Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Spatio-temporal trajectory estimation based on incomplete Wi-Fi probe data in urban rail transit network

Jinjing Gu a,b, Zhibin Jiang b,n, Yanshuo Sun c, Min Zhou b, Shenmeihui Liao b, Jingjing Chen d

a Department of Computer Science & Technology and Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai 201804, China
b College of Transportation Engineering, Tongji University, Key Laboratory of Road and Traffic Engineering of the State Ministry of Education, Key Laboratory of Rail Infrastructure Durability and System Safety, Shanghai 201804, China
c Department of Industrial and Manufacturing Engineering, FAMU-FSU College of Engineering, Florida State University, Tallahassee, FL 32310, USA
d Shanghai No. 4 Metro Operation Co., Ltd., Shanghai 200070, China

A R T I C L E   I N F O

Article history:
Received 26 June 2020
Received in revised form 6 October 2020
Accepted 12 October 2020
Available online 16 October 2020

Keywords:
Urban rail transit
Trajectory estimation
Spatio-temporal network
n-gram method
Wi-Fi probe data

A B S T R A C T

This study presents a methodology for estimating passenger’s spatio-temporal trajectory with personalization and timeliness by using incomplete Wi-Fi probe data in urban rail transit network. Unlike the automatic fare collection data that only records passenger’s entries and exits, the Wi-Fi probe data can capture more detailed passenger movements, such as riding a train or waiting on a platform. However, the estimation of spatio-temporal trajectories remains as a challenging task because a few unfavorable situations could result into deficient data. To address this problem, we first describe the Wi-Fi probe data and summarize their common defects. Then, the n-gram method is developed to infer missing spatio-temporal location information. Next, an estimation algorithm is designed to generate feasible spatio-temporal trajectories for each individual passenger by integrating multiple data sources, i.e., urban rail transit network topology, Wi-Fi probe data, train schedules, etc. This proposed method is tested on both simulated data in blind experiments and real-world data from a complex urban rail transit network. The results of case study show that 93% of passengers’ unique physical routes can be estimated. Then, for 80% of passengers, the number of feasible spatio-temporal trajectories can be reduced to one or two. Potential applications of the trajectory estimation approach are also identified.

1. Introduction

A spatio-temporal trajectory (STT) of passenger in urban rail transit (URT) network is the recording of one passenger’s motion, i.e., the route and train choice of each passenger at specific timestamps. Estimation of passengers’ route and train choices within a complex seamless transfer URT network is a challenging task due to the rapid expansion of network operation and high departure frequencies. In Shanghai, China, there are 415 stations on 16 URT lines, 53 of which are transfer stations as of October 2019 [1]. In general, there are multiple routes and trains between each origin–destination pair (OD). For example, from Disney Resort to Tongji University, there are at least 8 possible routes and the minimum headway of trains is 2 min and 30 s [2]. Obviously, it is critical for rail transit operators to understand how passengers choose routes in order to better plan and operate the services, especially on those dense networks with complex structures. For instance, the route choice is essential to allocating passenger flows to different rail transit lines, which further determine how fare revenues are distributed. The estimation results can also help operators to estimate the train loads over time and space. Furthermore, real-time demand management and application evaluation, e.g., personalized dynamic route guidance, flexible pricing strategy, and real-time passenger flow control, etc., can be conducted more efficient if passenger flows are assigned at the granularity of trains. Therefore, despite its technical intricacy, the STT estimation is required to support the day-to-day operations of large URT systems.

In previous studies [3,4], automatic fare collection (AFC) data have been used for inferring passenger route choices on UTR system. However, such transaction data only contain information on tap-in and tap-out timestamps and stations, without any information between the two trip ends. Therefore, it is very challenging to estimate the most likely route a passenger took based on the limited AFC data. On the other hand, although various complex methods are developed, such as stochastic assignment principle [5,6] and passenger flow assignment model [7–9], they still do not pay enough attention to personalization and timeliness when dealing with the problem of SST estimation. It would
result in neither the number of passengers left behind nor the real-time train load can be accurately calculated. It is desirable to introduce new technologies that could reveal intuitively detailed information about individual passengers’ STT during their trips, thus greatly reducing the set of feasible routes and trains.

Considering the very large-scale and complex URT networks in East Asia, especially in major Chinese metropolitans, it is necessary to explore new dataset and develop new approach for estimating rail transit route choices efficiently and relatively accurately. In this study, we use the emerging Wi-Fi probe data thanks to the recent deployment of Wi-Fi networks in major rail transit systems in China. But a few unfavorable situations could result into incomplete Wi-Fi probe data in URT network. Then, a spatio-temporal location inference model based on n-gram method is developed to analyze how an individual passenger travels on different rail transit lines and trains. Next, we present an estimation algorithm to generate feasible STT for each individual passenger by considering multiple data sources, i.e., metro network topology, Wi-Fi probe data, train schedules, and passenger walking times, etc. In general, this paper contributes to the URT route estimation literature by presenting the first study to estimate the STT of individual passenger using Wi-Fi probe data at the network level. The main contributions are summarized as:

1) This study extends personalization and timeliness into the STT estimation by integrating structurally heterogeneous data sources. This space–time network topology that is integrated by URT network’s physical topology and train diagram is set as boundary conditions of the individual passenger’s STT, and the real-time Wi-Fi probe data are adopted to reduce the STT searching space.

2) A new algorithm of estimating individual passenger’s STT that describes detailed information both on-board and at stations during the trip is developed. This methodology consists of three sub-tasks, i.e., constructing network topology, inferring spatio-temporal location, and matching Wi-Fi probe data to space–time network topology by considering the node–node, node–link and link–link relationships.

3) Real-world analyses are conducted for the URT network with the longest route length in China. This is also the first attempt to find out all the routes and trains that passengers actually choose using Wi-Fi probe data at the complex real-world URT network level.

This paper is organized as follows: Section 2 gives a review on route choice studies using AFC data and analyses of Wi-Fi probe data in URT networks. Then, Section 3 describes the Wi-Fi probe data and the algorithm outline. Afterwards, Section 4 explains in detail our proposal SST estimation methodology. Section 5 shows the results of blind experiment and real-world case study. And finally, Section 6 expounds the conclusions and future works.

2. Related work

The estimation of STT in URT networks is difficult, especially in the case of complex URT networks. Previous studies on STT estimation can be divided into two categories, i.e., route choice modeling using AFC transaction data and Wi-Fi probe data used in route-choice estimation.

2.1. Route choice model using AFC data

The AFC system is used for collecting fares when passengers tap-in or tap-out the URT. Furthermore, this system also records the station and time information which can be used for inferring the passenger’s route choice behavior [10,11]. In light of the AFC data, some researchers tackled this problem based on a stochastic assignment principle [5,6] and passenger flow assignment model [7–9]. Xu et al. [5] proposed a deletion algorithm for available routes based on the depth-first principle and calculated the corresponding proportions. But the model proposed in that paper is a static model. Moreover, train schedule connection networks (TSCN) were mainly used to estimate STT which was to assign passengers to each train. Zhou and Xu [12] connected rail network with train schedules. Sun et al. [13] named this network as TSCN and established a space–time trajectory estimation model which was a set generation and weighted assignment problem for feasible TSCN routes. The optimal trajectory problem using AFC data was also researched within the last few years. Chen et al. [14], Chen et al. [15] and Zhu et al. [16] concentrated on optimizing route estimation method and established models to estimate the most likely space–time trajectory. Comparison of previous studies on estimating metro passengers’ trajectories is listed in Table 1.

Although the modeling is considered rather comprehensive in theory, the input data and model parameters seriously restrict the validity of the model. On one hand, AFC data [3,4,17–19] could only provide passengers’ origin and destination information, which makes the estimation of STT and even route on the complex URT network become very difficult. On the other hand, there has no intermediate information between origin and destination, which means most constraints need to be established by setting model parameters. But these parameters are obtained through manual investigation and hardly take each individual’s differences into consideration. In addition, denied boarding at the station is a critical element that is neglected by [17]. Likewise, Hänseler et al. [18] assumed that passengers are able to board any desired train. This is also a critical limitation for estimating route choice using solely AFC in the congested metro systems. Without further information, it has a very limited way to differentiate a same 20 min journey observation is caused by denied boarding using a short route or no denied boarding with a long route. Therefore, the route and STT estimated by models could be inaccurate and the characteristics of individual cannot be reflected. Tang et al. [20] proposed a time geographic method to estimate the most likely space–time routes of passengers in vehicle level. Luo et al. [19] obtained load profiles of transit vehicles in the transit network consisting of 12 tram lines and 8 bus lines. The above-mentioned methods estimated metro passenger dynamics at the station- and vehicle-levels separately. Thus, there is an urgent need to develop a new algorithm to describe detailed STT information both on-board and at stations.

2.2. Analyses of Wi-Fi probe data in URT networks

The Wi-Fi probe technology has a relatively concentrated detection range than GPS data [21] and fast detection speed. So short-time passenger route restoration and monitoring can be realized. At present, most passengers are willing to open the Wi-Fi function of their cell phones when taking the metro, so the sampling rate can be guaranteed and the analysis results will be representative. With the comprehensive coverage of the Wi-Fi signal in the URT network, the quantity and position accuracy of the data are sufficient to support the STT estimation of the individual passenger.

The Wi-Fi probe data contains the interaction information between Access Points (APs) and the access device (e.g., mobile phones, computers). The route of the object carrying the Wi-Fi access device can be tracked dynamically. These studies about Wi-Fi technology mainly were applied in indoor positioning [22–24]. Zhang et al. [25] studied the AP location problem of smart phones by using automatic AP positioning and propagation parameter estimation of inertial navigation. Ma et al. [10]...
proposed a pedestrian trajectory estimation method to match the Wi-Fi indoor positioning system with the mapping building. Shang et al. [26] designed a space–time-state hyper network-based assignment approach to estimate the passenger flow state by integrating multiple data sources (i.e., AFC data, Wi-Fi data and flow data). Some researchers used media access control (MAC) addresses, which are unique identifiers of electronic devices, to estimate passengers’ travel times or waiting times in stations of urban rail or bus [27,28]. Abedi et al. [29] took pedestrians and cyclists as research objects, fixed Wi-Fi with Bluetooth to track people in train stations, airports, etc. Shlayan et al. [30] estimated time-dependent OD demands and station waiting time of bus and subway users by Bluetooth and Wi-Fi technologies to explore the potential benefits of improving the level of public transport services. Li et al. [31] focused on Received Signal Strength Indication (RSSI) technique to predict the trajectory of passengers within URT stations. To the best of our knowledge, none of the existing studies have used the Wi-Fi probe data to estimate the passenger’s STT on large-scale and complex networks.

3. Overview

3.1. Wi-Fi probe data

Over the past few years, indoor Wi-Fi technology has developed rapidly in China. Extensive Wi-Fi coverage in rail transit systems have been achieved in many cities. One of the largest URT networks in China, referred to as X, has been installed Wi-Fi AP devices in 277 stations and 464 trains across 12 URT lines. On each workday, about 70 million Wi-Fi probe data records are collected. The primarily purpose of these Wi-Fi devices is to provide better internet service for passengers. These AP infrastructures can also be harnessed to capture and archive bulk positioning data while the mobile devices of passengers are connected to corresponding AP device. After removing unrelated information, the Wi-Fi probe data attributes used in this study are listed in Table 2.

It is worth noting that the Wi-Fi probe data of this URT system is regulated by the public security department. The dataset used in this paper is only for research purposes. It has been authorized by the metro transit police department. What is more, the personal information has been desensitized through the privacy-preserving authentication protocol [32]. In an ideal situation, once an anonymous passenger with an access device enters the URT system, this passenger’s location information will be continuously collected over time. The location can be classified into station and link. Fig. 1 illustrates the Wi-Fi probe data collection process when one passenger travels from station A to station B via transfer station C.

As shown in Fig. 1, the blue arrows indicate the complete STT of this passenger. Depending on the positions of the APs, the passenger’s location can be detected in many places (such as station hall, platform, and train). If a passenger’s locations are detected regularly and frequently, e.g., every 30 s, it is not difficult to reconstruct the actual route and identify train the passenger chose. However, after preliminary analyses of the Wi-Fi probe data, only 5% of passenger’s complete routes can be reconstructed. For the rest of the passengers, their routes cannot be determined solely based on Wi-Fi probe data, because such data usually suffers from the following issues: (1) there is only one data record collection for an AD_MAC, (2) either the origin or the destination station is missing, or (3) although both the origin and destination station are identifiable, certain positioning data (i.e., transfer station, LN_ID and TRAIN_ID, etc.) are lost during the trip.

Therefore, this study aims to identify the most likely routes and trains that are taken by each individual passenger in term of AD_MAC. In particular, according to the historical trip information of an individual passenger, the station information can be complemented. Then, this problem is transformed into the third category issue proposed above. Considering that the information about trains or transfer stations of passenger is partially lost, additional data including walking time and actual train schedules are used to eliminate infeasible STT between the identified origins and destinations in this paper.

3.2. Algorithm outline

The STT estimation process of individual passenger is mainly composed of three steps. They can be respectively summarized as Wi-Fi probe data collection, path estimation and STT estimation. Fig. 2 illustrates the overall STT estimation process on a real-world URT network.

These points in Fig. 2(a) are Wi-Fi data points for one journey. First, we can generate an initial Wi-Fi probe data network by sequentially connecting the adjacent Wi-Fi probe data. Then, we can coordinate this Wi-Fi data network with the URT network topology. The purpose of path estimation is to identify these
Table 2
Attributes in Wi-Fi probe data in URT network.

| Attribute   | Explanation               | Type   | Sample                        |
|-------------|---------------------------|--------|-------------------------------|
| AD_MAC      | Anonymous MAC address     | String | 34-80-B3-3C-XX-XX            |
| CAPTURE_TIME| Time of detection         | Datetime | 2017-08-18 21:03:12          |
| STATION     | Station name              | String | Sipinglu Station             |
| POS         | Detected position         | String | 0: station hall; 1: platform; 2: train |
| LN_ID       | URT line number           | Number | 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 |
| TRAIN_ID    | Train number              | Number | 203                          |
| CAR_ID      | Carriage number           | String | 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 |

Fig. 1. Ideal case of passenger’s location identification based on Wi-Fi probe data.

crucial stations that passengers passed by coordinating the Wi-Fi data network with the URT network topology. These crucial stations include the original station, the destination station and the transfer station. As shown in Fig. 2(b), there are two feasible paths from station A to station B, namely A→C→B and A→D→E→B. According to the formed Wi-Fi data network, no data is recorded at station D, station E and URT line 3, so it can be sure that this passenger traveled from station A to station B by path A→C→B. To estimate the passenger’s STT along this estimated path, train timetable and transfer time parameters need to be used to generate the time when the passenger arrived at another four important positions. They are the station hall where passenger enters the URT system, the platform that passenger gets off a train, and the station hall where the passenger exits the URT system or transfers to another URT line. The estimated STT is connected with arrowheads in Fig. 2(c). And these four kinds of important positions are marked as red, yellow, green and blue, respectively. Finally, the SST of each individual passenger can be represented as a combination of the route and train choice at specific timestamps.

4. Methodology

The passenger STT estimation algorithm developed for this study includes three subtasks: (1) extracting a route skeleton from the collected Wi-Fi probe data, (2) inferring the spatiotemporal locations for each individual passenger according to the route skeleton and historical trajectories, and (3) estimating the STT based on the URT network’s physical topology and train diagram.

4.1. Network topology construction

(1) Space–time network topology

This space–time network topology is integrated by URT network’s physical topology and train diagram. An abstract URT network’s physical topology G that consists of stations and trains is shown in Fig. 3(a). The route from station A to station B by route A→C→B can be extended into such a STT, as shown in Fig. 3(b).

These four kinds important positions referred to above could be treated as nodes V without length and area, V = {v_{en}(S, T), v_{b}(S, T, K), v_{a}(S, T, K), v_{ex}(S, T)}. They are defined as entry node v_{en}(S, T), boarding node v_{b}(S, T, K), alighting node v_{a}(S, T, K), and exit node v_{ex}(S, T) respectively. Those four kinds nodes share the same spatiotemporal attribute: the metro station S and the timestamp T. Boarding and alighting nodes have an extra attribute: the train set K. Thus, according to different categories of nodes V, the spatial attributes of metro station can be expressed as S(v_{en}), S(v_{b}), S(v_{a}) and S(v_{ex}). The time attributes are respectively expressed as entry time T(v_{en}), boarding time/train departure time T(v_{b}), alighting time/train arriving time T(v_{a}) and exit time T(v_{ex}). The train set attributes of boarding and alighting nodes which are expressed as K(v_{b}) and K(v_{a}), come from train timetable data and indicate the possible train from which passengers may board or alight. Then the four processes can be
Fig. 3. A simple URT network and one STT. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

converted into links \( A \) in this network. Four types of links, namely access link \( a_{en} \in A_{en} \), egress link \( a_{ex} \in A_{ex} \), train link \( a_{t} \in A_{t} \) and transfer link \( a_{tr} \in A_{tr} \) are defined. Concretely, combining with the space–time network topology in Fig. 3(b), the route \( A \rightarrow C \rightarrow B \) can be described as \( \{v_{en}(A, T) \rightarrow v_{b}(A, T, K) \rightarrow v_{a}(C, T, K) \rightarrow v_{b}(C, T, K) \rightarrow v_{ex}(B, T)\} \).

(2) Wi-Fi probe data network

One passenger’s Wi-Fi probe data network \( G^{i} \) is generated by connecting the key Wi-Fi probe data points collected within a journey as the dashed line connected by red pentagrams marked in Fig. 3(b). The four corresponding processes of the passenger \( i \) can also be treated as nodes \( V^{i} \). These nodes are respectively
defined as origin node $v^o_t(S, T)$, train node $v^t_t(L, T)$, transfer node $v^p_t(S, T)$, and destination node $v^d_t(S, T)$. The spatial attributes of these nodes are defined as origin station $S(v^o_t)$, line number of the train $L(v^t_t)$, transfer station $S(v^p_t)$ and destination station $S(v^d_t)$. The time attributes $T(v^o_t)$, $T(v^t_t)$, $T(v^p_t)$, and $T(v^d_t)$ represent a period of time within each process. These time data are same type, but different values may be collected multiple times in each process. It is pretty important for route-estimation, in details, for positioning transfer stations or narrow the range of alternative trains. Then we further split these nodes into smaller discrete nodes which have only one time point and extract the nodes with critical time data. These nodes include the last record of the origin station $t(v^o_t)$, the first and last record on a train $(t_1(v^t_t), t_2(v^t_t))$, the first and last record of a transfer station $(t_1(v^p_t), t_2(v^p_t))$, and the first record of the destination station $t(v^d_t)$.

4.2. Spatio-temporal location inference

The collection frequency of passenger’s Wi-Fi probe data is 30 s when passengers coming into the detection area. It offers the possibility to address missed spatio-temporal locations attributes from the historical perspective. The n-gram model is commonly used to estimate the probability of different word sequences. It has supported a range of natural language processing applications such as speech recognition and machine translation. According to the chain rule, this n-gram model infers the next location attribute (i.e., timestamp and location) in a mobility sequence in the form of an $O(n−1)$ Markov model. This paper proposes a mobility n-gram model for location attribute inferring. The joint probability can be factorized into the product of two conditional probabilities and the target joint probability can be rewritten as:

$\pi_i (t_1, L | (t_2, \ldots, t_n), (L_1, \ldots, L_n))$

$= \pi_i (t_1 | (t_2, \ldots, t_n), (L_1, \ldots, L_n)) \pi_i (t_2 | (t_3, \ldots, t_n), (L_2, \ldots, L_n)) \ldots \pi_i (t_n | (t_1, \ldots, t_{n−1}), (L_n, \ldots, L_1))$

where $\pi_i (t_x | \ldots)$ is the joint probability of the xth location and timestamp according to the n trajecory attribute for passenger i. $t_x$ denotes the target timestamp variable. $x$ shows the number of timestamps. $L_x$ denotes the target location variable at the timestamp $x$. $(L_1, \ldots, L_n)$ represents the context location space variable of the n-gram. $\pi_i (t_1, L_1, \ldots, L_n)$ expresses the prior probability of target location according to the previous n location information. $(t_1, \ldots, t_n)$ indicates the context time variable of the n-gram. $\pi_i (t_1, L_1, \ldots, L_n)$ means the prior probability of target location’s time according to the previous n time information. $N_i(t_x, L_x, \ldots, L_n)$ expresses the number of times that $t_x$ occurs immediately after location sequences $(L_1, \ldots, L_n)$. $N_i(L_1, \ldots, L_n)$ denotes the number of times that location sequences $(L_1, \ldots, L_n)$ in historical trajectory dataset of passenger i. $N_i(t_x, L_1, \ldots, L_n)$ means the number of times that $t_x$ occurs immediately after time sequences $(t_1, \ldots, t_n)$. $N_i(t_1, \ldots, t_n)$ indices the number of times that time sequences $(t_1, \ldots, t_n)$.

4.3. Trajectory estimation algorithm

Given the URT network’s physical topology, Wi-Fi data network and train schedules, this method aims to estimate all the feasible STT to the passengers. Firstly, constraints at boarding nodes and alighting nodes are proposed to generate the feasible train set. Then, node−link relationships are used to narrow the range of optimal trains of each line. Finally, link−link relationships are considered to form a complete feasible trajectory. The main purpose of this method is to estimate real route and the precise trains that passengers take. Thus, the entry process and exit process are simplified. It is assumed that the last time recorded at origin station $t(v^o_t)$ and the first time recorded at destination station $t(v^d_t)$ are equal to entry time $T(v^o_t)$ and exit time $T(v^d_t)$.

The detailed assumptions are as follow:

$t(v^o_t) = T(v^o_t)$

$S(v^o_t) = S(v^o_t)$

$t(v^d_t) = T(v^d_t)$

$S(v^d_t) = S(v^d_t)$

Furthermore, Eqs. (8) and (9) should be held to avoid loops.

$t_2(v^t_t) \geq t_1(v^t_t)$

$t_2(v^p_t) \geq t_1(v^p_t)$

where $t_1(v^t_t)$ and $t_2(v^t_t)$ indicate the first time and last time of passenger i on a train. $t_1(v^p_t)$ and $t_2(v^p_t)$ mean the first and last time of passenger i on a transfer station.

(1) Boarding process

Boarding node $(S(T, K)$ is the tail node of access link $a_m$ or transfer link $a_q$, and the head node of train link $a_i$. Combined with Wi-Fi probe data, the boarding time $T(v^m_t)$ must be later than or equal to the last time $t(v^m_t)$ or $t(v^p_t)$ on the station $S(v^o_t)$ or $S(v^p_t)$ and earlier than or equal to the first time $t(v^m_t)$ on the train $L(v^m_t)$. In all circumstances, station constraint should be established to satisfy Eq. (10).

$S(v^o_t) = S(v^t_t), S(v^p_t) = S(v^o_t), S(v^d_t)$

(10)

For time constraint, four circumstances are proposed.

$T(v^o_t) \geq t_2(v^m_t), S(v^t_t) = S(v^o_t)$

$t(v^p_t) = T(v^m_t)$

(11)

$T(v^o_t) < t_1(v^m_t), S(v^t_t) = L(v^m_t)$

$t(v^p_t) = S(v^m_t)$

(12)

Assuming that $v^o(S, T, K)$ is the tail node of $a_m$. Both $v^m_t(S, T)$ and $v^t_t(L, T)$ are probed. Eqs. (11) and (12) should be established and $S(v^m_t) = L(v^m_t), S(v^t_t)$

Assuming that $v^o(S, T, K)$ is the tail node of $a_q$ and $v^p_t(L, T)$ is lost. Then, $t(v^o_t)$ can be used to replace $t_2(v^m_t)$ in Eq. (11). $S(v^o_t) = S(v^m_t)$

Assuming $v^p(L, T)$ is lost. Then, $t(v^o_t)$ is used to replace $t_1(v^m_t)$ and $S(v^t_t) = S(v^m_t)$ in Eq. (12).

Assuming that both $v^m_t(S, T)$ and $v^p_t(L, T)$ are lost, the origin and destination station are used to replace them respectively. In train timetable, $T(v^o_t)$ is the train departure time on station $S(v^o_t)$, so all the trains that satisfy Eqs. (10)–(12) at node $v^o(S, T, K)$ can form the train set $S(v^o_t)$.

(2) Alighting process

Alighting node $v^o(S, T, K)$ is the tail node of train link $a_i$ and the head node of egress link $a_e$ or transfer link $a_q$. Combined with Wi-Fi probe data, the alighting time $T(v^o_t)$ must be later than or equal to the last time $t_2(v^t_t)$ on the train $L(v^t_t)$ and earlier than or equal to the first time $t_1(v^o_t)$ or $t(v^m_t)$ on the station $S(v^o_t)$ or $S(v^p_t)$. Station constraint is shown in Eq. (13).

$S(v^o_t) = S(v^t_t), S(v^p_t) = S(v^o_t), S(v^d_t)$

(13)

Time constraint could also be divided into four circumstances.

$T(v^o_t) \geq t_2(v^m_t), S(v^t_t) = L(v^m_t)$

$t(v^p_t) = S(v^m_t)$

(14)
\[ T(v_a) < \begin{cases} 1 \cdot (v'_L, S(v') = L(v'_L) \\ t(v'_S) = S(v'_S) \end{cases} \tag{15} \]

Assuming that both \( v'_L (L, T) \) and \( v'_S (S, T) \) are recorded, Eqs. (11) and (12) are established and \( S(v) = L(v'_S) , S(v) = S(v'_S) \).

Assuming that \( v'_L (L, T) \) is lost, \( t(v'_m) \) can be used to replace \( t(v'_L) \) and \( S(v'_L) = S(v'_m) \) in Eq. (14).

Assuming that \( v'_S (S, T) \) is lost, \( t(v'_m) \) is used to replace \( t(v'_S) \) and \( S(v'_S) = S(v'_m) \) in Eq. (15).

Assuming that both the train link \( a_t \) and transfer link \( a_r \) are lost in Wi-Fi probe data, \( t(v'_m) \) and \( t(v'_m) \) are replaced by \( t(v'_m) \) and \( t(v'_m) \) respectively according to Eqs. (11) and (12). In timetable, \( T(v_a) \) is the train arriving time on station \( S(v_a) \) so all the trains that satisfy Eqs. (10)–(12) at node \( v_a (S, T, K) \) will form the train set \( K(v_a) \).

(3) Train link process

According to the above two steps, the feasible train set \( K(v_a) \) and \( K(v_a) \) at each \( v_a (S, T, K) \) and \( v_b (S, T, K) \) are formed. The final feasible train set of each train link is named as \( K_f \). Then, the boarding node and alighting node of the same train link \( a_t \) can be represented as \( v_b (a_t) \) and \( v_a (a_t) \). For the same line, trains in the intersection of the two train sets are kept, as expressed in Eq. (16).

\[ K_f = K(v_b(a_t)) \cap K(v_a(a_t)) \tag{16} \]

When generate a complete STT, Eq. (17) must be held to ensure the train attributes for the boarding node and alighting node of the train link are the same.

\[ K(v_b(a_t)) = K(v_a(a_t)) \tag{17} \]

(4) Transfer link process

For transfer links, \( S(v_b(a_t)) \) and \( v_b(a_t) \) represent the same station, and satisfy Eq. (15), while \( T(v_b(a_t)) \) and \( T(v_a(a_t)) \) should satisfy transfer time constraint Eqs. (18) and (19) to avoid train set errors.

\[ S(v_b(a_t)) = S(v_a(a_t)) \tag{18} \]

\[ T(v_b(a_t)) - T(v_a(a_t)) \geq T_c \tag{19} \]

\[ K(v_b(a_t)) \neq K(v_a(a_t)) \tag{20} \]

The main idea of passenger route generation is to estimate the location (i.e., hall, platform, train, etc.) that passengers may appear during the whole route in real time. It should consider the relationship between the time recorded by Wi-Fi probe data and the time of each node in URT space-time network. Once Wi-Fi probe data is lost in some sections, the time on origin station \( t(v'_m) \) and time on destination station \( t(v'_m) \) are used for replacement. Intuitively, the pseudocode of the trajectory estimation algorithm is provided in Algorithm 1.

5. Experiment and analysis

5.1. Blind experiments

So-called blind experiments [33] is used to test this proposed method. In such an experiment, “certain information which could introduce bias or otherwise skew the result is withheld from the participants, but the experimenter will be in full possession of the facts.” Two types of routes will be obtained using the proposed method: (1) the physical network route without any temporal attributes, and (2) the STT.

The blind experiment is conducted as follows: an experimenter generates a route with complete Wi-Fi probe data and thus has full knowledge of the experiment. Then the experimenter reveals partial Wi-Fi data to the observer according to different data missing rate. The main purpose of this experiment is to find the impacts of data missing rate on both route estimation and STT estimation.

The network topology used in this blind experiment is shown in Fig. 3(a). The train schedule of URT line 1 is as follows: the first train leaves Station A at 37957 and the last at 48397, with a fixed headway of 360 s during this time period. The travel time from Station A to Station D is fixed at 1719 s and from Station A to Station C is 2652 s. The first train of URT line 2 leaves Station C at 38298 and the last at 51138, with a fixed headway of 240 s. The travel time from Station C to Station B is 1479 s and from Station E to Station C is 946 s. The train schedule of URT line 3 starts at 37513 and ends at 50233, with a fixed headway of 270 s. The travel time from Station D to Station E is 1070. Firstly, we generate a sequence of Poisson arrival times in the interval [37959, 48397] at station A. Then, a trajectory for each passenger is selected and generated randomly from route A→C→B and A→D→E→B. Finally, interpolating time data into the generated trajectory with 120 s intervals to form a complete Wi-Fi data group for each passenger. On average, each trajectory is composed by 40 Wi-Fi probe data.

Within the experiment, each trajectory corresponds to one passenger. In order to derive feasible routes and trajectories, the experimenter will present all the train schedules, minimum transfer time and different sample rates to the observer. Using the method proposed above, no matter what the data missing rate is, for all the passengers, at least one feasible route/trajectory can be found. For clarity, Other detailed estimation results under different data missing rates are given in Table 3.

Table 3 shows that in this experiment, if the data missing rate is below 0.6, a unique route can be determined for 99% of passengers and a unique STT can be determined for 57% of passengers. Therefore, the method proposed in this paper is suitable for estimating feasible routes and trajectories unless the data missing rate is very high, e.g., 0.7. The method will be further demonstrated in a much more complex real-world network in the next section.

5.2. Case study

(1) Results of spatio-temporal location inference

This section presents a case study using the proposed spatio-temporal location inference method based on real-world URT network data. The Wi-Fi probe data is obtained from the Technology Center of Metro Co. Ltd. In addition, the Wi-Fi probe data has been eliminated outliers by referring to the approach mentioned in [34]. The dataset contains records during a period of 64 days (from November 29, 2017 to January 31, 2018). Each record contains the attributes that are listed in Table 2. We divide the dataset into two subsets. The training set consists of the records for the first 60 days from November 29, 2017 to January 27, 2018.
The last 4 days’ records from January 28, 2018 to January 31, 2018 are utilized as the testing dataset. In the training model, the training samples \((l_{x-1},\ldots,l_{x-n})\) and \((t_{x-1},\ldots,t_{x-n})\), \(x > n > 1\) of each passenger can be generated from his/her history trajectories when the value of \(n\) equals 1 to 3, which corresponds to the Bayes, 2-gram and 3-gram approach respectively. Secondly, the values of \(N_i(l_x|l_{x-1},\ldots,l_{x-n})\), \(N_i(l_{x-1},\ldots,l_{x-n})\), \(N_i(t_{x-1},\ldots,t_{x-n})\), \(x > n > 1\) for each passenger \(i\) are calculated. Thirdly, the values of \(p_i(l_x|l_{x-1},\ldots,l_{x-n})\) and \(p_i(t_x|t_{x-1},\ldots,t_{x-n})\), \(x > n > 1\) can be captured according to Eqs. (2) and (3). Finally, following the Eq. (1), the value of \(p_i(l_x|l_{x-1},\ldots,l_{x-n}),\ (l_{x-1},\ldots,l_{x-n})\), \(x > n > 1\) can be calculated in accordance with the training model. Furthermore, these training model parameters would be updated with the generation of individual trajectory data.

Notice that accuracy refers to the frequency of the true next spatio-temporal location occurring in the list of inferred locations. Let \(r(l_x,t_x)\) be 1 it does and 0 otherwise, then the accuracy value \(Acc\) can be calculated by \(1/m \sum_{m=1}^{m} r(l_m,t_m)\), where \(m\) indicates the number of location for each individual passenger. In order to certify that the next location would be affected by how many preceding locations, this paper studies the effect of the \(n\)-gram model by varying \(n\) from 1 to 3 (i.e., Bayes, 2-gram and 3-gram models). The example results of STT inference with the top-\(k\) rule when \(k\) equals 1, 3 and 5 are demonstrated in Fig. 4.

As can be observed from Fig. 4, the accuracy has an apparent improvement when the value of \(n\) increases from 1 to 2, and starts to decline when \(n\) equals 3. Apart from this, this 2-gram model achieves the best performance in these three models. The best accuracy values are marked in bold. It suggests that the next location may affected by the two preceding location mainly. The reason is that travel rule of individuals is pretty complex and may not travel following the historical trip always. In addition, we can find that these accuracy values improve as the value of \(k\) increases. This proposed algorithm gives a set/a sequence of locations, these top-\(k\) trajectories that best connect the given locations are retrieved from existing Wi-Fi trajectories. It is helpful to optimize the feasible scope of individual’s trajectory estimation in the next step.

(2) Results of trajectory estimation
One case consists of 7 stations from Siping Road to Xinjiang-wancheng on URT line 10 and 3 stations from Xiuyan Road to Disney Resort on line 11. We treat the 7 stations on line 10 as a station group and consider passengers entering from these stations have the same route-choice behavior because Siping Road is the only station they could choose to transfer to other lines. And these 3 stations on URT line 11 are regarded as the other station group for the same reason.

In theory, all the feasible routes between these two station groups should be set as the initial feasible routes. However, this sub-network is pretty complex and exist many feasible routes between these two station groups. Actually, passengers are highly sensitive to travel time and transfer times, which means most passengers’ route choice will be concentrated within a limited route set. Taking the Wi-Fi probe data collected at a transfer station with rail transit lines A and B as an example, the probability distributions of passengers’ transfer time are shown in Fig. 5.

In Fig. 5, the top-4 best fit distributions of passengers’ transfer time between line A and line B are listed. The value of average transfer time can be obtained according to the best fit distribution. Then, we treat the 8 paths as a set of feasible routes to improve the efficiency of our algorithm. In particular, train timetables and minimum transfer times using in this case study are obtained from the real-world URT system. Fig. 6. shows the spatial topology of the 8 paths in the real-world URT network of X.

As shown in Fig. 6, three typical circumstances occur within the estimation results: (1) the unique STT could be extracted; (2) the unique route could be found while two or more trajectories are possible; (3) more than one feasible route are found.

Here is one typical result about circumstance (2). The passenger whose AD_MAC is 90-3C-**-**-13-AD enters URT network at Xiuyan Road, and leaves at Guoquan Road. URT line 11, 2 and 10 have both been recorded in Wi-Fi probe data, while none of the transfer stations are collected. Fig. 7(b) illustrates the detailed information collected during the journey.

The red pentagrams marked in Fig. 7(b) represent the key Wi-Fi data we need. After route estimation, the only one route is determined that firstly transfers at Luoshan Road to line 16, then transfers to line 2 at Longyang Road. The third transfer occurs at East Nanjing Road when the passenger finally enters line 10 and arrives at Guoquan Road in Fig. 7(a). Because the information of line 16 is lost, both No. 163 train and No. 16-431 train are possible to be taken, so two feasible STT are extracted in this process as shown in Fig. 7(c).

The other typical route-choice estimation results about circumstance (3) is shown in Fig. 8.

According to original Wi-Fi probe data of the passenger whose AD_MAC is D8-63-**-**-4B-7F, he/she enters URT network from Disney Resort and the last time at this station is 20:10:18. The destination station is Guoquan Road, 21:27:31 is the first time recorded at this station. During the journey, one data on No. 1140 train is recorded, so the first and last time recorded on this train is calibrated as the same time, 20:32:00. Three data rows on No. 1020 train are recorded, calibrating the first time as 21:09:44 and the last time as 21:19:39. The records from 20:32:00 to 21:09:44 are lost, which need to be estimated by our method. In this case, two feasible routes are found: (1) transfer directly to line 10 at JiaoTong University; (2) transfer to line 8 at Oriental Sport Center, then transfer to line 10 at Laoximen. Fig. 8(a) shows the route estimation results in the real-world URT network. In route 1, two trajectories are found, while in route 2, six trajectories can be found. Fig. 8(b) shows the detailed STT estimation results in each route.

After data reprocessing, totally 345 valid data groups of passenger trajectories are recorded by Wi-Fi sensors. Their origin and destination information are complete and belong to the scope of the OD groups that we study. Then, sum of route and STT estimation results, as shown in Fig. 9.

In the total data, 97 original Wi-Fi data groups can directly reflect the real route, constituting 28.12% of total. After using our method, 343 data groups output feasible route estimation results, accounting for 99.4% of input data. As shown in Fig. 9, 316 data groups receive unique feasible route, constituting 92.9% of output data. So, we can obtain the conclusion that the route and STT estimation method using Wi-Fi probe data is pretty suitable to determine the unique route-choice in complex network structures. Within the 316 data, 152 data groups acquire only one feasible STT, which means 44.6% of passengers’ travel trajectory could be accurately restored using Wi-Fi probe data and the proposed method. 117 passengers gain two feasible trajectories and the rest of passengers get more than two feasible trajectories, and further study still need to be done to estimate the most likely trajectory.

5.3. Applications

The main purpose of this study is to design a methodology to reconstruct individual passenger’s actual route and train choice under incomplete Wi-Fi probe data situation. This proposed approach can be considered as one of the first steps toward the use of Wi-Fi probe data for the STT estimation at the network level. The solution to this problem facilitates a number of applications, for instance, the route choice is essential to allocating
Fig. 5. Probability distribution of transfer time in one transfer station.

(a) Transfer time from line A to line B.

(b) Transfer time from line B to line A.

Fig. 6. Set of available routes.

Direction: 11-10
- Path1 Origin-line 11-JiaoTong University-line 10-Destination
- Path2 Origin-line 11-Oriental Sport Center-line 8-Laoximen-line 10-Destination
- Path3 Origin-line 11-Oriental Sport Center-line 8-Siping Road-line 10-Destination
- Path4 Origin-line 11-Oriental Sport Center-line 8-Qufu Road-line 12-Tiantong Road-line 10-Destination
- Path5 Origin-line 11-Luoshan Road-line 16-Longyang Road-line 2-East Nanjing Road-line 10-Destination
- Path6 Origin-line 11-Luoshan Road-line 16-Longyang Road-line 2-Century Avenue-line 4-Hialun Road-line 10-Destination
- Path7 Origin-line 11-Longhua-line 12-Tiantong Road-line 10-Destination
- Path8 Origin-line 11-Longhua-line 12-South Shanxi Road-line 10-Destination
passenger flows to different rail transit lines, which further determine how fare revenues are distributed. The estimation results can also help operators to estimate the train loads over time and space. Furthermore, real-time demand management can be conducted more refined, e.g., detailed estimation of platform usage and number of passengers left behind for safety analysis, personalized dynamic route guidance, crowding information systems, real-time passenger flow control, and disruption management. Real-time application evaluation, such as analysis of car-specific on-board accumulation for comfort assessment and flexible pricing strategy and real-time train rescheduling strategies, can be optimized according these estimated SST results. It can also provide comprehensive and effective input data for trajectory tracking [35], trajectory pattern mining [36–38] and
6. Conclusions

This paper contributes to the route estimation research by presenting a general modeling framework for estimating passenger’s STT with personalization and timeliness using the Wi-Fi probe data at the URT network level. After reviewing relevant studies about route-choice models and Wi-Fi technology, several literature gaps are identified. First, at the present stage, most of the route-choice models take AFC transaction data as model input. Since no intermediate information between origin and destination is recorded, the accuracy of the estimation is difficult to determine. Second, most of studies use Wi-Fi technology to position and calculate time parameters in a very limited area, such as a station or an airport. Third, detailed information of Wi-Fi probe data has not been described specifically. Therefore, in this study, structure and problems of Wi-Fi probe data in URT system are introduced and useful Wi-Fi data are integrated in the estimation method. Blind experiments prove that our method can find feasible routes and STT in a simple network successfully under any data missing rates. In the case study, we reconstruct the real-world URT network into a topology network and choose a more complex OD pair which consisting of 8 feasible routes. Results show that (1) for 93% of passengers, a unique physical route can be estimated and (2) for 80% of passengers, the number of feasible STT is reduced to one or two. The STT estimation is capable to support the day-to-day operations of large URT systems. The route estimation results can be useful for metro corporation revenue allocations, the train load estimation over time and space, real-time demand management and application evaluation.

The proposed approach can be considered as a first step toward the utilization of Wi-Fi probe data for the STT estimation at the network level. This paper could thus stimulate further research in three aspects. First, combining data from more data sources, such as metro video surveillance, is helpful to improve the real-time STT estimation accuracy for individual passenger. Second, semantic STT patterns mining [40] is useful to find the most likely path and trajectory from the multiple feasible paths estimated, which is also the main point of our future study. Then, it is of great theoretical and practical significance for backtracking of COVID-19 spreading trajectory and tracking of close contacts of infected passenger in real time.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We greatly appreciate the very constructive comments made by the anonymous reviewers, who have helped significantly improve the quality and presentation of this paper. This work is supported by the National Natural Science Foundation of China (Grant No. 61473210), the Shanghai Science and Technology Committee (Grant No. 18DZ1041040) and the Shanghai Shentong Metro Group Co., Ltd. (Grant Nos. CX-GL20R014-WT-20037, JSKY19R024-WT-19005), the authors are grateful to these funding agencies. All authors approved the version of the manuscript to be published.

References

[1] Shanghai Metro, Wikipedia, 2020, https://en.wikipedia.org/wiki/Shanghai_Metro Accessed on January 26, 2020.
[2] Route & Fare Inquiry, Shanghai Metro website, 2020, http://service.shmetro.com/en/ Accessed on June 8, 2020.
[3] Y. Zhu, H.N. Koutsopoulos, N.H. Wilson, Inferring left behind passengers in congested metro systems from automated data, Transp. Res. Proced.: 23 (2017b) 362–370.
[4] Z. Ma, H.N. Koutsopoulos, Y. Chen, N.H. Wilson, Estimation of denied boarding in urban rail systems: alternative formulations and comparative analysis, Transp. Res. Rec. 2673 (11) (2019) 771–778.
[5] R.H. Xu, Q. Luo, P. Gao, Passenger flow distribution model and algorithm for urban rail transit network based on multi-route choice, J. China Railwa. Soc. 31 (2009) 110–114.
[6] S.G. Gao, W.U.J. Zhong, Modeling passenger flow distribution based on travel time of urban rail transit, J. Transp. Syst. Eng. Inf. Technol. 11 (2011) 124–130.
[7] E.V.D. Hurk, L. Kroon, G. Maroti, Deduction of passengers’ route choices from smart card data, IEEE Trans. Intell. Transp. Syst. 16 (2015) 430–440.
[8] L.J. Sun, Y. Lu, J.C. Jin, An integrated Bayesian approach for passenger flow assignment in metro networks, Transp. Res. C 52 (2015) 116–131.
[9] M.E. Ben-Akiva, S. Gao, Z. Wei, Y. Wen, A dynamic traffic assignment model for highly congested urban networks, Transp. Res. C 24 (2012) 62–82.
[10] L. Ma, Y. Fan, Y. Xu, Y. Cui, Pedestrian dead reckoning trajectory matching method for radio map crowdsourcing building in WiFi indoor positioning system, in: IEEE International Conference on Communications, 2017, pp. 1–5.
[11] Z. Zhao, H.N. Koutsopoulos, J. Zhao, Individual mobility prediction using transit smart card data, Transp. Res. C 89 (2018) 19–34.
[12] F. Zhou, R. Xu, Model of passenger flow assignment for urban rail transit based on entry and exit time constraints, Transp. Res. Rec. 2284 (2012) 57–61.
[13] Y. Sun, P. Schonfeld, Schedule-based rail transit route-choice estimation using automatic fare collection data, J. Transp. Eng. 142 (2016) 04015037.
[14] X. Chen, L. Zhou, J. Tang, H. Zhou, Estimating the most likely space-time route by mining automatic fare collection data, in: Presented at 97th Annual Meeting of Transportation Research Board, Washington, D. C. 2018.
[15] X. Chen, L. Zhou, Y. Yue, Y. Zhou, L. Liu, Data-driven method to estimate the maximum likelihood space–time trajectory in an urban rail transit system, Sustainability 10 (2018) 1752.
[16] Y. Zhu, H.N. Koutsopoulos, N.H. Wilson, A probabilistic passenger-to-train assignment model based on automated data, Transp. Res. B 104 (2017a) 522–542.
[17] Y. Zhang, E.J. Yao, J.Y. Zhang, K.N. Zheng, Estimating metro passengers’ route choices by combining self-reported revealed preference and smart card data, Transp. Res. C 52 (2017b) 76–89.
[18] F.S. Hanseler, J.P. van den Heuvel, O. Cats, W. Daamen, S.P. Hoogendoorn, A passenger-pedestrian model to assess platform and train usage from automated data, Transp. Res. A 132 (2020) 948–968.
[19] D. Luo, L. Bonnefruit, O. Cats, H. van Lint, Constructing spatiotemporal load profiles of transit vehicles with multiple data sources, Transp. Res. Rec. 2672 (8) (2018) 175–186.
[20] J.J. Tang, Y. Song, H.J. Miller, X.S. Zhou, Estimating the most likely space-time routes, dwell times and route uncertainties from vehicle trajectory data: A time geographic method, Transp. Res. C 66 (2016) 176–194.
[21] L.B. Luo, X.T. Hou, W.T. Cai, B. Guo, Incremental route inference from low-sampling GPS data: an opportunistic approach to online map matching, Inform. Sci. 512 (2020) 1407–1423.
[22] A. Alvarez-Alvarez, J.M. Alonso, G. Trivino, Human activity recognition in indoor environments by means of fusing information extracted from intensity of WiFi signal and accelerations, Inform. Sci. 233 (2013) 162–182.
[23] M. Kotaru, K. Joshi, D. Bharadia, S. Katti, Spotfi: decimeter level localization using wifi, in: ACM Conference on Special Interest Group on Data Communication, ACM, 2018, pp. 269–282.
[24] Q. Lu, X. Liao, S. Xu, W. Zhu, A hybrid indoor positioning algorithm based on WiFi fingerprinting and pedestrian dead reckoning, in: IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications, 2016, pp. 1–6.
[25] Y. Zhuang, Z. Syed, J. Gregory, N. ElSheimi, Autonomous smartphone-based WiFi positioning system by using access points localization and crowdsourcing, Pervasive Mob. Comput. 18 (2015) 118–136.
[26] P. Shang, R.M. Li, J.F. Guo, K. Xian, X.S. Zhou, Integrating Lagrangian and Eulerian observations for passenger flow state estimation in an urban rail transit network: a space–time–state hyper network-based assignment approach, Transp. Res. B 121 (2019) 135–167.
[27] A. Lesani, L.F. Miranda-Moreno, Development and testing of a real-time WiFi-Bluetooth system for pedestrian network monitoring and data extrapolation, in: Presented at 95th Annual Meeting of Transportation Research Board, Washington, D. C. 2016.
[28] S. El-Tawab, R. Oram, M. Garcia, B.B. Poster, Monitoring transit systems using low cost WiFi technology, in: Vehicular Networking Conference, IEEE, 2017, pp. 1–2.

[29] N. Abedi, A. Bhaskar, E. Chung, M. Misra, Assessment of antenna characteristic effects on pedestrian and cyclists travel-time estimation based on Bluetooth and WiFi MAC addresses, Transp. Res. C 60 (2015) 124–141.

[30] N. Shlayan, A. Kurkcu, K. Ozbay, Exploring pedestrian Bluetooth and WiFi detection at public transportation terminals, in: IEEE International Conference on Intelligent Transportation Systems, 2016, pp. 229–234.

[31] S. Li, W. Zhu, L. Guo, Characterizing travel space–time trajectory on urban rail transit network using Wi-Fi data, in: Presented at 95th Annual Meeting of Transportation Research Board, Washington, D.C. 2016.

[32] S.K. Zeng, Y. Mu, H.J. Zhang, M.X. He, A practical and communication-efficient deniable authentication with source-hiding and its application on Wi-Fi privacy, Inform. Sci. 516 (2020) 331–345.

[33] Y. Sun, R. Xu, Rail transit travel time reliability and estimation of passenger route choice behavior: Analysis using automatic fare collection data, Transp. Res. Rec. 2275 (2012) 58–67.

[34] J.J. Gu, Z.B. Jiang, W. Fan, J.M. Wu, J.J. Chen, Real-time passenger flow anomaly detection considering typical time series clustered characteristics at metro stations, J. Transp. Eng. A 146 (4) (2020) 04020015.

[35] T.-S. Wu, M. Karkoub, C.-C. Weng, W.-S. Yu, Trajectory tracking for uncertainty time delayed-state self-balancing train vehicles using observer-based adaptive fuzzy control, Inform. Sci. 324 (2015) 1–22.

[36] J. Zhu, C. Huang, M. Yang, G.P. Cheong Fung, Context-based prediction for road traffic state using trajectory pattern mining and recurrent convolutional neural networks, Inform. Sci. 473 (2019) 190–201.

[37] L. Bermingham, I. Lee, Mining distinct and contiguous sequential patterns from large vehicle trajectories, Knowl.-Based Syst. 189 (2020) 105076.

[38] M.Q. Lv, L. Chen, T. Chen, D. Zeng, B. Cao, Discovering individual movement patterns from cell-id trajectory data by exploiting handoff features, Inform. Sci. 474 (2019) 18–32.

[39] P.H. Wang, F.Y. Sun, D. Wang, J. Tao, X.H. Guan, B. Albert, Predicting attributes and friends of mobile users from ap-trajectories, Inform. Sci. 463–463 (2018) 110–128.

[40] D.Z. Zhang, K. Lee, I. Lee, Semantic periodic pattern mining from spatio-temporal trajectories, Inform. Sci. 502 (2019) 164–189.