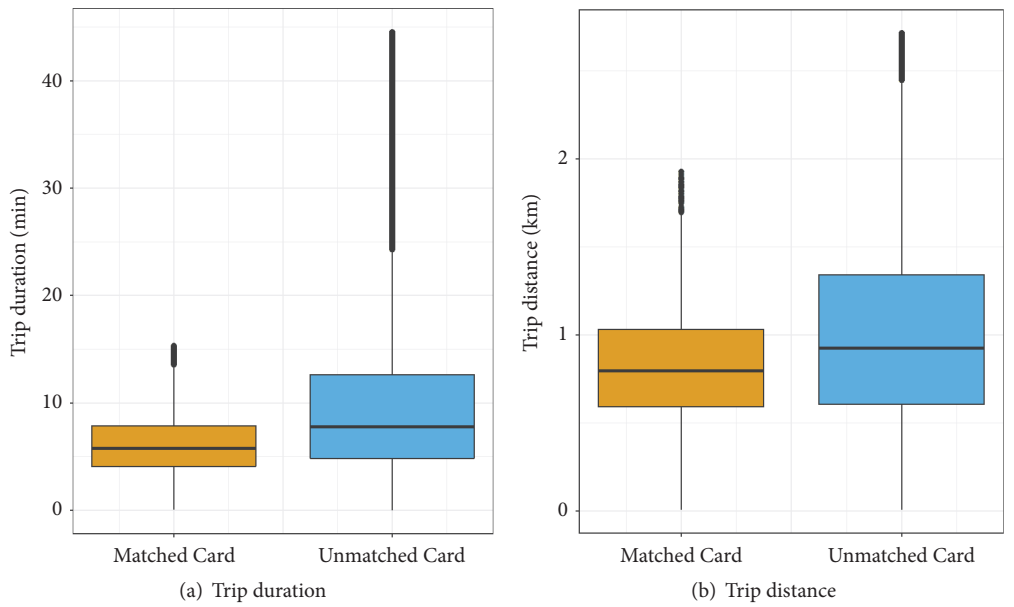
Our paper contributes to the state of knowledge by taking advantage of linked-SC data to analyze metro-bicycle transfers. We demonstrated that it is possible to match two cards of the same person based on historical SC data. Our proposed method successfully matched 573 pairs of smart cards with an accuracy of 100%, when setting three key



(a) Metro trip chain with last mile

(b) Metro trip chain with first mile

Figure 5: Duration of metro trip chain.



(a) Trip duration

(b) Trip distance

Figure 6: Travel patterns comparison between matched public bicycle SC and unmatched public bicycle SC.

parameters as TS = 10 min, MS = 0.00055, and MC = 0.4. TS determined the quality of conversion from SC data to transaction data, because TS is highly related to the metro-bicycle transfer time. However, even when transfer time is shorter than time slot (TS), the start time and end time of a transfer may still fall in two different transactions. Therefore, properly increasing the TS will help to reduce the rate of wrong grouping. Although the average transfer time is less than 1.8 min, the optimal TS of the proposed method is around 10 min.

Findings from this study suggest that around 2/3 of the transfers between metro and public bicycle in Nanjing occurred during peak hours. Although there are usually several public bicycle stops around metro station, users prefer to rent bicycles from only one or two of them, which of course exacerbates the burden of rebalancing. This paper also sheds light on the duration of the entire metro-bicycle trip chain. The connection time (public bicycle trip time and transfer time) takes around 27% of the total metro trip chain in Nanjing. When public bicycle was used as a first-mile mode, the trip duration in the evening peak is longer than morning peak or even nonpeak hours. This finding is consistent with common sense that people have to hurry to work in the morning, but they can take their time back home or dinner in the evening, while for both first-mile and last-mile mode, the public bicycle trip in nonpeak hours is even shorter than morning peak. This is probably because of the difficulty to find an unoccupied slot to return the bicycle during rush hours.

The proposed approach could be applied to other cities, where people also use different smart cards for different transportation modes. However, the parameters (MS, MC, and TS) in the model should be recalibrated based on actual SC data of the city. Because the city scale, transportation modes, and travel behaviors in different cities may greatly influence the standard of the parameters. There are several limitations of this study. Firstly, the total number of correctly matched cards were limited by the historical data. The dataset used in this research is not huge (only 24 days) due to the difficulty of obtaining both metro SC data and public bicycle SC data within the same period. Secondly, the proposed validation conditions are somewhat strict. This is because at least one metro-bicycle transfer (or bicycle-metro transfer) should occur in the remaining 4 days. Only frequent users of metro-bicycle transfer are likely to be successfully identified. However, this is the best way we can think of to validate the results without offense of individual privacy. More insightful findings about travel behaviors are expected to be found by harmonizing smart card technology across different transport modes. Speeding up the process of data integration and combining different smart cards into one will not only facilitate the traveler, but also provide data support for efficient transportation decision making.

Data Availability

The data that support the findings of this study are available from Nanjing Smart Card Company and Public Bicycle Company but restrictions apply to the availability of these data, which were used under license for the current study and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Nanjing Smart Card Company and Public Bicycle Company.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (71701047 and 51478112) and Projects of International Cooperation and Exchange of the National Natural Science Foundation of China (5151101143).

References

[1] Y. Ji, Y. Fan, A. Ermagun, X. Cao, W. Wang, and K. Das, "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*, vol. 11, no. 4, pp. 308–317, 2017.

[2] E. W. Martin and S. A. Shaheen, "Evaluating public transit modal shift dynamics in response to bikesharing: a tale of two U.S. cities," *Journal of Transport Geography*, vol. 41, pp. 315–324, 2014.

[3] J. Bachand-Marleau, J. Larsen, and A. M. El-Geneidy, "Much-anticipated marriage of cycling and transit: how will it work?" *Transportation Research Record*, vol. 2247, pp. 109–117, 2011.

[4] R. Nair, E. Miller-Hooks, R. C. Hampshire, and A. Bušić, "Large-scale vehicle sharing systems: analysis of Vélib'," *International Journal of Sustainable Transportation*, vol. 7, no. 1, pp. 85–106, 2012.

[5] Q. Zou, X. Yao, P. Zhao, H. Wei, and H. Ren, "Detecting home location and trip purposes for cardholders by mining smart card transaction data in Beijing subway," *Transportation*, vol. 45, no. 3, pp. 919–944, 2018.

[6] M. Trépanier, N. Tranchant, and R. Chapleau, "Individual trip destination estimation in a transit smart card automated fare collection system," *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, vol. 11, no. 1, pp. 1–14, 2007.

[7] M.-P. Pelletier, M. Trépanier, and C. Morency, "Smart card data use in public transit: a literature review," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 4, pp. 557–568, 2011.

[8] C. Morency, M. Trépanier, and B. Agard, "Measuring transit use variability with smart-card data," *Transport Policy*, vol. 14, no. 3, pp. 193–203, 2007.

[9] M. A. Munizaga and C. Palma, "Estimation of a disaggregate multimodal public transport Origin-Destination matrix from passive smartcard data from Santiago, Chile," *Transportation Research Part C: Emerging Technologies*, vol. 24, pp. 9–18, 2012.

[10] X. Ma, Y. J. Wu, Y. Wang, F. Chen, and J. Liu, "Mining smart card data for transit riders' travel patterns," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 1–12, 2013.

[11] X. Ma, C. Liu, H. Wen, Y. Wang, and Y. Wu, "Understanding commuting patterns using transit smart card data," *Journal of Transport Geography*, vol. 58, pp. 135–145, 2017.

[12] A.-S. Briand, E. Côme, M. Trépanier, and L. Oukhellou, "Analyzing year-to-year changes in public transport passenger behaviour using smart card data," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 274–289, 2017.

[13] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases," in *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB '94)*, Santiago, Chile, 1994.

[14] J. Zhao, J. Wang, and W. Deng, "Exploring bikesharing travel time and trip chain by gender and day of the week," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 251–264, 2015.

[15] S. Tao, D. Rohde, and J. Corcoran, "Examining the spatial-temporal dynamics of bus passenger travel behaviour using smart card data and the flow-comap," *Journal of Transport Geography*, vol. 41, pp. 21–36, 2014.

[16] M. Munizaga, F. Devillaine, C. Navarrete, and D. Silva, "Validating travel behavior estimated from smartcard data," *Transportation Research Part C: Emerging Technologies*, vol. 44, pp. 70–79, 2014.

[17] T. Kusakabe and Y. Asakura, "Behavioural data mining of transit smart card data: a data fusion approach," *Transportation Research Part C: Emerging Technologies*, vol. 46, pp. 179–191, 2014.

[18] X.-L. Ma, Y.-H. Wang, F. Chen, and J.-F. Liu, "Transit smart card data mining for passenger origin information extraction," *Journal of Zhejiang University SCIENCE C*, vol. 13, no. 10, pp. 750–760, 2012.

[19] J. Zhao, A. Rahbee, and N. H. M. Wilson, "Estimating a rail passenger trip origin-destination matrix using automatic data collection systems," *Computer-Aided Civil and Infrastructure Engineering*, vol. 22, no. 5, pp. 376–387, 2007.

[20] E. Fishman, "Bikeshare: a review of recent literature," *Transport Reviews*, vol. 36, no. 1, pp. 92–113, 2016.

[21] M. Ahillen, D. Mateo-Babiano, and J. Corcoran, "Dynamics of bike sharing in Washington, DC and Brisbane, Australia: implications for policy and planning," *International Journal of Sustainable Transportation*, vol. 10, no. 5, pp. 441–454, 2016.

[22] B. Caulfield, M. O'Mahony, W. Brazil, and P. Weldon, "Examining usage patterns of a bike-sharing scheme in a medium sized city," *Transportation Research Part A: Policy and Practice*, vol. 100, pp. 152–161, 2017.

[23] R. Beecham and J. Wood, "Exploring gendered cycling behaviours within a large-scale behavioural data-set," *Transportation Planning and Technology*, vol. 37, no. 1, pp. 83–97, 2014.

[24] M. Bordagaray, L. dell'Olio, A. Fonzone, and Á. Ibeas, "Capturing the conditions that introduce systematic variation in bike-sharing travel behavior using data mining techniques," *Transportation Research Part C: Emerging Technologies*, vol. 71, pp. 231–248, 2016.

[25] X. Ma and Y. Wang, "Development of a data-driven platform for transit performance measures using smart card and GPS data," *Journal of Transportation Engineering*, vol. 140, no. 12, Article ID 04014063, 2014.

[26] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 207–216, Washington, DC, USA, 1993.

[27] B. Keuleers, G. Wets, T. Arentze, and H. Timmermans, "Association rules in identification of spatial-temporal patterns in multiday activity diary data," *Transportation Research Record*, no. 1752, pp. 32–37, 2001.

[28] B. Keuleers, G. Wets, H. Timmermans, T. Arentze, and K. Vanhoof, "Stationary and time-varying patterns in activity diary panel data: explorative analysis with association rules," *Transportation Research Record*, vol. 1807, pp. 9–15, 2002.

[29] D. Kusumastuti, E. Hannes, D. Janssens, G. Wets, and B. G. C. Dellaert, "Scrutinizing individuals' leisure-shopping travel decisions to appraise activity-based models of travel demand," *Transportation*, vol. 37, no. 4, pp. 647–661, 2010.

[30] M. Diana, "Studying patterns of use of transport modes through data mining," *Transportation Research Record*, vol. 2308, pp. 1–9, 2012.

[31] K. K. A. Chu and R. Chapleau, "Augmenting transit trip characterization and travel behavior comprehension: multi-day location-stamped smart card transactions," *Transportation Research Record*, vol. 2183, pp. 29–40, 2010.

[32] Y. Zhang, T. Thomas, M. Brussel, and M. van Maarseveen, "Exploring the impact of built environment factors on the use of public bikes at bike stations: Case study in Zhongshan, China," *Journal of Transport Geography*, vol. 58, pp. 59–70, 2017.

[33] T.G.o. Nanjing, *Planning Standards for Public Facilities*, T.G.o. Nanjing, Nanjing, China, 2015.