European Transport Research Review

REVIEW



# Route choice modelling for an urban rail transit network: past, recent progress and future prospects

Yihan Tian[1,2], Wei Zhu[1,2*] and Fangqing Song[3]

## Abstract

Route choice modelling is a critical aspect of analysing urban rail transit (URT) networks and provides a foundation for URT planning and operation. Unlike in a free-flow road network, the consideration set for route choice decisions in a URT network does not depend purely on the physical connectivity of the network and decision makers'characteristics. Instead, it is also contingent on the train schedules. This paper delves into the evolution of research on route choices in URT networks, encompassing both probabilistic route choice modelling derived from utility maximisation theory and logit curve with physical connectivity, and retrospective route choice modelling based on travel time chaining along with comprehensive transport data. The former is noted for its conciseness, simplicity, and interpretability in real-world applications, even though the methodologies may not be cutting-edge. The latter incorporates dynamic temporal information to understand activities of passengers in URT networks. Enhancements of each genres are also examined. However, these improvements might not fully address the inherent limitations of models relating to a dependency on the quality of parameters, experience of experts, and calculation efficiency. In addition, novel research adopting contemporary data mining techniques instead of classical models are introduced. The historical development of research on URT network route choices underscores the importance of amalgamating independent information networks such as surveillance networks and social networks to establish a comprehensive multi-dimensional network. Such an approach integrates passenger attributes across networks, offering a multi-dimensional understanding of passengers' route choice behaviours. Our review work aims to present not only a systematic conceptual framework for route choices in URT networks but also a novel path for transport researchers and practitioners to decipher the travel behaviours of passengers.

**Keywords** Review, Probabilistic route choice modelling, Retrospective route choice modelling, Calibration and validation, Multi-dimensional information network, Urban rail transit

*Correspondence:
Wei Zhu
zhuweimail@tongji.edu.cn
[1]College of Transportation Engineering, Tongji University, Shanghai, China
[2]Shanghai Key Laboratory of Rail Infrastructure Durability and System Safety, Tongji University, Shanghai, China
[3]Institute for Transport Studies, University of Leeds, Leeds, UK

## 1 Introduction

As urban rail transit (URT) systems become increasingly prominent within urban public transportation, especially in highly populated areas, their availability and quality of service affect both the operations of cities and activities of citizens. Understanding the travel behaviours of passengers provides a foundation for subsequent transport operations, encompassing tasks such as train scheduling, network coordination, emergency responses, etc. An accurate and timely travel demand analysis is required to better support transport operations in a URT system. The route choice analysis of passengers is a key component of travel demand analysis. The restoration of accurate route choices and comprehension of travel patterns are thus important to making precise passenger flow assignments, forecasting transfer volumes, distributing fares, and shaping future passenger service customisation. Therefore, the development of route choice modelling for route choice analysis is essential to improving the quality of transport operations.

In most countries, such as Australia, the United States, the United Kingdom, Singapore, Japan, South Korea, China, etc., "seamless transfer" is applied to improve passengers' travel experience and alleviate congestion at bottleneck points. This means that passengers only need to tap in at the origin station and tap out at the destination station. Hence, only the tap-in and tap-out time stamps and locations are recorded, and the activities between tap-in and tap-out remain inaccessible to researchers. In this context, investigating route choice behaviour in a URT network becomes challenging due to three main reasons:

- *Increasing network complexity*: By the end of 2022, 545 cities in 78 countries or regions had introduced URT systems, among which 111 cities had a total operating distance exceeding 100 km [1]. Integrating individual lines into a connected and unified URT network may provide additional feasible routes for an origin-destination (OD) pair. The flexible train schedules further increase the difficulty of route choice estimation.
- *Unobserved choice outcome*: The "black-box" nature of route choice behaviour in a URT network means that in cases where multiple routes are available, the route choice outcome cannot be directly observed and instead has to be inferred based on other information [2].
- *Preference heterogeneity*: Route choice decision-making is affected by various factors, the influence of which might vary across individuals. Neglecting preference heterogeneity in route choice analysis would lead to biased estimation and impair the prediction of passenger flow assignment.

This paper aims to present not only a systematic conceptual framework for route choices in URT networks but also provide a novel path based on multi-dimensional perspectives for transport researchers and practitioners to decipher the route choice behaviours of passengers. A systematic review of route choice modelling for a URT system with seamless transfer is needed as extant review papers on route choice analysis do not cover recent advances such as restoring trajectories of individual passengers from transport big data and telecommunication information [3–5]. By tracing the development of modelling route choices in a URT network, we conduct a systematic review that sheds light on the current state of research in this domain. Our review provides a comprehensive understanding of route choices in URT networks and outlines future research directions.

The remainder of the paper is organised as follows. First, the review criteria and article classification are explained. Second, the approaches in the aforementioned two stages are introduced. Third, recent progress is reviewed from the perspective of improving existing models. Future research directions are then envisioned for better understanding the route choice behaviours of URT passengers. Finally, conclusions are presented.

## 2 Classification and review method

We adopt a systematic review methodology to extract pertinent papers from an extensive body of work. The detailed procedure is illustrated in Fig. 1. Initially, we scrutinise classic review papers on route choice modelling. Adopting the snowballing approach [6], we begin with a tentative set of papers sourced from Google Scholar. Three highly cited review papers written by Bovy [3], Prashker & Bekhor [7], and Prato [4] are included. Subsequently, we adopt forward snowballing and Boolean operations to refine the scope for manual filtering. In augmenting our data set, we incorporate the CNKI[1] data set to capture journals and thesis-type literature from mainland China, where there are a number of rapidly-developing URT systems and relative up-to-date research works. Notably, highly reputed transportation journals are included to uphold the quality of the review process. We implement filtering criteria, focusing on titles, abstracts, and even full papers, to exclude articles outside the scope. Adhering to these criteria, a total of 90 articles are retained for the subsequent analysis of route choices in URT networks.

Following Prato [4], we introduce the development of route choice modelling for a URT network. The research methods for the route choice problem in a URT network

[1] China National Knowledge Infrastructure (CNKI) owns numerous journals, doctoral and master's dissertations, proceedings, newspapers, yearbooks, statistical yearbooks, e-books, patents, and standards for researchers seeking Chinese academic materials.

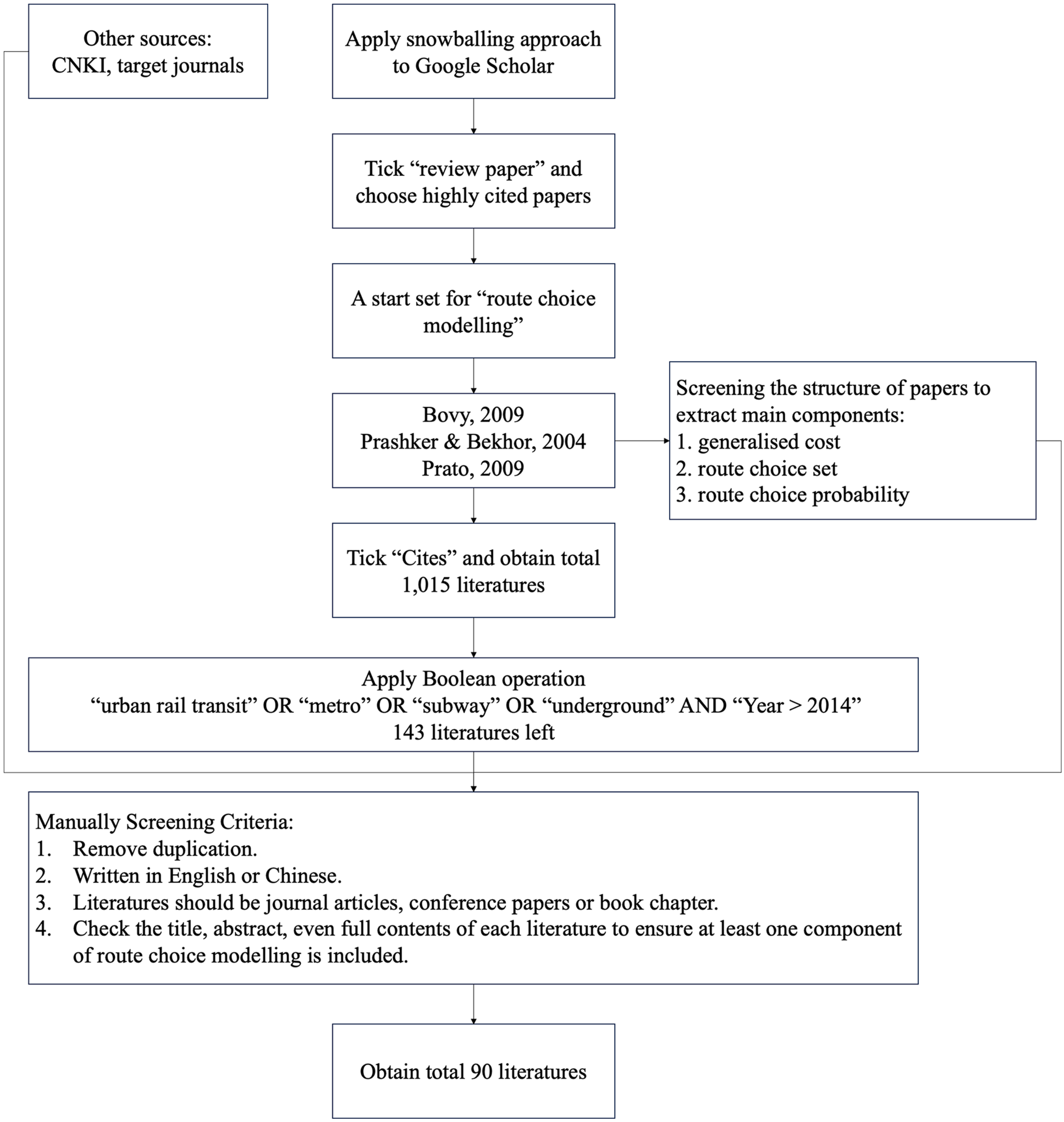


**Fig. 1** Procedure of literature selection

have undergone two main stages over the past few decades, which are adopting *probabilistic route choice modelling* and *retrospective route choice modelling*. Probabilistic route choice modelling assumes passengers have bounded rationality and align with the most economical route. Commonly adopted criteria include the shortest travel distance or the minimum travel time. However, these assumptions ignore heterogeneities in the perception of the route cost among independent passengers, and operational constraints such as train schedules, which might generate bias. Moreover, these indicators change dynamically through passenger–passenger and passenger–environment interactions. Retrospective route choice modelling dissects the whole travel procedure of a passenger into plenty of actons, such as tapping in at the fare gate, walking to the platform and boarding on the train. Each action corresponds to a time segment. Some time segments can be determined from automated fare collection (AFC) and automatic train supervision (ATS) data. Consequently, the trajectories of individual passengers can be deduced in a retrospective logic.

Both the probabilistic route choice modelling approach frequently adopted in route choice research and the retrospective route choice modelling are illustrated. The pros and cons of each approach are discussed, providing guidance for researchers and practitioners to select appropriate approaches that in line with the characteristics of the urban rail transit network and data availability. Furthermore, our review work also depicts an analysis framework with multi-dimensional network that incorporates the physical topology of the URT network, train schedules, surveillance video at stations and in trains, telecommunication data of passengers, and social relationships among passengers. It provides a novel path for further research on route choices in URT networks as well as for transport researchers and practitioners to decipher the travel behaviours of passengers. We believe our work will inspire transportation researchers and agencies working on route choice estimation for URT networks.

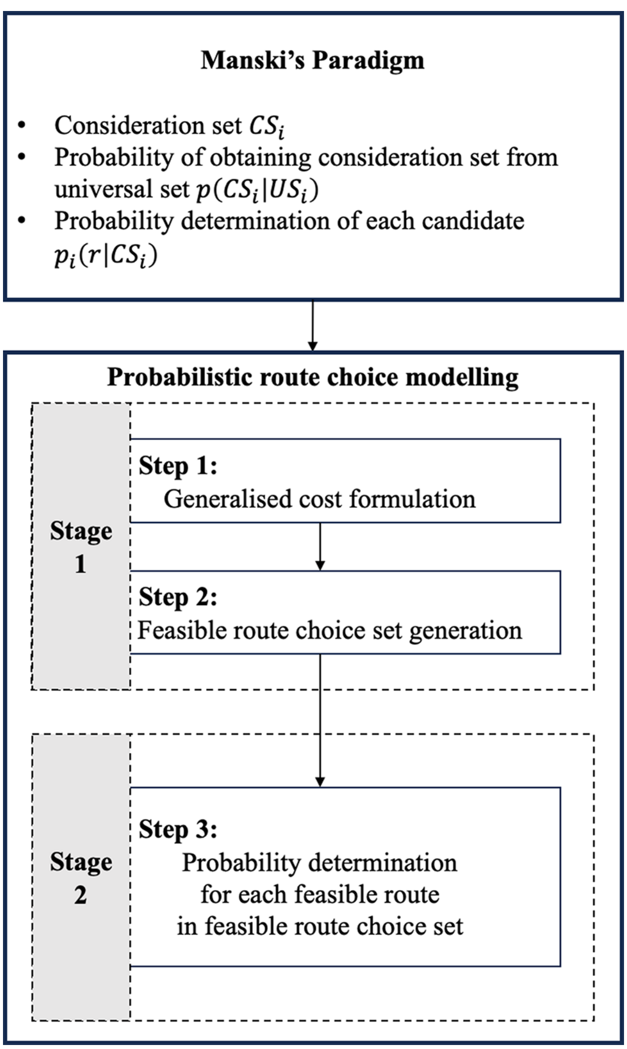


**Fig. 2** Framework of probabilistic route choice modelling

## 3 Probabilistic modelling of route choices in a URT network

Research on route choice analysis in the transport area can be traced back to an economic choice theory referred to as Manski's paradigm [8], which describes the probability of an actor *i* choosing alternative *r* from consideration set $CS_i$. In the transport research area, this theory is expressed as

$$p_i(r|US_i) = \sum_{CS_i \in PS_i} p_i(r|CS_i)\, p(CS_i|US_i), \tag{1}$$

where $p_i(r|US_i)$ is the probability that passenger *i* chooses route *r* from the universal route choice set $US_i$ from all alternatives available to *i*; $p_i(r|CS_i)$ is the conditional probability that passenger *i* chooses route *r* from his/her consideration route choice set $CS_i$, which is a subset of $US_i$; and $p(CS_i|US_i)$ is the probability that $CS_i$ is the consideration set of passenger *i* based on $US_i$. Manski's paradigm emphasises the generation of finite feasible routes to form consideration choice set $CS_i$ and the determination of the probability of each feasible route within the consideration set $CS_i$.

Route choice modelling is thus decomposed into two conceptual stages, namely feasible route choice set generation and probability determination for each feasible route in the consideration set. Generating a feasible route choice set $CS_i$ involves a generalised cost formulation and feasible route filtering. The feasible route choice set is assumed to be generic for all passengers to reduce the complexity of calculation. The probability of each feasible route $p_i(r|US_i)$ is then determined adopting a logit-based model. This process makes up a complete probabilistic route choice modelling framework for the URT network (Fig. 2).

### 3.1 Generalised cost formulation

In transport economics, the generalised cost is the weighted sum of the monetary cost and non-monetary cost of a journey [9]. Assuming all passengers are rational, they tend to minimise their moving effort. The generalised cost represents this effort incurred from one location to another along a specific route. The generalised cost is usually composed of components that reflect the level of service, for example, monetary cost, travel time, number of transfers, and crowding level. Given that the monetary cost remains consistent across different routes for the same OD pair in the URT system, it is excluded from the generalised cost formulation. Travel time can be further broken down into in-vehicle time, walking time, and waiting times at origin and transfer stations [10–12]. Psychological experiments have revealed that passengers are more sensitive to time spent outside vehicles [13]. Amplification factors have been widely introduced in calculating the perceived transfer time to transform the negative feelings of passengers into time elements. The number of transfers during a journey has also been shown to negatively affect the route choices of passengers [14, 15]. Furthermore, the role of latent factors in route choice behaviour is also increasingly recognised. For instance, the perceptions of crowding level might affect the route choice [16]. As the travel demand increases, passengers in URT networks confront crowding disutility, especially in megacities. The level of crowding both on the platform and in the train might result in subjective or passive alterations in the route choices of passengers. Shimamoto et al. [17] introduced boarding resistance to the generalised cost formulation to account for the impact of crowding. In addition, passengers' trip, socioeconomic and demographic characteristics, such as the travel purpose and income level, have also been included in the generalised cost formulation [18, 19].

### 3.2 Generation of a set of feasible route choices

The generation of a choice set is a process of generating alternatives for each decision maker, which could be an individual or a group of homogeneous passengers. Only a subset of the universal set is accessible in real situations. Moreover, constraints such as physical connectivity and dynamic train schedules narrow the set of passengers' options.

The size of the feasible route choice set affects the performance of route choice estimation [3]. In the case of a URT network, the generation of a feasible route choice set can be based on different methods, among which the shortest-path method is widely used. However, a validation experiment revealed that passengers did not always prefer the shortest path and detoured by approximately 13% for convenience [20]. To relax this rigorous assumption, the K shortest path method has been adopted to narrow the route choice set on account of its simplicity and acceptable precision, though it has also been criticised for purely relying on travel distance [21, 22]. Modified route searching methods based on additional filtering criteria such as time-saving, habit and level of service have been proposed to mitigate the sole-criterion problem [23, 24]. Other studies apply travel cost thresholds on top of the narrowed-down choice set to further refine the composition of the feasible route choice set [25–27]. Both absolute threshold and relative threshold are commonly used route filtering indicators that are determined through on-site investigation [28, 29]. Moreover, the availability of train service ought to be considered as a new constraint for the generation of a feasible route choice set [30].

### 3.3 Determination of the route choice probability

The determination of the route choice probability in a URT network primarily adopts logit models or probit models. Among logit curves, the multinomial logit model is the most well-known owing to its simple and operable characteristics. Nonetheless, the multinomial logit model neglects the issue of overlapping that arises for the route choice problem. Modified methods including the C-logit model and path size logit model have been shown to efficiently overcome the overlapping issue. Detailed expressions and performance comparisons of these models have been presented [7, 24, 31]. In analogy, probit models are also efficient in solving the route overlapping problem via introducing multivariate distribution [32, 33].

The above research on route choice analysis in a URT network has shown the good performance achieved in estimating the route choices of individual passengers. However, the trade-off between the computational efficiency and accuracy in practice ought to be carefully measured. A route probability determination method based on normal distribution is widely adopted in practical application [34]. The specific utility function of a feasible route, which is related to the generalised cost function, is calculated. The generalised travel cost difference between the shortest route and a feasible route from an origin to a destination is then obtained. The standard deviation of the normal distribution is considered a constant value for all OD pairs, derived from travel surveys of URT passengers. The proportion coefficient is also obtained in a similar way. Passengers wish to arrive at their destination with less travel time and travel cost to minimise their negative utilities. A larger utility value corresponds to a greater possibility of selection.

### 3.4 Limitations in probabilistic route choice modelling

Probabilistic route choice modelling does not perfectly adapt to the schedule-based behaviour of the URT system [35]. Taking the feasible route choice set generation as an example, diverse train operation schedules are released for a variety of passengers, which increases the complexity of generating a feasible route choice set. In addition, probabilistic route choice modelling faces challenges in accurately deducing route choices for OD pairs with minor passenger flow and abnormal travel, such as travelling backwards, group travelling, and an unreasonably long travel time. For the former problem, the route choice estimation for OD pairs with minor travel demand reveals a strong individual preference, which introduces huge uncertainty to the probabilistic route choice modelling. For the latter problem, improvements of the probabilistic route choice modelling method are one sided and locally effective. On the one hand, establishing comprehensive adapted probabilistic route choice modelling that considers all related attributes seems impossible. On the other hand, the reliability of probabilistic route choice modelling is contingent on the quality of prior knowledge. However, the nature and extent of abnormalities in travel behaviour can vary across different scenarios. Consequently, the reproducibility and transferability of probabilistic route choice modelling is the subject of controversy.

In view of the above inherent defects of probabilistic route choice modelling, a new modelling framework from the perspective of the individual passenger that considers both personal characteristics and a dynamic network service without reliance on expert experience is required for research on refined route choice deduction under diverse scenarios.

## 4 Retrospective route choice modelling for a URT network

In contrast to a flexible road traffic service, passengers in a URT network access the transportation service by adhering to train running timetables. Thus, the route choice of a passenger within the URT network is subject to not only the physical network topology but also the train operation information network, especially during peak hours, which may be a more important issue [36]. Estimating the train choice of the passenger is the key to analysing passenger flows on a schedule-based rail transit network for the following three reasons.

(1) As the premise for subsequent refined passenger flow analysis, the route choice of a passenger is essentially a sequence of train choices and hence requires the inference of the choice for each sequential train.
(2) The estimation of train choices can offer useful insights regarding individual passengers' spatial-temporal status and behavioural explanations. For example, probabilistic route choice modelling cannot explain the travel behaviour of a failure to board due to overcrowding especially during peak hours.
(3) A better understanding of passengers' train choices can be beneficial to the assessment and enhancement of URT services. For instance, rail transit agencies can collect the responses of passengers and compare the train selection before and after improvements to provide adaptive timetables [37].

The probabilistic route choice modelling method faces challenges in estimating passenger route choices under dynamic train operation constraints. These challenges include difficulty in generating a feasible route choice set that considers all possible train itineraries, handling increased uncertainty during a journey, dealing with increased computational complexity in determining train selection probability, and making complex validity judgments for feasible choice sets and probabilities of each feasible alternative. To overcome these difficulties and cater to booming big data technology, retrospective route choice modelling based on the travel time chaining method and transport big data has been established for train choice deduction and route choice estimation.

### 4.1 Retrospective logic based on the travel time chaining method

The proficient use of AFC data [38] has made spatial and temporal information at the origin and destination available. The total travel time of the individual passenger is made up of deterministic terms that are the same for all individuals and variable terms that need to be studied intensively. The travel time chain of each passenger is decomposed into time segments via in-station activities (Fig. 3), including walking to the platform at the origin station, waiting for trains on platforms, traveling on trains, walking and waiting at transfer stations, and walking out of the destination station from the platform. Among these segments, the travel time on train is the deterministic term. An accurate value is obtained from train running timetables once the train selection of passengers is available. A train choice inference method based on AFC data and the transfer network has been proposed using the travel time chain [39]. A set of routes were generated referring to two assumptions about the travel preferences of passengers, namely (1) the minimum waiting time at the origin and lost time at the destination and (2) the minimum frequency of transfer. Different routes that satisfy these two assumptions with the same transfer times are assigned equal probabilities. The effects of the waiting and walking times at different transfer stations have not been discussed in depth. According to the logic of travel time chaining [40], Bayesian probabilistic method was introduced for the calculation of the probability of candidate train selections on routes without transfer [41]. On the basis of that study, methods of estimating the egress and access time were improved and the probability of a train choice on routes with one transfer was confirmed [42]. Similar research has been conducted adopting Bayesian approach [43–45]. Sometimes, trains do not strictly stick to running schedules, and transport studies have introduced ATS data to account for feasible route filtering and route choice probability calibration [46]. ATS data store abundant and latest train operation information that is used in retrospective route choice modelling. In accordance with the passengers' behaviour of not hesitating at the destination station, the travel trajectory of an individual passenger can be backtracked from the destination step by step. The egress time can be estimated from the walking speed and distance,

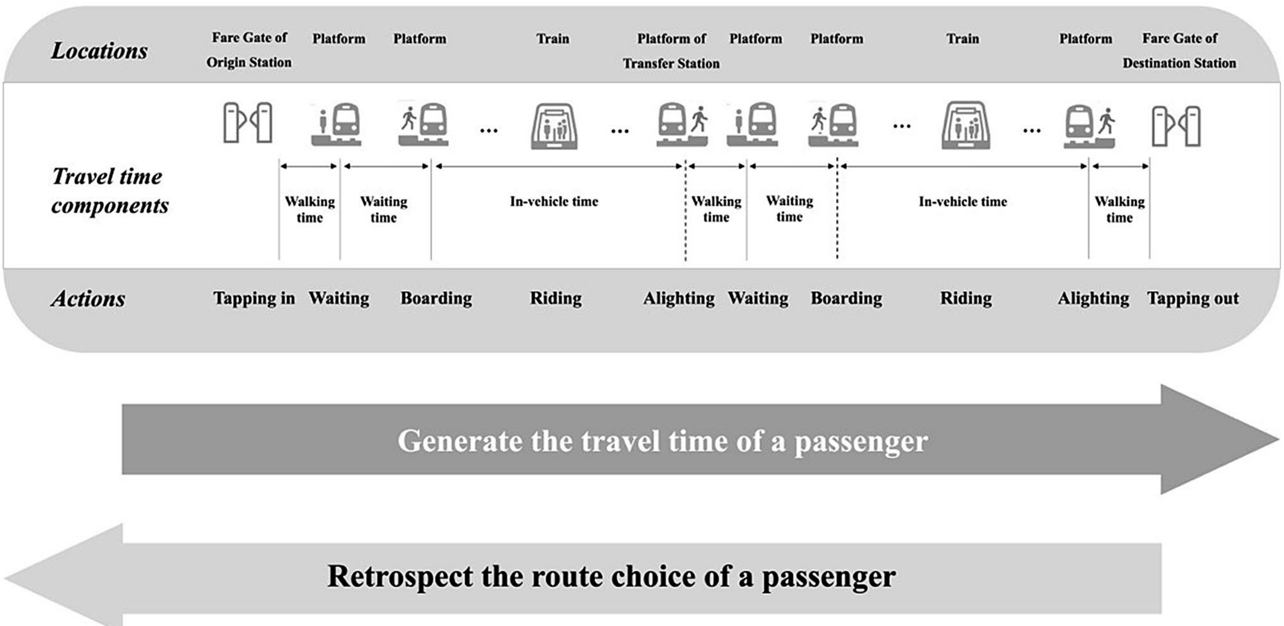


**Fig. 3** Retrospective logic based on travel time chaining

which are known values obtained from a walking speed survey and the layout information of each station. The last boarding train can then be estimated by comparing the estimated arrival time of the individual passenger and train arrival time in the ATS data. Once the train is confirmed, the in-vehicle time can be extracted from train running records of the ATS data. The behaviour of the individual passenger to walk consistently supports the estimation of the walking time at the transfer and origin stations. For travel without transfer, the waiting time at the origin station can be directly acquired. However, for journeys involving transfers, the waiting time at the transfer station is assumed to be identical to that of passengers entering the station without any transfer. All variable terms are transformed into deterministic terms reasonably and multi-source data are fully mined, integrated, and utilised.

The complete journey of an individual passenger can be clearly displayed by a spatial temporal graph (Fig. 4). In addition to the feasible route filtering via ATS data, the probability of each candidate route is calculated clearly according to the determination of the success to board, which distinguishes the probabilities of feasible routes with the same numbers of transfers. Considering unforeseen operational events such as a delay or malfunction, a retrospective route choice modelling method without a train running timetable has been generated by inferring train arrival times from passenger volume spikes at exit gates [47].

### 4.2 Retrospective logic based on the travel time chaining method and additional constraints

Building upon the work reviewed in Sect. 4.1, given the substantial volume of daily AFC data, the array of feasible train combination choices is extensive, which increases the redundancy of restoring a complete trajectory of passengers. To improve the efficiency of generating feasible train combinations without discarding the spatial and temporal details of transport big data, travel time thresholds are determined for each OD pair in different time spans based on a feasible route set and all feasible train combinations from ATS data in advance [48]. The number of transfers, walking features, and train numbers serve as additional constraints for the restoration of more accurate trajectories. Case studies indicate that 95% of route choices can be accurately estimated in common situations. Consequently, the integration of additional data sources as supplementary constraints aids transport researchers and agencies to gain a profound understanding of passengers' route choice behaviours.

### 4.3 Limitations of retrospective route choice modelling

In contrast to probabilistic route choice modelling, retrospective route choice modelling incorporates dynamic train running information and thus better matches URT operation and suits big data mining. Nonetheless, retrospective route choice modelling must obey rigorous assumptions, such as the walking consistency of the same passenger, the different waiting times of different passengers at the same location, no hesitation after alighting, and the passenger boarding on the first arriving train.

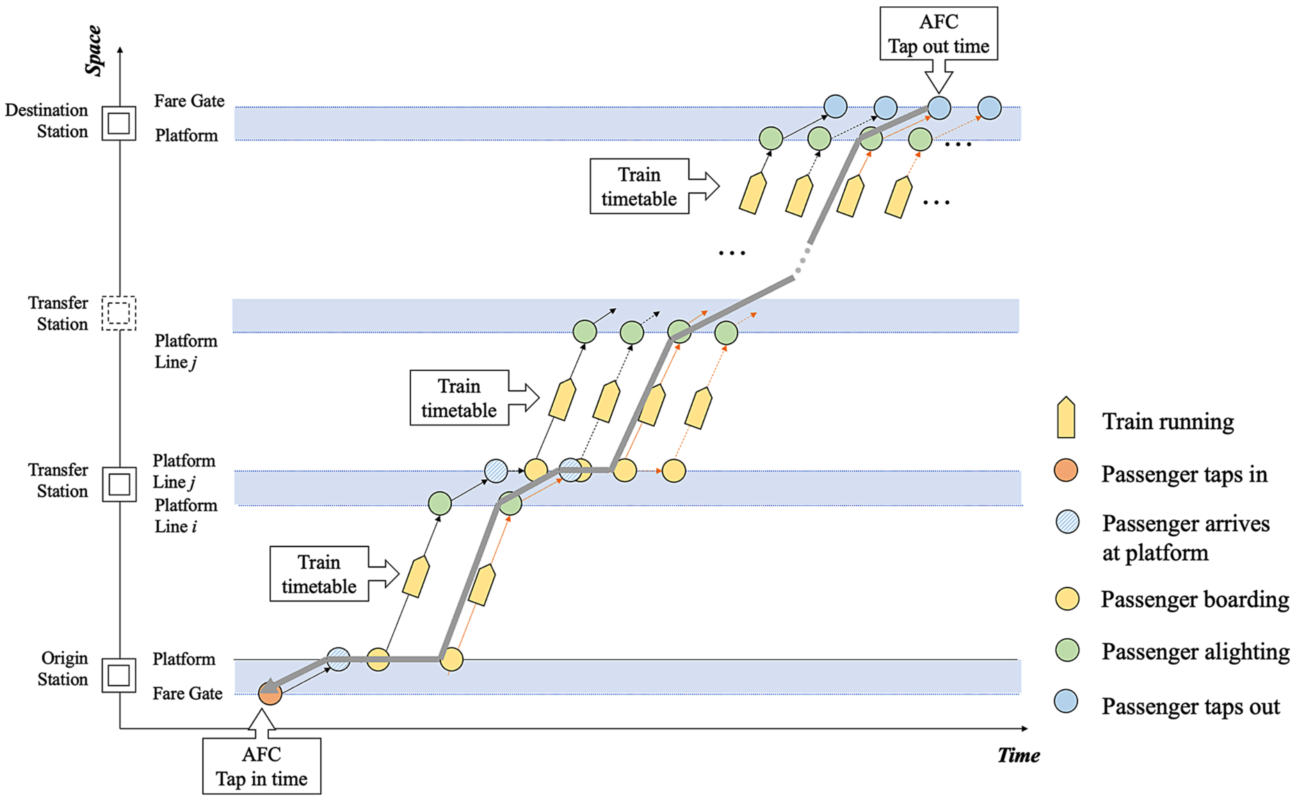


**Fig. 4** Illustration of the process of train choice deduction

Although trajectory backtracking at the individual level retains as much characteristic information as possible in contrast with probabilistic route choice modelling, some important decisive attributes are still missed, which might result in erroneous outcomes. For instance, some passengers might stay within stations for leisure purposes if there are enticing stores such as grocery stores and restaurants within toll zones [49]. These travel records are marked as abnormal in both probabilistic and retrospective route choice modelling, yet they are meaningful especially to future transit-oriented development. Additional information ought to be incorporated for comprehensive route choice and travel behaviour analysis.

## 5 Recent progress in route choice modelling for a URT network

In real URT operation, route choice methods should satisfy benchmarks of operability and accuracy. From the perspective of operability, as probabilistic route choice modelling has the prominent advantage of easy operation with acceptable accuracy, it dominates route choice deduction for URT networks in most cities. Retrospective route choice modelling preserves ample features of travel records, which improves the accuracy of route choice deduction. However, it increases the computational complexity and processing time while retaining features of data. To avoid redundancy and take advantages of the two methods, retrospective logic is retained as an auxiliary part of probabilistic route choice modelling. From the perspective of accuracy, the calibration and validation are meaningful processes by which to maintain the robustness of models. Calibration and validation undergo the upgrade of data collection from limited-scale manual investigation to abundant spatiotemporal transport big data, which enables rolling calibration and validation with a data-driven approach, and thus strengthens the robustness of route choice deduction.

According to the pros and cons of the above two existing methods, route choice modelling can be enhanced by improving present route choice models and incorporating additional information networks for comprehensive route choice analysis.

### 5.1 Improvement of existing models

Calibration and validation are main measures used in qualifying the feasibility of existing models. The former involves statistical or heuristic method to endow parameters of models with appropriate values. The latter implements models into scenarios with different locales or time spans to test the adaptability of models. The performance of models relies heavily on the quality of parameter calibration and model validation. Given that probabilistic route choice modelling dominates mainstream research in route choice estimation, this paper focuses on the calibration and validation of these models as the primary means of improving existing models. In the meantime, improvements to the retrospective route choice modelling are considered as a secondary aspect.

The overarching goal of model calibration and validation is to ensure the reasonableness of the model and the reproducibility of currently observed travel patterns. The integrity of influencing attributes, the quality of parameter calibration, and the performance validation of the probabilistic formula are the three decisive criteria used in measuring the overall performance of probabilistic route choice modelling.

Section 3.1 partially discussed the incorporation of attributes. However, researchers adopt the number of transfers to measure the disutility of transfer, which assumes all transfers during a journey have an equally negative impact on the passenger. Some researchers debate the applicability of this assumption, and interchange-oriented research has thus been conducted to analyse the effects of the interchange environment and order of interchange on passengers' route choices [50]. In addition, guidance information is likely to affect the route choice of an individual passenger. Passengers who are not familiar with a URT network choose the visually shortest path. In contrast to probabilistic route choice modelling based on the minimum travel cost, the deviation of a map from the real network topology leads to different route probabilities. Variables related to visual illusion on a network topology have been introduced to enrich previous probabilistic route choice models [51]. Distortions between a map and network topology have been shown to affect the decisions of passengers [52]. Analyses on other specific characteristics, including historical experience, information guidance, and travelling backwards, also have been presented [53–55].

In parameter calibration, adopting an appropriate methodology alongside transportation big data facilitates a timely update to explore the most suitable parameters at the present moment [56–59]. As influencing factors are dynamic, calibration ought to be conducted in a rolling manner to ensure the ongoing robustness of existing models. AFC data contain rich spatial and temporal travel information, which is widely used in parameter calibration of the probabilistic model and enriching the route choice set. A genetic-algorithm-based calibration method with nonparametric statistics techniques has been proposed to cyclically calibrate parameters [60]. Analogously, a data-driven automated calibration method based on the particle swarm algorithm and smart card data has been established [56]. Researchers have initially focused on improving the parameter calibration of models [61, 62]. The important role of feasible choice set generation has been neglected, and route omissions also have an impressive negative impact on the precision of probabilistic route choice modelling. The Rodriguez-Laio clustering method has been introduced to calibrate the feasible route choice set. On the basis of travel time clusters, new routes have been discovered through the automated update of the route choice set [63].

For retrospective route choice modelling, a synchronous clustering algorithm has been applied to AFC data trimmed by the train operation plan to reveal the travel behaviours of passengers [64]. Compared with the results of probabilistic route choice modelling, the results of the clustering method reveal that the probability of a feasible route varies across different periods of the day (i.e., morning-peak, evening-peak, and off-peak periods). From the perspective of passengers, the perception of crowding affects the train choice and route choice subjectively. The perceived crowding disutility generates an additional cost in the utility function. In addition, the trainload constraint results in passengers unwillingly remaining on a platform, which introduces uncertainty in route choice deduction. Past retrospective route choice research defaults to a URT system under normal travel demand, yet during peak hours, it can be impossible to board the first arriving train owing to the imbalance between the insufficient capacity and high travel demand. Hence, the estimation of denied boarding is crucial in modelling the real operational situation.

The fail-to-board sector has been explicitly modelled for the overcrowding scenario during peak hours [65]. Analogously, considering the trainload constraint, a left-behind model based on classical maximum likelihood calibration has been proposed [66] and added to retrospective route choice modelling [67]. Observations of denied boarding are required for calibrating the left-behind probability. To avoid the high cost of observations, a data-driven approach with the generalised EM algorithm has been developed for solving the mixture distribution framework [68]. In addition, simulation-based optimisations are effective for parameter calibration in route choice deduction [69, 70]. For a large and complex network, the simulation framework is adept at recognising rapid changes in the physical topology and the dynamic preferences of passengers.

Few studies have addressed the validation process in transport research. As conducting experiments in transport research are costly, the introduction of observation data is a popular approach for model validation. A review [71] of 226 transport research articles on discrete choice modelling between 2014 and 2018 showed that an over-reliance on goodness-of-fit measures rather than validation performance is unwise. Even though models show a high fitness in statistical examination, they might be unpersuasive in terms of guiding policy in real application. Only 18% papers incorporated a validation process, and most of these processes were internal validations owing to the limited observation data available. Transportation researchers and agencies should adopt validation processes to ensure the effectiveness of their models.

### 5.2 New data sources

The calibration and validation of existing models rely heavily on the quality and quantity of observation data. Inspired by the application of AFC and ATS data, the accuracy of route choice modelling can be improved using appropriate models and abundant transport data. Researchers should thus broaden their horizons beyond the rail transit network and the model itself.

Currently, additional networks, such as surveillance and telecommunication networks, are being explored in route choice analysis. Setting Wi-Fi collectors at stations enables the frequent capture of information on the interaction between the Wi-Fi base station and individual passengers. The intermediate locations of passengers are likely to be inferred from the latitude and longitude of a base station. This technology tends to be extended to analyse the waiting time and transfer time at stations, revealing the microscopic trajectory within the station and the macroscopic trajectory within the URT network [72, 73]. Addressing incompleteness, the overlapping and redundancy problems of raw Wi-Fi data have been explored [74, 75], providing a premise for using Wi-Fi data in travel behaviour analysis. However, information on passengers who have turned off the Wi-Fi function of their devices is not available. As an accompanying device, the cell phone is adept at tracking human movement by generating dense and sequential spatiotemporal data for the travel trajectory restoration of individual passengers. Comparing with a manual survey that includes daily and weekly travel behaviours, cell phone data tends to fill gaps in research on monthly and seasonal regularity and irregularity in human travel behaviour. In general, the conceptual theory of trajectory reconstruction via cell phone data is simple whereas in real application, noise data cleaning encounters technical barriers, including data redundancy, data error, and critical data loss. Researchers attempt to extract passengers' trajectories from subway base station cell phone data within a URT system and further improve the accuracy of trajectory by introducing ground base station cell phone data when encountering data omissions. There are other relevant areas of research on trajectory reconstruction in URT networks [76–79]. Furthermore, in the transportation research area, video data can be applied for the determination of passenger flow statuses and the early warning of massive passenger flows [80–82]. High-definition camera images are applied for coarse passenger flow detection in collaboration with Wi-Fi probe data for passenger flow forecasting adopting a convolutional neural network [83]. However, matching the same individual accurately across different surveillance cameras is a technical bottleneck in backtracking route choices. Current person re-ID methods generate multiple top candidates whereas matching an individual passenger across multiple cameras requires the identification of the best-fit candidate, which places higher requirements on the accuracy of the matching algorithm [84]. In addition, researchers integrate Global Positioning System (GPS) data and survey data for trajectory and time matching rather than utility maximisation [85]. Integrating GPS data as a reference for trajectory restoration has been confirmed to provide better performance than using purely cell phone data [86]. There is no doubt that integrating multi-source data increases the performance of route choice deduction, but the technical deficiency of data fusion hinders the progress of research. Recently, advanced data integration techniques such as those of the Internet of Things (IoT) and blockchains have shown promise in overcoming data barriers and unleashing the potential of data integration. Data integration heralds a new era for URT route choice analysis.

### 5.3 Research gap in recent progress in route choice modelling

According to the above review of recent progress in URT route choice modelling, processes of calibration and validation are seriously underestimated. Although research on model calibration has attracted the attention of transport researchers in recent years, the contributions are scattered and insufficiently convincing. The problem of data sparsity motivates researchers to broaden their horizons and use different data sources. With the development of communication technology, advanced information such as Wi-Fi data, cell phone data, and video data has proven effective in addressing the black-box problem and enhancing refined route choice analysis of a URT network. As each data source has pros and cons, integrating multiple data sources tends to provide better performance in route choice deduction. Breaking the barriers among data sources deserves specific attention. Furthermore, integrating data sources will help transportation researchers and practitioners to mine the route choice behaviours of passengers from a multidimensional perspective. For example, presently available supportive data do not reveal whether the comfort consideration or the perceived travel time is the more important indicator of the route choice [87]. Additional networks such as social networks, complete information and communication technology networks, and surveillance networks tend to reveal the comprehensive route choice behaviours of passengers.

## 6 Future prospects of route choice modelling for a URT network

The development from probabilistic route choice modelling to retrospective route choice modelling reveals the importance of incorporating multi-dimensional networks. In probabilistic route choice modelling, the route choice analysis mainly relies on the physical topology of the URT network along with the prior knowledge of individual characteristics obtained from on-site investigations or AFC data. In retrospective route choice modelling, the train running network provides abundant spatial and temporal information of trains within the overall URT network, which indirectly narrows the set of feasible route choices and reveals hidden travel behaviours of passengers.

Moreover, attempts at integrating video data from surveillance networks or cell phone data, Wi-Fi probe data, or Bluetooth data that belong to the information and communication technology network contribute to the mining of the latent attributes of route choice. Therefore, discovering related information networks, integrating these isolated networks into a multi-dimensional network, and interpreting the activity of an individual passenger will help researchers comprehend the passenger's route choice and make precise deductions. Figure 5 presents two scenarios of gradually incorporating additional information networks. Scenario 1 shows that the number of possible routes reduces with the enrichment of data sources. Sometimes, additional information networks not only narrow the set of possible routes but also reveal omitted feasible routes. For example, the cell phone data presents offset positions in Scenario 2. To further confirm whether the cell phone records are effective, the social network and point of interest (POI) assist researchers to understand such "abnormal choices". In an ideal situation, all travel is reasonable under abundant information, and the tendencies of passengers are inferable without the limitations of scale and time span.

Last but not least, recent research has presented the merging of graph theory with machine learning or deep learning methods as a possible solution to URT route choice problems. A knowledge graph (KG) provides an integration platform for multi-source information networks and is adept at storing relations and searching for indirect relations between different databases. Normally, the graph convolutional network of the deep learning method is applied in extracting and learning features of a knowledge graph, with the network adapting to synchronously reveal the attributes of route choices via a data-driven approach. Furthermore, this framework can be applied to mobility prediction [88], short-term passenger flow estimation [89, 90], and emergency prediction [91] in a URT network.

## 7 Summary

This paper reviewed and discussed progress and prospects in the research area of URT route choice modelling.

The developments of *probabilistic route choice modelling* were introduced and discussed. Two critical stages of probabilistic route choice modelling, namely feasible route choice set generation and probability determination in the URT network, were interpreted in detail. The feasibility of a route relies on not only the connectivity of physical routes but also the availability of trains. *Retrospective route choice modelling* based on travel time chaining was then discussed for the trajectory restoration of an individual passenger. The route choice problem transforms into a series of train selections adopting AFC data and ATS data.

*Recent progress* in route choice modelling was reviewed from two perspectives, namely the improvement of existing models via rolling calibration and validation and the incorporation of data sources. The review of recent progress highlighted shortcomings. First, although research on calibration has drawn the attention of researchers in recent years, most researchers focused only on a part of the modelling framework, such as the calibration of parameters or distribution formats in the probability deterministic procedure. The calibration of feasible route choices, which greatly affects the accuracy of results, has been seriously undervalued. Second, the connection between calibration and validation has been overlooked. Validation plays an important role in ensuring the robustness of models yet few studies have emphasised the importance of validation. The development of data science is making possible data-driven rolling calibration and validation that improves the accuracy and extends the lifecycle of route choice models. Third, advanced communication and information technology enable refined route choice deduction without additional expenditure for specific transport research. The incorporation of these technologies is capable of compensating the limitation of AFC data by providing intermediate information between origins and destinations. However, technical barriers of data integration have not yet been overcome.

Inspired by the developments of two genres of route choice modelling and recent progress, integrating data sources to construct a multi-dimensional network is expected to reveal the behaviours of individual passengers thoroughly. The latest research introduces POI information from the field of sociology and achieves better performance into travel behaviour analysis. We believe with the integration of *interdisciplinary information networks*, all travel by passengers will be reasonable and inferable, which will enable refined transport operations.

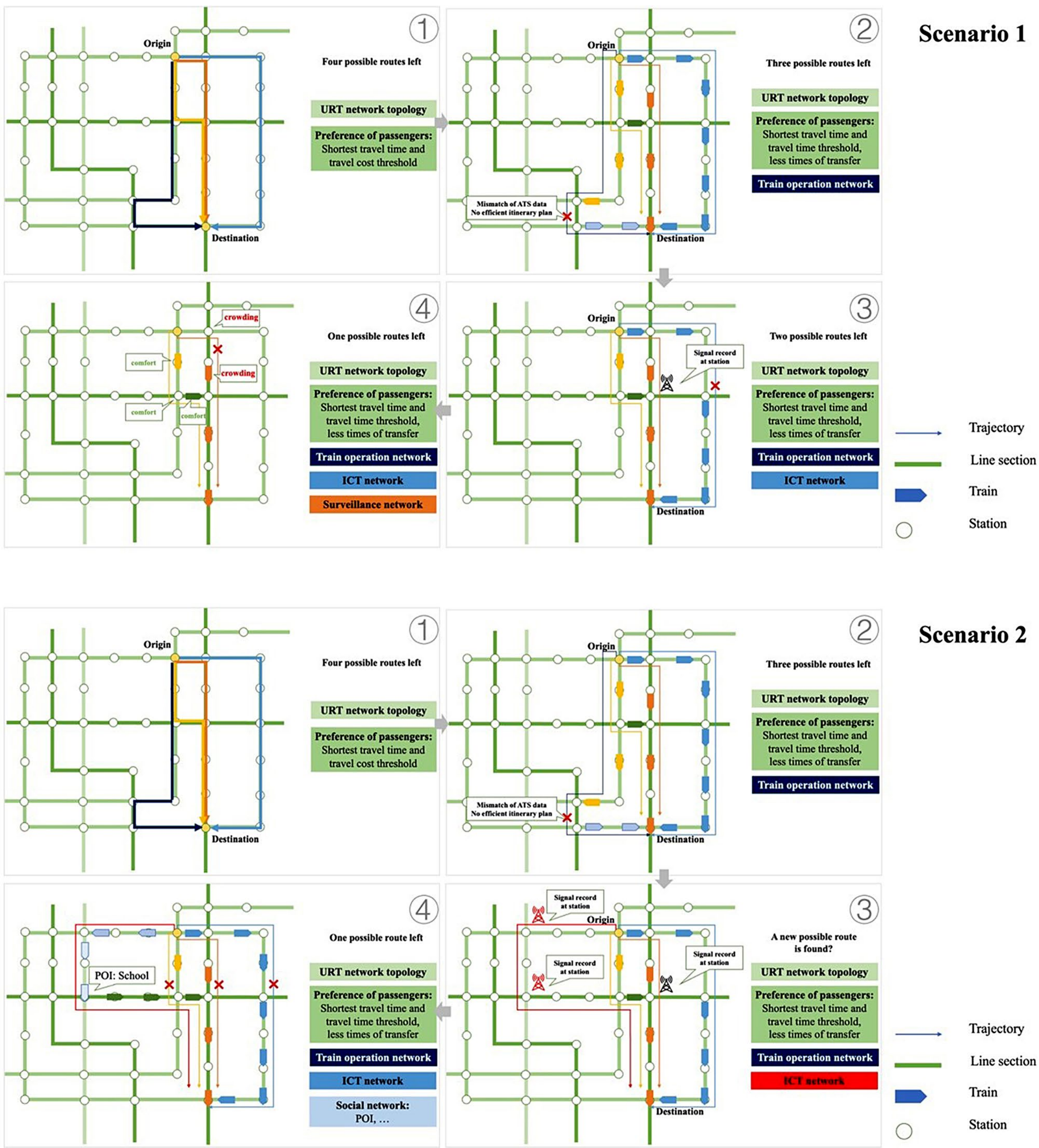


**Fig. 5** Logic of multi-dimensional network-based route choice inference

**Abbreviations**

| | |
|---|---|
| URT | Urban rail transit |
| OD pair | Origin-destination pair |
| AFC | Automatic fare collection |
| ATS | Automatic train supervision |
| POI | Point of interest |
| KG | Knowledge graph |
| IoT | Internet of thing |
| GPS | Global positioning system |

**Acknowledgements**
Not applicable.

**Author contributions**
YT: Conceptualisation, Literature search and assessment, Writing – original draft. WZ: Conceptualisation, Literature search and assessment, Writing – review and editing. FS: Writing – review and editing. All authors read and approved the final manuscript.

**Funding**
This work was supported by the [the National Natural Science Foundation of China] under Grant [number 72071147]; [the Fundamental Research Funds for the Central Universities of China] under Grant [number 22120220628] and [the Research Program of Nanchang Metro Co., Ltd.] under Grant [number 2021HGKYC005].

**Data availability**
Not applicable.

## Declarations

**Competing interests**
The authors report there are no competing interests to declare.

Received: 6 February 2024 / Accepted: 9 September 2024
Published online: 18 September 2024

**References**

1. Han, B. M., Xi, Z., Sun, Y. J., et al. (2023). Statistical Analysis of Urban Rail Transit Operation in the World in 2022: A review. *Urban Rapid Rail Transit*, *36*, 1–8.
2. Zhou, W., Han, B., & Feng, C. (2017). A review of passenger flow assignment model and algorithm for urban rail transit network. *Systems Engineering-Theory & Practice*, *37*, 440–451.
3. Bovy, P. H. L. On modelling route choice sets in transportation networks: A synthesis. Transp Rev 29:43–68. https://doi.org/10.1080/01441640802078673
4. Prato, C. G. (2009) Route choice modelling: Past, present and future research directions. Journal of Choice Modelling 2:65–100. https://doi.org/10.1016/S1755-5345(13)70005-8
5. Liu, Bunker, Y., & Ferreira, J. (2009). L (2010) Transit users' route-choice modelling in transit assignment: A review. Transp Rev 30:753–769. https://doi.org/10.1080/01441641003744261.
6. Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: ACM International Conference Proceeding Series. Association for Computing Machinery.
7. Prashker, J. N., & Bekhor, S. (2004). Route choice models used in the stochastic user equilibrium problem: A review. *Transp Rev*, *24*, 437–463. https://doi.org/10.1080/0144164042000181707
8. Manski, C. F. (1977). The structure of random utility models. *Theory Decis*, *8*, 229–254. https://doi.org/10.1007/BF00133443
9. Parson, P. S. (1972). Generalised costs and the estimation of movement costs and benefits in transport planning. *Highway Research Board*. Highway Research Board.
10. Jánošíkova, L., Slavík, J., & Koháni, M. (2014). Estimation of a route choice model for urban public transport using smart card data. *Transportation Planning and Technology*, *37*, 638–648. https://doi.org/10.1080/03081060.2014.935570
11. Ansari Esfeh, M., Wirasinghe, S. C., Saidi, S., & Kattan, L. (2021). Waiting time and headway modelling for urban transit systems–a critical review and proposed approach. *Transp Rev*, *41*, 141–163. https://doi.org/10.1080/01441647.2020.1806942
12. Wardman, M. (2004). Public transport values of time. *Transp Policy (Oxf)*, *11*, 363–377. https://doi.org/10.1016/j.tranpol.2004.05.001
13. Göransson, J., Andersson, H. Factors that make public transport systems attractive: a review of travel preferences and travel mode choices. European Transport Research Review 15.
14. Qian, K., Chen, Y., & Mao, B. H. (2023). (2015) Route Choice Behaviour for Urban Rail Transit Considering Transfer Time. Journal of Transportation Systems Engineering & Information Technology 15:116–121.
15. Guo, Z., & Wilson, N. H. M. (2011). Assessing the cost of transfer inconvenience in public transport systems: A case study of the London Underground. *Transp Res Part A Policy Pract*, *45*, 91–104. https://doi.org/10.1016/j.tra.2010.11.002
16. Koopmans, C., Groot, W., Warffemius, P., et al. (2013). Measuring generalised transport costs as an indicator of accessibility changes over time. *Transp Policy (Oxf)*, *29*, 154–159. https://doi.org/10.1016/j.tranpol.2013.05.005
17. Shimamoto, H., Kurauchi, F., Schmöcker, J. D., & Bell, M. G. H. (2008). Evaluating critical lines and stations considering the impact of the consequence using transit assignment model -case study of London's underground network. In: Journal of Advanced Transportation. Institute for Transportation, pp 291–310.
18. Mackie, P. J., Jara-D Õaz, S., & Fowkes, A. S. (2001). The value of travel time savings in evaluation. *Transp Res E Logist Transp Rev*, *37*, 91–106.
19. Parsons, G. R. (2017). Travel cost models. A primer on nonmarket valuation 187–233.
20. Lee, E. H., Kim, K., Kho, S. Y. (2022). Exploring for Route preferences of Subway passengers using Smart Card and Train Log Data. J Adv Transp 2022:. https://doi.org/10.1155/2022/6657486
21. Yen, J. Y. (1971). Finding the K Shortest Loopless paths in a network. *Management Science*, *17*, 712–716.
22. Eppstein, D. (1998). Finding the k Shortest paths. *SIAM Journal on Computing*, *28*, 652–673.
23. Kato, H., Kaneko, Y., Inoue, M. Comparative analysis of transit assignment: Evidence from urban railway system in the Tokyo Metropolitan Area. Transportation (Amst) 37:775–799. https://doi.org/10.1007/s11116-010-9295-8
24. Prato, C. G., & Bekhor, S. (2010). (2007) Modelling route choice behaviour: How relevant is the composition of choice set? Transp Res Rec 64–73. https://doi.org/10.3141/2003-09.
25. Van Der Hurk, E., Kroon, L., Maróti, G., & Vervest, P. (2015). Deduction of passengers' route choices from smart card data. *IEEE Transactions on Intelligent Transportation Systems*, *16*, 430–440. https://doi.org/10.1109/TITS.2014.2333583
26. Chen, H-K., & Feng, G. (2000). Heuristics for the stochastic/dynamic user-optimal route choice problem. *European Journal of Operational Research*, *126*, 13–30.
27. Zhao, J. J., Zhang, F., Tu, L. (2017). Estimation of Passenger Route Choice Pattern Using Smart Card Data for Complex Metro Systems. *IEEE Transactions on Intelligent Transportation Systems* 18:790–801. https://doi.org/10.1109/TITS.2016.2587864.
28. Schuessler, N., Zurich, E., Axhausen, K. W. Accounting for route overlap in urban and suburban route choice decisions derived from GPS observations. Arbeitsberichte Verkehrs-und Raumplanung. https://doi.org/10.3929/ethz-a-005916981
29. Kaplan, S., & Prato, C. G. (2009). (2010) Joint modelling of constrained path enumeration and path choice behaviour: A semi-compensatory approach. In: Proceedings of European Transport Conference.
30. Zhu, W., & Xu, R. H. (2016). Generating route choice sets with operation information on Metro networks. *Journal of Traffic and Transportation Engineering (English Edition)*, *3*, 243–252. https://doi.org/10.1016/j.jtte.2016.05.001
31. Cascetta, E., Russo, F., Viola, F. A., Vitetta, A. A model of route perception in urban road networks. Transportation Research Part B: Methodological 36:577–592.
32. Zhang, Y. S., Yao, E. J., Wei, H., (2017) Constrained multinomial Probit route choice modelling for passengers in large-scaled metro networks in China. Transportation Research Procedia 25:2385–2395. https://doi.org/10.1016/j.trpro.2017.05.269
33. Yai, T., Iwakura, S., Morichi, S. (2002). (1997) Multinomial Probit with Structured Covariance for Route Choice Behaviour. Transportation Research Part B: Methodological 31:195–207.
34. Xu, R. H., Luo, Q., & Gao, P. (2009). Passenger Flow distribution model and Algorithm for Urban Rail Transit Network based on multi-route choice. *Journal of the China Railway Society*, *31*, 110–114.
35. Bast, H., Delling, D., Goldberg, A. (2015). Route Planning in Transportation Networks. In: Algorithm Engineering. pp 19–80.
36. Lee, S., & Hickman, M. D. (2011). Travel Pattern Analysis Using Smart Card Data of Regular Users. In: Transportation Research Board 90th Annual Meeting.
37. Asakura, Y., Iryo, T., Nakajima, Y., & Kusakabe, T. (2012). Estimation of behavioural change of railway passengers using smart card data. *Public Transport*, *4*, 1–16. https://doi.org/10.1007/s12469-011-0050-0
38. Pelletier, M. P., Trépanier, M., Morency, C. Smart card data use in public transit: A literature review. Transp Res Part C Emerg Technol 19:557–568 https://doi.org/10.1016/j.trc.2010.12.003
39. Kusakabe, T., Iryo, T., & Asakura, Y. (2011). (2010) Estimation method for railway passengers' train choice behaviour with smart card transaction data. Transportation (Amst) 37:731–749. https://doi.org/10.1007/s11116-010-9290-0.
40. Sun, L. J., Lee, D-H., Erath, A., & Huang, X. F. (2012). Using smart card data to extract passenger's spatio-temporal density and train's trajectory of MRT system. In: Proceedings of the ACM SIGKDD International Workshop on Urban Computing. ACM, pp 142–148.
41. Zhu, Y., Koutsopoulos, H. N., & Wilson, N. H. M. (2017). A probabilistic passenger-to-train assignment Model based on automated data. *Transportation Research Part B: Methodological*, *104*, 522–542. https://doi.org/10.1016/j.trb.2017.04.012

42. Xiong, S., Li, C., Sun, X., et al. (2022). Statistical estimation in passenger-to-train assignment models based on automated data. *Applied Stochastic Models in Business and Industry*, *38*, 287–307. https://doi.org/10.1002/asmb.2660
43. Sun, L., Lu, Y., Jin, J. G., et al. (2015). An integrated bayesian approach for passenger flow assignment in Metro networks. *Transp Res Part C Emerg Technol*, *52*, 116–131. https://doi.org/10.1016/j.trc.2015.01.001
44. Hong, S. P., Min, Y. H., Park, M. J., et al. (2016). Precise estimation of connections of Metro passengers from Smart Card data. *Transportation (Amst)*, *43*, 749–769. https://doi.org/10.1007/s11116-015-9617-y
45. Leurent, F., & Xie, X. (2017). Exploiting smartcard data to estimate distributions of passengers' walking speed and distances along an urban rail transit line. In: Transportation Research Procedia. Elsevier B.V., pp 45–54.
46. Zhu, W., Wang, W., & Huang, Z. (2017). Estimating train choices of rail transit passengers with real timetable and automatic fare collection data. J Adv Transp 2017:. https://doi.org/10.1155/2017/5824051
47. Tiam-Lee, T. J., & Henriques, R. (2022). Route choice estimation in rail transit systems using smart card data: Handling vehicle schedule and walking time uncertainties. *European Transport Research Review*, *14*. https://doi.org/10.1186/s12544-022-00558-x
48. Zhu, W., Fan, W. L., Xu, C. Y., & Xu, R. H. (2022). Deduction method of passengers' train choices in Rail Transit Network based on Travel Time thresholds. *Journal of Tongji University Natural Science*, *50*, 87–95.
49. Zhou, F., & Xu, R. H. (2012). Model of passenger flow assignment for Urban rail transit based on entry and exit time constraints. *Transportation Research Record*, 57–61. https://doi.org/10.3141/2284-07
50. Cheng, Y., Ye, X. F., & Fujiyama, T. (2022). How does interchange affect passengers' route choices in urban rail transit?–a case study of the Shanghai Metro. *Transportation Letters*, *14*, 416–426. https://doi.org/10.1080/19427867.2021.1883803
51. Raveau, S., Muñoz, J. C., & de Grange, L. (2011). A topological route choice model for metro. *Transp Res Part A Policy Pract*, *45*, 138–147. https://doi.org/10.1016/j.tra.2010.12.004
52. Raveau, S., Guo, Z., Muñoz, J. C., Wilson, N. H. M. A behavioural comparison of route choice on metro networks: Time, transfers, crowding, topology and socio-demographics. Transp Res Part A Policy Pract 66:185–195 https://doi.org/10.1016/j.tra.2014.05.010
53. Bekhor, S., Albert, G. (2014) Accounting for sensation seeking in route choice behaviour with travel time information. Transp Res Part F Traffic Psychol Behav 22:39–49. https://doi.org/10.1016/j.trf.2013.10.009
54. Yu, C., Li, H. Y., Xu, X. Y., & Liu, J. (2014). (2020) Data-driven approach for solving the route choice problem with travelling backward behaviour in congested metro systems. Transp Res E Logist Transp Rev 142:. https://doi.org/10.1016/j.tre.2020.102037.
55. Tirachini, A., Hurtubia, R., Dekker, T., & Daziano, R. A. (2017). Estimation of crowding discomfort in public transport: Results from Santiago De Chile. *Transp Res Part A Policy Pract*, *103*, 311–326. https://doi.org/10.1016/J.TRA.2017.06.008
56. Tavassoli, A., Mesbah, M., & Hickman, M. (2020). Calibrating a transit assignment model using smart card data in a large-scale multi-modal transit network. *Transportation (Amst)*, *47*, 2133–2156. https://doi.org/10.1007/s11116-019-10004-y
57. Kim, I., Kim, H. C., Seo, D. J., & Kim, J. I. (2020). Calibration of a transit route choice model using revealed population data of smartcard in a multimodal transit network. *Transportation (Amst)*, *47*, 2179–2202. https://doi.org/10.1007/s11116-019-10008-8
58. Nassir, N., Hickman, M., & Ma, Z. L. (2019). A strategy-based recursive path choice model for public transit smart card data. *Transportation Research Part B: Methodological*, *126*, 528–548. https://doi.org/10.1016/j.trb.2018.01.002
59. Kusakabe, T., & Asakura, Y. (2014). Behavioural data mining of transit smart card data: A data fusion approach. *Transp Res Part C Emerg Technol*, *46*, 179–191. https://doi.org/10.1016/j.trc.2014.05.012
60. Zhu, W., Hu, H., & Huang, Z. D. (2014). Calibrating Rail Transit Assignment Models with Genetic Algorithm and Automated Fare Collection Data. *Computer-Aided Civil and Infrastructure Engineering* 29:518–530. https://doi.org/10.1111/mice.12075.
61. Wu, J., Qu, Y., Sun, H., et al. (2019). Data-driven model for passenger route choice in urban metro network. *Physica A: Statistical Mechanics and its Applications*, *524*, 787–798. https://doi.org/10.1016/j.physa.2019.04.231
62. Shi, J. G., Zhou, F., Zhu, W., & Xu, R. H. (2015). Estimation method of passenger route choice proportion in urban rail transit based on AFC data. *Journal of Southeast University Natural Science Edition*, *45*, 184–188.
63. Zhu, W., Fan, W. L., Wahaballa, A. M., & Wei, J. (2020). Calibrating travel time thresholds with cluster analysis and AFC data for passenger reasonable route generation on an urban rail transit network. *Transportation (Amst)*, *47*, 3069–3090. https://doi.org/10.1007/s11116-019-10040-8
64. Li, W., Luo, Q., Cai, Q., & Zhang, X. (2018). Using Smart Card Data trimmed by train schedule to Analyze Metro Passenger Route Choice with Synchronous Clustering. *J Adv Transp*, *2018*. https://doi.org/10.1155/2018/2710608
65. Sun, Y., & Schonfeld, P. M. (2016). Schedule-based rail transit path-choice estimation using automatic fare collection data. *J Transp Eng*, *142*. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000812
66. Zhu, Y. W., Koutsopoulos, H. N., & Wilson, N. H. M. (2018). Inferring left behind passengers in congested metro systems from automated data. *Transp Res Part C Emerg Technol*, *94*, 323–337. https://doi.org/10.1016/j.trc.2017.10.002
67. Zhu, Y. W., Koutsopoulos, H. N., & Wilson, N. H. M. (2021). Passenger itinerary inference model for congested urban rail networks. *Transp Res Part C Emerg Technol*, *123*. https://doi.org/10.1016/j.trc.2020.102896
68. Ma, Z., Koutsopoulos, H. N., Chen, Y., & Wilson, N. H. M. (2019). Estimation of Denied Boarding in Urban Rail Systems: Alternative formulations and comparative analysis. *Transportation Research Record*. https://doi.org/10.1177/0361198119857034
69. Mo, B. C., Ma, Z. L., Zhao, J. H., & Koutsopoulos, H. N. (2020). Calibrating Route Choice for Urban Rail System: A Comparative Analysis Using Simulation-based Optimization Methods. In: Transportation Research Board 99th Annual Meeting.
70. Mo, B. C., Ma, Z. L., Koutsopoulos, H. N., & Zhao, J. H. (2021). Calibrating path choices and Train Capacities for Urban Rail Transit Simulation models using Smart Card and Train Movement Data. J Adv Transp 2021:. https://doi.org/10.1155/2021/5597130
71. Parady, G., Ory, D., & Walker, J. (2021). The overreliance on statistical goodness-of-fit and under-reliance on model validation in discrete choice models: A review of validation practices in the transportation academic literature. *Journal of Choice Modelling* 38.
72. Li, Y., Nan, S. R., Guo, Y., et al. (2022). Detection and analysis of transfer time in urban rail transit system using WIFI data. *Transportation Letters*. https://doi.org/10.1080/19427867.2022.2079173
73. Reichl, P., Oh, B., Ravitharan, R., & Stafford, M. (2019). Using Wifi Technologies to Count Passengers in Real-time around Rail Infrastructure. In: 2018 International Conference on Intelligent Rail Transportation, ICIRT 2018. Institute of Electrical and Electronics Engineers Inc.
74. Gu, J. J., Jiang, Z., Bin, Sun, Y. S., et al. (2021). Spatio-temporal trajectory estimation based on incomplete Wi-Fi probe data in urban rail transit network. *Knowledge-Based Systems*, *211*. https://doi.org/10.1016/j.knosys.2020.106528
75. Zhao, J. J., Zhang, L. T., Ye, K. J. (2022). GLTC: A Metro Passenger Identification Method Across AFC Data and Sparse WiFi Data. *IEEE Transactions on Intelligent Transportation Systems* 23:18337–18351. https://doi.org/10.1109/TITS.2022.3171332.
76. Forghani, M., Karimipour, F., & Claramunt, C. (2020). From cellular positioning data to trajectories: Steps towards a more accurate mobility exploration. *Transp Res Part C Emerg Technol*, *117*. https://doi.org/10.1016/j.trc.2020.102666
77. Li, M. X., Gao, S., Lu, F., & Zhang, H. C. (2019). Reconstruction of human movement trajectories from large-scale low-frequency mobile phone data. *Computers, Environment and Urban Systems*, *77*. https://doi.org/10.1016/j.compenvurbsys.2019.101346
78. Zhang, J. W., Wu, W., Cheng, Q. X., Extracting the Complete Travel Trajectory of Subway Passengers Based on Mobile Phone Data. J Adv Transp 2022 https://doi.org/10.1155/2022/8151520
79. Servizi, V., Pereira, F. C., Anderson, M. K., Nielsen, O. A. (2022). (2021) Transport behaviour-mining from smartphones: a review. European Transport Research Review 13.
80. Sreenu, G., & Saleem Durai, M. A. (2019). Intelligent video surveillance: A review through deep learning techniques for crowd analysis. *J Big Data* 6.
81. Yan, X. L., Xie, Z. Y., & Wang, A. L. (2020). *Real-time monitoring system for passenger flow information of metro stations based on intelligent video surveillance* (pp. 329–335). Springer. In: Lecture Notes in Electrical Engineering.
82. Qian, X., Yu, X., & Fa, C. (2017). The passenger flow counting research of subway video based on image processing. In: Proceedings of the 29th Chinese Control and Decision Conference, CCDC 2017. Institute of Electrical and Electronics Engineers Inc., pp 5195–5198.
83. Ding, X., Liu, Z., & Xu, H. (2019). The passenger flow status identification based on image and WiFi detection for urban rail transit stations. *Journal of Visual Communication and Image Representation*, *58*, 119–129. https://doi.org/10.1016/j.jvcir.2018.11.033

84. Yang, H. (Frank), Cai, J., Liu, C., et al. (Eds.). (2023). Cooperative multi-camera vehicle tracking and traffic surveillance with edge artificial intelligence and representation learning. Transp Res Part C Emerg Technol 148:. https://doi.org/10.1016/j.trc.2022.103982
85. Yao, R., & Bekhor, S. (2020). Data-driven choice set generation and estimation of route choice models. *Transp Res Part C Emerg Technol*, *121*. https://doi.org/10.1016/j.trc.2020.102832
86. Du, Y., Luo, J. T., Cheng, K. F., et al. (2018). Recognition of urban travel method based on cell phone signaling and navigation map data. *Application Research of Computers*, *35*, 2311–2314.
87. Monterola, C., Legara, E. F., Pan, D. (2016). Non-invasive procedure to probe the route choices of commuters in rail transit systems. In: Procedia Computer Science. Elsevier B.V., pp 2387–2391.
88. Chi, H. N., Wang, B. Y., Ge, Q., & Bin, Huo, G. Y. (2022). Knowledge graph-based enhanced transformer for Metro Individual Travel Destination Prediction. J Adv Transp 2022:. https://doi.org/10.1155/2022/8030690
89. Wang, S., Lv, Y. M., Peng, Y. (2022). Metro Traffic Flow Prediction via knowledge graph and spatiotemporal graph neural network. J Adv Transp 2022:. https://doi.org/10.1155/2022/2348375
90. Zeng, J., & Tang, J. J. (2023). Combining knowledge graph into Metro passenger flow prediction: A split-attention relational graph convolutional network. *Expert Systems with Applications*, *213*. https://doi.org/10.1016/j.eswa.2022.118790
91. Zhu, G. Y., Zhang, M., & Yi, Y. (2022). Prediction of evolution results of Urban Rail Transit emergencies based on knowledge graph. *Journal of Electronic & Information Technology*, *44*. https://doi.org/10.11999/JEIT211594

## Publisher's note