A Review of Big Data Applications in Urban Transit Systems

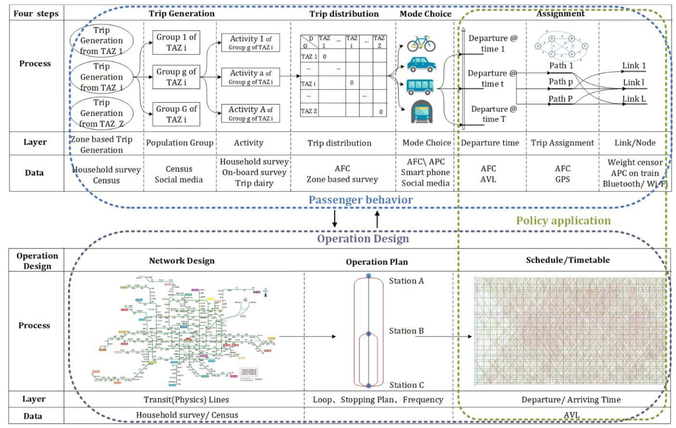
Kai Lu, Jiangtao Liu, Xuesong Zhou, *Member, IEEE*, and Baoming Han

***Abstract*— Operations, management and planning of urban transit systems have evolved substantially since the application of transit data collection technologies, such as, automated fare collection (AFC), Global Position System (GPS), smartphones and face identification. A diversity of detailed sensor data in urban transit systems are being used as fundamental data sources to observe passenger travel behavior, reschedule operation plans and adjust policy decisions from the daily operations to the long-term network planning. This review aims to summarize and analyze those related challenges and data-driven applications. Firstly, we review the data collecting technologies since the late 1990s by classifying the various technologies into two groups: traditional technologies and advanced technologies. A vast body of literature has been developed in this area given the wide range of problems addressed under the transit data label. A summary diagram is proposed to demonstrate the transit data applications and research topics. The data applications are classified into three branches: passenger behavior, operation optimization, and policy application. For each branch, the hot research direction and dimension shown as sub-branches are represented by reviewing the highly cited and the latest literature. As a result, this article discussed the concept and characteristics of transit data and its collection technologies, and further summarized the methodology and potential for each transit data application and suggested a few promising implications for future efforts.**

***Index Terms*— Transit big data application, summary tree diagram, transit passenger behavior analysis, transit operation optimization, transit policy application.**

I. INTRODUCTION

TRANSPORTATION system analysis and optimization have evolved substantially with the employment of ubiquitous types of sensors in the recent three decades. Particularly, there is an increasing research trend on big data applications for the public transit systems with high passenger-carrying capacity and low environmental impacts, which is providing a great opportunity to completely unveil the inner system working mechanism, while creating unclear impacts on passenger travel behavior, system operations and planning, and policy making [1].



TAZ-Traffic Analysis Zone; AFC-Automatic Fare Collection; APC-Automatic Passenger Counters; AVL-Automatic Vehicle Location

Fig. 1. Layered system schema for passenger behavior, operation planning and policy application.

Transit system is composed of demand (passenger behavior) and supply (transit service). Transit service is usually generated based on the requests of passengers, such as, trip origin and destination, departure time, and path choices, which are also cyclically influenced by transit timetable and networks. Theoretically, this cycle can finally reach an equilibrium between demand and supply. However, any unpredicted elements can destroy this stability in reality, especially in peak hours when the travel demand is substantially exceeds the transit system supply. Therefore, it results in a number of challenges in demand management, operation optimization and policy making, which are displayed in the layered system schema shown in Fig. 1.

Specifically, from the perspective of demand management, an eight-layer transportation system analysis in the left cycle in Fig. 1 is first introduced to represent the process of passenger performing their trips, ranging from trip generation, trip distribution, mode choice and final assignment, to determine trip origin, destination, departure time, mode choice and route choice. The required data sources for each layer are also listed in Fig. 1 to better calibrate and validate the real-world modeling outputs. Recently, a novel deep-learning-based framework

Manuscript received June 28, 2019; revised December 7, 2019; accepted February 5, 2020. Date of publication February 19, 2020; date of current version May 3, 2021. This work was supported in part by the Beijing Postdoctoral Research Foundation under Grant ZZ2019-118, in part by the National Natural Science Foundation of China through the Project titled Research on Advanced Theories for Urban Transportation Governance under Project 71734004, and in part by the National Science Foundation (NSF) of USA through the Collaborative Research: Improving Spatial Observability of Dynamic Traffic Systems through Active Mobile Sensor Networks and Crowdsourced Data under Grant CMMI 1538105. The Associate Editor for this article was C. G. Claudel. *(Corresponding author: Jiangtao Liu.)*

Kai Lu is with the School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China, and also with Traffic Control Technology Co., Ltd., Beijing 100070, China (e-mail: lukai_bjtu@163.com).

Jiangtao Liu is with Supply Chain Analytics, Walmart Inc., Bentonville, AR 72712 USA (e-mail: jliu215@asu.edu).

Xuesong Zhou is with the School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ 85287 USA (e-mail: xzhou74@asu.edu).

Baoming Han is with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China (e-mail: bmhan@bjtu.edu.cn).

Digital Object Identifier 10.1109/TITS.2020.2973365

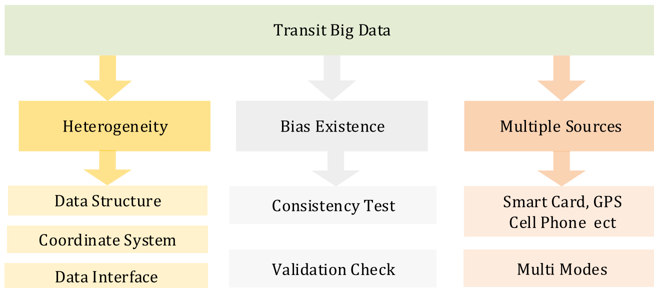


Fig. 2. The scatter emerging topics to urban transit big data.

is proposed to redesign the traditional four-step forecasting model by constructing a multi-layered Hierarchical Flow Network (HFN) mapped with household travel surveys, smart phone type devices, global position systems, and sensors [2]. Actually, in terminal level, the transit data collection is capable of providing more accurate trip information such as, the boarding and/or alighting information for each passenger in smart cards, passenger count observations from video system/vehicle weight sensors in stations/vehicles, for passenger behavior modeling and system operation optimizations.

The right cycle in Fig. 1 provides a three-layer chart for transit system operation design and optimization, ranging from physical network design, operation plan and timetable design. Usually, the transit lines/network design and operation plan are determined by the general transit trip generation and distribution through travel survey and smart card data. The operation plan includes the operation loop, stopping scheme and the loop frequency, which further provides necessary inputs for timetable design to decide the final vehicle departure and arrival time at each stop as the final transit service product, which could be reflected by the real-time Automatic Vehicle Location (AVL) data. In addition, the middle cycle in Fig. 1 combines the travel demand and network supply in transit systems for further policy making and applications, which will be illustrated further in the Section 4.3.

To further utilize the multi-source big data in transit applications, a number of research topics have been focused as shown in Fig. 2, ranging from data heterogeneity, data bias to multiple sources data infusion. Specifically, several key questions need to be carefully addressed.

(i) Heterogeneity:

• How to quickly pre-process those data with different data structures, data loss, coordinate systems and data interfaces?

• How to protect users' privacy without loss of useful information?

(ii) Bias existence:

• How to pre-process those data to ensure observations consistency from different sources?

• How to cross validate those corrected data/ information?

(iii) Multiple sources:

• How to incorporate those multi-source data in a unified modeling framework to model different transit requests at different management levels?

• How to extract the passenger-based characteristics to provide personalized services?

• How to design corresponding efficient and scalable algorithms to fast utilize those data and communicate with the physical transit system?

• How to evaluate the value of different kinds of data under different goals/performances?

• How to consider the impacts from other emerging mode changes (e.g., bike sharing, autonomous vehicles)?

Recently, Liu and Zhou [3] first adopted the concept of information space from control theory to connect the multi-source data with different transit states under a unified modeling framework, and the data consistency is also considered as a data conciliation problem. Currently, there have been a number of great review studies on smart card data, focusing on its pros and cons [4], technology patent [5] and data privacy policy [6]. However, this paper will review the existing research on transit data-driven applications by specifically focusing on data collection technologies, methodologies in passenger behavior analysis, operation optimization and policy applications, and future research opportunities, while always incorporating the characteristics of possible used multi-source data.

In addition, this study mainly focuses on the traditional urban transit systems where vehicles (bus and train) follow the pre-set schedule and timetable. The emerging customized urban transit systems are out of this research scope and will be discussed finally. The remainder of this paper is organized as follows. Section 2 presents the data sources that have been applied in transit data analysis. In Section 3, we summarize the application of transit data in different fields to propose the research branches in our focus. Section 4 reviews the existing specific literature on each branch, correspondingly, which covers total 95 cases in 17 countries around the world. Finally, future research directions are discussed in Section 5.

II. Data Collecting Technologies for Urban Transit Systems

Data collecting technologies are reviewed from the late 1990s, and are classified into two general groups: traditional data collecting technologies and advanced data collecting technologies. The former mainly relates to technologies designed with transit characteristics and focused on collecting user travel information and vehicle operation information, particularly, such as, Automatic Fare Collection (AFC), Automatic Vehicle Location (AVL), Automatic Passenger Counters (APC) and General Transit Feed Specification (GTFS). The latter is not purposely designed for transit systems but can provide useful passenger-level travel information and transit system-level states, such as, smart phone location inquiry, Bluetooth technology, Wi-Fi technology, biometric face recognition, and social media information sharing.

A. Traditional Data Collecting Technologies

1) Automatic Fare Collection (AFC): Smart card technology, as the foundation technology for the AFC implementation, was introduced to transit systems in the late 1990s such as Washington (Smartrip) and Tokyo (Suica). Since then, the smart card system is applied in many cities, such as Santiago [7], New York [8], Beijing [9]–[11], Quebec [12].

Generally, when tapping the card at the station, the passenger's location and time are recorded in the AFC system. In some transit systems, such as Beijing metro and Shanghai metro, passengers need to tap the card when boarding and alighting, which provides accurate OD information and this kind of systems is called the closed system. However, in some cities, passengers only tap the card when entering or exiting the system [13]. The origin or destination information is lost and we call it an open system. The trip chaining approach is usually applied to estimate complete trips in the open system, which will be discussed in Section 4.

2) Automatic Vehicle Location (AVL): In response to growing passenger demands for operation reliability, many transit operators are seeking to improve transit vehicle operations by investing in the AVL [14] technology. These systems collect the location of vehicles usually by broadcasting the sensors' values using an interval of 10-30 seconds depending on the radio capacity. Typically, AVL systems are based on GPS measurements [15].

At the very beginning, the offline AVL systems in which the data couldn't be transmitted to the main server in time can now produce continuous data streams with the development of online technology. Specifically, each vehicle transmits the data with a very short (but certain) periodicity to the main server, namely real-time AVL system [16]. Based on the real-time data, many researchers provide real-time decision model to support the operation control such as travel time and dwell time prediction [16], [17] and real-time rescheduling [18].

3) Automatic Passenger Collection (APC): APC data is an important passenger information supplement for AVL. It relies on estimation techniques based on door loop counts or weight sensors installed in vehicles [15]. The APC system records passenger activities such as boarding and alighting [14]. Based on the APC and AVL data, researchers could estimate the passenger O-D matrix [14], dwell time estimation [19] and passenger assignment [20].

4) General Transit Feed Specification (GTFS): GTFS provides a common format standard for open public transportation schedules and associated geographic information by over 150 cities around the world, especially covering the information on trip, route, stop times and stop location etc. Some recent research [21]–[23] have used GTFS to analyze transit accessibility by estimating the point-to-point travel times at different time periods of day. Fortin *et al.* [24] also import the GTFS data to perform transit network analysis for dynamic network connectivity, service frequency at stops and service speed at routes. In addition, a few papers also use GTFS data for schedule-based transit system modeling and optimization [3], [25], [26]. Recently, a real-time version, named GTFS-realtime (GTFS-rt), has begun to emerge to allow agencies to update real-time trip information, vehicle location, and service alters. To address the prediction errors of real-time GTFS data, Barbeau [27] developed an open-source tool to monitor and validate data and further produce statistics for all validations. Additionally, the Transportation Research Board at the US had awarded a grant to improve the quality of GTFS real-time feeds.

B. Advanced Data Collecting Technologies

1) Smart Phone Data (SPD): The smartphone is an integration of GPS, Wi-Fi, and accelerometers. It provides a new way to track the individual long-term travel data, which could help track the passenger mobility and behaviors in the transit systems [28], [29]. Generally, mobile phone data have high penetration rates, so their anonymized flow data has been mined to perform transit analysis and optimization [30].

In order to record users' travel data, some researchers have developed several cell phone applications. Once the applications are activated, the cell phone will send periodic and anonymized location updates to a central tracking server. Those data could help distinguish whether the passenger is on a vehicle or not. The GPS trajectory data are used as the input for the route matching algorithm that determines whether the user is in a bus or another vehicle [31].

Another advantage of smartphone data is the collection of the passengers' attitudes for real-time information. Watkins *et al.* [32] using OneBusAway application, observed passengers arriving at Seattle-area bus stops to measure their waiting time while asking a series of questions, including how long they believed they had waited for. It is found that for passengers without real-time information, the perceived waiting time is greater than the measured waiting time. However, when passengers have the real-time information, their perceived time is no longer more than the experienced waiting time. Moreover, the real-time information users wait almost 2 minutes less than those using traditional schedule information. Also, mobile real-time information has the ability to improve the experience of transit passengers by providing available transit system information in their pre-trips.

2) Bluetooth Technology: Bluetooth is a wireless technology standard for exchanging data over short distances, which provides a new way of analyzing the passenger behavior at some specific locations of the stations, such as stairs or elevators. Wu *et al.* [33] represented one of the earliest attempts to use the Bluetooth technology as an information collection application in an Android-based smartphone for the Beijing Metro in peak hours. From the total of 41,806 records during 120 days, they performed extensive analysis on a variety of statistics for the number of neighbors, the lifetime of the neighbors, flow speed of the nearby passengers and battery usage rate. It shows that the Bluetooth is a perfect technology to build up a relatively small multi-hop wireless network with the network size of four nodes, and it is applicable and may promote new applications in an underground environment during the peak hour. Meanwhile, taking advantage of the Bluetooth short distance connection, the Bluetooth technology has also been used to estimate the crowdedness in the stairs and elevators [34].

3) Wi-Fi Technology: Wi-Fi is a technology for wireless local area networking with location information. The passenger's OD information could be detected by backward tracking Wi-Fi signals of mobile devices carried by transit passengers. The O-D flows, determined directly from the Wi-Fi data for a specific bus trip, do not correspond well to the ground truth flows, but the results demonstrate the promise of using

TABLE I
THE APPLICATION FOR DIFFERENT DATA COLLECTION TECHNOLOGIES

Data Source	Passenger travel time	Passenger distribution (spatial and temporal)	Trip Chaining	Data validation	Demand Forecast	User Density	Dwell time	Reschedule (Real-time)
AFC	√	√	√	√	√			
AVL							√	√
APC		√		√	√		√	
SPD	√							√
Bluetooth				√		√		
Wi-Fi		√			√	√		
Social Media	√				√			√
GTFS				√			√	√

Wi-Fi signal data for O-D flow determination when the data are aggregated across multiple bus trips for a time-of-day period, especially when being used in conjunction with APC data [35].

4) Biometric Face Recognition: It is vital to detect the passenger distribution in large-scale transportation systems. One of the main difficulties in tracking passengers is the lack of detailed individual information in the transit system. Biometric facial recognition technologies make it possible to obtain the individual position of each transit user [36]. Biometric facial recognition technologies contain automated methods for verifying or recognizing the identity of a person on the basis of a facial image. The basic structure of an automated facial recognition system consists of four fundamental blocks: face detector, feature extractor, database and classifier [37]. Mikłasz *et al.* [38] first attempted to use the facial recognition to obtain pedestrian distribution in the interchange station. Passenger transfer matrices obtained from optical analysis and traditional survey proves the effectiveness and potential of image analysis methods. In addition to obtaining pedestrian distribution, this technology has been used in the metro and railway security check-in [39].

5) Social Media Data: Twitter, Facebook, and WeChat are widely-used social media applications. They collect social interactions of a large number of people, thereby, they are a valuable resource for predicting various large-scale trends [40]. Social media data have been used to improve models for predicting flu trends and detect seismic activity after earthquakes [41], [42]. Similar research in the field of transportation monitoring has been carried out as well. Sentiment analysis were used to reveal public opinions regarding transit agencies [43].

Table I summarizes the aspects of the application of different data collection technologies, from different research directions, such as the passenger travel time estimation, passenger trip distribution, demand forecast and timetable rescheduling. It ranges from passenger behavior to system optimal operations. As observed in Table I, one specific research problem can be analyzed or optimized by utilizing multi-source sensor data. Therefore, how those different data can be incorporated and evaluated is important for model validation, system performance optimization and long-term system planning.

III. THE SYSTEMATIC IMPLEMENTATION OF TRANSIT BIG DATA

In order to derive valuable insights for decision support and automation from huge volumes of heterogenous sensor data, a few of big data technologies have been focused and invested by different enterprises and organizations. Those technologies mainly belong to three categories, including data storage (such as, Hadoop, Data Lakes, NoSQL Databases, etc.), data processing (such as, Spark, Hadoop, data governance, etc.), and data analytics (Spark, cloud computing, edge computing, artificial intelligence, modeling and optimization, etc.). Specifically, the Hadoop ecosystem has been widely recognized due to its reliable, efficient and scalable distributed processing of large data sets [44]. In intelligent transportation systems, different frameworks for big data analytics on transportation management and operations are introduced by using Hadoop with MapReduce or Spark [45]–[48]. Wang *et al.* [44] shows that the performance of processing mass GPS data via Hadoop platform is improved by 4000% compared with the serial program. Focusing on the urban rail transit system, historical passenger travel data from AFC are processed to derive the passenger arrival rate, passenger alighting proportion, and travel patterns in Hadoop big data platform [49], [50]. In addition, Hadoop platform is also applied in the Bus Rapid Transit (BRT) system to analyze passenger travel pattern by improving the K-means++ clustering performance on large datasets scalability [51] and identify fraud using transaction profiling from 165 million records [52]. However, our paper will mainly focus on different methodologies used in big data analytics on statistics, optimization, simulation and machine learning, etc., which will be explained in detail in the following sections.

Specifically, this paper will focus on the review on the big data applications and analytics in transit systems classified as three aspects, including passenger behavior analysis, operation planning and policy making, whose connection is clearly illustrated in Fig.3 by a tree-based graph.

The words in blue represent the three applications of transit data. The words in green represent the research branches and subtopic of each data application. The words in purple represent the objectives of each subtopic. The words in black represents the constraints and methodologies.

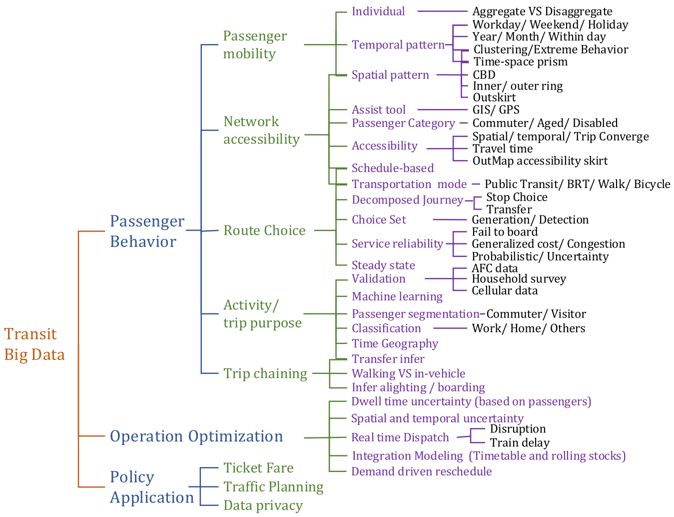


Fig. 3. The tree diagram for different data applications.

• Passenger behavior: Trip purpose, trip start time, transportation mode choice, trip frequency, activity duration, and route choice, are all aspects that represent the passenger behavior differently. Emerging AFC and APC examine more types or aspects of passenger trips over more time horizons and in larger sample sizes [4]. According to the research related to the passenger behavior with transit data, this section classified the research into 5 categories: the temporal and spatial trip distribution, network accessibility and mobility, activity and trip purpose, trip chaining, travel time reliability and route choice. Due to the data diversity, there are several hot branch points for each topic such as aggregated and disaggregated model. Meanwhile, some topics share the same research branches, such as schedule-based analysis and spatial-temporal trip distribution analysis.

• Operation optimization: Schedule is another main subject in the transit network operations and provides its services for the public, and each transit vehicle follows the sole timetable (schedule) during the operation time. Depending on the temporal travel demand pattern, the schedules usually are classified into peak and off-peak strategies for a regular day. Transit systems with physical constraints (vehicle tight capacity and given travel demand) are the main research focus in the previous scheduling models. When travel demand is much greater than the network capacity, passengers may fail to board the vehicle and have to wait for the next available one. In those cases, transit operators should provide the dynamic demand-sensitive timetables to meet greater demands with capacity-limited vehicle services [53]–[55]. AFC and AVL system data flows that provide in-time operation data in the network and operator could be used to adjust the operation plan quickly.

• Policy applications: Policy applications belong to a part of the strategic transit planning which contains numerous topics, such as, land use, ticket fare, and transit data privacy [56]. It is conventional that transit big data are providing mass information about passenger travel patterns and habits, which are the fundamental and basic inputs for long-term planning and land use [57]. Although the potential of public transit smart card systems for commercial applications has not yet received much attention, some experiments are currently underway around the world [56]. Some attempts to adjust the fare scheme have also been tested, such as fares with individual characteristics and possible reservation and tradable credits [10].

TABLE II

THE SCALE OF DATA SETS AT DIFFERENT URBAN RAIL TRANSIT SYSTEMS

City Name	Year	Operations Time	Network Scale /km	No. of Lines	Ridership
Beijing	2018	5:00am-1:00 am	637	22	12.41 million per day
Shanghai	2018	4:55am-0:50 am	705	16	10.15 million per day
Singapore	2017	5:16am-11:50pm	199.3	5	3.12 million per day
Tokyo	2014	05:00am-0:30am	195.1	9	9.69 million per day
London	2017	05:00am-01:00am	402	11	1.357 billion per year
New York	2017	0:00am-0:00 am	1370	36	1.727 billion per year
Chicago	2018	0:00am-0:00 am	165.4	8	226.08 million per year
Seoul	2017	5:30am-12:00pm	331.5	23	3.07 billion per year
Amsterdam	2018	6am-12:30am	42.7	5	24.74 thousand per day

IV. Data-Driven Applications in Transit Systems

In order to cover the urban transit systems from all over the world, this section explored total 94 case studies, covering 17 countries from the perspective of passenger behavior analysis, operations optimization and policy applications, respectively. A few representative data sets used in the following literatures are selected in Table II to generally show how big of data need to addressed in the urban rail transit systems.

A. Passenger Behavior Analysis

Passenger behavior analysis is the foundation of operation plan, and passenger spatial and temporal trip distribution determines the schedule planning and the network design. Meanwhile, unlike car users, transit passengers' accessibility is highly dependent on the accessibility of transit service network. Furthermore, passenger behavior is also influenced by factors such as activity and trip purpose, transfer, travel time reliability, etc. Therefore, in this research, passenger behavior is divided into three levels: mobility, accessibility and miscellaneous. There are three further aspects in the miscellaneous level: activity and trip purpose, trip chaining and travel time reliability.

1) Passenger Mobility: Temporal and Spatial Trip Distribution: The design of transit operation plans and physical networks mainly depends on the passenger spatial and temporal trip distribution. Temporal and spatial features

TABLE III
REVIEW OF STUDIES ON TEMPORAL AND SPATIAL DISTRIBUTION USING TRANSIT BIG DATA

Authors	Geographic Location	Data Source	Aspects	Methodology	Potential
Munizaga and Palma (2012) [7]	Santiago, Chile	AFC GPS	OD matrix; Multimodal transit	Reconstruct trip chains with generalized time	Recover the trip distribution structure - Help to generate the load profiles of bus routes.
Nishiuchi, King and Todoroki (2013) [59]	Japan	AFC	Travel behavior Trip frequency	Statistical tools (Analysis of variance)	Improve individual operation schemes.
Briand *et al*. (2016) [61]	Rennes France	AFC	Weekly Temporal analysis, Travel variability;	Two-level (temporal clustering and spatial clustering) generative mixture model	- Travel variability at individual level; - Passenger demand classification; - Capable of detecting subtle trends such as the presence of nightlife activity
Ma *et al*. (2013) [47], Kieu *et al*. (2015) [63]	Beijing, China Queensland, Australia	AFC	Travel pattern top level; Daily changes Clustering	Density-based Spatial Clustering of Applications with Noise (DBSCAN) algorithm	- Estimate an individual transit rider's OD using that rider's historical travel pattern - Reduce the analysis sample size and reveals temporal travel pattern.
Zhong *et al*. (2015) [64]	Singapore	AFC	variability of temporal and spatial patterns; Clustering	Statistical analysis, correlation matrix and network-based clustering methods	- Establish a generic framework with well-organized analytical methods.
Briand *et al*. (2017) [65]	Gatineau, Canada.	AFC	Regular/ diffuse activities; Temporal clustering	Two level (card membership clustering and Gaussian mixture model for temporal clustering) generative model	- Highlight the close link between the type of statistical modelling and temporal habits. - Cross-city data application.
Zhong *et al*. (2016) [66]	London Singapore, Beijing	AFC	mobility patterns	variability statistics	- Examining regularities and variations in trip generation by station by different periods of time across cities
Ma *et al*. (2017) [67]	Beijing, China	AFC	Spatial-temporal Commuters pattern	Density-Based Scanning Algorithm with Noise, iterative self-organizing data analysis technique	- Reduce car dependency, shorten excess commute, and alleviate traffic congestion.
Tao *et al*. (2014) [26]	Brisbane, Australia	AFC GTFS	Spatial-temporal dynamics pattern	Classification and Flow-comap visualization	- Provide capacity information for public. - Improve service and infrastructure.
Long *et al*. (2015) [70]	Beijing, China;	AFC Survey	Extreme passenger Aggregated analysis	Statistical analysis (such as, kernel density estimation)	- Improve public transit services to extreme transit rider
Cui *et al*. (2015) [71]	Beijing, China	AFC	Extreme passenger Behavior regularities	Extreme Index (EI)-based mixture Gaussian model	Improve the EI model with respect to the variation of trips' temporal information and entropy of the occurrence of travel activities.
Sun *et al*. (2013) [72]	Singapore	AFC	Reproducible patterns, Individual regularities	Diffusion/spreading processes.	Collective human behaviors, dynamical evolution of social networks
Luo, Cats and Lint (2017) [73]	Haaglanden, Netherlands	AFC	Aggregation of the demand of spatially close stations	k-mean-based station aggregation based on flow and spatial distance	Short-term transit demand prediction and demand assignment.
Zhou *et al*. (2017) [74]	Queensland, Australia	AFC	Monitor transit-served area	Metrics or measurement analysis	- Improve the land use to shorten average travel distance of transit riders Services adjustments and optimization.
Sun *et al*. (2014) [75]	Singapore	AFC AVL	One-week activity Boarding/alighting	Regression analysis	Identify optimal vehicle type for a particular route and modelling transit service reliability.
Ma *et al*. (2012) [76]	Beijing, China	AFC AVL	Trip origin estimation	Markov chain - Bayesian decision tree algorithm	Transit origin data are highly valuable for transit system planning and route optimization.
Ali, Kim and Lee (2016) [77]	Seoul, South Korea	AFC	Travel pattern Transfer detection	Transit simulation tool MATSim	Transit vehicles are optimization in managing the frequency and headway of transit vehicles.
Chakirov and Erath (2011) [78]	Singapore	AFC	Travel pattern (mode choice, waiting time, trip purpose and location);	Statistical analysis	Infer travel purposes and locations of regular activities to improve public transit service.

are the fundamental elements of travel pattern (see Table III) [58]–[60]. Clustering is the basic method for mobility analysis, including two-level models with temporal trip clustering and spatial trip clustering [61], density-based spatial clustering to overcome the huge amount of data [62], [63], network-based clustering methods [64], two-level model with membership clustering and Gaussian mixture model for temporal trip clustering [65].

Daily habitual travel patterns possess regularity as a basis, but more considerations on long-term trends are required for transportation planners., the variability and correlation of travel patterns and time changes for transit network planning have gained a great attention, such as, Analysis of variance (ANOVA) in trip patterns for one month [59], variability statistics for regularity analysis of demand pattern [66], statistical analysis and correlation matrix [64]. Other approaches for

TABLE IV
REVIEW OF STUDIES ON NETWORK ACCESSIBILITY USING TRANSIT BIG DATA

Authors	Geographic Location	Data Source	Aspects	Methodology	Potential
Owen and Levinson (2015) [79]	Twin cities, United States	publicly-a vailable data	Share of transit/ Auto, Job accessibility	A binomial logic model	Implement transit accessibility at low cost. Improve the performance of ridership and mode share.
Widener *et al*. (2015) [80]	Cincinnati, United States	Transit data	Food Deserts	Statistical analysis	Link interaction potential scores with individual-level behavioral.
Delmelle and Casas (2012) [81]	Cali, Colombia	Socio-eco nomic data	Bus Rapid Transit; activities access	Statistical analysis	Walking access to the BRT system is greatest for middle income groups and most limited for groups in the highest and lowest income.
Hadas (2013) [82]	Auckland, United States	Google Transit data	GIS-based Connectivity, Network coverage	Connectivity and accessibility indicators	Demonstrate dynamic properties of the network, specifically the time of day and day of the week.
Lin *et al*. (2014) [83]	Perth, Australia	Survey	Aging population; Stop level	A composite index model	Improvements of the walking service quality, walking route directness and land-use diversity
Badland *et al*. (2014) [84]	Perth, Australia	Self report	Commuter; Trip purpose	Statistical analysis	Improve the transit access for commuters.
Pendyala, Yamamoto and Kitamura (2002) [86]	San Francisco, Miami, United States	Activity data sets	time-space prisms; activity-travel pattern	Stochastic frontier modeling	Temporal vertices of time-space prisms.
Chen, Claramunt and Ray (2014) [87]	Guangzhou, China.	AFC	Connectivity	Accessibility indices	Evaluate the nodes accessibility and characterizes the spatial-temporal evolution of the metro network.
Fransen *et al*. (2015) [22]	Flanders, Belgium.	Demograp hics	Population segments; Schedule-based	Time-continuous schedule-based public transport accessibility model	Improve the time variability of public transport frequencies.
Kitamura *et al*. (2001) [88]	KOK, Japan; California, United States	Smart Phone AFC	Long and short term passenger behavior Time availability	Accessibility indices	Time availability is more closely associated with engagement in activities than accessibility
Mavoa *et al*. (2012) [89]	Auckland New Zealand	AFC	public transit access	Walking accessibility index; transit frequency measure	A complete and realistic picture of transit access and explore the potential for mode substitution and accessibility for non-car user.
Ford *et al*. (2015) [90]	London, UK	Network	Accessibility pattern; Transportation mode	GIS-based tool	Increase the use of low-carbon forms of transport.

demand patter analysis include iterative self-organizing data analysis [67], flow-comap-based visualization [26].

While commuters make the majority of daily journeys [67], extreme travelers are increasingly considered by both the academia and mass media recently due to the increased numbers of the unemployed, self-employed and part-timers, the rise of telecommuters and low-paying jobs relocated to cheaper places inside or outside a region/country [68]. At first, extreme travelers are defined as passengers who take excessively long trips [69]. There are three more types of extreme travelers in the context of the Chinese society, including the public transit passengers who (1) make significantly more trips ('recurring itinerants'), (2) travel significantly earlier than average passengers (the 'early birds') during weekdays, and (3) ride in unusually late hours (the 'night owls') during weekdays [68], and the applied methods have Extreme Index (EI)-based mixture Gaussian model [71] and kernel density estimation for four extreme transit behaviors [70].

Expect for the demand patterns, Sun *et al.* [72] applied a simulation tool (MATSim) for transfer behavior detection, Luo et al. aggregated the demand of spatially close stations for transit demand construction by k-mean-based station aggregation based on passenger flow and spatial station distance [73], and four corresponding metrics (the minimum, actual, random and maximum travels) are calculated to reflect transit riders' elasticity of distance travelled (EDT) relative to the cost of travel to monitor the "transit-served areas (TSAs) [74], Sun *et al.* applied regression analysis on the dynamics of boarding/alighting activities and its impact on bus dwell times [75], and Ma *et al.* developed Markov chain based Bayesian decision tree algorithm to estimate passengers' origin in Beijing's flat-rate bus system [76].

2) Network Accessibility: Passenger accessibility highly relies on the accessibility of transit network and its schedule. In the current literature, accessibility can be calculated for different categories of opportunities (See Table IV). The range of the classic models indicates its complexity and significance. At the disaggregate level, every user can be viewed as the representative of a unique class, which is viewed as agent-based models. In addition, a classification based on the trip purpose, vehicle type, and transportation mode would provide reasonable realism to minimize the classes/flows in the models. A number of research focus on work accessibility with gravity models [79], food accessibility for residents who rely on public transit [80] and activities accessibility for Bus Rapid Transit (BRT) [81]. Initially, researchers only consider the spatial facility accessibilities [82], such as walking service quality [83], [84]. The mobility is measured by spatial accessibility [82], [85], time-space prism [86], and connectivity [87]. However, for the schedule-based transit network, especially

TABLE V
REVIEW OF STUDIES ON TRIP PURPOSE AND PASSENGER ACTIVITY ANALYZING TRANSIT BIG DATA

Authors	Geographic Location	Data Source	Aspects	Methodology	Potential
Devillaine, Munizaga, and Trépanier. (2012) [91]	Santiago, Chile, Quebec, Canada,	AFC	Location, Time duration Land use User behavior	Discrete choice model	Discover differences in behavioral activity patterns due to sociological, cultural, and geopolitical differences.
Chakirov and Erath (2012) [92]	Singapore	AFC	Location	Discrete choice modelling with activity duration	Improve the accessibility of the stations which are close to the workplaces.
Langlois, Koutsopoulos and Zhao (2016) [93]	London	AFC	passenger heterogeneity	Clustering with longitudinal representation	Provide special service for different group of users with the significant connections exist between the demographic attributes of users and activity patterns
Jiang *et al*. (2017) [94]	Singapore	Smart Phone	Individual	Data mining framework	Understand the human activity patterns in cities, and provide targeted plans for future sustainable development.
Xue *et al*. (2014) [95]	Singapore	transit data	Travel pattern, Tourists	Machine learning	Gain insight into tourists' travelling behaviors and make the tour policy for tourists.
Lee and Hickman (2014) [96]	Twin Cities United States	AFC	Travel pattern; Travel symmetry	Stop aggregating Transfer detection model	Infer the purpose or activity
Bohte and Maat (2009) [97]	Netherland	Travel Survey	Travel mode	GPS logs, GIS technology and web-based validation	Help collect trips that are missed by paper diary methods.
Wang *et al*. (2017) [98]	Shanghai, China	AFC	After-work activity; Travel impendence	Proxy indicators	Commuter detection and trip chain method

for the low-frequency system, it is necessary to evaluate the accessibility based on the timetable [22], [88].

The analysis methodologies are mostly depended on the statistical analysis [80], [81], [84], stochastic frontier modeling [86], accessibility indicators [22], [79], [82], [83], [87]–[89] and GIS-based tools for visualization [90].

3) Activity and Trip Purpose: Trip purpose is one of main aspects when passengers determine their behaviors in the transit network. For work trips, passengers prefer to choose the shortest path to save time, but passengers with shopping trips usually prefer to have better comfortability and may choose the path that is less crowded. It is possible to obtain the purpose-based behavior characteristics when aggregated passenger behavior includes different trip purposes, which are then used for the customized service. Trip purpose is one of the main topics for analyzing passenger behavior (see Table V). Passengers, especially the commuters, may change their choices at different times of the day. In the morning peak hour, passengers pay more attention to the service reliability and prefer the metro. In the off-peak hour, bus or taxi may be preferred by passengers to finish their trips given the amount of travel time. The discrete choice model [91], [92], clustering method [93] and data mining and machine learning [94], [95] are the most popular algorithm for analyzing trip activity and trip purpose.

Generally, the trip purpose can be classified into 3 categories: work, home, and others [91]. Activity duration time, land use of station location and station frequency are the basic parameters to determine the trip purpose [91]–[93]. In addition, the daily trip symmetry characters are also applied to determine the trip purpose [96]. The work and home trip purpose usually appear in the first and last trip within one day, and the passenger may have other activities after work. To obtain more detailed personal information, household survey and GPS logs are also taken as supplements for AFC data [97]. Some new data analysis methods such as machine learning have also been used to analyze those emerging data [95].

4) Trip Chaining: Trip chaining is an important research topic for researchers to obtain the destination location and transfer stations for each passenger. These results could help designers optimize operation schedule for successive legs to save transfer time and optimize the stop locations. In open AFC systems, only the boarding information like boarding time and location is available, so the alighting information is not available, such as in the London bus system and Metro Transit in Minneapolis-St Paul [99]. Based on other support data resources such as the automated vehicle location (AVL) and automated data collection (ADC), more detailed information about boarding and alighting can be inferred. Researchers have worked on the individual trip destination estimation for tap-in only transit system (see Table VI) in the past decades. There are two basic assumptions for the trip-chaining algorithms.

- A high percentage of passengers returns to the destination station of their previous trip to begin their next trip.
- A high percentage of passengers ends their last trip of the day at the station where they began their first trip of the day.

Some other parameters, such as walking distance and the time interval between two trip legs, were studied when matching separate journeys for an individual cardholder, especially for the transfer determination [100]. When applying the trip-chaining methodology to infer the alighting station for the current trip, each cardholder should have more than one trip in the transit system and there is no private transportation mode trip segment such as car or bicycle between consecutive transit trip segments in a daily trip sequence [13]. In some research, the single trip destination is inferred based on other days' records, and if it only appears once, the alighting station is invalid. The sensitivity analysis is applied with

TABLE VI
REVIEW OF STUDIES ON TRIP CHAINING USING TRANSIT BIG DATA

Authors	Geographic Location	Data sources	Methodology	Potential
Munizaga *et al.* (2014) [7]	Santiago, United States	AFC GPS	Position-time alighting estimate - Bus- minimize the generalized time with the next boarding position-time Metro- walking distance threshold with next boarding station	- Simple rules used to identify activities can be replaced by more sophisticated rules that include additional information such as the frequency and land use. - The destination estimation for once trip is missing.
Barry *et al.* (2002) [8]	New York, United States	AFC	Trip-chaining methodology	97% of the cardholder travel by bus or metro for more than once.
Trépanier *et al* (2007) [12]	Gatineau, Australia	AFC	Transportation Object-oriented modeling Vanishing route set One trip passenger analysis	- Some passenger information such as single ticket user is missing. - Combining other data sources such as APC and onboard survey data to improve the estimation.
Zhao *et al.* (2007) [13]	Chicago, United States	AFC	Database management systems(DBMS) Geographic information system (GIS)	The model was just focused on the bus and rail station
Chu and Chapleau (2008) [60]	Canada	AFC	The linear interpolation and extrapolation are used to infer the vehicle position when the GPS data is not available.	Multiday analyses that study the dynamics of the demand, can also serve to improve the results of trip purpose and destination inference.
Nassir *et al.* (2011) [99]	Twin Cities, United States	AFC	OD estimation algorithm	The transfer time threshold is fixed, and it should be possible to adjust the static boundary in accordance with network configuration and transit passenger behavior.
Alsger *et al*. (2015) [100]	Queensland, Australia	AFC	OD estimation algorithm with 15 min transfer time interval	This work should be extended to compare other estimation methods that use data from smart card systems in other cities.
Wang, Attanucci and Wilson (2011) [101]	London, UK	AFC	Trip-chaining methodology mostly based on next trip is bus or rail	By linking system usage to home addresses, access behavior could be better understood

onboard survey data to validate the feasibility of the method. However, the onboard survey is expensive and data samples are usually limited. The used methodologies mainly depend on the logics of trip chain based on available AFC, AVL, ADC and onboard survey data, OD estimation approach, sampling analysis, sensitivity analysis, etc.

5) Travel Time Reliability and Route Choice: Passengers are normally pretty sensitive to the waiting time in their trips, which could be represented by the reliability of travel time. They may change their routes based on the path travel time reliability, especially when some accidents occur. These reliability-based route choice characteristics and the real-time information-based behavior are important references for accident rescheduling. Passenger route choice and travel time reliability is an ongoing research endeavour (see Table VII). Travel time is the foundation for route choice, which could be estimated by link travel time or trip time from AFC data. Generally, the link travel time and trip time share a closed-from time distribution based on the additive property [102]. Reliability indicator calculation [103], Guassian Mixture [104], [105] Markov Chain [76], and Bayesian inference [106] are the most popular method to estimate the travel time distribution and trip time. The key for route choice model is to find the probability for different paths. Logit model is the basic one [107]–[109]. Considering the passenger preference and travel strategy, the Nested and Mixed logit model is also applied in research [110]. Kim *et al.* performed regression analysis on route stickiness [111] and Nassir *et al.* developed a statistical inference to deduce the set of attractive routes for public transit passengers [112]. More detailed information is discussed as below. Generally, researchers assume that passengers made their decisions by a maximum utility function, which is usually the generalized path cost [76]. The generalized cost contains the total travel time and the penalty for transfers and crowdedness. As such, the crowdedness plays a crucial role in passenger route choice behavior [25], [107]. In busy transit systems such as at Beijing and Shanghai, some passengers fail to board on the incoming train and need to wait for the next one in the morning peak hour. Sun and Schonfeld [108] proposed a method to estimate the ‘failing to board’ phenomena. Actually, Liu and Zhou [25] shows how this kind of failing to board under tight capacity constraints can invoke the bounded rationality behavior. In addition, a number of generalized user equilibriums are also studied by considering different route choice assumptions, such as, the experienced least-cost path selection [113], optimal strategy for expected least-cost path set selection [114], path selection with perceived random error [115], and the non- cooperative behavior with a number of exogenous priority loading rules [116].

Modelling travel time is fundamental for the route choice model. Before performing passenger assignment, a path set is usually generated for each passenger. The number of passenger paths and link uncertainty matter for the efficiency of the assignment algorithm. Passengers generally pay more attention to the travel time reliability (service reliability) rather than the total travel time [102], [117], especially for public transit and other transit modes [110], [118]. The algorithms will become much more complex when mapping the timetable to the physical map. An onboard survey showed that some of the passengers always selected the same route (high stickiness) compared with those with more varied patterns of route selection (low stickiness). The number of the feasible paths will decrease with the stickiness index [111]. Another major

TABLE VII
REVIEW OF STUDIES ON ROUTE CHOICE AND TRAVEL TIME RELIABILITY USING TRANSIT BIG DATA

Authors	Geographic Location	Data sources	Aspects	Methodology	Potential
Ma *et al*. (2017) [76]	Netherlands Railways	AVL	link level travel time	A generalized Markov chain; Gaussian based heuristic clustering	Calculate the correlations within multimodal distributions and multimodal sharing.
Sun and Xu (2012) [102]	Beijing, China	AFC	link uncertainty; platform elapsed time; elapsed time-transfer	Travel time distribution; Mixture density of *n* component distributions	Facilitate analysis of transit service reliability and passenger flow assignment in daily operations.
Oort (2014) [103]	30 cities	Survey	Service reliability; international survey	International and Dutch survey on travel time for public transit, average additional travel time for service reliability indicators	Improve the design of networks and timetables and improve service reliability.
Yin *et al*. (2015) [104]	Sioux Falls	AFC	Additive property	Gaussian mixture model and Expectation Maximization algorithm	Provide faster solutions for link travel time estimation.
Zhao *et al*. (2017) [105]	Shenzhen, China	AFC	Route and train choice	Probabilistic model with maximum-likelihood estimation.	The measured results provide us with useful input when building the passenger path choice model.
Zhang and Yao (2015) [106]	Guangzhou, China	AFC	link time distribution	Bayesian inference; Markov Chain and Monte Carlo	The superiority of the proposed model over the existing model in estimation and forecasting accuracy is also demonstrated.
Kim *et al*. (2015) [107]	Seoul, South Korea	AFC	delay of crowding	Logit discrete model	Service adjustment.
Sun and Schonfeld (2016) [108]	Beijing, China	AFC	Train schedule connection; Schedule-based choice, Fail to board (FtB)	The logit model with the FtB estimating method	Locate network capacity bottlenecks at the train level and evaluate the effects of adjusting train timetables.
Xu *et al* (2018) [109]	Beijing, China	AFC	Link travel time In vehicle crowdedness	Bayesian inference and Metropolis-Hasting sampling for logit model	Calibrate parameters of the logit model.
Hassan *et al*. (2016) [110]	Queensland Australia	Travel survey AFC	Transportation mode; Access to stop Stop choice strategy	Nested and Mixed nested Logit model	The choice of access stop is affected by transit path between the journey ends and the attributes of the departure stop.
Kim *et al.* (2017) [111]	Brisbane, Australia	AFC	Individual and aggregated Habitual behavior	Regression analysis of route stickiness	Re-design and schedule public transit systems through unveiling the complexities of transit behavior.
Nassir, Hickman and Ma (2015, 2017) [112, 118]	Brisbane Australia	AFC	Deduce the set of attractive routes; Boarding strategy	Statistical inference and multinomial probability	Deduce the trip purpose for those repeated boarding trips. And provide better service for those stations.
Ma,Ferreira and Mesbah (2014) [117]	Brisbane Australia	AVL	Reliability Buffer time for passenger	A Gaussian mixture models	Improve the trip planning with travel reliability.
Hurk *et al*. (2015) [119]	Netherlands Railways	AFC	Disruption	Route generation and selection with the conductor check data	Analyze and evaluate passenger service. Improve the transit service.
Ceapa, Smith and Capra (2012) [120]	London UK	AFC	Station crowding pattern	Sensitivity station crowdedness	Predict the crowding in the network and incorporated within ATIS to provide crowd information for passengers.
Jang (2010) [121]	Seoul, South Korea	AFC	Transfer pattern	Travel time map	Improve the felicities for the transfer station.
Zhu, Wang and Huang (2017) [122]	Shanghai, China	AFC AVL	individual choice	Manski's paradigm	Deduce other important operational indicators like route choices, passenger flows and load factor.

application of the route choice is in the cases of disruption occurrence. When accidents or disruptions happen in the system, the operation agencies need to adjust the train schedule and provide a quick response service for passengers to dispatch the stranded passengers [119].

B. Operation Optimization

Operation optimization is one of the most important topics in transit data application (see Table VIII). Operation optimization contains operation plan, schedule and the fare scheme. In this part, we mainly focus on the plan and schedule. The ticket scheme and fare optimization will be discussed in next policy part in detail.

Due to the model complexity and some concave constrains, Genetic Algorithm (GA) [123]–[125] is the most popular algorithm for operation optimization. To accelerate the solution searching time, some updated algorithms are further applied, such as, Branch-and-bound GA [126], Simulation-based GA [127] and non-dominated sorting genetic (NSGA-II) based algorithm [128]. In addition, other heuristic algorithms, such as, hybrid artificial bee colony algorithm [129], Tabu search [130], are also considered.

TABLE VIII
REVIEW OF STUDIES ON TRANSIT BIG DATA FOR OPERATION OPTIMIZATION

Authors	Geographic Location	Aspects	Aim/ Objects	Methodology/ Algorithm
Sun *et al*. (2014) [53]	Singapore	Dynamic timetable Capacity constraints	Minimize the passenger total waiting time	Standard solvers (e.g., CPLEX)
Niu, Zhou and Gao (2015) [54]	Shanghai-Hangzhou	Skip-stop; Time-varying demand	Minimize the total passenger waiting time	General purpose high-level optimization solvers
Yang *et al*. (2017) [123]	Beijing, China	Cyclic operation, Dwell time uncertainty	Minimize net energy consumption and total travel time with provision for dwell time uncertainty	A genetic algorithm framework to determine the Pareto optimal solutions.
Shi *et al*. (2016) [124]	Beijing, China	Loop line	Minimize the average waiting time of access passengers and transfer passengers	Genetic Algorithm
Zhen and Jing (2017) [125]	Beijing, China	Reschedule with train delay	Minimize the negative effect of train delay to passengers	Genetic Algorithm
Albrecht (2009) [126]	Dresden German	Elastic demand; Optimal train frequency and departure time	Minimize operation cost	Branch-and-bound and Genetic Algorithms
Yang *et al*. (2016) [127]	Beijing, China	Integrated timetable and speed profile	Minimize the total tractive energy consumption	Simulation-based genetic algorithm
Wu *et al*. (2016) [128]	Shenyang, China.	Re-synchronizing timetable	Maximize the number of smooth transfers; minimize the maximal deviation from the departure times of the existing timetable	Non-dominated sorting genetic (NSGA-II) based algorithm
Chen *et al*. (2015) [129]	Suzhou, China	Skip-stop; Vehicle capacity and stochastic travel time	Minimize passenger travel time Minimize operation cost	A hybrid artificial bee colony (ABC) and Monte Carlo method
Cao *et al*. (2016) [130]	Beijing, China	Stop-skipping; Elasticity of train operations under demand variations	Minimize total passenger total travel time	A tabu algorithm,
Yin *et al*. (2017) [131]	Beijing, China	Optimize timetable	Minimize the operational costs and passenger waiting time	Lagrangian relaxation (LR)-based heuristic algorithm
Tong *et al*. (2017) [132]	Beijing, China	Internet-based customized bus services	Minimize the number of unserved passengers as well as the routing costs	Lagrangian decomposition and space -time prism
Gao, Yang and Gao (2017) [134]	Beijing, China	Real-time automatic rescheduling strategy,	Minimize the deviation between the rescheduled timetable and the planned timetable	an iterative algorithm with discrete events/ ordered time simulation
Jiang *et al*. (2016) [135]	Shanghai, China	Evaluate train timetable	Evaluate train timetable	Time-driven passenger microscopic simulation
Gao *et al*. (2016) [146]	Beijing, China	Disruption recover	Minimize travel time and passenger flow	2 step iterative algorithm for existing service and future service
D'Ariano *et al*. (2008) [147]	Dutch	Real-time traffic management system	Minimize train delay	Branch-and-bound algorithm; a local search algorithm
Sels *et al*. (2016) [149]	Belgium	Dynamic and uncertain demand; Robustness against uncertainty; Cyclic railway timetabling.	Minimize total passenger travel time	OnTime of TrafIT GmbH and VIA Consulting & Development GmbH
Hassannayebi *et al*. (2016) [150]	Tehran, Iran	Robustness against uncertainty	Minimize the expected value of the passenger waiting times	AUGMECON2 by GAMS to solve robust stochastic model

Besides those heuristic approaches, Lagrangian relaxation is also widely used in timetable optimization [131]–[133]. Different simulation methods, such as discrete events/ ordered time simulation [134] and Time-driven passenger microscopic simulation [135], are also adopted to evaluate the timetable. More detailed information is discussed as below.

During daily operations, transit vehicle follows the timetable which has been set before the operation. Despite the high fixed headway and dwell time in the peak hour, the service may still not meet the high passenger demand. Ignoring such demand dynamics may result in minor disruptions and poor service reliability. Therefore, it is necessary to understand the spatial-temporal travel pattern of the passengers and to design demand sensitive timetables to meet the demand uncertainty [53]–[55], [126]. The objectives of the research on demand-driven timetables are to minimize the total passenger waiting time or operation costs. Some researchers applied a more flexible stopping plan with the elastic demands, such as skip-stop schedule [129], [130].

Rescheduling and dispatching addressing external disruptions (such as, traffic congestion, road accidents) and internal system delays (such as, dynamic vehicle running time, stochastic passenger demand) in the transit system are much more difficult than determining the daily dynamic schedule. Based on the real-time AVL and APC data, a number of studies have focused on the real-time vehicle bunching and control to reduce headway deviation and passenger waiting time. Adaptive control schemes are proposed by dynamically determines bus holding times at a route's control points [136] and adjusting a bus cruising speed [137] based on real-time headway information. In addition, two holding methods were investigated, including threshold-based control logic and real-time control based on preceding and following headways, to determine the locations and numbers of control points and optimal control strength [138]. A combination of different real-time data-driven prediction approaches was also proposed to optimize holding control strategies and station skipping [139]–[142]. Other approaches, such as, dynamic bus propagation model [142], boarding person limit [143], are also proposed to improve the system efficiency. Recently, the performance of different holding methods was compared in terms of headway instability and mean holding time with and without real-time predictions [144], and real-time transfer synchronization was considered to reconcile single-line regularity and inter-line arrivals [145]. Most of performance evaluations are conducted in different simulation-based environments. In addition, some adjusted Genetic algorithms were proposed to address the train delay in metro systems [125], [146], and some researchers also conducted tests on real-time automatic rescheduling strategies to make the dynamic train timetable more flexible [134], [147].

In addition, incorporating travel behaviors in timetable design is still a challenging issue. The travelers' response to the adjusted transit schedules should be considered in a network-level transit system. Particular studies conducted by Liu and Zhou is firstly to propose an agent-based modeling framework to consider the optimal design in transit schedule network with boundedly rational travelers in the operational planning [125].

Another topic of interest is the study of timetable evaluation. The evaluation could be time-driven microscopic simulation method [148] and demand uncertainty for robustness [149], [150]. Other works are related to the special cases in timetable optimization, such as, the cyclic railway timetabling [124] and the schedule for loop line [123].

C. Policy Applications

In this section, we collect various works related to land use, financial applications and data privacy with transit big data.

The public transit services could affect the long-term land use pattern, such as, how people select their home and business locations; on the other hand, land use patterns further influence the demand for transportation related to travel distances. By recognizing that the public transportation/land-use relationship is extremely complex, Polzin [151] conducted deep analysis to better understanding of this relationship in terms of accessibility improvements, complementary policies, and momentum and promotion. Johnson [152] pointed out that there are three main approaches to enhance transit ridership by land use planning near transit corridors, including increasing residential density near transit corridors, mix-use for land, and retail development. In addition, Ratner and Goetz [153] focused on how the transit-oriented development (TOD) reshapes the land use and urban form throughout the entire Denver region and finds that the transit stations attract different type of land use and development based on their urban locations. For the sustainability of mass rail transit (MRT), Li *et al.* [154] presents a TOD planning model to find the optimal schemes of land-use type and land-use density in planned region around Shenzhen Metro line 3 in China. After reviewing a number of papers related to BRT development and land use, Stokenberga [155] concluded that it is still not clear whether BRT has improved accessibility for the population around the corridors or the desired land value increases have resulted in significant population displacement. Kim *et al.* [156] investigated the relationship between transit investment and urban land use change in a parcel-level land use in Southern California by developing a multinomial logistic regression model, which shows that vacant parcels within the vicinity of new transit stations are more likely to be developed not only for residential but also for other urban purposes. Hu *et al.* [157] proposed three machine-learning models to quantify the interdependencies between land use and public transport ridership and further provided the guidance of development of city regional center and amenity resource allocations in Singapore.

The smart cards had been used for collecting the fare and improving the profit of the operators. Some operators changed the fare scheme to make more profit. In 2014, Beijing metro substituted a flat-fare policy for a distance-based fare policy. Mining the passenger response to the fare change is significant for the next round of fare adjustment. Lots of work have been working on the transit fare estimation based on the fare elasticity. Generally, the elasticity is taken as -0.3 regardless of case-by-case difference [158]. The variety of passengers' responses to fare change exists at a station level and three fare increase alternatives (high, medium, and low) were evaluated in terms of their impacts on ridership and revenue [10]. For each alternative, the majority of the total trips with a length of around 15 km are the most sensitive to fare increases. Meanwhile, travel responses are influenced by many factors, not only price but also age, gender, income, day of week, time of day and trip purpose [159]–[161]. The new ticket scheme should consider serval affects and provide more flexible fare scheme for different purpose, such as the accumulate discount for commuters, early bird discount for elder passenger and daily pass for visitors.

The advances in technology make it possible to share the detailed individual travel information. This information provides great benefits, including improved services for customers and increased revenues and decreased costs for businesses. However, it has also raised important issues such as the misuse of personal information and loss of privacy [162]. Chen, Fung and Desai proposed an efficient data-dependent differentially private transit data sanitization approach based on a hybrid-granularity prefix tree structure [163].

As a short summary, there are a few common research topics between bus transit and rail transit, such as, network accessibility, route choice, activity purpose analysis, and timetable optimization and rescheduling. However, the big data applications still vary from each mode due to their specific characteristics. In Metro systems, the one-pay tickets with origin and destination in smart cards makes more research focus on the temporal and spatial demand pattern analysis, and the seamless transfer leads to more studies on passenger transfer estimation and First/Last train connection. On the other hand, in the urban bus systems, the unrecorded trip destination results in more research on trip origin-destination demand estimation and trip chain analysis.

V. Future Research Directions

The transit big data application has evolved rapidly over the past three decades, fueled by the diverse technologies ranging from the traditional smart card technique to smartphone data, while providing rich data resources for research. Characterized by inherent applications and detailed passenger travel patterns, it has nevertheless spawned a vast body of literature that encompasses a broad gamut of research directions. While the previous research has led to rapid studies in the understanding of passenger behavior and operation adjustment, there are still a number of challenges and trends worthy to be conducted in the future.

• Accurate transit demand acquisition with induced demand: One of the public transit target markets is the commuter, but some passengers may choose to use car or shared mobility for their commute due to the possible low-quality transit services. In this case, transit data sets only record the passengers who take transit, which only accounts for part of the total transit demand. Once the multi transportation datasets are merged, it is possible to obtain those potential and attractive passengers from other transportation modes by analyzing the spatial and temporal distribution within the completed trip chain individually. Meanwhile, the smartphone communication datasets also provide access to obtain the individual passenger who works at weekends or works very late. The potential transit demand can be obtained by considering the social characteristics of passengers such as the income and work place. In addition, it is worthy to study the interaction of travel demand and transit network structure as a closed loop [164].

• Data visualization: Compared with the traditional statistical analysis, data visualization provides a more direct view about the passenger travel pattern and passenger travel clustering [165], [148]. The long-term cumulative multi-source data make it possible to obtain the variance for the passenger and vehicles during the week or within one year [166], [167]. Visualizations clarify the spatial and temporal changes for the passenger flow and loading factor in the network under different conditions (e.g a regular day versus a national holiday, sunny day versus a rainy day). In this term, as stated above, the data visualization tool with GIS and GPS will become a hot topic in the next few years. Real-time information and reschedule: The AFC and AVL data can be packaged and sent to the main server every 15 min to satisfy the requirements of data analysis and operation optimization. In addition, when a signal failure or accident happened in the transit system, it becomes feasible to inform the passenger and obtain the real-time station and network operation conditions to reschedule the timetable [168], [169]. Meanwhile, the real-time operation information will help passengers choose a better pre-trip or en-route travel strategy.

• Cross-validation and data sample size: Most transit data, containing a lot of detailed passenger travel information, are not designed to obtain all the information about passengers such as socioeconomic information and trip purpose. To better understand the passenger behavior, some researchers have already merged household survey data with transit data to detect passengers' social group and travel pattern [68], [170], [171]. Despite a household survey, there will be more interaction with land use [172], [173] and census data to better reveal the passenger behavior for long-term and short-term public policy decisions and city planning. At the same time, the number of data samples required to analyze passenger behaviors is also another tough research question.

• New analysis methodology: More and more research focus on the agent-based behavior and disaggregated models and approaches due to the availability of individual travel information. Some new analysis methods from computer science and statistics have been applied to transmit data analysis, such as machine learning and data fusion clustering [174]. There will be more new analysis methods, especially for variation and clustering studies along with the correlation with other majors and fields, such as data mining and artificial intelligence.

• Shared Mobility: Transit system only provide stop-level service in the city. The transportation service from trip origin to transit stop is also necessary from the perspective of multi-modal transportation systems. The First Mile Last Mile (FMLM) challenge garners significant attention as a means to assess the accessibility of the first leg to public transit and the last leg from transit [175]. In recent years, shared mobility, including car sharing, personal vehicle sharing (peer-to-peer car sharing and fractional ownership), bike sharing and customized bus, has proliferated in global cities not only as an innovative transportation mode enhancing urban mobility but also as a potential solution to address first- and last-mile connectivity with public transit [176]. The research challenges and hot topics of shred mobility are as follows.

1) Bike sharing: In recent years, bicycle-sharing programs, such as CityCycle [177] and NiceRide, have received increasing attention with initiatives to increase bike usage, better meet the demand of a more mobile demand, and lessen the environmental impacts of our transportation activities [178]–[180]. China is taking a leading role in both public bike share and private electric bike (e-bike) growth. Based on the bike trajectories and the impacts of bike share demand such as distance, temperature and user heterogeneities, Campbell *et al.* [181] analyzed the viability of deploying large-scale shared e-bike systems in China. While the shared bikes provide better service for accessibility to transit system, it caused some serious social problems such as the unregulated bike parking and road occupancy of broken bikes. It is possible to collaborate with the available transit data to optimize the parking spot to normalize user parking behavior and forecast the proper bike demand to avoid capacity waste.

2) Customized transit service: Feeder bus or Customized transit service provides the personalized and flexibility that travelers need to access or egress from a bus or rail 'trunk line', especially for the commuter; whereas public transit is often constrained by fixed routes, driver availability, and vehicle scheduling [132], [182], [183]. Big data in transit systems are providing the available travel information such as demand, activity, passenger category and travel pattern. These characteristics offer better pictures for customer persona, which helps researchers optimize the bus stop locations, routes, timetables, and passenger-to-vehicle assignments to provide better feeder transit. Tong *et al.* developed a joint optimization model, providing flexible public transportation

services for major origin-destination (OD) pairs to/from inner cities with limited physical road infrastructure [132].

3) Self-driving transit: Technology is transforming transportation. Self-driving or driverless vehicles and buses are serving passengers in some cities [184]. As a new mode for future transportation system, the analysis of the self-driving vehicle as a connected mode for public transit and the correlation with other modes is a potential research topic [133].

As an emerging transportation mode, shared mobility has made a strong influence on the traditional public transit systems, and some research revealed that it has reduced the public transit ridership, to some extent [185]–[187]. How the public transit system should cooperate with shared mobility and what kinds of role public transit should represent have been paid great attention recently. Three scenarios are discussed, including the market-driven, public controlled and public-private. The scope, usage, access and business model of public transit was specifically analyzed [187]. Meanwhile, the public transit also needs to adjust itself to satisfy the future mobility system. The micro-transit may displace, evolve and even replace the fixed route public transit to fill the gap in service to compete with the private car [188]. How to combine the public transit with the new mobility mode in a better way and update the tradition public transit to provide a better transit service, is definitely a significant and potential topic for the future study.

Acknowledgment

The authors would like to thank Prof. Z. Wang from Beijing Jiaotong University, R. Liao, an international student from Beijing Jiaotong University for helping them to edit and polish the language of this article.

References

[1] J. Liu, "Passenger-focused schedule transportation systems: From increased observability to shared mobility," Ph.D. Dissertation, School Sustain. Eng. Built Environ., Arizona State Univ., Tempe, AZ, USA, 2018.

[2] X. Wu, J. Guo, K. Xian, and X. Zhou, "Hierarchical travel demand estimation using multiple data sources: A forward and backward propagation algorithmic framework on a layered computational graph," *Transp. Res. C, Emerg. Technol.*, vol. 96, pp. 321–346, Nov. 2018.

[3] J. Liu and X. Zhou, "Observability quantification of public transportation systems with heterogeneous data sources: An information-space projection approach based on discretized space-time network flow models," *Transp. Res. B, Methodol.*, vol. 128, pp. 302–323, Oct. 2019.

[4] M. Bagchi and P. R. White, "The potential of public transport smart card data," *Transp. Policy*, vol. 12, no. 5, pp. 464–474, Sep. 2005.

[5] V. Lockton and R. S. Rosenberg, "RFID: The next serious threat to privacy," *Ethics Inf. Technol.*, vol. 7, no. 4, pp. 221–231, Dec. 2005.

[6] C. D. Cottrill, "Approaches to privacy preservation in intelligent transportation systems and vehicle–infrastructure integration initiative," *Transp. Res. Rec.*, vol. 2129, no. 1, pp. 9–15, Jan. 2009.

[7] M. Munizaga, F. Devillaine, C. Navarrete, and D. Silva, "Validating travel behavior estimated from smartcard data," *Transp. Res. C, Emerg. Technol.*, vol. 44, no. 4, pp. 70–79, 2014.

[8] J. J. Barry, R. Newhouser, A. Rahbee, and S. Sayeda, "Origin and destination estimation in New York City with automated fare system data," *Transp. Res. Rec.*, vol. 1817, no. 1, pp. 183–187, Jan. 2002.

[9] K. Lu, B. Han, and X. Zhou, "Smart urban transit systems: From integrated framework to interdisciplinary perspective," *Urban Rail Transit*, vol. 4, no. 2, pp. 49–67, Jun. 2018.

[10] Z. J. Wang, X. H. Li, and F. Chen, "Impact evaluation of a mass transit fare change on demand and revenue utilizing smart card data," *Transp. Res. A, Policy Pract.*, vol. 77, pp. 213–224, Jul. 2015.

[11] K. Lu, B. Han, F. Lu, and Z. Wang, "Urban rail transit in China: Progress report and analysis (2008–2015)," *Urban Rail Transit*, vol. 2, nos. 3–4, pp. 93–105, Dec. 2016.

[12] M. Trépanier, N. Tranchant, and R. Chapleau, "Individual trip destination estimation in a transit smart card automated fare collection system," *J. Intell. Transp. Syst.*, vol. 11, no. 1, pp. 1–14, Apr. 2007.

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

[14] P. G. Furth, B. Hemily, T. Müller, and J. G. Strathman, "Uses of archived AVL-APC data to improve transit performance and management: Review and potential," in *Proc. Transp. Res. Board Annu. Meeting*, Jan. 2003, pp. 1–167.

[15] L. Moreira-Matias, J. Mendes-Moreira, J. F. De Sousa, and J. Gama, "Improving mass transit operations by using AVL-based systems: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 1636–1653, Aug. 2015.

[16] W.-H. Lin and J. Zeng, "Experimental study of real-time bus arrival time prediction with GPS data," *Transp. Res. Rec.*, vol. 1666, no. 1, pp. 101–109, Jan. 1999.

[17] A. Shalaby and A. Farhan, "Prediction model of bus arrival and departure times using AVL and APC data," *J. Public Transp.*, vol. 7, no. 1, pp. 41–61, Apr. 2015.

[18] J. G. Strathman, T. J. Kimpel, and S. Callas, "Headway deviation effects on bus passenger loads: Analysis of tri-met's archived AVL-APC data," Citeseer, State College, PA, USA, Tech. Rep. 62, 2003.

[19] R. Rajbhandari, S. I. Chien, and J. R. Daniel, "Estimation of bus dwell times with automatic passenger counter information," *Transp. Res. Rec.*, vol. 1841, no. 1, pp. 120–127, Jan. 2003.

[20] M. Frumin and J. Zhao, "Analyzing passenger incidence behavior in heterogeneous transit services using smartcard data and schedule-based assignment," *Transp. Res. Rec.*, vol. 2274, no. 1, pp. 52–60, Jan. 2012.

[21] S. Farber, M. Z. Morang, and M. J. Widener, "Temporal variability in transit-based accessibility to supermarkets," *Appl. Geogr.*, vol. 53, pp. 149–159, Sep. 2014.

[22] K. Fransen, T. Neutens, S. Farber, P. De Maeyer, G. Deruyter, and F. Witlox, "Identifying public transport gaps using time-dependent accessibility levels," *J. Transp. Geogr.*, vol. 48, pp. 176–187, Oct. 2015.

[23] S. K. S. Fayyaz, X. C. Liu, and G. Zhang, "An efficient general transit feed specification (GTFS) enabled algorithm for dynamic transit accessibility analysis," *PLoS ONE*, vol. 12, no. 10, Oct. 2017, Art. no. e0185333.

[24] P. Fortin, C. Montréal, C. Morency, M. Trépanier, C. Montréal, and C. Montréal, "Innovative GTFS data application for transit network analysis using a graph-oriented method," *J. Public Transp.*, vol. 19, no. 4, pp. 18–37, Oct. 2016.

[25] J. Liu and X. Zhou, "Capacitated transit service network design with boundedly rational agents," *Transp. Res. B, Methodol.*, vol. 93, pp. 225–250, Nov. 2016.

[26] S. Tao, J. Corcoran, I. Mateo-Babiano, and D. Rohde, "Exploring bus rapid transit passenger travel behaviour using big data," *Appl. Geogr.*, vol. 53, pp. 90–104, Sep. 2014.

[27] S. J. Barbeau, "Quality control-lessons learned from the deployment and evaluation of GTFS-realtime feeds," in *Proc. Transp. Res. Board Annu. Meeting*, Jan. 2018, pp. 1–20.

[28] O. Järv, R. Ahas, and F. Witlox, "Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records," *Transp. Res. C, Emerg. Technol.*, vol. 38, pp. 122–135, Jan. 2014.

[29] M. G. Demissie, S. Phithakkitnukoon, T. Sukhvibul, F. Antunes, R. Gomes, and C. Bento, "Inferring passenger travel demand to improve urban mobility in developing countries using cell phone data: A case study of senegal," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 9, pp. 2466–2478, Sep. 2016.

[30] M. Berlingerio *et al.*, "AllAboard: A system for exploring urban mobility and optimizing public transport using cellphone data," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*. Berlin, Germany: Springer, 2013, pp. 663–666.

[31] A. Thiagarajan *et al.*, "Cooperative transit tracking using smartphones," in *Proc. ACM SenSys*. Zürich, Switzerland, Nov. 2010, pp. 85–98.

[32] K. E. Watkins, B. Ferris, A. Borning, G. S. Rutherford, and D. Layton, "Where is my bus? Impact of mobile real-time information on the perceived and actual wait time of transit riders," *Transp. Res. A, Policy Pract.*, vol. 45, no. 8, pp. 839–848, 2011.

[33] R. Wu, Y. Cao, C. H. Liu, P. Hui, L. Li, and E. Liu, "Exploring passenger dynamics and connectivities in Beijing underground via Bluetooth networks," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, Apr. 2012, pp. 208–213.

[34] J. van den Heuvel *et al.*, "Using Bluetooth to estimate the impact of congestion on pedestrian route choice at train stations," in *Traffic and Granular Flow*, M. Chraibi, M. Boltes, A. Schadschneider, and A. Seyfried, Eds. Cham, Switzerland: Springer, 2015.

[35] R. G. Mishalani, M. R. Mccord, and T. Reinhold, "Use of mobile device wireless signals to determine transit route-level passenger origin–destination flows: Methodology and empirical evaluation," *Transp. Res. Rec.*, vol. 2544, no. 1, pp. 123–130, Jan. 2016.

[36] M.-H. Yang, D. J. Kriegman, and N. Ahuja, "Detecting faces in images: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 1, pp. 34–58, Jan. 2002.

[37] G. Kukharev, A. Kuzminski, and A. Nowosielski, "Structure and characteristics of face recognition," *Comput., Multimedia Intell. Techn.*, vol. 1, no. 1, pp. 111–124, 2005.

[38] M. Mikłasz, p. Olszewski, A. Nowosielski, and G. Kawka, "Pedestrian traffic distribution analysis using face recognition technology," in *Activities of Transport Telematics*, vol. 395. Berlin, Germany: Springer, 2013, pp. 303–312.

[39] N. Ran and H. L. Wang, "Application research on face recognition system in safety area of railway stations," *Railway Comput. Appl.*, vol. 21, no. 9, pp. 21–24, 2012.

[40] E. Mai and R. Hranac, "Twitter interactions as a data source for transportation incidents," in *Proc. Transp. Res. Board Annu. Meeting*, Jan. 2013, p. 1636.

[41] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes Twitter users: Real-time event detection by social sensors," in *Proc. 19th Int. Conf. World Wide Web*, Raleigh, NC, USA, 2010, pp. 851–860.

[42] H. Achrekar, A. Gandhe, R. Lazarus, S.-H. Yu, and B. Liu, "Predicting flu trends using Twitter data," in *Proc. IEEE Conf. INFOCOM WKSHPS*, Apr. 2011, pp. 702–707.

[43] C. Collins, S. Hasan, and S. Ukkusuri, "A novel transit rider satisfaction metric: Rider sentiments measured from online social media data," *J. Public Transp.*, vol. 16, no. 2, pp. 21–45, 2012.

[44] Y. Wang, C. Jiang, and H. Ren, "Model of delay prediction for signalized intersection based on GPS data," in *Proc. AMTIA*, Shanghai, China, Sep. 2016, pp. 1–8.

[45] L. Zhu, F. R. Yu, Y. Wang, B. Ning, and T. Tang, "Big data analytics in intelligent transportation systems: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 1, pp. 383–398, Jan. 2019.

[46] V. Vidya and N. Deepa, "Big data analytics in intelligent transportation systems using Hadoop," *Int. J. Recent Technol. Eng.*, vol. 7, no. 6S4, pp. 75–80, 2019.

[47] G. Zeng, "Application of big data in intelligent traffic system," *IOSR J. Comput. Eng.*, vol. 17, no. 1, pp. 1–4, 2015.

[48] H. Khazaei, S. Zareian, R. Veleda, and M. Litoiu, "Sipresk: A big data analytic platform for smart transportation," in *International Summit, Smart City 360°*. Berlin, Germany: Springer, 2016, pp. 419–430.

[49] Y. Wang, L. Zhu, Q. Lin, and L. Zhang, "Leveraging big data analytics for train schedule optimization in urban rail transit systems," in *Proc. IEEE Conf. ITSC*, Nov. 2018, pp. 1928–1932.

[50] J. Maktoubian *et al.*, "Analyzing large-scale smart card data to investigate public transport travel behaviour using big data analytics," *J. Inf. Technol. Softw. Eng.*, vol. 7, no. 4, pp. 1–3, 2017.

[51] F. Dzikrullah, N. A. Setiaawan, and S. Sulistyo, "Implementation of scalable K-means plus clustering for passengers temporal pattern analysis in public transportation system (BRT trans jogja case study)," in *Proc. InAES*, Yogyakarta, Indonesia, 2016, pp. 78–83.

[52] M. Riasetiawan, I. M. Harwanto, A. I. Falakh, A. K. Harryajie, J. Munjazi, and T. B. Adji, "Profiling and clustering methods for transaction profiling in BRT transaction," in *Proc. IEEE Conf. ICITEE*, Oct. 2017, pp. 1–6.

[53] L. Sun, J. G. Jin, D.-H. Lee, K. W. Axhausen, and A. Erath, "Demand-driven timetable design for metro services," *Transp. Res. C, Emerg. Technol.*, vol. 6, pp. 284–299, Sep. 2014.

[54] H. Niu, X. Zhou, and R. Gao, "Train scheduling for minimizing passenger waiting time with time-dependent demand and skip-stop patterns: Nonlinear integer programming models with linear constraints," *Transp. Res. B, Methodol.*, vol. 76, pp. 117–135, Jun. 2015.

[55] H. Niu and X. Zhou, "Optimizing urban rail timetable under time-dependent demand and oversaturated conditions," *Transp. Res. C, Emerg. Technol.*, vol. 36, no. 11, pp. 212–230, 2013.

[56] M.-P. Pelletier, M. Trépanier, and C. Morency, "Smart card data use in public transit: A literature review," *Transp. Res. C, Emerg. Technol.*, vol. 19, no. 4, pp. 557–568, 2011.

[57] J. Morris and F. Wang, "Planning for public transport in the future: Challenges of a changing metropolitan Melbourne," in *Proc. Australas. Transp. Res. Forum*, Canberra, ACT, Australia, 2002, pp. 2–4.

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

[59] H. Nishiuchi, J. King, and T. Todoroki, "Spatial-temporal daily frequent trip pattern of public transport passengers using smart card data," *Int. J. Intell. Transp. Syst. Res.*, vol. 11, no. 1, pp. 1–10, Jan. 2013.

[60] K. K. A. Chu and R. Chapleau, "Enriching archived smart card transaction data for transit demand modeling," *Transp. Res. Rec.*, vol. 2063, no. 1, pp. 63–72, Jan. 2008.

[61] A.-S. Briand, E. Côme, M. K. El Mahrsi, and L. Oukhellou, "A mixture model clustering approach for temporal passenger pattern characterization in public transport," *Int. J. Data Sci. Anal.*, vol. 1, no. 1, pp. 37–50, Apr. 2016.

[62] X. Ma *et al.*, "Mining smart card data for transit riders' travel patterns," *Transp. Res. C, Emerg. Technol.*, vol. 36, pp. 1–12 Nov. 2013.

[63] 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," *Transp. Res. C, Emerg. Technol.*, vol. 58, pp. 193–207, Sep. 2015.

[64] C. Zhong, E. Manley, S. M. Arisona, M. Batty, and G. Schmitt, "Measuring variability of mobility patterns from multiday smart-card data," *J. Comput. Sci.*, vol. 9, pp. 125–130, Jul. 2015.

[65] A.-S. Briand, E. Côme, M. Trépanier, and L. Oukhellou, "Analyzing year-to-year changes in public transport passenger behaviour using smart card data," *Transp. Res. C, Emerg. Technol.*, vol. 79, pp. 274–289, Jun. 2017.

[66] C. Zhong *et al.*, "Variability in regularity: Mining temporal mobility patterns in London, Singapore and Beijing using smart-card data," *PLoS ONE*, vol. 11, no. 2, Feb. 2016, Art. no. e0149222.

[67] X. Ma, C. Liu, H. Wen, Y. Wang, and Y.-J. Wu, "Understanding commuting patterns using transit smart card data," *J. Transp. Geogr.*, vol. 58, pp. 135–145, Jan. 2017.

[68] Y. Long and J. C. Thill, "Combining smart card data and household travel survey to analyze jobs-housing relationships in Beijing," *Comput. Environ. Urban Syst.*, vol. 53, pp. 19–35, 2015.

[69] B. Marion and M. W. Horner, "Comparison of socioeconomic and demographic profiles of extreme commuters in several U.S. metropolitan statistical areas," *Transp. Res. Rec.*, vol. 2013, no. 1, pp. 38–45, Jan. 2007.

[70] Y. Long, X. Liu, J. Zhou, and Y. Chai, "Early birds, night owls, and tireless/recurring itinerants: An exploratory analysis of extreme transit behaviors in Beijing, China," *Habitat Int.*, vol. 57, pp. 223–232, Oct. 2016.

[71] Z. Cui, Y. Long, R. Ke, and Y. Wang, "Characterizing evolution of extreme public transit behavior using smart card data," in *Proc. IEEE IS*, vol. 2, Oct. 2015, pp. 1–6.

[72] L. Sun, K. W. Axhausen, D.-H. Lee, and X. Huang, "Understanding metropolitan patterns of daily encounters," *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 34, pp. 13774–13779, 2013.

[73] D. Luo, O. Cats, and H. V. Lint, "Constructing transit origin–destination matrices with spatial clustering," *Transp. Res. Rec.*, vol. 2652, pp. 39–49, Jan. 2017.

[74] J. Zhou, N. Sipe, Z. Ma, D. Mateo-Babiano, and S. Darchen, "Monitoring transit-served areas with smartcard data: A Brisbane case study," *J. Transp. Geogr.*, vol. 76, pp. 265–275, Apr. 2019.

[75] L. Sun *et al.*, "Models of bus boarding and alighting dynamics," *Transp. Res. A, Policy Pract.*, vol. 69, pp. 46–447, Nov. 2014.

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

[77] A. Ali, J. Kim, and S. Lee, "Travel behavior analysis using smart card data," *KSCE J.-Civil Eng.*, vol. 20, no. 4, pp. 1532–1539, May 2016.

[78] A. Chakirov and A. Erath, "Use of public transport smart card fare payment data for travel behaviour analysis in Singapore," in *Proc. 16th Int. Conf. Hong Kong Soc. Transp. Stud.*, Hong Kong, 2011, pp. 1–12.

[79] A. Owen and D. M. Levinson, "Modeling the commute mode share of transit using continuous accessibility to jobs," *Transp. Res. A, Policy Pract.*, vol. 74, pp. 110–122, Apr. 2015.

[80] M. J. Widener, S. Farber, T. Neutens, and M. Horner, "Spatiotemporal accessibility to supermarkets using public transit: An interaction potential approach in Cincinnati, Ohio," *J. Transp. Geogr.*, vol. 42, pp. 72–83, Jan. 2015.

[81] E. C. Delmelle and I. Casas, "Evaluating the spatial equity of bus rapid transit-based accessibility patterns in a developing country: The case of Cali, Colombia," *Transp. Policy*, vol. 20, pp. 36–46, Mar. 2012.

[82] Y. Hadas, "Assessing public transport systems connectivity based on Google transit data," *J. Transp. Geogr.*, vol. 33, pp. 105–116, Dec. 2013.

[83] T. Lin *et al.*, "Spatial analysis of access to and accessibility surrounding train stations: A case study of accessibility for the elderly in Perth, Western Australia," *J. Transp. Geogr.*, vol. 39, pp. 111–120, Jul. 2014.

[84] H. Badland, S. Hickey, F. Bull, and B. Giles-Corti, "Public transport access and availability in the RESIDE study: Is it taking us where we want to go?" *J. Transp. Health*, vol. 1, no. 1, pp. 45–49, Mar. 2014.

[85] A. Aklilu and T. Necha, "Analysis of the spatial accessibility of addis Ababa's light rail transit: The case of East–West corridor," *Urban Rail Transit*, vol. 4, no. 1, pp. 35–48, Mar. 2018.

[86] R. M. Pendyala, T. Yamamoto, and R. Kitamura, "On the formulation of time-space prisms to model constraints on personal activity-travel engagement," *Transportation*, vol. 29, no. 1, pp. 73–94, 2002.

[87] S. Chen, C. Claramunt, and C. Ray, "A spatio-temporal modelling approach for the study of the connectivity and accessibility of the Guangzhou metropolitan network," *J. Transp. Geogr.*, vol. 36, pp. 12–23, Apr. 2014.

[88] R. Kitamura, T. Akiyama, T. Yamamoto, and T. F. Golob, "Accessibility in a metropolis: Toward a better understanding of land use and travel," *Transp. Res. Rec.*, vol. 1780, no. 1, pp. 64–75, Jan. 2001.

[89] S. Mavoa, K. Witten, T. Mccreanor, and D. O'Sullivan, "GIS based destination accessibility via public transit and walking in Auckland, New Zealand," *J. Transp. Geogr.*, vol. 20, no. 1, pp. 15–22, Jan. 2012.

[90] A. C. Ford, S. L. Barr, R. J. Dawson, and P. James, "Transport accessibility analysis using GIS: Assessing sustainable transport in London," *ISPRS Int. Geo-Inf.*, vol. 4, no. 1, pp. 124–149, 2015.

[91] F. Devillaine, M. Munizaga, and M. Trépanier, "Detection of activities of public transport users by analyzing smart card data," *Transp. Res. Rec.*, vol. 2276, no. 1, pp. 48–55, Jan. 2012.

[92] A. Chakirov and A. Erath, "Activity identification and primary location modelling based on smart card payment data for public transport," in *Proc. 13th Int. Conf. Travel Behav. Res.*, 2012, pp. 1–24.

[93] G. G. Langlois, H. N. Koutsopoulos, and J. Zhao, "Inferring patterns in the multi-week activity sequences of public transport users," *Transp. Res. C, Emerg. Technol.*, vol. 64, pp. 1–16, Mar. 2016.

[94] S. Jiang, J. Ferreira, and M. C. Gonzalez, "Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore," *IEEE Trans. Big Data*, vol. 3, no. 2, pp. 208–219, Jun. 2017.

[95] M. Xue, H. Wu, W. Chen, W. S. Ng, and G. H. Goh, "Identifying tourists from public transport commuters," in *Proc. ACM KDD*, New York, NY, USA, Aug. 2014, pp. 1779–1788.

[96] S. G. Lee and M. Hickman, "Trip purpose inference using automated fare collection data," *Public Transp.*, vol. 6, nos. 1–2, pp. 1–20, Apr. 2014.

[97] W. Bohte and K. Maat, "Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in The Netherlands," *Transp. Res. C, Emerg. Technol.*, vol. 17, no. 3, pp. 285–297, 2009.

[98] Y. Wang, G. H. D. A. Correia, E. De Romph, and H. J. P. Timmermans, "Using metro smart card data to model location choice of after-work activities: An application to Shanghai," *J. Transp. Geogr.*, vol. 63, pp. 40–47, Jul. 2017.

[99] N. Nassir, A. Khani, S. G. Lee, H. Noh, and M. Hickman, "Transit stop-level origin–destination estimation through use of transit schedule and automated data collection system," *Transp. Res. Rec.*, vol. 2263, no. 1, pp. 140–150, Jan. 2011.

[100] A. A. Alsger, M. Mesbah, L. Ferreira, and H. Safi, "Use of smart card fare data to estimate public transport origin–destination matrix," *Transp. Res. Rec.*, vol. 2535, no. 1, pp. 88–96, 2015.

[101] W. Wang, J. Attanucci, and N. Wilson, "Bus passenger origin-destination estimation and related analyses using automated data collection systems," *J. Public Transp.*, vol. 14, no. 4, pp. 131–150, Mar. 2015.

[102] Y. Sun and R. Xu, "Rail transit travel time reliability and estimation of passenger route choice behavior: Analysis using automatic fare collection data," *Transp. Res. Rec.*, vol. 2275, no. 1, pp. 58–67, Jan. 2012.

[103] N. Van Oort, "Incorporating service reliability in public transport design and performance requirements: International survey results and recommendations," *Res. Transp. Econ.*, vol. 48, pp. 92–100, Dec. 2014.

[104] K. Yin *et al.*, "Link travel time inference using entry/exit information of trips on a network," *Transp. Res. B, Methodol.*, vol. 80, pp. 303–321, Oct. 2015.

[105] J. Zhao *et al.*, "Estimation of passenger route choice pattern using smart card data for complex metro systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 4, pp. 790–801, Apr. 2017.

[106] Y.-S. Zhang and E.-J. Yao, "Splitting travel time based on AFC data: Estimating walking, waiting, transfer, and in-vehicle travel times in metro system," *Discrete Dyn. Nature Soc.*, vol. 2015, pp. 1–11, Dec. 2015.

[107] K. M. Kim *et al.*, "Does crowding affect the path choice of metro passengers?" *Transp. Res. A, Policy Pract.*, vol. 77, pp. 292–304, Jul. 2015.

[108] Y. Sun and P. M. Schonfeld, "Schedule-based rail transit path-choice estimation using automatic fare collection data," *J. Transp. Eng.*, vol. 142, no. 1, Jan. 2016, Art. no. 04015037.

[109] X. Xu, L. Xie, H. Li, and L. Qin, "Learning the route choice behavior of subway passengers from AFC data," *Expert Syst. Appl.*, vol. 95, pp. 324–332, Apr. 2018.

[110] M. Hassan *et al.*, "Modeling transit users stop choice behavior: Do travelers strategize?" *J. Public Transp.*, vol. 19, no. 3, pp. 98–116, Aug. 2016.

[111] J. Kim, J. Corcoran, and M. Papamanolis, "Route choice stickiness of public transport passengers: Measuring habitual bus ridership behavior using smart card data," *Transp. Res. C, Emerg. Technol.*, vol. 83, pp. 146–164, Oct. 2017.

[112] N. Nassir, M. Hickman, and Z. Ma, "Statistical inference of transit passenger boarding strategies from farecard data," *Transp. Res. Rec.*, vol. 2652, no. 1, pp. 8–18, Jan. 2017.

[113] M. H. Poon, S. C. Wong, and C. O. Tong, "A dynamic schedule-based model for congested transit networks," *Transp. Res. B, Methodol.*, vol. 38, pp. 343–368, May 2004.

[114] Y. Hamdouch, P. Marcotte, and S. Nguyen, "Capacitated transit assignment with loading priorities," *Math. Program.*, vol. 101, no. 1, pp. 205–230, 2004.

[115] A. Nuzzolo, U. Crisalli, and L. Rosati, "A schedule-based assignment model with explicit capacity constraints for congested transit networks," *Transp. Res. C, Emerg. Technol.*, vol. 20, no. 1, pp. 16–33, 2012.

[116] S. Binder, Y. Maknoon, and M. Bierlaire, "Exogenous priority rules for the capacitated passenger assignment problem," *Transp. Res. B, Methodol.*, vol. 105, pp. 19–42, Nov. 2017.

[117] Z. Ma, L. Ferreira, and M. Mesbah, "Measuring service reliability using automatic vehicle location data," *Math. Problems Eng.*, vol. 2014, pp. 1–12, Apr. 2014.

[118] N. Nassir, M. Hickman, and Z. Ma, "Behavioral findings from observed transit route choice strategies in the farecard data of Brisbane," in *Proc. 37th ATRF*, Sydney, NSW, Australia, 2015, pp. 1–11.

[119] E. Van Der Hurk, L. Kroon, G. Maroti, and P. Vervest, "Deduction of passengers' route choices from smart card data," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 1, pp. 430–440, Feb. 2015.

[120] I. Ceapa, C. Smith, and L. Capra, "Avoiding the crowds: Understanding Tube station congestion patterns from trip data," in *Proc. ACM SIGKDD Int. Workshop Urban Comput. (UrbComp)*, vol. 12, 2012, pp. 134–141.

[121] W. Jang, "Travel time and transfer analysis using transit smart card data," *Transp. Res. Rec.*, vol. 2144, no. 1, pp. 142–149, Jan. 2010.

[122] W. Zhu, W. Wang, and Z. Huang, "Estimating train choices of rail transit passengers with real timetable and automatic fare collection data," *J. Adv. Transp.*, vol. 2017, pp. 1–12, Aug. 2017.

[123] X. Yang, A. Chen, B. Ning, and T. Tang, "Bi-objective programming approach for solving the metro timetable optimization problem with dwell time uncertainty," *Transp. Res. E, Logistics Transp. Rev.*, vol. 97, pp. 22–37, Jan. 2017.

[124] R.-J. Shi, B.-H. Mao, Y. Ding, Y. Bai, and Y. Chen, "Timetable optimization of rail transit loop line with transfer coordination," *Discrete Dyn. Nature Soc.*, vol. 2016, pp. 1–11, Aug. 2016.

[125] Q. Zhen and S. Jing, "Train rescheduling model with train delay and passenger impatience time in urban subway network," *J. Adv. Transp.*, vol. 50, no. 8, pp. 1990–2014, Dec. 2016.

[126] T. Albrecht, "Automated timetable design for demand-oriented service on suburban railways," *Public Transp.*, vol. 1, no. 1, pp. 5–20, May 2009.

[127] X. Yang, A. Chen, B. Ning, and T. Tang, “A stochastic model for the integrated optimization on metro timetable and speed profile with uncertain train mass,” *Transp. Res. B, Methodol.*, vol. 91, pp. 424–445, Sep. 2016.

[128] Y. Wu, H. Yang, J. Tang, and Y. Yu, “Multi-objective re-synchronizing of bus timetable: Model, complexity and solution,” *Transp. Res. C, Emerg. Technol.*, vol. 67, pp. 149–168, Jun. 2016.

[129] J. Chen, Z. Liu, S. Zhu, and W. Wang, “Design of limited-stop bus service with capacity constraint and stochastic travel time,” *Transp. Res. E, Logistics Transp. Rev.*, vol. 83, pp. 1–15, Nov. 2015.

[130] Z. Cao, Z. Yuan, and S. Zhang, “Performance analysis of stop-skipping scheduling plans in rail transit under time-dependent demand,” *Int. J. Environ. Res. Public Health*, vol. 13, no. 7, p. 707, Jul. 2016.

[131] J. Yin, L. Yang, T. Tang, Z. Gao, and B. Ran, “Dynamic passenger demand oriented metro train scheduling with energy-efficiency and waiting time minimization: Mixed-integer linear programming approaches,” *Transp. Res. B, Methodol.*, vol. 97, pp. 182–213, Mar. 2017.

[132] L. Tong *et al.*, “Customized bus service design for jointly optimizing passenger-to-vehicle assignment and vehicle routing,” *Transp. Res. C, Emerg. Technol.*, vol. 85, pp. 451–475, Dec. 2017.

[133] M. D. Yap, G. Correia, and B. van Arem, “Preferences of travellers for using automated vehicles as last mile public transport of multimodal train trips,” *Transp. Res. A, Policy Pract.*, vol. 94, pp. 1–16, Dec. 2016.

[134] Y. Gao, L. Yang, and Z. Gao, “Real-time automatic rescheduling strategy for an urban rail line by integrating the information of fault handling,” *Transp. Res. C, Emerg. Technol.*, vol. 81, pp. 246–267, Aug. 2017.

[135] Z. Jiang, C.-H. Hsu, D. Zhang, and X. Zou, “Evaluating rail transit timetable using big passengers’ data,” *J. Comput. Syst. Sci.*, vol. 82, no. 1, pp. 144–155, Feb. 2016.

[136] C. F. Daganzo, “A headway-based approach to eliminate bus bunching: Systematic analysis and comparisons,” *Transp. Res. B, Methodol.*, vol. 43, no. 10, pp. 913–921, 2009.

[137] C. F. Daganzo and J. Pilachowski, “Reducing bunching with bus-to-bus cooperation,” *Transp. Res. B, Methodol.*, vol. 45, no. 1, pp. 267–277, Jan. 2011.

[138] L. Fu and X. Yang, “Design and implementation of bus–holding control strategies with real-time information,” *Transp. Res. Rec.*, vol. 1791, no. 1, pp. 6–12, Jan. 2002.

[139] G. E. Sánchez-Martínez, H. N. Koutsopoulos, and N. H. M. Wilson, “Real-time holding control for high-frequency transit with dynamics,” *Transp. Res. B, Methodol.*, vol. 83, pp. 1–19, Jan. 2016.

[140] L. Moreira-Matias, O. Cats, J. Gama, J. Mendes-Moreira, and J. F. De Sousa, “An online learning approach to eliminate bus bunching in real-time,” *Appl. Soft Comput.*, vol. 47, pp. 460–482, Oct. 2016.

[141] M. Andres and R. Nair, “A predictive-control framework to address bus bunching,” *Transp. Res. B, Methodol.*, vol. 104, pp. 123–148, 2017.

[142] W. Wu, R. Liu, and W. Jin, “Modelling bus bunching and holding control with vehicle overtaking and distributed passenger boarding behaviour,” *Transp. Res. B, Methodol.*, vol. 104, pp. 175–197, Oct. 2017.

[143] S. Zhao, C. Lu, S. Liang, and H. Liu, “A self-adjusting method to resist bus bunching based on boarding limits,” *Math. Problems Eng.*, vol. 2016, pp. 1–7, May 2016.

[144] S. J. Berrebi, E. Hans, N. Chiabaut, J. A. Laval, L. Leclercq, and K. E. Watkins, “Comparing bus holding methods with and without real-time predictions,” *Transp. Res. C, Emerg. Technol.*, vol. 87, pp. 197–211, Feb. 2017.

[145] A. Gavriilidou and O. Cats, “Reconciling transfer synchronization and service regularity: Real-time control strategies using passenger data,” *Transportmetrica A, Transp. Sci.*, vol. 15, no. 2, pp. 215–243, 2018.

[146] Y. Gao, L. Kroon, M. Schmidt, and L. Yang, “Rescheduling a metro line in an over-crowded situation after disruptions,” *Transp. Res. B, Methodol.*, vol. 93, pp. 425–449, Nov. 2016.

[147] A. D’Ariano, F. Corman, D. Pacciarelli, and M. Pranzo, “Reordering and local rerouting strategies to manage train traffic in real time,” *Transp. Sci.*, vol. 42, no. 4, pp. 405–419, Nov. 2008.

[148] N. Andrienko and G. Andrienko, “Visual analytics of movement: An overview of methods, tools and procedures,” *Inf. Vis.*, vol. 12, no. 1, pp. 3–24, Jan. 2013.

[149] P. Sels *et al.*, “Reducing the passenger travel time in practice by the automated construction of a robust railway timetable,” *Transp. Res. B, Methodol.*, vol. 84, pp. 124–156, Feb. 2016.

[150] E. Hassannayebi, S. H. Zegordi, M. R. Amin-Naseri, and M. Yaghini, “Train timetabling at rapid rail transit lines: A robust multi-objective stochastic programming approach,” *Oper. Res.*, vol. 17, no. 2, pp. 435–477, Jul. 2017.

[151] S. E. Polzin, “Transportation/land-use relationship: Public transit’s impact on land use,” *J. Urban Planning Develop.*, vol. 125, no. 4, pp. 135–151, Dec. 1999.

[152] A. Johnson, “Bus transit and land use: Illuminating the interaction,” *J. Public Transp.*, vol. 6, no. 4, pp. 21–39, Apr. 2015.

[153] K. A. Ratner and A. R. Goetz, “The reshaping of land use and urban form in Denver through transit-oriented development,” *Cities*, vol. 30, pp. 31–46, Feb. 2013.

[154] Y. Li, H. L. Guo, H. Li, G. H. Xu, Z. R. Wang, and C. W. Kong, “Transit-oriented land planning model considering sustainability of mass rail transit,” *J. Urban Planning Develop.*, vol. 136, no. 3, pp. 243–248, Sep. 2010.

[155] A. Stokenberga, “Does bus rapid transit influence urban land development and property values: A review of the literature,” *Transp. Rev.*, vol. 34, no. 3, pp. 276–296, 2014.

[156] J. H. Kim *et al.*, “Infill dynamics in rail transit corridors: Challenges and prospects for integrating transportation and land use planning,” Dept. Transp. Division Res. Innov., Irvine, CA, USA, Tech. Rep. CA16-2641, 2016.

[157] N. Hu, E. F. Legara, K. K. Lee, G. G. Hung, and C. Monterola, “Impacts of land use and amenities on public transport use, urban planning and design,” *Land Use Policy*, vol. 57, pp. 356–367, Nov. 2016.

[158] G. Bresson *et al.*, “The main determinants of the demand for public transport: A comparative analysis of England and France using shrinkage estimators,” *Transp. Res. A, Policy Pract.*, vol. 37, no. 7, pp. 605–627, 2003.

[159] B. E. Mccollom and R. H. Pratt, “Traveler response to transportation system changes. Chapter 12-transit pricing and fares,” in *Proc. Transp. Res. Board Annu. Meeting*, Jan. 2004, p. 69.

[160] N. Paulley *et al.*, “The demand for public transport: The effects of fares, quality of service, income and car ownership,” *Transp. Policy*, vol. 13, no. 4, pp. 295–306, Jul. 2006.

[161] T. Litman, “Transit price elasticities and cross-elasticities,” *J. Public Transp.*, vol. 7, no. 2, pp. 37–58, 2004.

[162] S. P. Hong and S. Kang, “Ensuring privacy in smartcard-based payment systems: A case study of public metro transit systems,” in *Communications and Multimedia Security*, vol. 4237. Berlin, Germany: Springer, 2006, pp. 206–215.

[163] R. Chen, B. C. M. Fung, B. C. Desai, and N. M. Sossou, “Differentially private transit data publication: A case study on the montreal transportation system,” in *Proc. ACM KDD*, 2012, pp. 213–221.

[164] H. Badia, J. Argote-Cabanero, and C. F. Daganzo, “How network structure can boost and shape the demand for bus transit,” *Transp. Res. A, Policy Pract.*, vol. 103, pp. 83–94, Sep. 2017.

[165] B. Dewulf *et al.*, “Examining commuting patterns using floating car data and circular statistics: exploring the use of new methods and visualizations to study travel times,” *J. Transp. Geogr.*, vol. 48, pp. 41–51, Oct. 2015.

[166] S. Ghaemi *et al.*, “A visual segmentation method for temporal smart card data,” *Transportmetrica A, Transp. Sci.*, vol. 13, no. 5, pp. 381–404, 2017.

[167] M. Mesbah, G. Currie, C. Lennon, and T. Northcott, “Spatial and temporal visualization of transit operations performance data at a network level,” *J. Transp. Geogr.*, vol. 25, pp. 15–26, Nov. 2012.

[168] M. M. Rahman, S. C. Wirasinghe, and L. Kattan, “Users’ views on current and future real-time bus information systems,” *J. Adv. Transp.*, vol. 47, no. 3, pp. 336–354, Apr. 2013.

[169] C. G. Walker, J. N. Snowdon, and D. M. Ryan, “Simultaneous disruption recovery of a train timetable and crew roster in real time,” *Comput. Oper. Res.*, vol. 32, no. 8, pp. 2077–2094, Aug. 2005.

[170] A. Alsger *et al.*, “Validating and improving public transport origin–destination estimation algorithm using smart card fare data,” *Transp. Res. C, Emerg. Technol.*, vol. 68, pp. 490–506, Jul. 2016.

[171] A. Vij and K. Shankari, “When is big data big enough? Implications of using GPS-based surveys for travel demand analysis,” *Transp. Res. C, Emerg. Technol.*, vol. 56, pp. 446–462, Jul. 2015.

[172] H. Wei, T. Zuo, H. Liu, and Y. J. Yang, “Integrating land use and socioeconomic factors into scenario-based travel demand and carbon emission impact study,” *Urban Rail Transit*, vol. 3, no. 1, pp. 3–14, Mar. 2017.

[173] S. A. Mckinley and H. Wei, "Viability assessment of light rail line planning: Case study of cincinnati eastern corridor," *Urban Rail Transit*, vol. 3, no. 1, pp. 34–44, Mar. 2017.
[174] J. P. Locquiao, "Multifaceted analysis of transit station accessibility characteristics based on first mile last mile," M.S. thesis, Univ. Utah, Salt Lake City, UT, USA, 2016.
[175] J. Mendes-Moreira, L. Moreira-Matias, J. Gama, and J. F. De Sousa, "Validating the coverage of bus schedules: A machine learning approach," *Inf. Sci.*, vol. 293, pp. 299–313, Feb. 2015.
[176] S. Shaheen and N. Chan, "Mobility and the sharing economy: Potential to facilitate the first- and last-mile public transit connections," *Built Environ.*, vol. 42, no. 4, pp. 573–588, Dec. 2016.
[177] E. Fishman, S. Washington, and N. Haworth, "Barriers and facilitators to public bicycle scheme use: A qualitative approach," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 15, no. 6, pp. 686–698, Nov. 2012.
[178] J.-R. Lin and T.-H. Yang, "Strategic design of public bicycle sharing systems with service level constraints," *Transp. Res. E, Logistics Transp. Rev.*, vol. 47, no. 2, pp. 284–294, 2011.
[179] A. Khani, V. Livshits, and A. Dutta, "Modeling regional bicycle travel in Phoenix Metropolitan area," in *Proc. Transp. Res. Board Annu. Meeting*, Jan. 2014, p. 18.
[180] N. Tilahun, P. Thakuriah, M. Li, and Y. Keita, "Transit use and the work commute: Analyzing the role of last mile issues," *J. Transp. Geogr.*, vol. 54, pp. 359–368, Jun. 2016.
[181] A. A. Campbell, C. R. Cherry, M. S. Ryerson, and X. M. Yang, "Factors influencing the choice of shared bicycles and shared electric bikes in Beijing," *Transp. Res. C, Emerg. Technol.*, vol. 67, pp. 399–414, Jun. 2016.
[182] T. Liu and A. Ceder, "Analysis of a new public-transport-service concept: Customized bus in China," *Transp. Policy*, vol. 39, pp. 63–76, Apr. 2015.
[183] J de Oña, R de Oña, and G López, "Transit service quality analysis using cluster analysis and decision trees: a step forward to personalized marketing in public transportation," *Transportation*, vol. 43, no. 5, pp. 725–747, Sep. 2016.
[184] P. Bos, "Self-driving bus to improve accesibility of rural areas in The Netherlands. Peoplemover as a first- and last mile solution," M.S. thesis, Nijmegen School Manage., Radoud Univ., Nijmegen, Holland, 2017.
[185] J. Sochor, I. C. M. Karlsson, and H. Strömberg, "Trying out mobility as a service: Experiences from a field trial and implications for understanding demand," *Transp. Res. Rec.*, vol. 2542, no. 1, pp. 57–64, Jan. 2016.
[186] I. M. Karlsson, J. Sochor, and H. Strömberg, "Developing the 'service'in mobility as a service: Experiences from a field trial of an innovative travel brokerage," *Transp. Res. Procedia*, vol. 14, pp. 3265–3273, Jan. 2016.
[187] G. Smith, J. Sochor, and M. Karlssona, "Mobility as a service: Implications for future mainstream public transport," in *Proc. ITLS*, vol. 15, 2017, pp. 1–15.
[188] Y. Z. Wong, D. A. Hensher, and C. Mulley, "Emerging transport technologies and the modal efficiency framework: A case for mobility as a service (MaaS)," in *Proc. ITLS Working Papers*, 2018, pp. 1–28.

Kai Lu received the Ph.D. degree in traffic and transportation management from Beijing Jiaotong University, Beijing, China, in 2019.

She is currently a Post-Doctoral Fellow with Beijing Jiaotong University. She is also working with Traffic Control Technology Co., Ltd., Beijing. Her research interests include transit passenger behavior, dynamic urban rail transit assignment, and traffic estimation and prediction.

Jiangtao Liu received the Ph.D. degree in civil engineering from the School of Sustainable Engineering and the Built Environment, Arizona State University, in 2018.

He is currently a Data Scientist with Supply Chain Analytics, Walmart Inc. His research focused on scheduled transportation system modeling, supply chain analytics, shared mobility optimization and simulation, and big data applications.

Xuesong Zhou (Member, IEEE) received the Ph.D. degree in civil engineering from the University of Maryland, College Park, MD, USA, in 2004.

He is currently an Associate Professor with the School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ, USA. His research work focuses on dynamic traffic assignment, traffic estimation and prediction, large-scale routing, and rail scheduling.

Dr. Zhou is the Co-Chair of the IEEE ITS Society Technical Committee on Traffic and Travel Management and the Co-Vice Chair of the Railway Applications Section (RAS), Institute for Operations Research and the Management Sciences.

Baoming Han received the Ph.D. degree in civil engineering from the University of Liege, Belgium, in 1997.

He is currently a Professor with the School of Traffic and Transportation, Beijing Jiaotong University. His research interests include urban rail transit operation management, dynamic transit assignment, and high-speed rail network planning, scheduling, and optimization.