


Article

Fine-Grained Metro-Trip Detection from Cellular Trajectory Data Using Local and Global Spatial–Temporal Characteristics

Guanyao Li ^{1,2,3,4,†} , Ruyu Xu ^{5,†}, Tingyan Shi ⁶, Xingdong Deng ^{2,3,4,*}, Yang Liu ^{2,3,4}, Deshi Di ^{2,3,4}, Chuanbao Zhao ^{2,3,4} and Guochao Liu ^{2,3,4}

- ¹ School of Civil Engineering and Transportation, South China University of Technology, Guangzhou 510641, China; gyli@gzpi.com.cn
² Guangzhou Urban Planning & Design Survey Research Institute Co., Ltd., Guangzhou 510060, China; liuyang@gzpi.com.cn (Y.L.); deshi.di@gzpi.com.cn (D.D.); zhaochuanbao@gzpi.com.cn (C.Z.); liu.guochao@gzpi.com.cn (G.L.)
³ Collaborative Innovation Center for Natural Resources Planning and Marine Technology of Guangzhou, Guangzhou 510060, China
⁴ Guangdong Enterprise Key Laboratory for Urban Sensing, Monitoring and Early Warning, Guangzhou 510060, China
⁵ Transportation College, Jilin University, Changchun 130000, China; xury1721@mails.jlu.edu.cn
⁶ College of Art and Science, New York University, New York, NY 10012, USA; ts4340@nyu.edu
* Correspondence: dengxingdong@gzpi.com.cn
† These authors contributed equally to this work.

Abstract: A fine-grained metro trip contains complete information on user mobility, including the original station, destination station, departure time, arrival time, transfer station(s), and corresponding transfer time during the metro journey. Understanding such detailed trip information within a city is crucial for various smart city applications, such as effective urban planning and public transportation system optimization. In this work, we study the problem of detecting fine-grained metro trips from cellular trajectory data. Existing trip-detection approaches designed for GPS trajectories are often not applicable to cellular data due to the issues of location noise and irregular data sampling in cellular data. Moreover, most cellular data-based methods focus on identifying coarse-grained transportation modes, failing to detect fine-grained metro trips accurately. To address the limitations of existing works, we propose a novel and efficient fine-grained metro-trip detection (FGMTD) model in this work. By considering both the local and global spatial–temporal characteristics of a trajectory and the metro network, FGMTD can effectively mitigate the effects of location noise and irregular data sampling, ultimately improving the accuracy and reliability of the detection process. In particular, FGMTD employs a spatial–temporal hidden Markov model with efficient index strategies to capture local spatial–temporal characteristics from individual positions and metro stations, and a weighted trip-route similarity measure to consider global spatial–temporal characteristics from the entire trajectory and metro route. We conduct extensive experiments on two real datasets to evaluate the effectiveness and efficiency of our proposed approaches. The first dataset contains cellular data from 30 volunteers, including their actual trip details, while the second dataset consists of data from 4 million users. The experiments illustrate the significant accuracy of our approach (with a precision of 87.80% and a recall of 84.28%). Moreover, we demonstrate that FGMTD is efficient in detecting fine-grained trips from a large amount of cellular data, achieving this task within 90 min of processing a day’s data from 4 million users.

Keywords: transportation mode detection; fine-grained metro-trip detection; cellular trajectory; mobile computing; user-mobility analysis



Citation: Li, G.; Xu, R.; Shi, T.; Deng, X.; Liu, Y.; Di, D.; Zhao, C.; Liu, G. Fine-Grained Metro-Trip Detection from Cellular Trajectory Data Using Local and Global Spatial–Temporal Characteristics. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 314. <https://doi.org/10.3390/ijgi13090314>

Academic Editor: Hartwig H. Hochmair and Wolfgang Kainz

Received: 10 July 2024
 Revised: 22 August 2024
 Accepted: 29 August 2024
 Published: 30 August 2024



Copyright: © 2024 by the authors. Published by MDPI on behalf of the International Society for Photogrammetry and Remote Sensing. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>)

1. Introduction

The metro system plays a vital role in modern cities, offering a range of benefits, including safety, efficiency, and punctuality [1]. It provides a reliable and efficient means of

transportation, effectively alleviating traffic congestion within urban areas. Furthermore, metro systems make significant contributions to reducing air pollution and greenhouse gas emissions. Encouraging the use of a metro helps mitigate the environmental impact of urban transportation systems.

A fine-grained metro trip contains comprehensive information on user mobility, including the original station, destination station, original time of departure, destination time of arrival, transfer station(s), and corresponding transfer time during the journey by metro. Understanding individuals' fine-grained metro trips in a city is of utmost importance. It facilitates travel-demand management, improves transportation and land-use planning, enhances safety measures, and enables the provision of better services to the residents, commuters, and visitors of a city [2–5].

One commonly used method to obtain metro-trip information uses smart card data, as users tap in and out when using the metro system. However, this data has limitations in analyzing fine-grained metro trips, as it only captures the information on origins and destinations for a user's journey without providing other information, such as transfer stations and corresponding transfer times. As a result, the exact passenger number for those transfer stations cannot be estimated since the number of transfer passengers is unknown.

When users make calls, send messages, or access the internet on their mobile phones, the phones establish connections with nearby cell towers. Even when the phone is not in active use, it periodically establishes connections at regular intervals, such as half an hour [6]. As a result, a vast amount of cellular trajectory data has been collected, comprising a sequence of positions indicating the locations of connected cellular towers along with their corresponding timestamps. Differing from metro card-swipe data, these data present an invaluable opportunity to extract information on fine-grained metro trips for passengers.

In this work, we focus on studying fine-grained metro-trip detection using cellular data. The problem is challenging due to the issues of location noise and irregular data sampling in cellular data.

- **Location noise:** We use the locations of cell towers to represent user locations in a cellular trajectory. However, such location data can be noisy due to the extensive coverage area of cellular towers, which can range from tens to hundreds of meters. Moreover, it is important to consider the oscillation problem, which introduces additional challenges in analyzing the cellular trajectory data. The oscillation problem arises when the coverage areas of multiple cell towers intersect, causing mobile phones to rapidly switch between different towers, sometimes within intervals as short as one second. Consequently, the positions within a cellular trajectory may not always accurately reflect the exact locations of users. This lack of precision leads to the ineffective identification of metro trips.
- **Irregular data sampling:** The data-sampling rates in a cellular trajectory are irregular, influenced by factors such as signal strength and the frequency of mobile phone usage by passengers. Thus, the time intervals between successive locations in cellular trajectories can vary significantly, sometimes extending up to tens of minutes. During these periods, passenger locations remain unobserved, leading to sparse observations of their positions. This sparsity introduces uncertainty when attempting to accurately detect metro trips.

There have been studies focused on transportation-mode detection and trip identification using cellular data. Some of these studies have successfully identified coarse-grained transportation modes, such as public transportation or private car [7,8], bus or car [9,10], and on-foot or motorized [11]. However, they cannot provide detailed metro-trip information as we do in our work [12–15]. This limitation restricts their applicability in scenarios where a more comprehensive understanding of individual trips is required. Moreover, other studies [16–20] first divided a trajectory into segments and then inferred transportation modes for each segment. However, these approaches rely on the assumption that each segment corresponds to a single transportation mode, which is too rigid when dealing

with cellular trajectories. The issues of location noise and irregular data sampling make it difficult to accurately ensure that a segment solely contains a single transportation mode.

To overcome the limitations of existing approaches, we propose the novel and efficient **Fine-Grained Metro Trip Detection (FGMTD)** model. It employs a spatial–temporal hidden Markov model (ST-HMM) to consider the characteristics of individual cell towers in a trajectory and individual stations on a metro route (i.e., local characteristics). Moreover, FGMTD analyzes the shape similarity between the entire trajectory and the metro route (i.e., global characteristics) to enhance detection accuracy. Specifically, it mitigates the location noise of cell towers by introducing an emission probability to estimate the likelihood of observing a particular tower. To address irregular data sampling, FGMTD leverages travel-time information to estimate the probability of users traveling between two locations via metro, even with irregular time intervals. In addition, it considers the weights of trajectory segments for a trip-route similarity measure, which is computed based on their irregular time interval.

An overview of our approach is presented in Figure 1. We first employ a series of data-processing approaches for trajectory denoising and merging. Then, we propose a spatial–temporal hidden Markov model (ST-HMM) to detect a candidate metro trip, considering the local spatial–temporal characteristics of individual positions in cellular trajectories. Specifically, we propose a spatial proximity-based approach to estimating emission probabilities, enabling the detection of candidate metro stations for observed cellular towers. Based on that, we develop a transition-probability estimation method based on travel time to determine whether a trip between two towers likely involves the metro. When the likelihood is minimal, we assign the two consecutive towers to separate trips. This method enables us to simultaneously identify transportation modes and segment trajectories. We then utilize the Viterbi algorithm to efficiently detect candidate metro trips. Furthermore, we propose a weighted trip-route similarity measure that considers global spatial–temporal characteristics to filter out dissimilar candidate trips. This involves defining the station-route distance and the segment-route distance, as well as introducing a trip-route similarity measure to quantify the similarity between metro routes and candidate trips.

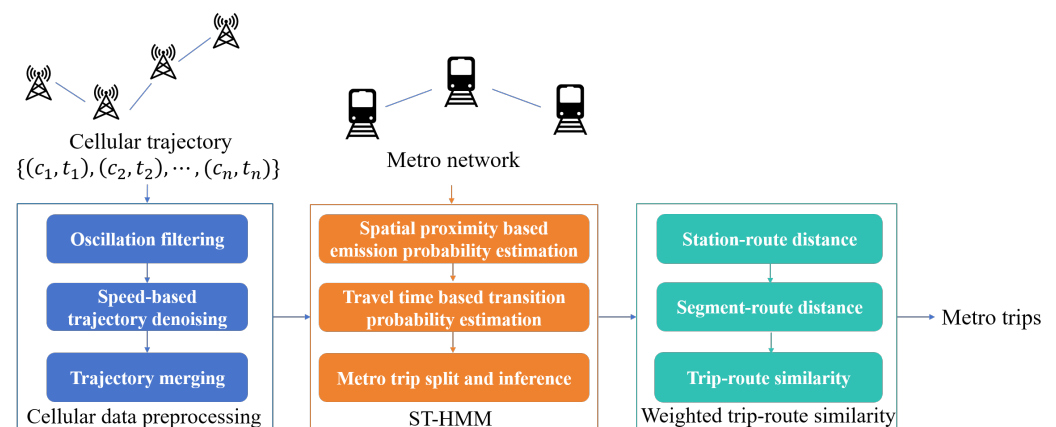


Figure 1. Overview of FGMTD.

Our contributions are summarized as follows.

- **Efficient ST-HMM** considers the spatial–temporal characteristic of individual positions in cellular trajectories: We propose an efficient ST-HMM to consider the proximity between individual positions (i.e., a cellular tower) and metro stations, as well as the transition probability between consecutive positions in a trajectory. In contrast to existing methods that decouple trajectory segmentation and transportation mode identification [16–20], our method adopts a joint approach. By avoiding the initial segmentation step and the assumption of single-mode segments, we effectively mitigate

location noise and irregular data-sampling issues. Moreover, we design efficient index strategies to significantly improve the computation efficiency.

- Trip-route similarity focuses on the spatial-temporal characteristics of a whole trajectory: We propose a novel route-trip similarity measure to evaluate the spatial proximity between a whole cellular trajectory and a metro route. Because of the irregular sampling issue, the time interval of segments (i.e., two consecutive positions in a trajectory) vary, resulting in differences in their importance for evaluating trip-route similarity. To mitigate the issue, we determine segment weights by considering the time intervals between consecutive positions when assessing trip-route similarity.
- Extensive experiments and case studies to validate the effectiveness of FGMTD: We conducted extensive experiments, and we provide case studies using two datasets. The first dataset was collected from 30 volunteers with their real trip information, while the second dataset consists of data from 4 million users, provided by a telecommunications company in China. The results demonstrate the high accuracy and efficiency of our proposed approach (with a precision of 87.80% and a recall of 84.28%), outperforming the existing baseline approaches. Moreover, the experiments show that our approach is significantly efficient, which makes it suitable for dealing with large-scale data.

The remainder of this paper is organized as follows. Section 2 discusses related works. Section 3 presents some relevant definitions. Section 4 introduces the detail of data processing for cellular data. The details of ST-HMM and weighted trip-route similarity are presented in Sections 5 and 6, respectively. The experimental setting and experimental results are shown in Section 7. Section 8 concludes this paper.

2. Related Work

Transportation-mode detection and trip identification have attracted much attention due to their importance in both academia and industry. Depending on the data used, we categorized existing works into three groups: sensor-based, GPS-based, and cellular-based approaches.

2.1. Sensor-Based Approaches

Sensor-based approaches utilize data from various smartphone sensors, including accelerometers, gyroscopes, magnetometers, linear acceleration, gravity, orientation, and ambient pressure, to detect transportation modes. These approaches employ machine learning techniques such as naive Bayes, support vector machines (SVMs), and decision trees, as demonstrated in studies [12–15], to infer activities such as stillness, walking, running, biking, driving (car), bus riding, train commuting, and subway travel. Additionally, deep learning methods like convolution neural networks (CNNs) [21], LSTM [22], and transformers [23] have been utilized for transportation-mode detection. However, these approaches solely provide coarse-grained transportation modes and do not offer detailed routes, rendering them inadequate for scenarios requiring fine-grained trips.

2.2. GPS-Based Approaches

GPS-based studies initially segment a trajectory into sub-trajectories, and they extract informative features for each sub-trajectory, such as speed, heading changes, acceleration, distance, and more [16–18]. Then, they employ machine learning approaches to infer transportation modes for each sub-trajectory, such as neural networks (NNs) [24], random forests [25], CRF-based inference [16], the LightGBM classifier [17], decision trees [26], and hidden Markov models [26,27]. Furthermore, some studies combine GPS trajectory data with extensive GIS data, such as road networks, subway networks, railway networks, and real-time bus locations, to infer transportation modes [25,28,29]. Compared with sensor-based approaches, GPS-based approaches could provide fine-grained trips.

Additionally, some studies utilize both GPS data and accelerometer data to determine transportation modes [26,30]. While these efforts are commendable, they are not suitable for identifying metro trips from cellular trajectories. Firstly, GPS signals tend to be weak

indoors, making them unsuitable for metro environments. Moreover, the issues of location noise and irregular data sampling results in challenges in extracting features because it is difficult to directly infer precise user location, movement direction, and speed from cellular data. Additionally, these approaches rely on trajectory segmentation and assume that each segment represents a single transportation mode. Yet, accurately ensuring that a segment solely contains a single transportation mode proves challenging due to the impact of location noise and irregular data sampling.

2.3. Cellular-Based Approaches

Some cellular-based methods extract mobility features (e.g., velocity and acceleration) from cellular data [31–33] or use handover and received signal strength (RSS) information from the serving cell tower as features [11]. Then, they apply classification techniques to identify transportation modes. These methods include convolutional neural network (CNNs) [11], the gated recurrent unit (GRU) neural network [31], LSTM [32], and random forests [33]. However, existing classification methods designed for GPS trajectories often lack applicability to cellular network data due to their lower spatio-temporal granularity compared to GPS data [34]. Consequently, these cellular-based approaches usually focus on classifying coarse-grained transportation modes, such as motor or non-motor [7], air or ground [35], and public transportation or private car [7,8,35–38], failing to provide information on fine-grained trips.

Some works detect metro trips based on indoor cell towers. For example, in the context of Singapore, indoor metro stations' platforms and tunnels are exclusively served via dedicated indoor cell towers, limiting the connection of cell phones outside the metro network to these towers. Exploiting this characteristic, the work [39] utilized the connection to these indoor cell towers to detect metro trips. However, the specific characteristics rely on dedicated indoor cell towers and may not be suitable for other cities. Moreover, some approaches partition a trajectory into segments and utilize external data sources such as smart card data [19] or real-time bus locations [20] to infer different transportation modes and corresponding trips. While these methods are impressive, the availability of such external data may be limited in certain scenarios, constraining their applicability. Our proposed approach takes into account the characteristics of individual positions and metro stations, as well as the overall trajectory and metro route. This comprehensive consideration enables the accurate detection of metro trips, making our approach applicable to a wide range of scenarios.

3. Preliminaries

When users make calls, send messages, or access the internet on their mobile devices, their devices establish connections with nearby cell towers. Thus, we can use the location of a cell tower to approximate a user's location. Over time, the sequence of connected cell towers forms a trajectory, which reflects the user's mobility patterns. We define a cellular trajectory as follows and present an example in Table 1:

Table 1. An example of cellular trajectory data: each entry comprises location coordinates in terms of latitude and longitude, accompanied by a corresponding timestamp.

	Latitude	Longitude	Time
1	23.1394	113.3794	19 October 2023 23:50:01
2	23.1409	113.3814	19 October 2023 23:50:02
3	23.1394	113.3794	19 October 2023 23:50:03
4	23.1394	113.3794	19 October 2023 23:52:23
...
21	23.1338	113.3795	19 October 2023 23:55:07
22	23.1338	113.3795	19 October 2023 23:58:54

Definition 1. (Cellular trajectory) A cellular trajectory is represented as a sequence of locations with the timestamps $Tra = \{(c_1, t_1), (c_2, t_2), \dots, (c_i, t_i), \dots, (c_n, t_n)\}$, where $c_i = (lat_i, lon_i)$ is the cell tower that the user connected to at time t_i , and (lat_i, lon_i) is the latitude and longitude of the cell tower.

In contrast to prior works that primarily focused on coarse-grained transportation-mode detection, our approach goes a step further by extracting fine-grained trip details from cellular trajectory data. We introduce the problem of fine-grained metro-trip detection as follows:

Definition 2. (Fine-grained metro trip detection) Given a cellular trajectory and a metro network, the problem of fine-grained metro-trip detection is to determine whether there are any metro trips and extract relevant trip details, including the original station, the destination station, the original time of departure, the destination time of arrival, any encountered transfer station(s), and the corresponding transfer time during the metro journey.

4. Cellular Data Processing

In this section, we introduce a series of cellular data-processing techniques to mitigate the oscillation problem and data error in cellular trajectories. These techniques are designed to enhance the reliability and accuracy of cellular data.

The oscillation problem is a significant challenge in analyzing cellular data [40]. It refers to the rapid and repeat switching of mobile phones between different cell towers, even when users remain stationary. This phenomenon occurs due to the overlapping coverage areas of cell towers, resulting in swift transitions within short intervals, such as one second. For example, the No.2 record in Table 1 was caused by the oscillation problem. This oscillation phenomenon hinders accurate trajectory analysis.

To address the oscillation problem, we consider the moving angle of a trajectory. Given $\{(c_{i-1}, t_{i-1}), (c_i, t_i), (c_{i+1}, t_{i+1})\}$ in a cellular trajectory, we denote the angle formed by the segments $\overline{c_{i-1}c_i}$ and $\overline{c_i c_{i+1}}$ as α_i . In the oscillation problem, if $c_{i-1} = c_{i+1}$ and $c_{i-1} \neq c_i$, the angle would be 0. Moreover, as illustrated in Figure 2, rapid switches between far cell towers would lead to a small angle. Thus, if we observe that α_i is smaller than a predefined threshold value, β , and $t_{i+1} - t_i < \tau$, we identify the location c_i as a noisy data point affected by the oscillation problem. Consequently, we remove c_i from the trajectory. In this work, β was set to 15, and τ was set to 5 s.

In addition to the oscillation problem, cellular data can be affected by various sources of errors, including measurement error, network instability, and signal interference. We present an example of data error in Figure 2. As it is highly unlikely for an individual to traverse the city at an excessively high speed (e.g., hundreds of km per h), we consider the moving speed to effectively identify and filter out potentially erroneous data points. Given $\{(c_{i-1}, t_{i-1}), (c_i, t_i), (c_{i+1}, t_{i+1})\}$ in a cellular trajectory, we denote the speed from c_{i-1} to c_i as $v_{i-1,i}$. If $v_{i-1,i} > \theta$, $v_{i,i+1} > \theta$, while $v_{i-1,i+1} < \theta$, we identify the cell tower c_i as an error data point, and we remove c_i from the trajectory. In this work, θ was set to 200 km/h.

As shown in Table 1, user devices may remain connected to the same cell tower for extended periods due to the large signal range of the cell tower, resulting in redundant data (e.g., No.1, No.3, and No.4 records in Table 1). This redundancy may increase computational complexity and hamper efficient detection. To mitigate redundancy within cellular trajectories, we implemented a merging strategy specifically designed for consecutive data records associated with the same cellular tower. Whenever two consecutive data records are linked to the same tower, we merge them, and additionally, we use two extra attributes to indicate the start time and end time of the association. This merging process results in a refined representation of the cellular trajectory, denoted as $\hat{Tra} = \{(c_1, t_1^s, t_1^e), (c_2, t_2^s, t_2^e), \dots, (c_i, t_i^s, t_i^e), \dots, (c_m, t_m^s, t_m^e)\}$, where c_i is the location of the cellular tower that the device connects to from the start time t_i^s to the end time t_i^e .

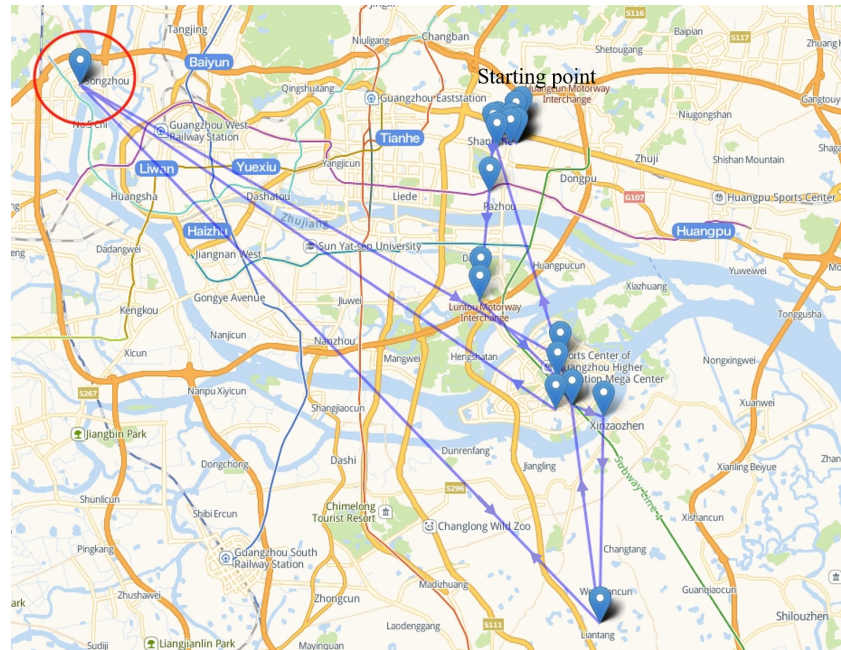


Figure 2. An example of a data error in a cellular trajectory, which could be detected by evaluating the moving speed or angle. The arrows indicate the moving direction.

An example of a processed cellular trajectory is presented in Table 2. The second record in Table 1 was removed, as it resulted from the oscillation problem. Moreover, the records of No.1, No.3, and No.4 in Table 1 were merged as the first record in Table 2, and the records of No.21 and No.22 were merged as the No.6 record in Table 2.

Table 2. An example of processed cellular trajectory data: each data point includes latitude and longitude coordinates along with the start and end times.

	Latitude	Longitude	Start Time	End Time
1	23.1394	113.3794	19 October 2023 23:50:01	19 October 2023 23:52:23
...
6	23.1338	113.3795	19 October 2023 23:55:07	19 October 2023 23:58:54

5. Spatial–Temporal Hidden Markov Model

In our work, we propose a spatial–temporal hidden Markov model (ST-HMM) designed specifically for detecting candidate metro trips from cellular trajectories. Furthermore, we propose efficient data-index strategies to enhance the efficiency of the detection process.

The ST-HMM considers the characteristics of individual positions within a trajectory and metro stations in a metro network. In the ST-HMM, a cellular trajectory is treated as a sequence of cell-tower observations influenced by hidden metro trips. Consequently, the hidden state space consists of all metro stations within a city, while the observation space includes all cell towers within the same area. The emission probability in the ST-HMM is defined based on the spatial proximity between the cell towers and metro stations (Section 5.1), whereas the transition probability is estimated using temporal information regarding the travel time between two cell towers (Section 5.2). By leveraging the results obtained from the emission and transition probabilities, we are able to jointly segment a trajectory into candidate trips and utilize the Viterbi algorithm to detect metro trips (see Section 5.3).

5.1. Emission Probability Estimation

We propose an emission-probability estimation approach to indicate the likelihood that a user was on a metro route. Moreover, we propose efficient index strategies to avoid redundant computation and improve computation efficiency.

Given an observation (c_i, t_i^s, t_i^e) in a trajectory of a user and a set of metro stations, S , the emission probability $P(s_j|c_i)$ represents the likelihood of c_i being observed if the user is located at a metro station, $s_j \in S$. A higher emission probability is associated with c_i if s_j is closer to c_i . Following some prior works on HMM [41–43], we use Gaussian distribution to model the emission probability:

$$P(s_j|c_i) = \frac{1}{\sqrt{2\pi}\delta} e^{-\frac{dis(c_i, s_j)^2}{2\delta^2}} \quad (1)$$

where δ is the standard deviation of positioning measurement noise, and $dis(c_i, s_j)$ represents the distance between the cellular tower c_i and the metro station s_j .

Considering the limited signal-coverage range of cellular towers, it is unnecessary to calculate the probability for all metro stations in the entire metro network. For the Gaussian distribution, the values less than three standard deviations from the mean account for 99.73% of the set. Thus, we focus on a subset of metro stations whose distance to c_i is less than the distance $3 \times \delta$. In a cellular trajectory, we consider a tower as a candidate tower if there exists a metro station, $s_j \in S$, within a distance threshold $3 \times \delta$ from the tower's location. This approach allows us to exclusively focus on calculating the emission probability for candidate towers, thereby avoiding redundant computations. Moreover, since trajectory locations are represented by cell towers within a city, we can precompute and store the emission probabilities associated with the candidate towers. This optimization further enhances efficiency and eliminates the necessity of redundant calculations. By precomputing and storing these probabilities, we eliminate the need to repeatedly calculate them for each position in the cellular trajectory. Consequently, the computational complexity of calculating emission probability becomes constant at $O(1)$.

Based on the precomputed emission probabilities above, a cellular trajectory could then be denoted as $\{(\hat{S}_1, \hat{t}_1^s, \hat{t}_1^e), (\hat{S}_2, \hat{t}_2^s, \hat{t}_2^e), \dots, (\hat{S}_p, \hat{t}_p^s, \hat{t}_p^e)\}$, where $\hat{S}_i \subset S$ represents a subset of metro stations and $\hat{S}_i \neq \emptyset$.

5.2. Transition-Probability Estimation

Given the results of the emission-probability estimation, we then compute the transition probability $P(\hat{s}_i^u, \hat{s}_{i+1}^v | \hat{t}_i^e, \hat{t}_{i+1}^s)$, where $\hat{s}_i^u \in \hat{S}_i$, $\hat{s}_{i+1}^v \in \hat{S}_{i+1}$. The probability represents the likelihood of moving from a candidate metro station, \hat{s}_i^u , to \hat{s}_{i+1}^v during the time interval between \hat{t}_i^e and \hat{t}_{i+1}^s .

To calculate this transition probability, we consider two important factors: the time interval $\mathcal{I}_{i,j} = \hat{t}_{i+1}^s - \hat{t}_i^e$ between the observations and the real travel time, $\delta_{i,j}^{u,v}$, from \hat{s}_i^u to \hat{s}_{i+1}^v . Unlike other transportation modes, metro travel times between stations are generally stable and reliable, making them readily available from official sources.

Following prior works [41–43], we formulate the transition probability based on the exponential probability distribution:

$$P(\hat{s}_i^u, \hat{s}_{i+1}^v | \hat{t}_i^e, \hat{t}_{i+1}^s) = e^{-|\mathcal{I}_{i,j} - \delta_{i,j}^{u,v}|}. \quad (2)$$

The absolute difference between the observed time interval and the real travel time is used to determine the probability, with smaller differences indicating a higher likelihood of transitioning between the metro stations. Given a trajectory with T time steps and N hidden states for each time step, the computational complexity is $O(T \times N^2)$.

5.3. Candidate Metro-Trip Inference

Based on the results of emission and transition probabilities, we can identify a candidate metro trip by finding the most likely sequence of hidden states (i.e., metro stations) that generates a given sequence of observations (i.e., cell towers). To speed up the computation, we employ the Viterbi algorithm [44] in our work; it is a dynamic programming algorithm to find the most likely sequence for a hidden Markov model.

During the computation, a low transition probability, $P(\hat{s}_i^u, \hat{s}_{i+1}^v | \hat{t}_i^e, \hat{t}_{i+1}^s)$, suggests that there is less likelihood that \hat{s}_i^u and \hat{s}_{i+1}^v are in the same metro trip for the user. If $P(\hat{s}_i^u, \hat{s}_{i+1}^v | \hat{t}_i^e, \hat{t}_{i+1}^s) < \gamma$ for all \hat{s}_i^u in \hat{S}_i and all \hat{s}_{i+1}^v in \hat{S}_{i+1} , we define \hat{S}_i and \hat{S}_{i+1} as belonging to two distinct trips, in which \hat{S}_i is the destination of a candidate trip, and \hat{S}_{i+1} is the origination for the consecutive candidate trip. γ was set to 0.05 in our work.

Then, for each candidate trip, we find the most likely sequence of hidden states (i.e., metro stations) in the ST-HMM to detect a metro trip. If we consider and evaluate all possible state sequences to find the optimal sequence, the time complexity is up to $O(N^T)$, where T is the number of time steps, and N is the number of candidate metro stations. To improve the efficiency, we employ the Viterbi algorithm [44] to identify candidate metro trips based on the most likely sequence of metro stations. The Viterbi algorithm is a dynamic programming algorithm used to find the most likely sequence of hidden states (metro stations) in an HMM that generates a sequence of observed events (cell towers in a trajectory). With the Viterbi algorithm, the computational complexity of detecting a metro trip is reduced to $O(T \times N^2)$. Since the extra computational complexity for calculating emission probability and transition probability is $O(1)$ and $O(T \times N^2)$, respectively, the computation complexity of ST-HMM is $O(T \times N^2)$. A candidate metro trip from the ST-HMM is represented as $\{(s_1, \hat{t}_1^s, \hat{t}_1^e), (s_2, \hat{t}_2^s, \hat{t}_2^e), \dots, (s_q, \hat{t}_q^s, \hat{t}_q^e)\}$, where s_i is a metro station.

6. Weighted Trip-Route Similarity

ST-HMM focuses on cell towers in a trajectory that are near metro stations (within a distance of less than $3 \times \delta$), and it compares their time interval with the actual travel time. However, it overlooks the position data in the trajectory that are not near metro stations. To address the limitation, we propose the weighted trip-route similarity to further evaluate the spatial similarity between the trajectory of a candidate metro trip and its corresponding metro route. This method considers all positions within a trajectory when assessing the similarity between the trajectory and a route, thereby potentially improving detection accuracy.

Given a trajectory, $\hat{T}ra = \{(c_1, t_1^s, t_1^e), (c_2, t_2^s, t_2^e), \dots, (c_i, t_i^s, t_i^e), \dots, (c_n, t_n^s, t_n^e)\}$, and a metro route, $R = \{s_1, s_2, \dots, s_j, s_{j+1}, \dots, s_m\}$, where s_j is the j -th metro station in the metro route, we first define the tower-route distance and segment-route distance for the trajectory and route. By utilizing these measures, we can then determine the overall trip-route similarity, which provides a comprehensive evaluation of the similarity between a given trip and its corresponding route.

Given a position, c_i , (i.e., a cell tower) in a trajectory and a metro route, R , the tower-route distance is computed as the distance between the cell tower and its nearest metro station within the metro route:

$$\hat{dis}(c_i, R) = \min\{dis(c_i, s_j), j = 1, 2, \dots, m\} \quad (3)$$

where $dis(c_i, s_j)$ is the physical distance between c_i and s_j . Based on this approach, we determine the segment-route distance by measuring the maximum distance between the endpoints of the segment and the corresponding route. We further define the distance between a segment, $\overline{c_i c_{i+1}}$, and the metro route as follows:

$$dis'(\overline{c_i c_{i+1}}, R) = \max(\hat{dis}(c_i, R), \hat{dis}(c_{i+1}, R)). \quad (4)$$

Then, the trip-route similarity can be calculated by taking the average of the segment-route distances. However, due to irregular data sampling, certain segments may have large time intervals between them, while others may have smaller time intervals. Hence, the contribution of each segment is not equal. To address this issue, we introduce a novel approach called temporal weighted trip-route similarity, which takes into account the varying time intervals between segments. The calculation of this similarity measure is defined as follows:

$$Sim(\hat{Tra}, R) = \frac{1}{\sum_{i=1}^{i=n-1} w_{i,i+1} \times dis'(\overline{c_i c_{i+1}}, R)} \quad (5)$$

where $w_{i,i+1}$ is the weight of the segment $\overline{c_i c_{i+1}}$. The weight, $w_{i,i+1}$, of the segment $\overline{c_i c_{i+1}}$ is calculated based on the time interval between them:

$$w_{i,i+1} = \frac{t_{i+1}^s - t_i^s}{t_n^s - t_1^s}. \quad (6)$$

If the similarity is larger than a threshold, the candidate trip is then detected as a metro trip. In this work, the threshold was set to 1. Given a trajectory with N positions and a metro route with M metro stations, the computation complexity is $O(M \times N)$.

After that, we complement a metro trip, $\{(s_1, \hat{t}_1^s, \hat{t}_1^e), (s_2, \hat{t}_2^s, \hat{t}_2^e), \dots, (s_q, \hat{t}_q^s, \hat{t}_q^e)\}$, by aligning it with the metro network. If two consecutive stations, s_i and s_{i+1} , are not consecutive stations in the metro network, we incorporate intermediate stations between s_i and s_{i+1} into the trip. When multiple routes exist between s_i and s_{i+1} within the metro network, we choose the route with the travel time closest to the given time interval. For those stations complemented in this step, we approximate their timestamps based on actual travel time from a station with a recorded timestamp. This method enables us to obtain a fine-grained metro trip, including information on the origin station, destination station, departure time, arrival time, transfer stations, and corresponding transfer time.

7. Illustrative Experimental Results

In this section, we first introduce the datasets, baseline approaches, and evaluation metrics used for experiments in Section 7.1. Then, we compare the detection accuracy of our proposed approach with baseline approaches in Section 7.2. Moreover, we present the ablation study in Section 7.3 and discuss the effect of a hyperparameter in Section 7.4. In addition, we present the evaluation results for the efficiency and scalability of our approach in Section 7.5. A case study is presented in Section 7.6.

7.1. Datasets and Baselines

We conducted experiments on two datasets to evaluate our approach, the details of which are summarized in Table 3. The first dataset comprises cellular data from 30 volunteers in which the volunteers labeled the ground truth of metro trips. Among them, 10 volunteers collected data for 15 days, while the remaining 20 volunteers collected data for 5 days each. The second dataset consists of cellular data from 4,089,902 users collected over a single day in a city. The datasets were provided by a telecommunication company in China. All data are anonymous, and no personal information was used. The actual metro trip information was provided by volunteers who consented to participate in this evaluation. Moreover, all data operations were conducted on a computer that was not connected to the internet. We confirm that the data do not involve any privacy concerns. The difference between the two datasets is that the first dataset includes actual trip information, while the second lacks such information. In our experiments, we used the first dataset to evaluate the detection accuracy due to its inclusion of real trip information, and we used the second dataset to evaluate the detection efficiency because of its substantial volume of data.

Table 3. Dataset overview.

	Users	Average Points per Day	Duration	Actual Trip Information
The first dataset	30	77.55	15 days/5 days	Yes
The second dataset	4,089,902	68.35	1 days	No

We evaluated the detection accuracy by comparing the detected trips with the actual trip information provided by the volunteers. Specifically, we used two evaluation metrics, precision and recall, defined in previous works [6,20], to evaluate the effectiveness of metro trip detection.

Definition 3. (Precision of a detected trip) Given a detected trip, T_i , with $|T_i|$ stations, if the trip exists and the real trip, T'_j , consists of $|T'_j|$ stations, then the precision of the trip is $p_i = \frac{(|T_i \cap T'_j|)}{|T_i|}$. Otherwise, $p_i = 0$.

Definition 4. (Recall of a detected trip) Given a real trip, T'_j , with $|T'_j|$ stations, if the trip is detected and the detected trip, T_i , has $|T_i|$ stations, the recall is $r_j = \frac{(|T_i \cap T'_j|)}{|T'_j|}$. Otherwise, $r_j = 0$.

Based on the above definition, given a set of detected trips, $\{T_1, T_2, \dots, T_n\}$, and the corresponding real trip, $\{T'_1, T'_2, \dots, T'_m\}$, the precision is calculated as

$$Precision = \frac{\sum_{i=1}^n p_i}{n}. \quad (7)$$

The recall is calculated as

$$Recall = \frac{\sum_{j=1}^m r_j}{m}. \quad (8)$$

We compared our approach with two representative works.

- Supervised learning-based approach (SL) [17]: SL is a transportation-mode classification method based on a light gradient boosting machine (LightGBM). It first divides original trajectories into some sub-trajectories and assumes that there is only one transportation mode in each sub-trajectory. Then, it employs various features to train the model, including a distance feature, five velocity-related features, two acceleration-related features, a heading-change rate, a stop rate, and a velocity-change rate.
- Unsupervised learning-based approach (UL) [7]: Based on the results of data cleansing and trajectory segmentation, UL uses an electronic navigation service to obtain the travel time of the origin-destination (OD) and compare the travel time for travel-mode identification.

We implemented our approach and the baseline approaches in Python. The LightGBM algorithm was implemented using scikit-learn (<https://scikit-learn.org/stable/index.html> (accessed on 10 July 2024)). Moreover, we used the official navigation service to obtain the travel time (<https://www.gzmtr.com/> (accessed on 10 July 2024)).

7.2. Accuracy of Trip Detection

To evaluate the accuracy of trip identification, we compared our proposed FGMTD with the baseline approaches using the labeled dataset. Precision and recall were used as evaluation metrics in the experiments. The comparison results are presented in Figure 3. Notably, FGMTD achieved significantly higher precision (87.80%) and recall (84.28%) than the baseline approaches, highlighting the effectiveness of our approach.

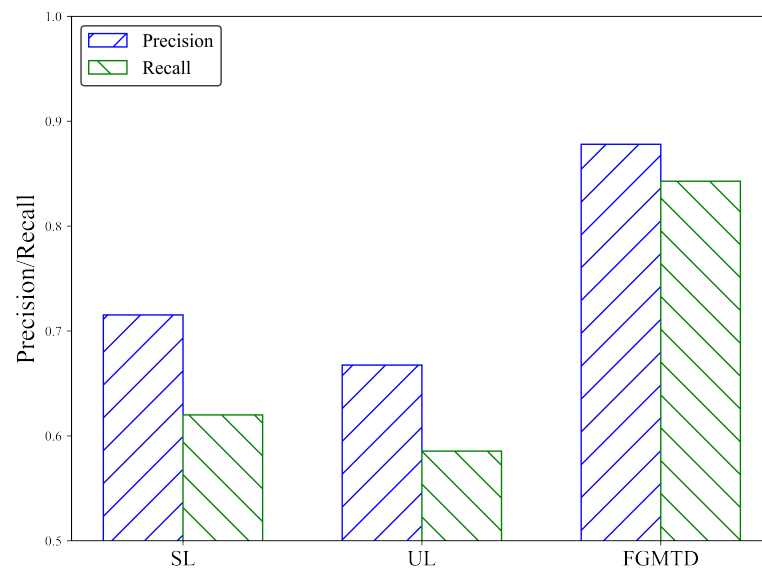


Figure 3. Comparison results of precision and recall.

Moreover, our results indicate that the unsupervised learning approach outperforms the supervised learning approach. This can be attributed to two main factors. Firstly, the supervised learning approach heavily relies on a substantial amount of training data. However, the limited availability of such data restricts its detection performance. Secondly, the supervised learning approach depends on accurate extracted features, such as distance, velocity, and acceleration. Nevertheless, due to issues of location noise and irregular data sampling, these extracted features become less accurate, resulting in inferior detection performance.

7.3. Ablation Study

FGMTD uses the ST-HMM to consider the local characteristic of an individual position in a trajectory and the weighted trip-route similarity to consider the global characteristic of the whole trajectory. We compared FGMTD with its variants to evaluate the effectiveness of the proposed modules. Precision and recall were used as evaluation metrics. In our experiment, the following variants were discussed:

- **ST-HMM:** We removed the weighted trip-route similarity from FGMTD. Only the ST-HMM module was used to detect a metro trip. The sequence of metro stations with the highest probability were identified as metro trips.
- **Weighted trip-route similarity (WTRS):** We removed the ST-HMM from FGMTD, and we only considered the similarity between a cellular trajectory and metro routes. We first employed a segmentation approach to divide a trajectory into segments. Then, we evaluated the similarity between the segments and metro routes. If a metro route with the highest similarity was greater than 1, it was identified as a metro trip.

The results of precision and recall for different variants are presented in Figure 4. Notably, FGMTD outperformed the sole use of ST-HMM or WTRS in terms of precision and recall. This improvement was achieved by combining ST-HMM and WTRS, highlighting the effectiveness of considering the characteristics of both individual positions and the entire trajectory.

ST-HMM exhibits superior performance compared to WTRS due to its consideration of not only the proximity of a position to a metro station but also the transition characteristics between consecutive positions. It captures the movement patterns between positions, leading to improved detection accuracy. Moreover, ST-HMM demonstrates the same recall value as FGMTD. This is because WTRS is primarily utilized to filter out dissimilar candidate metro trips, thereby not directly contributing to the improvement in detection recall.

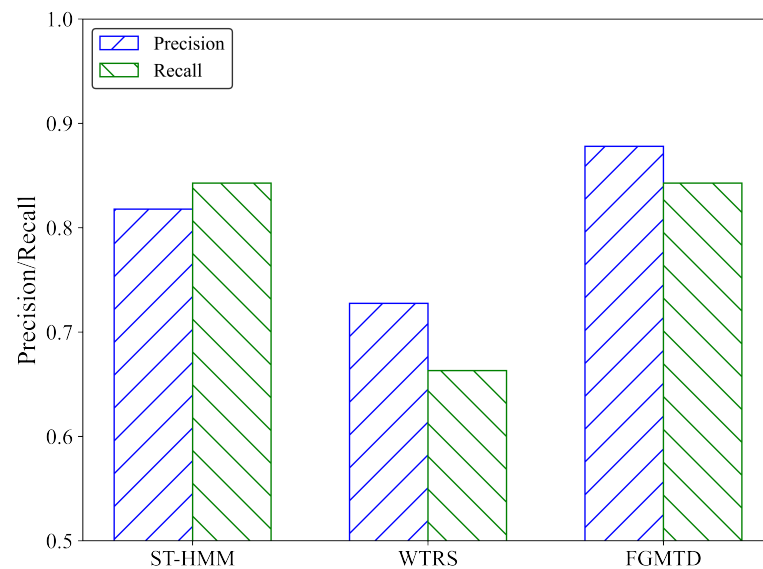


Figure 4. Precision and recall versus different variants.

7.4. Effect of Distance Threshold δ

When calculating the emission probability, we employ a distance threshold, δ , to filter out irrelevant cell towers. A smaller δ value can enhance computational efficiency, but it may also mistakenly filter out some relevant cell towers, leading to a decline in performance. To evaluate the impact of this threshold, we conducted experiments using δ values of 50 m, 100 m, 150 m, 200 m, and 250 m. The results of precision and recall versus different δ values are shown in Figure 5. As the threshold δ increases, precision and recall initially demonstrate improvement before eventually declining. When δ is set to 100 m, our approach attains optimal performance with respect to both precision and recall. This choice strikes a balance: when δ is too low, only a few cell towers are considered for trip identification, resulting in the loss of valuable information. Conversely, when δ is excessively high, irrelevant cell towers are included, introducing noise into the identification process. Therefore, the setting of 100 m strikes a balance by including relevant towers while minimizing the inclusion of irrelevant ones.

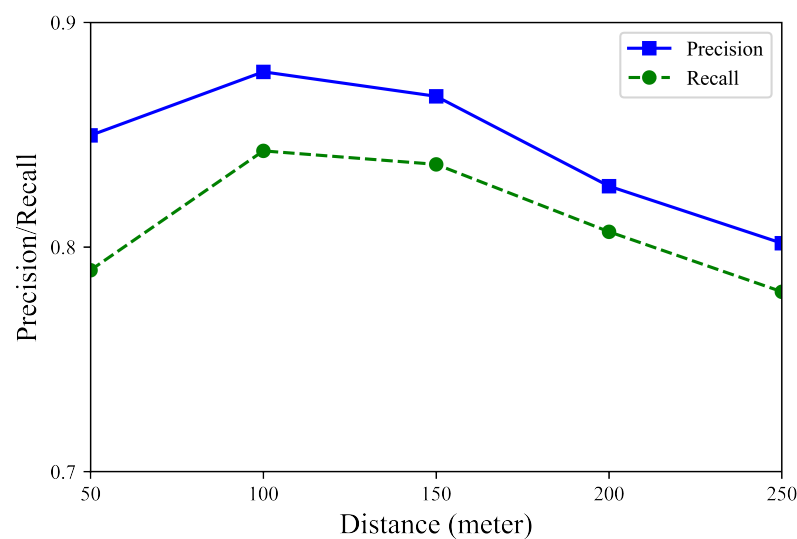


Figure 5. Precision and recall versus different δ values.

7.5. Computation Efficiency of Trip Detection

We used the large dataset to evaluate the efficiency of our proposed approach. Moreover, we used different amount of trajectories to evaluate the efficiency and scalability of our proposed approach. The trajectory amount was set to 1 million, 2 million, 3 million, and 4 million in our experiments. The running time of FGMTD versus different trajectory amounts is presented in Figure 6. Impressively, our approach demonstrated remarkable efficiency by successfully identifying metro trips from the trajectories of 4 million users in approximately 3200 s. Moreover, as the volume of data increased, the processing time scaled linearly, demonstrating the scalability of our model to large datasets.

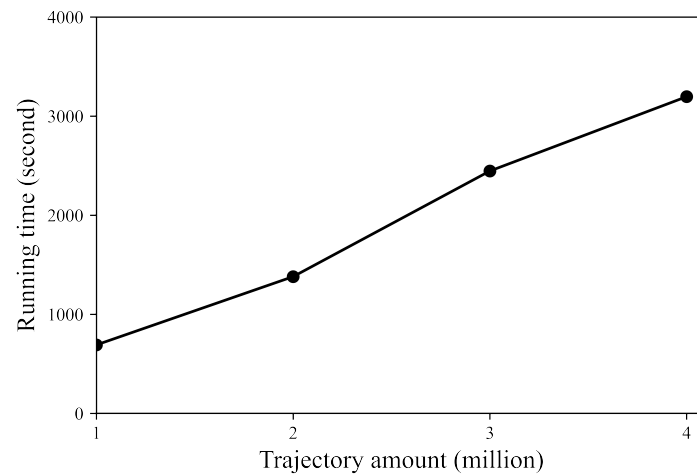


Figure 6. Running time versus different trajectory amounts.

As discussed in Section 5.1, we propose efficient index strategies to improve computational efficiency. To showcase the advantages of these strategies, we conducted additional experiments to compare the running time of our approach with and without the index. We varied the number of trajectories in our experiments, including {5000, 10,000, 15,000, 20,000} trajectories.

The results of the comparison between using the index and not using the index are presented in Figure 7. Remarkably, FGMTD with the index significantly outperformed the version without the index, achieving approximately 100 times greater efficiency. This result illustrates the importance of employing the index for efficient detection. The substantial efficiency gain achieved by utilizing the index demonstrates that our approach is well suited for handling large volumes of data, making it a valuable asset in practical applications.

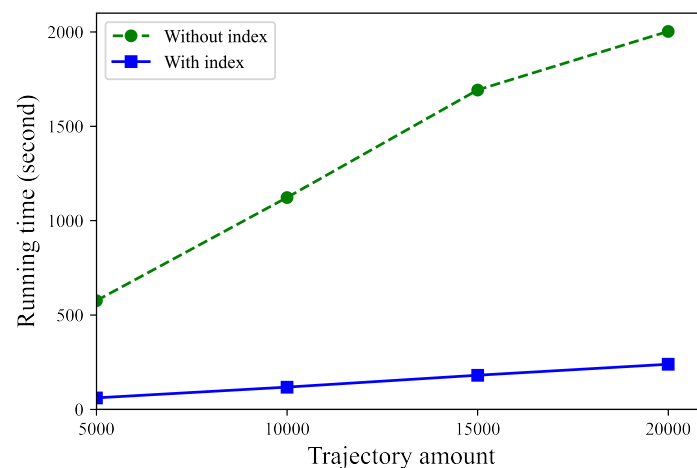


Figure 7. Comparison of running time between using the index and not using the index.

7.6. Case Study

We provide a case study using real data to demonstrate metro-trip detection from an individual's cellular trajectory. Figure 8 shows a raw cellular trajectory of an individual, and an outlier in the top left corner is evident due to a data error. Given such a noisy cellular trajectory, after the operations of data preprocessing, ST-HMM, and WTRS, we present the detection results in Figure 9.

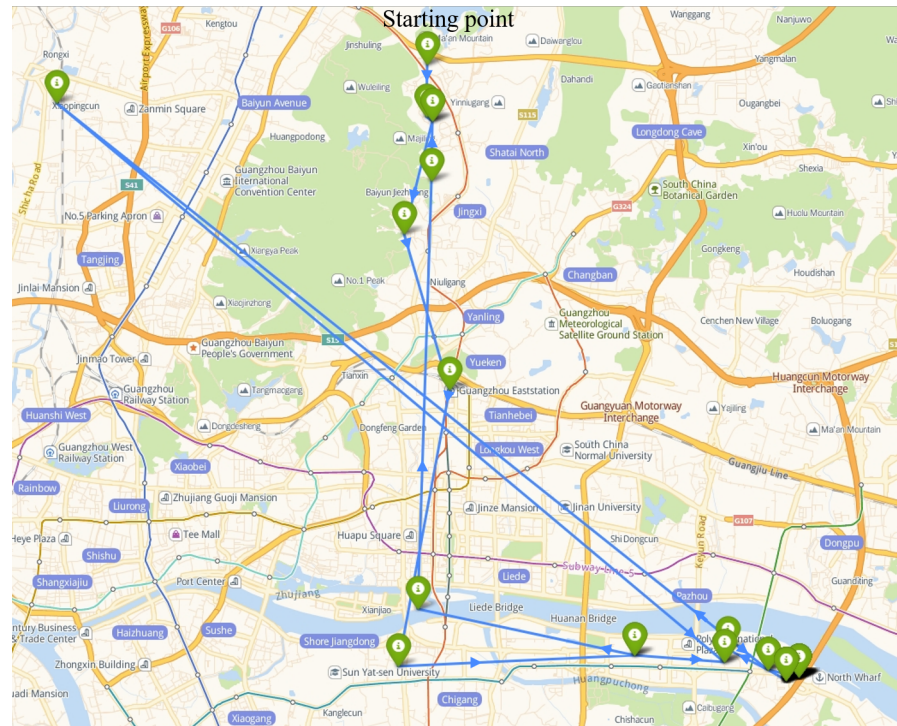


Figure 8. An illustration of an individual's cellular trajectory. The arrows indicate the moving direction.

As shown in Figure 9, the outlier in Figure 8 has been removed, and the detected metro trips are highlighted using different colors. Specifically, a red line and a green line represent two detected metro trips. We further detail the detection results in Table 4, which reflects the user's commuting patterns. In the morning, the user took Metro Line 3 from Tonghe Station, transferred to Line 8 at Kecun Station, and finally arrived at Wanshengwei Station. In the evening, the user returned by taking Line 8 from Wanshengwei Station, transferring to Line 3 at Kecun Station, and ultimately reaching Tonghe Station. This case study effectively demonstrates the accuracy and utility of our method in detecting fine-grained metro trips.

Table 4. Results of fine-grained metro trip detection: One trip was from Tonghe Station on Line 3 to Wangshengwei Station on Line 8, with a transfer at Kecun Station. Another trip was recorded from Wangshengwei Station back to Tonghe Station.

	Original Station	Departure Time	Destination Station	Arrival Time	Transfer Station	Transfer Time
1	Tonghe (Line 3)	08:07:04	Wanshengwei (Line 8)	08:52:29	Kecun (Lines 3 and 8)	08:40:38
2	Wanshengwei (Line 8)	18:30:29	Tonghe (Line 3)	19:20:44	Kecun (Lines 3 and 8)	18:49:41

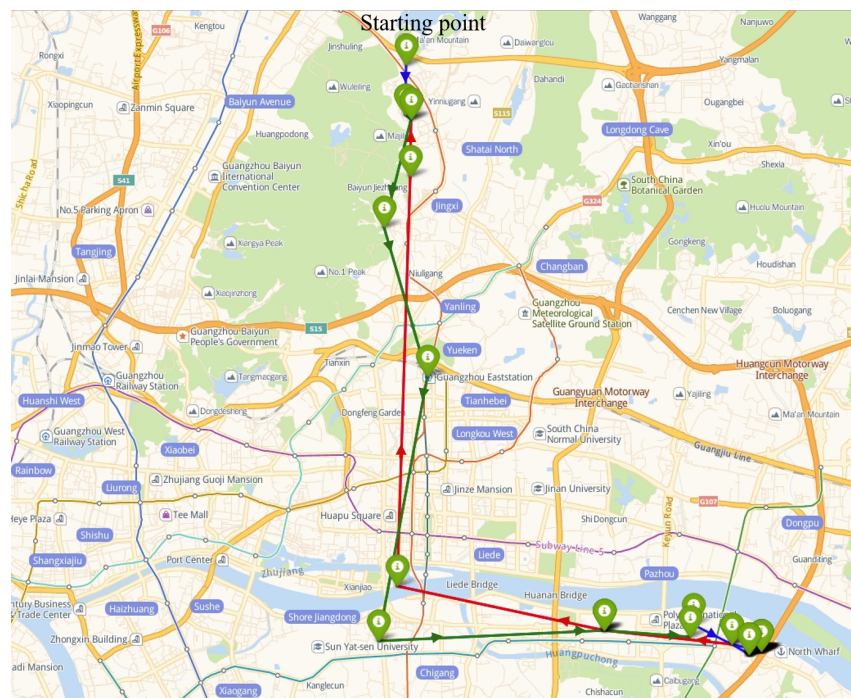


Figure 9. A case study of metro-trip detection results: The red line and the green line represent two distinct metro trips, while the blue lines refer to trips with other transportation modes. The arrows indicate the moving direction.

8. Conclusions and Future Works

In this paper, we have proposed a novel and efficient fine-grained metro-trip detection (FGMTD) model to extract detailed metro-trip information from cellular data. This information consists of crucial details such as the original station, destination station, departure time, arrival time, transfer station(s), and corresponding transfer time during a metro journey. In particular, FGMTD employs ST-HMM to identify candidate cell towers and determine whether a segment between two towers likely involves metro travel. This method enables us to simultaneously identify transportation modes and segment trajectories. Unlike existing methods, we skip the initial segmentation step and the assumption that each segment contains only one mode, allowing us to sidestep location noise and irregular data-sampling issues effectively. To further improve the detection precision, FGMTD uses WTRS to assess the similarity between a trajectory and a metro route. We conducted extensive experiments on two real datasets to validate the effectiveness and efficiency of our approach. Experiments on the dataset with actual trip information showed a substantial performance improvement in terms of precision and recall compared to previous works. Moreover, our findings highlight that, while ST-HMM exhibits commendable performance in metro detection, the integration of trip-route similarity assessments through WTRS leads to further enhancements in detection precision. Notably, experiments conducted on a substantial dataset with over 4 million users illustrate the efficiency and scalability of our proposed approach.

Below, we discuss the limitations and possible future directions of our work. One limitation stems from our reliance on static travel-time data. Our proposed approach determines the likelihood of a segment taking place via metro by comparing the actual travel time between metro stations to the segment's time interval. Given the metro system's exceptional punctuality, boasting an on-time reliability of 99.9% [45], we utilized static travel time data provided by the metro company in our work. However, disruptions within a metro system can lead to fluctuations in travel times, potentially impacting the efficacy of our approach. Furthermore, the utilization of static travel-time data restricts the adaptability of our method to detect other modes of transportation, such as buses, which

often experience significant variations in travel times due to factors like traffic congestion and weather conditions.

Therefore, integrating our work with dynamic travel-time data could be a future direction to enhance the performance of our approach and broaden its applicability to identifying various transportation modes. To achieve this, one way is to collect the dynamic data over time from metro and bus operators or navigation platforms like Amap and Google Maps. Another direction is to infer the dynamic travel times of different transportation modes for given origin–destination (OD) pairs from a large amount of cellular data. Identifying OD pairs within the dataset and employing clustering techniques on their travel-time information makes it feasible to categorize the dynamic travel times associated with diverse modes of transportation.

Author Contributions: Conceptualization, Guanyao Li, Ruyi Xu, and Xingdong Deng; funding acquisition, Xingdong Deng and Yang Liu; methodology, Guanyao Li, Ruyi Xu, Tingyan Shi, Xingdong Deng, Yang Liu, Deshi Di, Chuanbao Zhao, and Guochao Liu; project administration, Xingdong Deng and Yang Liu; software, Guanyao Li and Ruyi Xu; supervision, Xingdong Deng and Yang Liu; validation, Guanyao Li, Ruyi Xu, Tingyan Shi, Deshi Di, Chuanbao Zhao, and Guochao Liu; visualization, Tingyan Shi, Deshi Di, Chuanbao Zhao, and Guochao Liu; writing—original draft, Guanyao Li, Ruyi Xu, Tingyan Shi, and Deshi Di; writing—review and editing, Guanyao Li, Ruyi Xu, Tingyan Shi, Xingdong Deng, Yang Liu, Chuanbao Zhao, and Guochao Liu. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Collaborative Innovation Center for Natural Resources Planning and Marine Technology of Guangzhou (No. 2023B04J0301), the Guangdong Enterprise Key Laboratory for Urban Sensing, Monitoring and Early Warning (No. 2020B12120219), and the National Key R&D Program of China (2022YFB3904105).

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors Guanyao Li, Xingdong Deng, Yang Liu, Deshi Di, Chuanbao Zhao, and Guochao Liu are employed by the company Guangzhou Urban Planning and Design Survey Research Institute Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Abbreviations

ST-HMM	Spatial–temporal hidden Markov model
WTRS	Weighted trip-route similarity

References

1. Cai, Z.; Wang, J.; Li, T.; Yang, B.; Su, X.; Guo, L.; Ding, Z. A novel trajectory based prediction method for urban subway design. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 126. [[CrossRef](#)]
2. Deng, X.; Gao, F.; Liao, S.; Liu, Y.; Chen, W. Spatiotemporal evolution patterns of urban heat island and its relationship with urbanization in Guangdong–Hong Kong–Macao greater bay area of China from 2000 to 2020. *Ecol. Indic.* **2023**, *146*, 109817. [[CrossRef](#)]
3. Deng, X.; Gao, F.; Liao, S.; Li, S. Unraveling the association between the built environment and air pollution from a geospatial perspective. *J. Clean. Prod.* **2023**, *386*, 135768. [[CrossRef](#)]
4. Huang, J.; Liu, X.; Zhao, P.; Zhang, J.; Kwan, M.P. Interactions between bus, metro, and taxi use before and after the Chinese Spring Festival. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 445. [[CrossRef](#)]
5. Xi, Y.; Hou, Q.; Duan, Y.; Lei, K.; Wu, Y.; Cheng, Q. Exploring the Spatiotemporal Effects of the Built Environment on the Nonlinear Impacts of Metro Ridership: Evidence from Xi’an, China. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 105. [[CrossRef](#)]
6. Li, G.; Chen, C.J.; Peng, W.C.; Yi, C.W. Estimating crowd flow and crowd density from cellular data for mass rapid transit. In Proceedings of the 6th International Workshop on Urban Computing, Halifax, NS, Canada, 14 August 2017; pp. 18–30.
7. Chen, J.; Xiong, C.; Cai, M. A travel mode identification framework based on cellular signaling data. *Mob. Inf. Syst.* **2022**, *2022*, 1. [[CrossRef](#)]
8. Zhao, Y.; Wang, X.; Li, J.; Zhang, D.; Yang, Z. CellTrans: Private Car or Public Transportation? Infer Users’ Main Transportation Modes at Urban Scale with Cellular Data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2019**, *3*, 1–26. [[CrossRef](#)]

9. Chen, H.K.; Ho, H.C.; Wu, L.Y.; Lee, I.; Chou, H.W. Two-stage procedure for transportation mode detection based on sighting data. *Transp. A Transp. Sci.* **2024**, *20*, 2118558. [[CrossRef](#)]
10. Zeng, J.; Yu, Y.; Chen, Y.; Yang, D.; Zhang, L.; Wang, D. Trajectory-as-a-Sequence: A novel travel mode identification framework. *Transp. Res. Part C Emerg. Technol.* **2023**, *146*, 103957. [[CrossRef](#)]
11. Mostafa, S.; Harras, K.A.; Youssef, M. Ubiquitous Transportation Mode Estimation using Limited Cell Tower Information. In Proceedings of the 2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring), Florence, Italy, 20–23 June 2023; pp. 1–5.
12. Drosouli, I.; Voulodimos, A.; Miaoulis, G. Transportation mode detection using machine learning techniques on mobile phone sensor data. In Proceedings of the 13th ACM International Conference on Pervasive Technologies Related to Assistive Environments, Corfu, Greece, 30 June–3 July 2020; pp. 1–8.
13. Alaoui, F.T.; Fourati, H.; Kibangou, A.; Robu, B.; Vuillerme, N. Urban transportation mode detection from inertial and barometric data in pedestrian mobility. *IEEE Sens. J.* **2021**, *22*, 4772–4780. [[CrossRef](#)]
14. Yu, M.C.; Yu, T.; Wang, S.C.; Lin, C.J.; Chang, E.Y. Big data small footprint: The design of a low-power classifier for detecting transportation modes. *Proc. VLDB Endow.* **2014**, *7*, 1429–1440. [[CrossRef](#)]
15. Jahangiri, A.; Rakha, H.A. Applying machine learning techniques to transportation mode recognition using mobile phone sensor data. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 2406–2417. [[CrossRef](#)]
16. Zheng, Y.; Liu, L.; Wang, L.; Xie, X. Learning transportation mode from raw gps data for geographic applications on the web. In Proceedings of the 17th International Conference on World Wide Web, Beijing, China, 21–25 April 2008; pp. 247–256.
17. Wang, B.; Wang, Y.; Qin, K.; Xia, Q. Detecting transportation modes based on LightGBM classifier from GPS trajectory data. In Proceedings of the 2018 26th International Conference on Geoinformatics, Kunming, China, 28–30 June 2018; pp. 1–7.
18. Biljecki, F.; Ledoux, H.; Van Oosterom, P. Transportation mode-based segmentation and classification of movement trajectories. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 385–407. [[CrossRef](#)]
19. Poonawala, H.; Kolar, V.; Blandin, S.; Wynter, L.; Sahu, S. Singapore in motion: Insights on public transport service level through farecard and mobile data analytics. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 589–598.
20. Li, G.; Chen, C.J.; Huang, S.Y.; Chou, A.J.; Gou, X.; Peng, W.C.; Yi, C.W. Public transportation mode detection from cellular data. In Proceedings of the 2017 ACM Conference on Information and Knowledge Management, Singapore, 6–10 November 2017; pp. 2499–2502.
21. Tambi, R.; Li, P.; Yang, J. An efficient CNN model for transportation mode sensing. In Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems, Raleigh, NC, USA, 5–7 November 2018; pp. 315–316.
22. Jeyakumar, J.V.; Lee, E.S.; Xia, Z.; Sandha, S.S.; Tausik, N.; Srivastava, M. Deep convolutional bidirectional LSTM based transportation mode recognition. In Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, Singapore, 8–11 October 2018; pp. 1606–1615.
23. Tian, Y.; Hettiarachchi, D.; Kamijo, S. Transportation mode detection combining CNN and vision transformer with sensors recalibration using smartphone built-in sensors. *Sensors* **2022**, *22*, 6453. [[CrossRef](#)]
24. Byon, Y.J.; Abdulhai, B.; Shalaby, A. Real-time transportation mode detection via tracking global positioning system mobile devices. *J. Intell. Transp. Syst.* **2009**, *13*, 161–170. [[CrossRef](#)]
25. Stopher, P.; FitzGerald, C.; Zhang, J. Search for a global positioning system device to measure person travel. *Transp. Res. Part C Emerg. Technol.* **2008**, *16*, 350–369. [[CrossRef](#)]
26. Reddy, S.; Mun, M.; Burke, J.; Estrin, D.; Hansen, M.; Srivastava, M. Using mobile phones to determine transportation modes. *ACM Trans. Sens. Netw. (TOSN)* **2010**, *6*, 1–27. [[CrossRef](#)]
27. Bantis, T.; Haworth, J. Who you are is how you travel: A framework for transportation mode detection using individual and environmental characteristics. *Transp. Res. Part C Emerg. Technol.* **2017**, *80*, 286–309. [[CrossRef](#)]
28. Stenneth, L.; Wolfson, O.; Yu, P.S.; Xu, B. Transportation mode detection using mobile phones and GIS information. In Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Chicago, IL, USA, 1–4 November 2011; pp. 54–63.
29. Li, J.; Pei, X.; Wang, X.; Yao, D.; Zhang, Y.; Yue, Y. Transportation mode identification with GPS trajectory data and GIS information. *Tsinghua Sci. Technol.* **2021**, *26*, 403–416. [[CrossRef](#)]
30. Feng, T.; Timmermans, H.J. Transportation mode recognition using GPS and accelerometer data. *Transp. Res. Part C Emerg. Technol.* **2013**, *37*, 118–130. [[CrossRef](#)]
31. Wang, Y.; Yang, F.; He, L.; Liu, H.; Tan, L.; Wang, C. Inferring travel modes from cellular signaling data based on the gated recurrent unit neural network. *J. Adv. Transp.* **2023**, *2023*, 1987210. [[CrossRef](#)]
32. Gou, X.; Hung, C.C.; Li, G.; Peng, W.C. PTGF: Public Transport General Framework for Identifying Transport Modes Based on Cellular Data. In Proceedings of the 2019 20th IEEE International Conference on Mobile Data Management (MDM), Hong Kong, China, 10–13 June 2019; pp. 563–568.
33. Chin, K.; Huang, H.; Horn, C.; Kasanicky, I.; Weibel, R. Inferring fine-grained transport modes from mobile phone cellular signaling data. *Comput. Environ. Urban Syst.* **2019**, *77*, 101348. [[CrossRef](#)]
34. Breyer, N.; Gundlegård, D.; Rydberg, C. Travel mode classification of intercity trips using cellular network data. *Transp. Res. Procedia* **2021**, *52*, 211–218. [[CrossRef](#)]

35. Hui, K.T.Y.; Wang, C.; Kim, A.; Qiu, T.Z. Investigating the use of anonymous cellular phone data to determine intercity travel volumes and modes. In Proceedings of the Transportation Research Board 96th Annual Meeting, Washington, DC, USA, 8–12 January 2017; No. 17-03652.
36. Wang, H.; Calabrese, F.; Di Lorenzo, G.; Ratti, C. Transportation mode inference from anonymized and aggregated mobile phone call detail records. In Proceedings of the 13th International IEEE Conference on Intelligent Transportation Systems, Madeira, Portugal, 19–22 September 2010; pp. 318–323.
37. Kalatian, A.; Shafahi, Y. Travel mode detection exploiting cellular network data. In Proceedings of the 5th International Conference on Transportation and Traffic Engineering—EI Compendex, Lucerne, Switzerland, 6–10 July 2016; Volume 81, p. 03008.
38. Qu, Y.; Gong, H.; Wang, P. Transportation mode split with mobile phone data. In Proceedings of the 2015 IEEE 18th International Conference on Intelligent Transportation Systems, Gran Canaria, Spain, 15–18 September 2015; pp. 285–289.
39. Holleczeck, T.; The, Anh, D.; Yin, S.; Jin, Y.; Antonatos, S.; Goh, H.L.; Shi-Nash, A. Traffic measurement and route recommendation system for mass rapid transit (mrt). In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia, 10–13 August 2015; pp. 1859–1868.
40. Wu, W.; Wang, Y.; Gomes, J.B.; Anh, D.T.; Antonatos, S.; Xue, M.; Shi-Nash, A. Oscillation resolution for mobile phone cellular tower data to enable mobility modelling. In Proceedings of the 2014 IEEE 15th International Conference on Mobile Data Management, Brisbane, QLD, Australia, 17–18 July 2014; Volume 1, pp. 321–328.
41. Newson, P.; Krumm, J. Hidden Markov map matching through noise and sparseness. In Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Seattle, WA, USA, 4–6 November 2009; pp. 336–343.
42. Koller, H.; Widhalm, P.; Dragaschnig, M.; Graser, A. Fast hidden Markov model map-matching for sparse and noisy trajectories. In Proceedings of the 2015 IEEE 18th International Conference on Intelligent Transportation Systems, Gran Canaria, Spain, 15–18 September 2015; pp. 2557–2561.
43. Zhou, X.; Ding, Y.; Tan, H.; Luo, Q.; Ni, L.M. HMM: An HMM-based interactive map-matching system. In Proceedings of the Database Systems for Advanced Applications: 22nd International Conference, DASFAA 2017, Suzhou, China, 27–30 March 2017; Proceedings, Part II 22, pp. 3–18.
44. Forney, G.D. The viterbi algorithm. *Proc. IEEE* **1973**, *61*, 268–278. [[CrossRef](#)]
45. Zhang, S.; Lo H.K.; Ng, K.F.; Chen, G. Metro system disruption management and substitute bus service: A systematic review and future directions. *Transp. Rev.* **2021**, *41*, 230–251. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.