

Research Article

Identifying Public Transit Commuters Based on Both the Smartcard Data and Survey Data: A Case Study in Xiamen, China

Shichao Sun ¹ and Dongyuan Yang ²

¹College of Transportation Engineering, Dalian Maritime University, Dalian 116026, China

²Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China

Correspondence should be addressed to Shichao Sun; sunshichao1988@hotmail.com

Received 23 May 2018; Revised 3 October 2018; Accepted 17 October 2018; Published 1 November 2018

Guest Editor: Hamzeh Khazaei

Copyright © 2018 Shichao Sun and Dongyuan Yang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Understanding the travel patterns of public transit commuters was important to the efforts towards improving the service quality, promoting public transit use, and better planning the public transit system. Smartcard data, with its wide coverage and relative abundance, could provide new opportunities to study public transit riders' behaviors and travel patterns with much less cost than conventional data source. However, the major limitation of smartcard data is the absence of social attributes of the cardholders, so that it cannot clearly extract public transit commuters and explain the mechanism of their travel behaviors. This study employed a machine learning approach called Naive Bayesian Classifier (NBC) to identify public transit commuters based on both the smartcard data and survey data, demonstrated in Xiamen, China. Compared with existing methods which were plagued by the validation of the accuracy of the identification results, the adopted approach was a machine learning algorithm with functions of accuracy checking. The classifier was trained and tested by survey data obtained from 532 valid questionnaires. The accuracy rate for identification of public transit commuters was 92% in the test instances. Then, under a low calculation load, it identified the objectives in smartcard data without requiring travel regularity assumptions of public transit commuters. Nearly 290,000 cardholders were classified as public transit commuters. Statistics such as average first boarding time and travel frequency of workdays during peak hours were obtained. Finally, the smartcard data were fused with bus location data to reveal the spatial distributions of the home and work locations of these public transit commuters, which could be utilized to improve public transit planning and operations.

1. Introduction

Public transit systems have long been regarded as an effective way to mitigate the growing urban congestion, exhaust emissions, and energy consumption caused by the excessive use of private automobiles [1, 2]. The effect is very significant especially in China where the level of car ownership has developed rapidly in the past decade [3, 4]. In order to unceasingly improve the performance and promote public transit use, the authorities in metropolises of China have been working for years to obtain a better understanding of passengers' travel characteristics [5–7]. In this context, mining the travel patterns of public transit commuters has received much attention of researchers [1, 4, 8–11]. It results

from the fact that the commuters not only represent the frequent public transit users, but also represent the ones who comprise the major component of travel demand during peak hours. The accurate knowledge of their demands and travel patterns may help the agencies adjust route plans, provide appropriate policies to retain the loyal riders and enhance the attractiveness of public transit systems [10]. However, acquiring the precise travel information of public transit commuters used to be challenging. The traditional way to analyze the public transit demand largely depended on travel diary surveys, which was often plagued by its high cost, low data accuracy, low response rates, and also privacy concern. Fortunately, Automatic Fare Collection (AFC) systems have been widely implemented in China as a more efficient way of

managing fare over the manual collection method [12]. At the same time, smartcard data can provide more abundant and higher quality travel data with less cost, through which travel patterns could be analyzed based on precise observations of individuals' smartcard usage in a time period [10, 13–18].

Nevertheless, the archived smartcard data collected from AFC systems also have its limitations [19]. For instance, smartcards are usually anonymous in China. Thus, social-demographic attributes of the cardholders were absent, such as gender, age, education, and trip purpose. Therefore, identifying public transit commuters from smartcard data was not a straightforward task [5]. To this end, spatial and temporal regularities of recurring trips on the public transit networks were usually incorporated to determine the public transit commuters among the cardholders [4, 6, 9], while, even if the spatial and temporal characteristics of each cardholder could be mined from the smartcard data, how the identified spatiotemporal travel pattern tied to a real commuter was still not clear [5]. To demonstrate it, a case study in Xiamen, China was given, where the smartcard dataset harvested from 1st to 26th June, 2015 was employed. Specifically, for each cardholder, we calculated the number of workdays where he/she swiped cards during both the morning peak and the evening peak (on a maximum of 20 workdays). Then, we obtained the distribution diagram of cardholders according to the calculated number of workdays, as illustrated in Figure 1. In this context, a large number of related works usually adopted a hypothetical threshold or clustering methods to segment the commuters and noncommuters (e.g., cardholders traveling for at least a number of days are identified as transit commuters) [4, 12, 20]. However, there seemed to be no obvious fluctuation between the number of cardholders corresponding to “9 workdays” and the subsequent counterparts. Thus, it was really hard and confused to determine an exact threshold value for the identification according to the statistics in Figure 1. Moreover, even if a threshold value could be set up, it was still challenging to test the accuracy of results without any further circumstantial evidence. Therefore, in the case study of Xiamen, it was difficult to identify the commuters among cardholders by using the traditional approaches derived from the literature.

To this end, an original machine learning algorithm called Naïve Bayesian Classifier (NBC) was adopted in this paper to identify public transit commuters based on both the smartcard data and survey data. Compared with existing methods which were plagued by the validation of the accuracy of the identification results, the employed approach was a machine learning algorithm with functions of accuracy checking. The rest of the paper was organized as follows. In Section 2, a review of literature on smartcard data application was drawn, and the existing approaches of identifying commuters in smartcard data were discussed. Then, the Naïve Bayesian Classifier was introduced, and the classifier was built, trained, and tested in Section 3. Subsequently, the attribute of “commuter or not” for each cardholder was estimated based on the tested classifier, and travel characteristics and regularities of the identified commuters were analyzed in Sections 4 and 5. In the end, conclusions were drawn, and future directions were under discussion.

2. Literature Review

In recent years, the application of smartcard data mining in the planning and management of public transit systems has received much research attention. The previous works can be grouped into three categories, which can be, respectively, described as tactical-level studies, operational-level studies, and strategic-level studies [10].

In the tactical level, the trip pattern analysis and market segment were frequently addressed, contributing to the operational adjustment in public transit [14, 15]. Zhao et al. [21] developed a methodology to detect individuals' travel pattern changes by using smartcard data from London, UK, over two years. They specified one distribution for each of the three dimensions of travel behavior, and a Bayesian method was developed to identify significant change points in travel patterns. The results show that, compared to the traditional generalized likelihood ratio (GLR) approach, the Bayesian method required less predefined parameters and was more robust. Ma et al. [6] employed a Density-Based Scanning Algorithm with Noise (DBSCAN) to segment the smartcard data by analyzing different travel patterns. Then, spatial and temporal regularity of each cluster was derived through a continuous long-term observation period. The original DBSCAN method was improved by Kieu et al. [8, 9] and then applied to divide the passengers under a much lower calculation complexity. The heterogeneity of public transit riders was analyzed by Langlois et al. [22] through four-week smartcard data, and 11 travel patterns were generated. Legara and Monterola [16] developed a new classification method with promising accuracy by using the concept of eigentravel matrices which captured a commuter's characteristic travel routine, while, in operational-level studies, precise performance indicators such as schedule adherence, the number of transfers, and vehicle-kilometers on a public transit network were discussed [1, 10, 13]. Among them, the study of origin-destination (OD) and interchange inference was a hot topic and received much research attention. Trepanier et al. [23] and Ma et al. [24] combined AFC data and AVL data to infer boarding locations. For the alighting location estimation, Munizaga and Palma [7] set up a disutility function which took account of time and distance and then bus destinations were inferred by minimizing the generalized cost of the function. Sanchez-Martinez [17] extended the disutility function approach to rail networks, and a dynamic programming algorithm was designed to infer destinations. Wang et al. [25] applied the trip-chaining to infer bus passenger OD from AFC and AVL data, while Carrel et al. [26] inferred OD matrices by using smartphone and AVL data.

Regarding the applications on the strategic side, although previous researches have adopted smartcard data for the assessment and planning of public transit systems, however, the data source was somehow fragmentary for the behavior analysis [19]. For example, most of the previous literature on the analysis of cardholders' commuting characteristics oversimplified the definition of a public transit commuter [5]. In this context, Kusakabe and Asakura [19] developed a data fusion method based on the Naïve Bayesian Classifier (NBC) by integrating the smartcard data with the personal travel

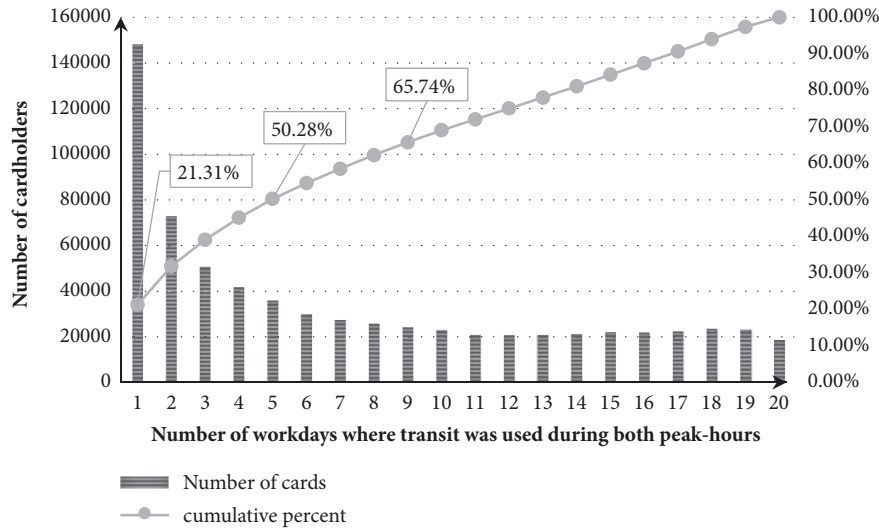


FIGURE 1: Basic statistics of smartcard data in Xiamen, China.

TABLE 1: Examples of smartcard transaction records.

Card id	Date	Transaction Time	Bus Route	Bus Plate Number
326352926	2015/6/22	09:43:47	Bus Route 815	D3789
154380354	2015/6/23	06:40:50	Bus Route 301	D3775
323446469	2015/6/23	20:40:13	Bus Route 713	D9342
051397471	2015/6/24	11:40:23	Bus Route 301	D3764
360122282	2015/6/24	08:42:30	Bus Route 512	D6453

survey data. The approach was intended to estimate absent attributes of trip purpose in smartcard data and enhance the understanding of travelers' behavior. Nevertheless, their study was implemented at only one railway station run by a private company, and the proportion of cardholders only accounted for about 10% of the total ridership. Thus, the applicability, portability, and representativeness of the proposed approach may be affected by the sample size. Moreover, other than estimating the travel purpose for each trip, it was also not clear whether the data fusion method could be applied to infer the attribute of cardholders. Since achieving the identification of public transit commuters in smartcard data was not an easy task in most metropolis of China, this paper extended the application of NBC approach and employed the method to estimate the attribute of cardholders instead of trip purpose. Besides, as an extension to the work of Kusakabe and Asakura [19], the approach was adopted and applied to the whole public transit network in Xiamen, China, with 85% of the transactions completed through smartcards. The spatial distribution of commuters' home and work locations in the case study were also revealed by fusing together smartcard and Automatic Vehicle Location (AVL) data.

3. The Smartcard Data

3.1. A Brief Introduction of Xiamen, China. Xiamen is administered as a subprovincial city of Fujian province with an area

of 1,699.39 square kilometers. The population in Xiamen was nearly 3.5 million according to the Census in 2014. Regular bus systems and the Bus Rapid Transit (BRT) are the principal public transportation modes for local residents. There are nearly 320 regular bus routes and 5 BRT lines currently in service in Xiamen.

3.2. Data Preparation. "E-Tong Card" is the only contactless smartcard used for electronic payments in the public transportation system of Xiamen, China, and it started to come into use in 2006. Through the smartcards, passengers can access all the bus routes and BRT lines. Until 2015, the circulation of E-Tong cards has exceeded 6 million, and over 85% of the transactions in the public transit network were completed through the automated fare system. Thus, it provides a good opportunity for researchers to obtain the users' travel activities through smartcard data mining.

The smartcard data analyzed in this paper were harvested from the archived data in the AFC system of "E-Tong Card". The original dataset contained whole transaction records on the public transit network, collected from 22nd to 26th June, 2015 (except elder cards and student cards; 5 consecutive workdays were covered). Since the pricing of the bus system was not distance-based in Xiamen, the passenger only swiped the card once when boarding. Each record included the information of card id, date, transaction time, and bus route as well as plate number, as shown in Table 1.

4. Methods

4.1. The Naïve Bayesian Classifier. The Naïve Bayes classifier was a machine learning technique efficiently utilized for classification/identification applications in data mining environment. The reliability in predicting and decision-making constructed the statistical nature of the NBC. In practice, NBC could estimate the absent attributes of the data by predicting probabilities of class membership. To classify a new instance, the algorithm used the Bayesian rule to calculate the conditional probability of each class value and took the class with the maximum probability as the identified class. The collected survey data were usually split into two parts: training part and test part. The algorithm used the training data to build the classifier and then estimated the required probability values in the test data to test the accuracy of the built classifier.

Let vector $\mathbf{F} = \{f_1, f_2, \dots, f_N\}$ be a set of behavioral attributes/features, as shown in Figure 2. Each element of vector \mathbf{F} represented a common attribute or travel characteristic which existed in both survey and smartcard data, such as travel frequency or boarding time. Let C be a variable of the absent classification item in the smartcard data but could be measured by survey data [19]. It was assumed that elements of vector \mathbf{F} were conditionally independent when C was given. In this study, classification item C and elements of vector \mathbf{F} were treated as discrete variables. Then, $p(C | \mathbf{F})$ could be trained through Bayes' theorem by using the survey data.

$$\begin{aligned} p(C | \mathbf{F}) &= \frac{1}{p(\mathbf{F})} p(C) p(\mathbf{F} | C) \\ &= \frac{1}{p(\mathbf{F})} p(C) \prod_{N=1}^{N'} p(f_N | C) \end{aligned} \quad (1)$$

In (1), $p(C)$, $p(\mathbf{F})$, and $p(f_N | C)$ could be estimated by survey data. Specifically, $p(C)$ and $p(\mathbf{F})$ were the probabilities derived from the proportion of interviewees having classification item C and vector \mathbf{F} , respectively. The conditional probability distribution $p(f_N | C)$ represented the composition rate of interviewees having attribute f_N corresponding to classification item C .

The attribute vector \mathbf{F} could also be measured for each cardholder in the smartcard dataset, since it was shared in both of the two data sources. Then, the trained classifier could be adopted to estimate the absent attribute C of cardholders. It could be expressed as the following equation:

$$\hat{c}(\mathbf{F}) = \arg \max_{c \in C} p(c | \mathbf{F}), \quad C = \{c_1, c_2, \dots, c_M\} \quad (2)$$

where c was one item of the classification attribute C .

4.2. Survey Design. The flow chart of the employed NBC method was illustrated in Figure 3. The core of machine learning in this study was to estimate absent elements of smartcard data by using the survey data to train and test the classifier. The survey was designed as shown in Figure 3. The classification item C could only be observed in the survey data, and it could be measured through the

interviewees' response to "whether the bus service is your first choice when commute". The attribute vector \mathbf{F} represented behavioral attributes that could be observed by both of the two data sources. In this study, the vector \mathbf{F} contained three independent attributes, and it was relatively easy to measure them by using any one of the two datasets: (f_1) "average first boarding time on the workdays in last week"; (f_2) "average last boarding time on the workdays in last week"; and (f_3) "the number of workdays where commuting by bus in the last week", which represented the number of workdays where the cardholders swiped cards during both of the two peak-hours in the last week.

5. Experimental Analysis

5.1. The Trained and Tested NBC. The NBC model was built and tested by the use of survey data harvested from Xiamen, China. The research area in this paper was not broad with about 150 square kilometers; it can be assumed that the characteristics of public transit users were homogenous all over the city. The survey was implemented in the major industrial parks, software parks, and CBDs of Xiamen on 29th, June, 2015. Since these areas nearly provided more than 70% of job positions in Xiamen [27], the samples could be used to serve as the representative of typical commuters in Xiamen. The interviewees were selected at their workplaces with random sampling by meeting the conditions: (1) permanent staff with smartcards and (2) usually commute between home and their workplaces. Thus, the definition of public transit commuters as well as its response indicator "whether the bus service is your first choice when commute" in the survey could be better understood by the selected interviewees in a standard way. The survey was designed as above mentioned, through which the three independent attributes and the classification item could be measured. A total of 900 questionnaires were issued. Eventually, 532 valid samples were collected with 62% (330) of which stated themselves as public transit commuters. The reliability of the survey data was acceptable, since the value of Cronbach's Alpha was greater than 0.8.

Regarding the descriptive statistics of the valid samples (Table 2), 54.1% of the interviewees were female corresponding with 45.9% of which were male. It was in line with the census data of Xiamen in 2014. The age of the valid interviewees mostly fell in the range between 18 and 60. Almost half of the passengers were between 21 and 30 years old (49.6%). Thus, it was also consistent with the characteristics of commuters and meet the feature of smartcard dataset excluding student cards and elder cards. Statistics related to the occupation and income of the interviewees were also calculated. It indicated that company employees (46.7%) accounted for almost half of the samples. Correspondingly, more than half of the interviewees earned less than 4500 yuan (RMB) per month, whereas there was still 38.6% of the ones had a monthly income of less than 3,000 yuan (RMB). The average monthly salary of Xiamen in 2015 was 3,508 yuan (RMB), according to the salary report of employees in 2016. Thus, it implied that there was no potential bias from the selected interviewees regarding the income.

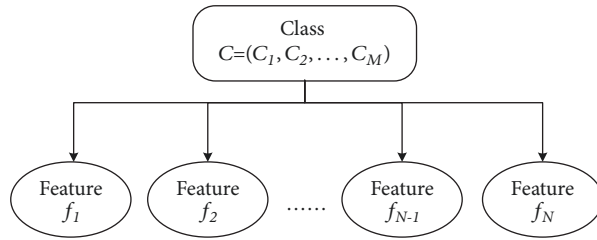


FIGURE 2: Basic structure of the NBC.

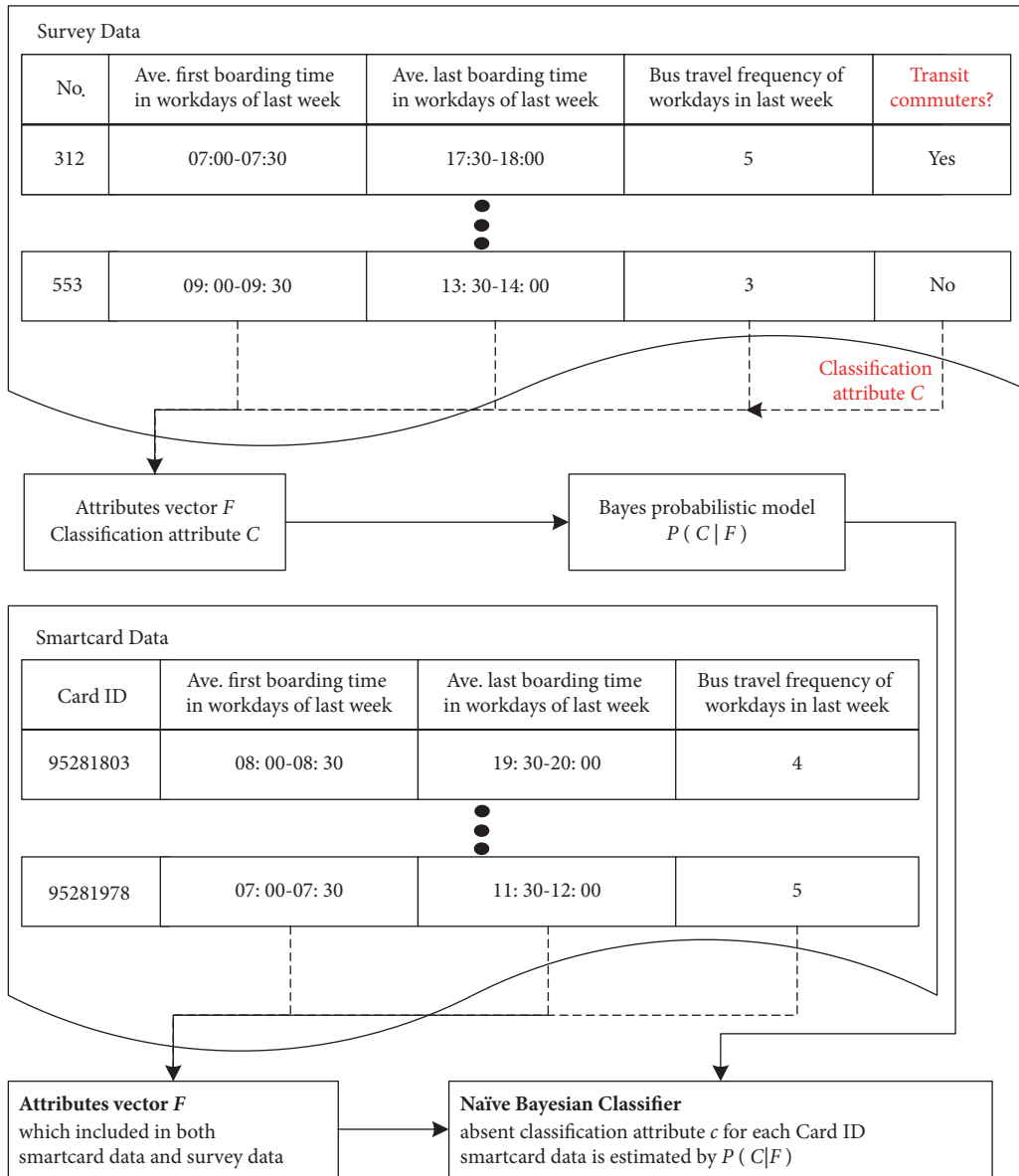


FIGURE 3: Flow chart of NBC for identifying commuters in smartcard data.

Then, the survey data were divided into two datasets: a training dataset and a testing dataset. Each dataset contained 266 valid samples, respectively, 165 of which were stated public transit commuters. By using the training dataset, the conditional probability distribution $p(f_N | C)$ could be

derived from the proportion of interviewees having attribute f_N ($N = 1, 2, 3$) corresponding to each classification C (public transit commuters or not), shown in Figures 4(a), 4(b), and 4(c). Since $P(F)$ did not depend on C , it could be regarded as a constant value when F was given. The distribution $P(C)$

TABLE 2: Descriptive statistics of valid samples.

Attributes	Statistics
Gender	Male (45.9%), female (54.1%)
Age (years)	<=20 (1.2%), 21-30 (49.7%), 31-40 (24.0%), 41-50 (11.7%), 51-60 (11.9%), ≥60 (1.5%)
Occupation	Company employees (46.7%), civil servant (6.8%), educator and researcher (8.8%), service worker (20.7%), self-employed worker (14.6%), other (2.4%)
Monthly income (Unit: RMB)	<=3000 (38.6%), 3001-4500 (20.1%), 4501-6000 (25.6%), >6000 (15.7%)

TABLE 3: Test Results of the trained model.

	Transit Commuters (165 samples)	Non-Transit Commuters (101 samples)
Correctly estimated	151	87
Failure number	14	14
Accuracy Rate	92%	86%

could also be easily derived from the training data, which was calculated here as follows: $P(C_{commuter}) = 62\%$.

Based on (1) and (2), when a new attribute vector \mathbf{F} was given, the conditional probability of each classification item could be calculated. The item with the maximum probability could be considered as the identified class. For example, the probability of an interviewee being a public transit commuter could be calculated as $p(C_{commuter} | \mathbf{F}) = 97.4\%$, if his/her attribute vector could be denoted as

$$\mathbf{F} = \{f_1 = 07 : 00 : 00 - 07 : 30 : 00, f_2 = 16 : 30 : 00 - 17 : 00 : 00, f_3 = 5\} \quad (3)$$

Then, the trained NBC model was tested by using the testing dataset. Firstly, only the attribute vector \mathbf{F} of each interviewee was utilized in the NBC model to predict their class values. Then the accuracy would be checked by comparing the estimated class membership with the actual classification of the interviewee [19]. Test results were reported in Table 3.

For a total of 266 samples in the testing dataset, the average accuracy rate could be calculated as $[1 - (14+14)/266] * 100\% = 90\%$, including 14 failures in each of classification items. Specifically, 92.0% of public transit commuters (151 samples) and 86.0% of nonpublic transit commuters (87 samples) were correctly identified. Since the Root Mean Square Error for the identification of public transit commuters was less than 0.3 (0.291), the accuracy rate of the trained NBC could be considered acceptable. Regarding a smartcard dataset which contained M cardholders, when P of them were identified as commuters by using the NBC ($P < M$), then the actual number of public transit commuters among cardholders could be calculated according to (4), taking the accuracy rate into account.

$$\text{Actual Number} = [P * 0.92 + (M - P) * 0.14] \quad (4)$$

5.2. Results and Discussions. The NBC was employed to identify the commuters in the smartcard dataset, which was trained and tested by using survey data. The identification

results and some other statistics were listed in Table 4. Eventually, nearly 290,000 cardholders (41.94% of the total) were identified as public transit commuters. According to the error analysis based on test instances in the last section, the accuracy rate for this result of commuter-identification should be around 92%, corresponding with the Root Mean Square Error of 0.291. Regarding the identified commuters, the statistics reflected that they averagely began their trips at 7:30-8:00 and swiped cards for their last-boarding at 17:30-18:00, 87% of which swiped cards twice in a typical workday. This was likely to represent a typical commuting trip chain, where public transit riders took a bus from home to their workplaces in the morning and then returned home in the evening [6].

Then, in order to better improve and test the accuracy of results, we cross-checked the bus route information of each identified commuter who swiped cards twice per workday. Due to the bus routes overlap, the demand of cardholders' first travel may be fulfilled by different bus routes in a period time. Nevertheless, during the 5-workday period, the statistical results indicated that the bus routes corresponding to commuters' first boarding were stable, more than 93% of which depended on the same bus route. In addition, we also examined whether the bus route corresponding to each commuter's last trip per workday was contained in his/her bus route dataset of first boarding. Since a typical commuting chain reflected a home-work-home journey, the bus route/metro line used in commuter's last trip on the workday should have a strong correlation with the ones of his/her first boarding. The results indicated that 96% of the assumed home-work-home journeys (2 trips per workday) satisfied the above conditions. Thus, the primary OD demand of public transit commuters can be obtained by connecting the first-boarding locations with the last-boarding locations of the identified commuters who swiped cards twice per workday. Since 87% of the identified commuters swiped cards twice per workday, it implied that the planning and the network structure of bus lines in Xiamen were really well to fulfill the demand of commuters. However, the average bus line overlap factor in Xiamen was 5.15 [27]. Thus, in

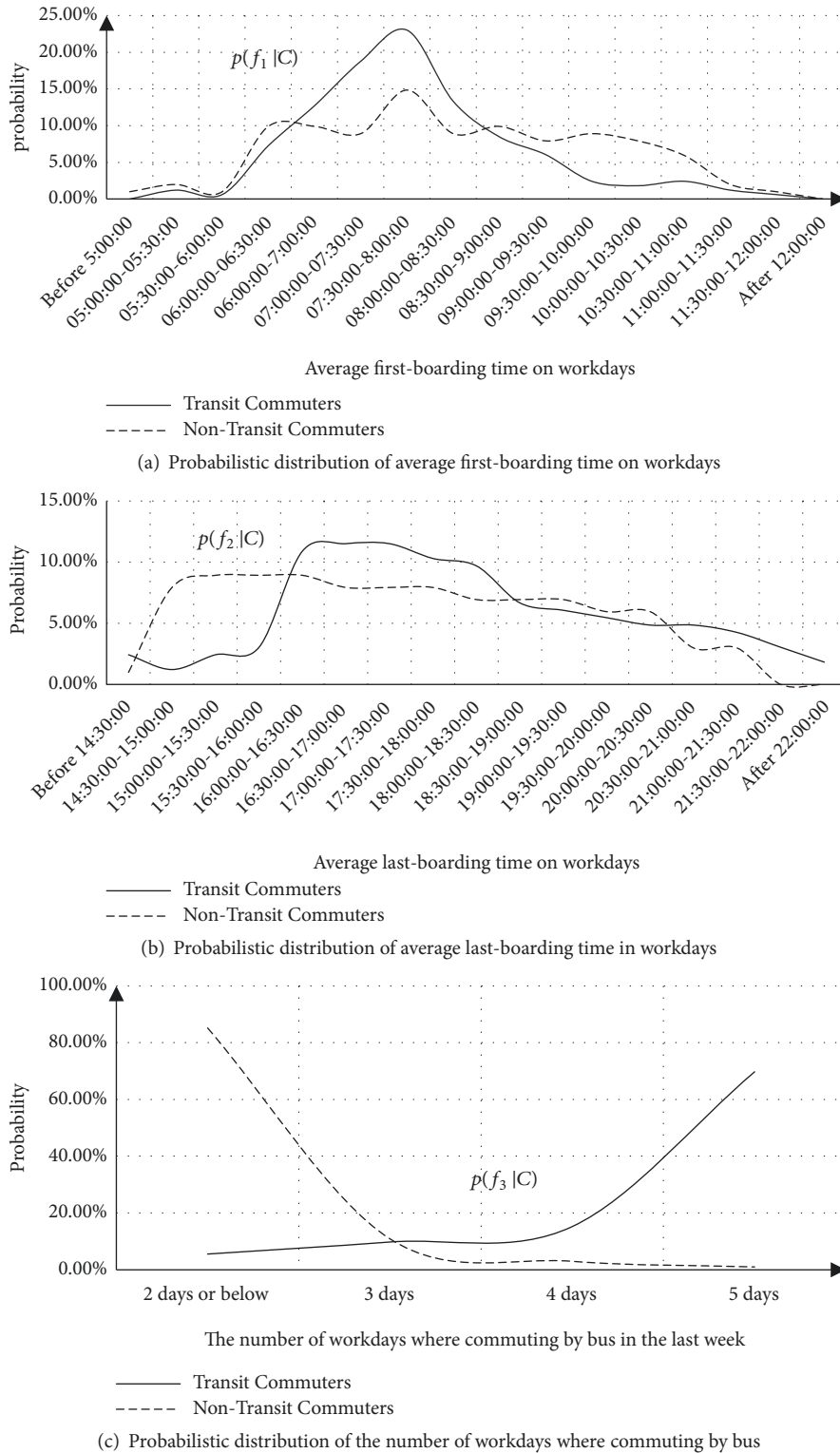
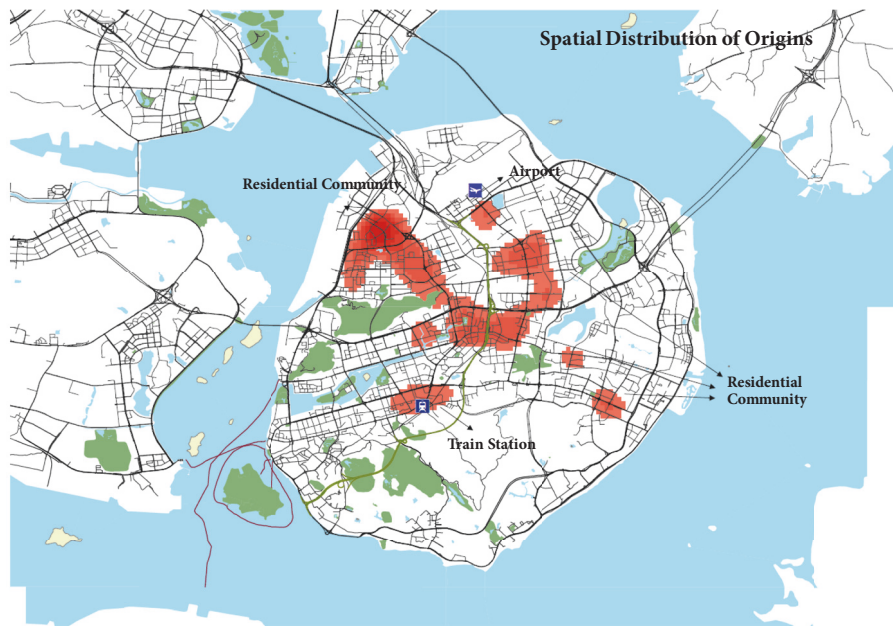


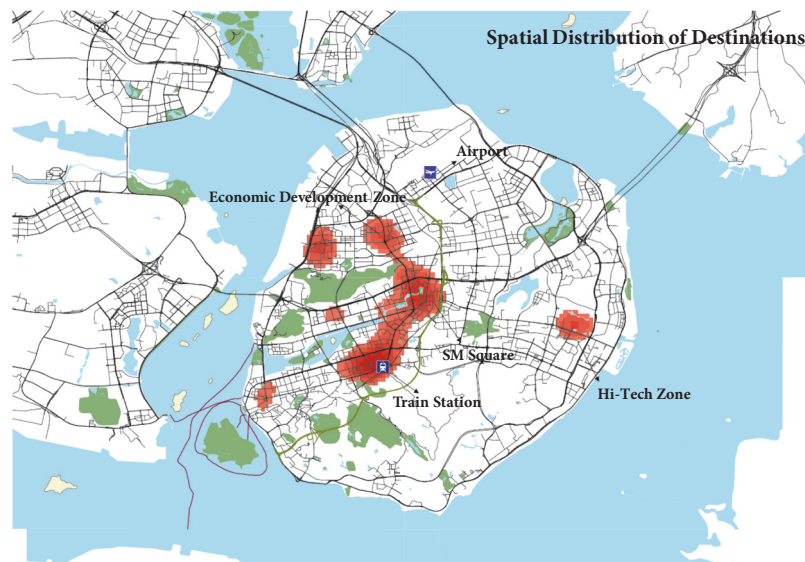
FIGURE 4

TABLE 4: Identification results of the NBC model.

Transit Commuters	Number of cards	Ave. first-boarding time	Ave. last-boarding time	Travel frequency of workdays in a week
	288153	7:30-8:00	17:30-18:00	4.2 days



(a) Spatial distribution of origins of commute activities



(b) Spatial distribution of destinations of commute activities

FIGURE 5

contrast, the above results also reflected that the existing bus overlap lines have not played an effective role in supporting the majority of commuting activities, which may lead to an extreme crowding of bus routes on main public transit corridors.

As for the analysis of primary OD demand of public transit commuters, however, the AFC system was not initially integrated with AVL system in Xiamen, China. Thus, the boarding-location should be estimated by integrating the two separate databases. In this paper, we employed the estimation method of boarding location proposed in the literature of Ma et al. [6]. It matched the transaction time of smartcard data with the time of GPS data generated from the bus with the same plate number and then inferred the location where he/she got on board. Finally, it obtained the station information by minimizing the distance between the boarding-location and bus stations on the same direction of the route. About 94% of the transaction records were finally matched with the GPS data by minimizing the time difference and then obtained the boarding-location information. Then, based on the boarding-location information, the distribution of commuters' origins and destinations of commuting trips could be obtained, shown in Figure 5. It was likely to depict a whole picture of the distributions of their residence and workplaces. Specifically, the results indicated that the spatial distribution of public transit commuters' origins was disperse when compared with the distribution of their destinations. Most of the public transit commuters lived in several core residential communities. However, workplaces of public transit commuters were more concentrated in downtown and several development zones. For instance, a large proportion of public transit commuters worked at the downtown area near the train station as well as the SM Square which was the CBD of Xiamen. Another two concentration areas of employees were located in the economic development zone in the northwest and the Hi-Tech parks in the east of Xiamen.

6. Conclusion and Future Researches

This paper employed the Naïve Bayesian Classifier to identify public transit commuters based on both the smartcard data and the survey data. A case study was given in Xiamen, China, and the classifier was trained and tested by related survey data collected from 532 valid samples. Then, it was applied to the smartcard cardholders for the identification of public transit commuters. The results indicated the following:

- (i) The success rate of the identification of public transit commuters was 92% in the test case. Nearly 290,000 cardholders were classified as public transit commuters. However, considering the error of the NBC for identification of commuters, the actual number of public transit commuters with smartcards in Xiamen should be around 323,000.
- (ii) Regarding the identified commuters, the statistics reflected that they averagely began their trips at 7:30-8:00 and swiped cards for their last-boarding at 17:30-18:00, 87% of which swiped cards twice in a typical workday.
- (iii) The travel pattern reflected by the commuters who swiped cards twice in a typical workday was likely to represent a typical commuting trip chain, where public transit riders took a bus from home to their workplaces in the morning and then returned home in the evening.
- (iv) Through few transfers, bus lines in Xiamen were really well to fulfill the demand of commuters. Conversely, it also reflected that the existing bus overlap lines have not played an effective role in supporting the majority of commuting activities, which may lead to an extreme crowding of bus routes on main public transit corridors.
- (v) The primary OD demand of public transit commuters was obtained. It was found that home locations of public transit commuters were more disperse than their work locations in Xiamen. Most of the public transit commuters lived in several core residential communities. However, workplaces of public transit commuters were more concentrated in downtown and several development zones.

Nevertheless, based on the identified public transit commuters, travel pattern analysis of commuting activities could be conducted in further studies. More and more smartcard data related challenges may need to be addressed when researching on deep analysis of public transit rider's travel behavior or fusion method with other data sources.

Data Availability

The data supporting the results in this paper was authorized by "Transport Commission of Xiamen Municipality". Thus, the database of residents' travel survey could only be accessible with the authorization and permission of the organization.

Conflicts of Interest

The authors declare that there are no conflicts of interest in this paper.

Acknowledgments

This research is supported by National Natural Science Foundation (Project: 51478350) and the Fundamental Research Funds for the Central Universities (3132018305; 3132016301).

References

- [1] J. Y. Park, D.-J. Kim, and Y. Lim, "Use of smart card data to define public transit use in Seoul, South Korea," *Transportation Research Record*, no. 2063, pp. 3-9, 2008.
- [2] Y. Shiftan, Y. Barlach, and D. Shefer, "Measuring passenger loyalty to public transport modes," *Journal of Public Transportation*, vol. 18, no. 1, pp. 1-16, 2015.
- [3] S. Sun, "Public Transit Loyalty Modeling Considering the Effect of Passengers' Emotional Value: A Case Study in Xiamen,

- China,” *Journal of Advanced Transportation*, vol. 2018, pp. 1–12, 2018.
- [4] J. Zhou, E. Murphy, and Y. Long, “Commuting efficiency in the Beijing metropolitan area: An exploration combining smart-card and travel survey data,” *Journal of Transport Geography*, vol. 41, pp. 175–183, 2014.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] L.-M. Kieu, A. Bhaskar, and E. Chung, “A modified density-based scanning algorithm with noise for spatial travel pattern analysis from smart card AFC data,” *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 193–207, 2015.
- [9] L. M. Kieu, A. Bhaskar, and E. Chung, “Passenger segmentation using smart card data,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 3, pp. 1537–1548, 2015.
- [10] 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.
- [11] S. Sun, Z. Duan, Q. Xu, and P. Chen, “School bus routing problem in the stochastic and time-dependent transportation network,” *PLoS ONE*, vol. 13, no. 8, p. e0202618, 2018.
- [12] J. Chen, Y. K. Lv, and M. L. Cui, “Estimating alighting stops of smart card public transportation passengers based on travel patterns,” *Xi’an Univ. of Arch. & Tech I*, vol. 1, pp. 23–29, 2018.
- [13] M. Bagchi and P. R. White, “The potential of public transport smart card data,” *Transport Policy*, vol. 12, no. 5, pp. 464–474, 2005.
- [14] 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.
- [15] M. K. El Mahrsi, E. Côme, L. Oukhellou, and M. Verleysen, “Clustering Smart Card Data for Urban Mobility Analysis,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 3, 2017.
- [16] E. F. T. Legara and C. P. Monterola, “Inferring passenger types from commuter eigentravel matrices,” *Transportmetrica B*, vol. 6, no. 3, pp. 230–250, 2018.
- [17] G. E. Sánchez-Martínez, “Inference of public transportation trip destinations by using fare transaction and vehicle location data: Dynamic programming approach,” *Transportation Research Record*, vol. 2652, pp. 1–7, 2017.
- [18] D. Tribone, D. Block-Schachter, F. Salvucci, J. Attanucci, and N. Wilson, “Automated, data-driven performance regime for operations management, planning, and control,” *Transportation Research Record*, vol. 2415, pp. 72–79, 2014.
- [19] 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.
- [20] J. Chen and D.-Y. Yang, “Identifying boarding stops of bus passengers with smart cards based on intelligent dispatching data,” *Jiaotong Yunshu Xitong Gongcheng Yu Xinxi/Journal of Transportation Systems Engineering and Information Technology*, vol. 13, no. 1, pp. 76–80, 2013.
- [21] Z. Zhao, H. N. Koutsopoulos, and J. Zhao, “Detecting pattern changes in individual travel behavior: A Bayesian approach,” *Transportation Research Part B: Methodological*, vol. 112, pp. 73–88, 2018.
- [22] G. Goulet Langlois, H. N. Koutsopoulos, and J. Zhao, “Inferring patterns in the multi-week activity sequences of public transport users,” *Transportation Research Part C: Emerging Technologies*, vol. 64, pp. 1–16, 2016.
- [23] 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.
- [24] 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.
- [25] W. Wang, J. P. Attanucci, and N. H. M. Wilson, “Bus passenger origin-destination estimation and related analyses using automated data collection systems,” *Journal of Public Transportation*, vol. 14, no. 4, pp. 131–150, 2011.
- [26] A. Carrel, P. S. C. Lau, R. G. Mishalani, R. Sengupta, and J. L. Walker, “Quantifying transit travel experiences from the users’ perspective with high-resolution smartphone and vehicle location data: Methodologies, validation, and example analyses,” *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 224–239, 2015.
- [27] S. Zhang, M. Ding, and Z. F. Shi, “Transport Commission of Xiamen Municipality,” Xiamen Transport Annual Report in 2014, Xiamen, China, 2015.



Hindawi

Submit your manuscripts at
www.hindawi.com

