Research Article

A Hierarchical Passenger Mobility Prediction Model Applicable to Large Crowding Events

Bao Guo,1 Hu Yang,1 Fan Zhang,2 and Pu Wang1

1School of Traffic and Transportation Engineering, Rail Data Research and Application Key Laboratory of Hunan Province, Central South University, Changsha 410000, China
2Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518000, China

Correspondence should be addressed to Pu Wang; wangpu@csu.edu.cn

Received 7 January 2022; Revised 1 May 2022; Accepted 13 May 2022; Published 1 June 2022

Academic Editor: Luigi Dell’Olio

Copyright © 2022 Bao Guo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Predicting individual mobility of subway passengers in large crowding events is crucial for subway safety management and crowd control. However, most previous models focused on individual mobility prediction under ordinary conditions. Here, we develop a passenger mobility prediction model, which is also applicable to large crowding events. The developed model includes the trip-making prediction part and the trip attribute prediction part. For trip-making prediction, we develop a regularized logistic regression model that employs the proposed individual and cumulative mobility features, the number of potential trips, and the trip generation index. For trip attribute prediction, we develop an n-gram model incorporating a new feature, the trip attraction index, for each cluster of subway passengers. The incorporation of the three new features and the clustering of passengers considerably improves the accuracy of passenger mobility prediction, especially in large crowding events.

1. Introduction

Large crowding events, which may be triggered by religious, commercial, or recreational activities [1–3], are observed around the world. In large crowding events, crowd density often exceeds its safety threshold [1], posing tremendous pressure to the safety of crowd. If crowd density cannot be effectively controlled, crowd disasters may occur. Over the last decade, crowd disasters have caused more than 2,000 fatalities around the world [4]. Just on April 30, 2021, a total of 45 people were killed and about 150 people were injured in Israel when celebrating the Jewish holiday of Lag BaOmer. Thus, an in-depth understanding of mobility patterns in large crowding events is crucial for mobility anomaly detection and public safety management.

Human mobility patterns in large crowding events are different from those under ordinary conditions [5]. Remarkable efforts have been devoted to investigate the collective human mobility behaviors [6, 7], such as in large crowding events [8]. Fan et al. [9] predicted the short-term citizen movement using multiple random Markov chains. Using the mobile phone GPS log data, Fan et al. predicted the number of people went to the New Year’s Eve activities and the Comiket 80 (C80) show. Dong et al. [10] collected five months of call detail records (CDR) data of the mobile phone users in Ivory Coast. The authors employed the cylindrical cluster detection algorithm to compare the travel patterns of residents in unusual activities and daily routines. Zhou et al. [11] introduced the gradient boosting decision tree (GBDT) algorithm to assess the risk of crowding events. After studying the crowding events occurred in Beijing, Shanghai, and Shenzhen, Zhou et al. found that early warning of potential crowd events could be released 1–3 hours in advance by using the Baidu Map query data. Lam et al. [12] used bike sharing data collected in Washington DC and Philadelphia to identify the unusual events. The authors successfully identified unusual activities near Drexel University through spectral clustering and threshold-based detection. Taken together, collective human mobility in large crowding events has been comprehensively studied in recent years. However, individual human mobility in large crowding events is still not sufficiently explored.
With increasingly abundant big data, individual human mobility research has experienced rapid developments over the last decade. Mobile phone data [13, 14], social media data [15], taxi GPS data [16], automatic fare collection (AFC) data [17], and online-map location data [18] were used to study and discover the laws of human mobility. In individual mobility prediction area, Gambos et al. [19] explored the next location prediction problem using the Markov model early in 2012. In the recent decade, three further research topics have received more attention. In the first topic, various methods are employed to model individual mobility behaviors and predict an individual’s next location, such as hidden Markov model [20], suffix trees [21], association rule mining [22], logistic regression [23], and recurrent neural network [24]. It is noted that abundant travel records are essential for modeling individual mobility patterns, and researchers explored hybrid models that integrated individual historical mobility data and collective mobility data of a group of individuals, which formed the second research topic. For example, Asahara et al. [25] divided individuals into different groups and proposed the mixed Markov model (MMM) to predict the mobility patterns of individuals in each group; Alhasoun et al. [26] identified the “similar strangers” of each individual based on their historical mobility patterns and predicted an individual’s mobility by integrating his/her historical mobility information and the mobility patterns of his/her similar strangers. Noticing the importance of time information of a trip, in the third research topic, researchers predict not only the location of the later trip but also the time when the later trip will take place. For example, Hsieh et al. [27] extracted a location time distribution (LTD) for each location and a transition time distribution (TTD) for each location pair, and individual mobility prediction is made based on how well the location sequence matches the TTD and LTD; using smart card data, Zhao et al. [28] predicted the three trip attributes for each trip, that is, the departure time, the origin, and the destination of the trip. Trip attributes of the first trip in a day were predicted based on the date factors, while trip attributes of the other trips were predicted based on the attributes of the previous trip. However, only a few works investigated human mobility patterns in nonordinary conditions [29, 30]. Individual human mobility in large crowding events has not been well understood and modeled [31].

In summary, previous works focused on collective mobility patterns in crowding events [9–12] or individual mobility patterns in ordinary condition [19–28]. Individual mobility patterns in crowding events are not sufficiently understood and modeled. The anomalous passenger flow during crowding events could not be well reproduced if we use the individual mobility model established for ordinary condition. To fill the research gap, we employ the subway smart card data of 2.5 million passengers of Shenzhen Metro to study the passenger mobility patterns in a large crowding event and develop a hierarchical passenger mobility prediction model. We discovered two primary changes of individual mobility patterns in crowding events, that is, subway passengers make more trips in crowding events, and only the passenger flow directing to the crowding station increased prominently. Therefore, in the present study, the trip generation index is proposed to capture the increased trips in crowding events, and the trip attraction index is introduced to capture subway passengers’ selection of trip destination. By introducing the trip generation index and the trip attraction index, the proposed model can well predict the individual mobility patterns of passengers in crowding events and reproduce the anomalous passenger flow at the crowding station.

The main contributions of the present study could be summarized as follows:

(1) We introduce two new features, the number of potential trips and the trip generation index, in trip-making prediction, which improves the accuracy of trip-making prediction (whether a passenger will make a trip).

(2) We develop a hierarchical n-gram model by incorporating a new feature, the trip attraction index, for each group of clustered passengers, which improves the accuracy of trip attribute prediction, especially in large crowding events.

The remainder of this study is organized as follows: Section 2 introduces the data used in this study. Section 3 presents the model for predicting individual mobility of subway passengers. In Section 4, the proposed model is validated by comparing with several benchmark models. Section 5 concludes the findings of this study. Limitations of the research and future research directions are also discussed.

2. Data

The data used in this study were provided by Shenzhen Transportation Authority. The geographic information systems (GIS) data of the Shenzhen subway network were collected in 2014 when there were 5 lines and 118 stations (Figure 1). The smart card data were collected from October...
13 to October 31, 2014. When a subway passenger entered or exited a subway station, she was required to tap her card, and the passenger smart card ID, the station ID, and the time were recorded. The ID of each smart card holder was anonymized in the used smart card data. In addition, in the present study, only statistical results are analyzed and presented for privacy protection. During the observation period, 2,520,455 passengers generated a total of 30,669,097 smart card records. Shenzhen Metro uses the off-board ticketing system with proof-of-payment (BTTS-POpA), where fare evasion is almost absent [32]. Using the smart card data, we estimated the average number of trips of smart card users per day, which is 1.61 million. We also obtained that the total number of subway trips per day was approximately 2.84 million in Shenzhen in 2014 [33]. That is, the used smart card data recorded roughly 57% of the subway trips, and the other 43% passengers used one-way tickets. Therefore, the passenger demand was estimated based on an up-scaling factor $\beta$:

$$T(i, j, t) = N(i, j, t) \times \beta,$$

where $i$ is the trip origin, $j$ is the trip destination, and $t$ is the 1-hour time slot, $N(i, j, t)$ is the demand of smart card users, $T(i, j, t)$ is the estimated passenger demand, and the up-scaling factor is estimated to be $\beta = 1.77$. A crowding event caused by Halloween recreational activity took place near the Window of World station on October 31, 2014. This offers us the data for studying the passenger mobility patterns during large crowding events.

In this study, the smart card records collected from October 13 to October 26, 2014, are used as the training data, and the smart card records collected from October 27 to October 31, 2014, are used as the test data. The Window of World station is here denoted as the crowding station $d_c$. A large number of passengers took subway to the Window of World station during the time period of the crowding event $T_c$, which was from 4:00 p.m. to 11:30 p.m. (Figure 2(a)). We find that only the passenger flow directing to the crowding station increased prominently during $T_c$. The passengers who exited station $d_c$ during time period $T_c$ are defined as the event participants. In this study, only the passengers with more than 14 trips (571,171 passengers in total) during the observation period are used to guarantee sufficient historical individual mobility data.

3. Model

3.1 Modeling Framework. There are three modules in our proposed framework, that is, feature extraction, individual mobility prediction, and passenger clustering (Figure 3).

In the feature extraction module, multiple mobility features are extracted from smart card data to establish the individual mobility prediction model, including the number of potential trips, the trip generation index, and the trip attraction index.

The individual mobility prediction module is used to predict the individual mobility patterns of subway passengers. There are two ways to predict the location of an individual. One way is to predict the individual location for each time slot. The other way is to predict the departure time, the origin, and the destination of the next trip. In this study, the second way was employed given that the mobility information provided by the smart card data is sparsely distributed in time. Inspired by Zhao et al. [28], the individual mobility prediction problem is split into two parts: trip-making prediction (TM) and trip attribute prediction (TA) (Figure 3). Trip-making prediction determines if a passenger will make a trip. The number of potential trips and the trip generation index are used in trip-making prediction to quantify the traveling willingness of subway passengers. Trip attribute prediction determines the departure time $t$, the origin $o$, and the destination $d$ of the trip. The trip attraction index is used in trip attribute prediction to capture the attraction of subway stations to the passengers. The training and prediction procedures of the proposed individual mobility model are as follows: at the start of a day, trip-making prediction is first conducted to identify if the passenger will make the first trip in the day (TM1). If so, the trip attributes of the first trip are predicted (TA1); otherwise, prediction is over for the day. After that, whenever a trip record is observed, trip-making prediction determines if the later trip will occur (TML). If the later trip will not occur, prediction is over for the day; otherwise, the trip attributes of the later trip are predicted (TAL). TML and TAL will be repeatedly conducted until no new trip record is observed within the day.

The passenger clustering module is introduced to cluster the passengers with similar temporal mobility patterns. Note that individuals tend to travel like their “similar strangers” [26]. By generating the time series, generating the overlapped slots, and extracting the temporal mobility patterns, the studied passengers are clustered into groups based on their temporal mobility patterns. Consequently, the silhouette coefficient is employed to identify the optimal number of passenger clusters. When training the individual mobility model (TAF and TAL) for a cluster of passengers, only mobility data of passengers in the studied cluster are used, while the mobility data of other clusters of passengers are not considered. The parameter and variable notations are introduced in Table 1.

3.2 Trip-Making Prediction. At the start of a day, the probability that a passenger makes the first trip (TMF) is estimated using the conditional probability $P(y_1|F_t) \cdot \Pi$, where $y_1$ is a binary variable indicating if the passenger will make the first trip in the day; $F_t$ is the feature set, which includes the date information and the passenger’s historical mobility information; $\Pi$ is the trip generation index proposed in this study. Trip generation index quantifies passengers’ overall traveling willingness in the day, which is a cumulative mobility feature of all passengers. When the $(i-1)$th trip of the passenger is observed, the probability that a passenger makes the later trip (TML) is estimated using the conditional probability $P(y_i|t_{i-1}, o_{i-1}, d_{i-1}) \cdot \Pi$, where $i \geq 2$ represents the $(i)$th trip in the day, $y_i$ determines if the $(i)$th trip will occur in the day, and $t_{i-1}, o_{i-1}, d_{i-1}$ represent the departure time, the origin, and the destination of the observed $(i-1)$th trip in the day, respectively.
Trip generation index $\Pi$ is calculated using the total number of trips in the event day divided by the average number of trips in ordinary days. $\Pi$ is set to 1 for ordinary days. In the event day, the value of $\Pi$ is larger than 1, indicating that passengers are more willing to travel in that day. Given that passengers may go to the crowding location in their first trip or the later trips, trip generation index is considered in both TMF and TML.

### 3.3. Trip Attribute Prediction

For the first trip of a day, its attributes (TAF) are predicted using the conditional probability $P_k(t_1, o_1, d_1 | F_1) \cdot \Phi$, where $k$ is the $k$th cluster, $t_1, o_1, d_1$ represent the departure time, the origin, and the destination of the first trip, respectively, $F_1$ represents the date features, and $\Phi$ is the trip attraction index. Similar to TML, the attributes of the later trip (TAL) are predicted using the conditional probability $P_k(t_i, o_i, d_i | F_i) \cdot \Phi$, where $k$ is the $k$th cluster, $i \geq 2$ represents the ($i$)th trip in the day, and $t_i, o_i, d_i$ represent the departure time, the origin, and the destination of the ($i$)th trip, respectively. The combination of $t_i, o_i, d_i$ ($i \geq 1$) achieving the largest conditional probability is regarded as the predicted trip attributes. In this study, the resolution of the time is 1 hour.

The method to determine the trip attraction index $\Phi$ is as follows:

1. Calculate the attraction coefficient $\varphi$ of the crowding station for each passenger:
   \[
   \varphi = \frac{P(d_i | F)}{P(d_{\text{modal}} | F)}
   \]
   where $d_i$ is the crowding station, $P(d_i | F)$ is the conditional probability that the destination of the
Table 1: Parameter and variable notations.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Π</td>
<td>Trip generation index</td>
</tr>
<tr>
<td>Φ</td>
<td>Trip attraction index</td>
</tr>
<tr>
<td>y_i (i ∈ N^+)</td>
<td>If the passenger will make the i-th trip in the day</td>
</tr>
<tr>
<td>F_i</td>
<td>The feature set that includes the date information and the passenger’s historical mobility information</td>
</tr>
<tr>
<td>t_i, o_i, d_i (i ∈ N^+)</td>
<td>The departure time, the origin, and the destination of the i-th trip in the day</td>
</tr>
<tr>
<td>F_k</td>
<td>The date features</td>
</tr>
<tr>
<td>k</td>
<td>The k-th cluster</td>
</tr>
<tr>
<td>Variables</td>
<td>Description</td>
</tr>
<tr>
<td>P_k(t_i, o_i, d_i</td>
<td>The joint probability of the first trip for a passenger in the k-th cluster</td>
</tr>
<tr>
<td>P_k(t_i, o_i, d_i</td>
<td>The joint probability of the later trip for a passenger in the k-th cluster</td>
</tr>
<tr>
<td>P_k(o_i</td>
<td>F_2, t_i)</td>
</tr>
<tr>
<td>C(F_2, t_i, o_i)</td>
<td>The time-smoothed counting function of the passenger</td>
</tr>
<tr>
<td>m_k(o_i</td>
<td>F_2, t_i)</td>
</tr>
<tr>
<td>c(F_2, t_i, o_i)</td>
<td>The number of times that (F_2, t_i, o_i) emerges in the training data of the passenger</td>
</tr>
<tr>
<td>A(F_2, t_i)</td>
<td>The collection of contexts adjacent to context (F_2, t_i) of the passenger</td>
</tr>
<tr>
<td>P_0(o_i</td>
<td>F_2, t_i)</td>
</tr>
<tr>
<td>C_0(F_2, t_i, o_i)</td>
<td>The collective time-smoothed counting function for the passengers in the k-th cluster</td>
</tr>
<tr>
<td>c_0(F_2, t_i, o_i)</td>
<td>The number of times that (F_2, t_i, o_i) emerges in the training set for the passengers in the k-th cluster</td>
</tr>
<tr>
<td>A_0(F_2, t_i)</td>
<td>The collection of contexts adjacent to context (F_2, t_i) for a passenger in the k-th cluster</td>
</tr>
</tbody>
</table>

trip is d_e under condition F, F includes the date features F_2 for TAF and the attributes of last trip (t_{i-1}, o_{i-1}, d_{i-1}) for TAL, d_{most} is the subway station where the passenger visited most under condition F in ordinary days, and P(d_{most}|F) is the conditional probability that the destination of the passenger is d_{most} under condition F. For the passengers whose most visited station is the crowding station, the attraction coefficient is set to ϕ = 1. While for other passengers, their attraction coefficient ϕ < 1. The smaller the attraction coefficient ϕ, the lower the probability that a passenger visits the crowding station. The attraction coefficient ϕ reflects the relative attraction of the crowding station compared with other stations.

(2) Arrange the attraction coefficients ϕ in a descending order for all studied passengers (ϕ_1 > ϕ_2 > … > ϕ_N, where N is the total number of passengers).

(3) The trip attraction index Φ of a crowding event is calculated as follows:

Φ = \frac{1}{ϕ_n} \quad (3)

where n is the number of event participants. For a passenger i (i ≤ n), if P(d_e|F) · Φ = P(d_e|F) · (1/ϕ_n) ≥ P(d_{most}|F), he/she is predicted to go to the crowding station (event). This type of passengers is identified as the predicted participants. While for a passenger i (i > n), if P(d_e|F) · Φ = P(d_e|F) · (1/ϕ_n) < P(d_{most}|F), he/she is predicted to go to the most visited location d_{most}. The number of predicted participants is equal to the number of participants n. The trip attraction index Φ > 1 only when destination of the trip d_e is the crowding station d_e and departure time of the trip t_e is within the time period of the crowding event T_e; in other cases, Φ = 1.

3.4. Clustering Subway Passengers into Groups. In trip attribute prediction, passengers with similar historical mobility patterns are clustered into the same group. The trip attributes of passengers in each group are predicted using the models specifically constructed for the group.

We cluster subway passengers based on their temporal mobility patterns [34]. First, a day is divided into 18 one-hour time slots since the operation hours of the Shenzhen Metro were from 6:00 a.m. to 12:00 a.m. For each passenger, a time series T = {T_1, T_2, …, T_18} is generated to extract his/her temporal mobility pattern (Figure 4), where T_j = 1 when the passenger swiped the card in the time slot with the highest proportion of active days D_j, which is the number of days in the training data.

We generate the overlap slots O = {O_1, O_2, …, O_j, …, O_{16}}, each of which lasts for 3 hours. The attributes of each overlap slot O_j are the average number of trips H_j and the proportion of active days D_j, both of which are calculated according to the time series T:

\[ H_j = \frac{1}{D} \sum_{i=1}^{D} (T_j^{i} + T_{j+1}^{i} + T_{j+2}^{i}), \]

\[ D_j = \frac{1}{D} \sum_{i=1}^{D} \text{MAX}(T_j^{i}, T_{j+1}^{i}, T_{j+2}^{i}), \]

where the operator MAX returns the maximum value. The overlap slots are generated to identify the typical temporal mobility patterns of subway passengers. There are two ways to split the time slots of a day. One way is to split a day into fixed intervals as tagged time slots, such as 0:00–2:59, 3:00–5:59, 6:00–8:59, …, the other way is to split a day into overlap slots: 0:00–2:59, 1:00–3:59, 2:00–4:59, …, If we use the first way to split time slots, passengers with similar temporal mobility patterns may be grouped into different clusters. For example, if
passenger $A$ takes subway during the period from 8:50 to 8:59 and passenger $B$ takes subway during the period from 9:01 to 9:10; the two passengers will be regarded as the data samples of different tagged time slots (i.e., 6:00–8:59 for passenger $A$ and 9:00–11:59 for passenger $B$), albeit that their temporal mobility patterns are rather similar. However, if we use the overlap slots, passenger $A$ and passenger $B$, which have pretty similar temporal mobility patterns, will be regarded as the data samples of the same overlap slot (i.e., 7:00–9:59 or 8:00–10:59). Therefore, the overlap slots can better capture the temporal mobility patterns of subway passengers.

We sort the overlap slots $O$ in descending order of $H_j$ and generate the sorted overlap slots $O' = \{O_1', O_2', \ldots, O_{16}'\}$. Then, we iteratively select $O_j'$ from $O_i'$ to $O_{16}'$ to generate the nonoverlap slots $O'' = \{O_1'', O_2'', \ldots, O_{16}''\}$, where the elements in $O''$ have no overlap in time. The first slot in the nonoverlap slots $O''$ ($O_1''$) is the first slot in the sorted overlap slots $O'$ ($O_1'$, 6:00–8:59). Given that there are overlapped time slots between $O_j'$ and $O_j$ (i.e., 7:00–8:59), $O_j$ is not selected to add into $O''$. Next, given that there are no overlapped time slots between $O_j'$ and $O_j$, $O_j$ is selected as the second slot in $O''$. We iteratively select $O_j'$ from $O_j'$ to $O_{16}'$ to generate the nonoverlap slots. Finally, $O_1'', O_3'', O_{14}''$, and $O_{15}''$ are selected to generate the nonoverlap slots $O''$. The proportions of active days $D_j$ of the top four elements in $O''$ are used as features for passenger clustering [34].

The $K$-means algorithm is used to cluster passengers into different groups. We use $\{D_{top1}, D_{top2}, D_{top3}, D_{top4}\}$ to generate the feature space, where $D_{top}$ is the proportion of active days of the top $i$ element in the nonoverlap slot $O''$. The silhouette coefficient [35] is employed to determine the number of optimal cluster number $K$. The number of passenger groups is tested from 2 to 17 (i.e., $K \in \{2, 3, \ldots, 17\}$), and the value of $K$ that achieves the highest silhouette coefficient is used.

3.5. Solving the Model. For trip-making prediction, a binary variable $y_i (i \geq 1)$ determines if the $i$th trip will occur in the day. The binary variable $y_i (i \geq 1)$ in TMF and TML can be solved using the logistic regression algorithm [28]. In order to avoid overfitting and reduce model complexity, L2 regularization [36] is employed in the logistic regression algorithm. Through inputting the feature set $F_i$ and the trip generation index $\Pi$, we train the L2 regularization model for TMF and predict if the passenger will make the first trip in a day. We input the features $t_{i-1}, o_{i-1}, d_{i-1}, p$, and $\Pi$ to train the L2 regularization model for TML and predict if the later trip will occur.

For trip attribute prediction, to reduce the feasible solution space and simplify the solution process of TAF and TAL, the joint probability $P_k(t_i, o_i, d_i|F_2)$ in TMF and the joint probability $P_k(t_i, o_i, d_i|t_{i-1}, o_{i-1}, d_{i-1})$ in TAL are respectively decomposed into the product of three conditional probabilities based on the chain rule [28]:

$$P_k(t_i, o_i, d_i|F_2) = P_k(t_i|F_2) \cdot P_k(o_i|F_2, t_1) \cdot P_k(d_i|F_2, t_1, o_1),$$

$$P_k(t_i, o_i, d_i|t_{i-1}, o_{i-1}, d_{i-1}) = P_k(t_i|t_{i-1}, o_{i-1}, d_{i-1}) \cdot P_k(o_i|t_{i-1}, o_{i-1}, d_{i-1}, t_1) \cdot P_k(d_i|t_{i-1}, o_{i-1}, d_{i-1}, t_1, o_1).$$

Figure 4: Illustration of the generation of nonoverlap slots.
In natural language processing, n-gram model is widely used to predict the next word or phrase based on given context and shows good performance in machine translation, speech recognition, and other applications [37]. We use the n-gram model [38] to solve the conditional probability obtained by the chain rule decomposition.

In equation (7), the origin of the first trip in a day \( o_i \) is the variable to be predicted, where \( F_2 \) is the date feature, \( t_i \) is the departure time of the first trip, and \( (F_2, t_i, o_i) \) represents that the origin of the first trip in a day is \( o_i \) with given context \((F_2, t_i)\). For all possible \( o_i \), we calculate the conditional probability \( P_k(o_i|F_2, t_i) \) using the given context \((F_2, t_i)\):

\[
P_k(o_i|F_2, t_i) = \frac{C(F_2, t_i, o_i) + \alpha \cdot m_k(o_i|F_2, t_i)}{\sum_{o_i \in V} C(F_2, t_i, o_i) + \alpha},
\]

where \( C(F_2, t_i, o_i) \) is the time-smoothed counting function, \( \alpha \) is a scalar concentration parameter of Dirichlet prior, \( m_k(o_i|F_2, t_i) \) is the expectation of prior probability \( P_k(o_i|F_2, t_i) \), and \( V \) is the feasible solution set of \( o_i \). The time-smoothed counting function \( C(F_2, t_i, o_i) \) returns the estimated frequency of \((F_2, t_i, o_i)\) and is calculated using the following equation:

\[
C(F_2, t_i, o_i) = \frac{c(F_2, t_i, o_i) + \sum_{(F_2', t_i') \in A(F_2, t_i)} C(F_2', t_i', o_i)}{1 + |A(F_2, t_i)|}
\]

where \( c(F_2, t_i, o_i) \) is the number of times that \((F_2, t_i, o_i)\) emerges in the training set, and \( A(F_2, t_i) \) is the collection of contexts “adjacent” to context \((F_2, t_i)\). Context \((F_2, t_i)\) is adjacent to context \((F_2, t_i)\) if \( t_i - t_i' = 1 \) or \( t_i - t_i' = -1 \).

The expected prior probability \( m_k(o_i|F_2, t_i) \) is estimated using the following equation:

\[
m_k(o_i|F_2, t_i) = \beta \cdot P_k(o_i|t_i) + (1 - \beta) \cdot P^o_k(o_i|F_2, t_i),
\]

where \( P_k(o_i|t_i) \) is the back-off prior of \( P_k(o_i|F_2, t_i) \) [39], \( P^o_k(o_i|F_2, t_i) \) is a collective prior of \( P_k(o_i|F_2, t_i) \), \((n-1)\)-gram model \( P_k(o_i|t_i) \) and collective \( n\)-gram model \( P^o_k(o_i|F_2, t_i) \) are balanced using the parameter \( \beta \). The collective probability \( P^o_k(o_i|F_2, t_i) \) is calculated as follows:

\[
P^o_k(o_i|F_2, t_i) = \frac{C^o_k(F_2, t_i, o_i) + \alpha_0 \cdot P^o_k(o_i|t_i)}{\sum_{o_i' \in V} C^o_k(F_2, t_i, o_i') + \alpha_0}
\]

where \( C^o_k(F_2, t_i, o_i) \) is a counting function representing the travel behaviors of all passengers in the k-th cluster, \( P^o_k(o_i|t_i) \) is a collective prior of \( P_k(o_i|t_i) \), and \( \alpha_0 \) is a scalar concentration parameter of Dirichlet prior and is used to represent the effect of collective behaviors on individual mobility prediction. The counting function \( C^o_k(F_2, t_i, o_i) \) is estimated using the following equation:

\[
C^o_k(F_2, t_i, o_i) = \frac{c^o_k(F_2, t_i, o_i) + \sum_{(F_2', t_i') \in A^o_k(F_2, t_i)} C^o_k(F_2', t_i', o_i)}{1 + |A^o_k(F_2, t_i)|}
\]

where \( c^o_k(F_2, t_i, o_i) \) represents the frequency of \((F_2, t_i, o_i)\) in the training data and \( A^o_k(F_2, t_i) \) is the collection of contexts “adjacent” to context \((F_2, t_i)\) for all passengers in the k-th cluster.

Note that the conditional probabilities \( P_k(F_2, t_i, o_i) \) and \( P^o_k(F_2, t_i, o_i) \) are calculated recursively. When the number of context variables \((F_2, t_i)\) is zero, the individual probability that the first trip starts from \( o_i \) and \( (P_k(o_i)) \) and the collective probability that the first trip starts from \( o_i \) \((P^o_k(o_i)) \) are calculated using the following equations:

\[
P_k(o_i) = \frac{C(o_i) + \alpha \cdot P(o_i)}{\sum_{o_i \in V} C(o_i) + \alpha}
\]
\[
P^o_k(o_i) = \frac{C^o(o_i) + \alpha_0 (1/|V|)}{\sum_{o_i' \in V} C^o(o_i') + \alpha_0}
\]

In order to improve the model performance, parameters \( \alpha \) and \( \beta \) are determined for each individual [28]. The training data are further divided into training part (from October 13 to October 19) and validation part (from October 20 to October 26). The value of parameters \( \alpha \) and \( \beta \) are tested from 0 to 1 with a tolerance of 0.1. The values of \( \alpha \) and \( \beta \) that achieve the highest average prediction accuracy rates are selected.

Next, the conditional probabilities \( P_k(t_i|F_2) \), \( P_k(d_i|F_2, t_i, o_i) \), \( P_k(t_i|t_{i-1}, o_{i-1}, d_{i-1}) \), \( P_k(o_i|t_{i-1}, o_{i-1}, d_{i-1}, t_i) \), and \( P_k(d_i|t_{i-1}, o_{i-1}, d_{i-1}, t_i, o_i) \) can be calculated, which are further used to estimate the joint probability \( P^o_k(t_i, o_i, d_i|F_2) \) in TAF equation (5) and the joint probability \( P_k(t_i, o_i, d_i|F_2) \) in TAL (equation 6).

### 3.6. Prediction Performance Indices

There are three indices used to measure the prediction performance of trip-making prediction, that is, prediction accuracy rate, recall, and precision. The prediction accuracy rate of a passenger is defined as the number of correctly predicted binary variables \( y_i \) divided by the total number of binary variables \( y_i \). The recall of a passenger is defined as the number of correctly predicted positive binary variables \( y_1 \) divided by the total number of positive binary variables \( y_1 \). The precision of a passenger is defined as the number of correctly predicted positive binary variables \( y_1 \) divided by the total number of predicted positive binary variables \( y_1 \).

Meanwhile, four indices are used to measure the prediction performance of trip attribute prediction, that is, prediction accuracy rate for the prediction of individual mobility, and mean absolute percentage error (MAPE), mean absolute error (MAE), and root mean squared error (RMSE) for the prediction of the number of event participants from the passenger sources of crowding station. The prediction accuracy rate is defined as the number of trips with trip attribute correctly predicted divided by the total number of trips in the test data. The indices MAPE, MAE, and RMSE are defined as follows:
3.7. Comparative Models. In this study, three benchmark models are established to validate the proposed nfg-gram model, that is, a first-order Markov chain model (MC model), a basic n-gram model, and a n-gram model incorporating the proposed indices (nfg-gram model).

In the MC model, the transition probabilities \( P(t_i|t_{i-1}) \), \( P(o_i|d_{i-1}) \), and \( P(d_i|d_{i-1}) \) are obtained based on the mobility patterns in the training data:

\[
P(t_i|t_{i-1}) = \frac{c(t_{i-1}, t_i) + \alpha}{\sum_{t \in T} c(t_{i-1}, t) + \alpha},
\]

\[
P(o_i|d_{i-1}) = \frac{c(d_{i-1}, o_i) + \alpha}{\sum_{o \in O} c(d_{i-1}, o) + \alpha},
\]

\[
P(d_i|d_{i-1}) = \frac{c(d_{i-1}, d_i) + \alpha}{\sum_{d \in D} c(d_{i-1}, d) + \alpha},
\]

where \( c(t_{i-1}, t_i) \) is the number of times that the last trip starts in time slot \( t_{i-1} \) and the later trip starts in time slot \( t_i \), \( c(d_{i-1}, o_i) \) is the number of times that the destination of the last trip is \( d_{i-1} \) and the origin of the later trip is \( o_i \), \( c(d_{i-1}, d_i) \) is the number of times that the destination of the last trip is \( d_{i-1} \) and the destination of the later trip is \( d_i \), \( |T| \) is the number of possible time slots that a passenger may start a trip, \( |L| \) is the number of possible subway stations that the passenger may enter or exit, and parameter \( \alpha \) is also used for smoothing. Next, the mobility probability of the later trip \((t_i, o_i, d_i)\) can be calculated based on the observed trip \((t_{i-1}, o_{i-1}, d_{i-1})\).

In the n-gram model, the trip-making prediction is estimated using the conditional probability \( P(y_i|F_i) \) for TMF and \( P(y_i|F_1, F_2) \) for TML, and the trip attribute prediction is estimated using the conditional probability \( P(t_i, o_i, d_i|F_2) \) for TAF or \( P(t_i, o_i, d_i|F_1, F_2) \) for TAL.

In the nfg-gram model, the trip-making prediction is estimated using the conditional probability \( P(y_i|F_1) \) · \( P(t_i, o_i, d_i|F_2) \) · \( P(\pi) \) for TMF and \( P(y_i|F_1, F_2) \) · \( P(\pi) \) for TML, where the proposed trip generation index \( \pi \) is used to capture the increased trips in crowding events. The trip attribute prediction is estimated using the conditional probability \( P(t_i, o_i, d_i|F_2) \) · \( P(\pi) \) for TAF and \( P(t_i, o_i, d_i|F_1, F_2) \) · \( P(\pi) \) for TAL, where the proposed trip attribute index \( \pi \) is used to capture subway passengers’ selection of trip destination. The conditional probabilities of n-gram model and nfg-gram model are obtained using the mobility data of all studied passengers.

4. Results

4.1. Trip-Making Prediction. The problem of trip-making prediction includes TMF and TML, which are solved by estimating the conditional probabilities \( P(y_i|F_1) \) · \( P(t_i, o_i, d_i|F_2) \) · \( P(\pi) \) and \( P(y_i|F_1, F_2) \) · \( P(\pi) \). The average number of trips is 1,601,737 trips per day in ordinary days, while there were 1,837,826 trips in the event day. The trip generation index \( \pi \) is therefore set to 1.15 in the event day. Feature set \( F_1 \) contains the date features and the mobility characteristics of the passenger (Table 2). The Category A features include the day of week, the day is a holiday or not, and passenger travelled in the last day or not, which are common features in individual mobility prediction. The feature, day of week, is categorical variables and converted into a series of binary variables (e.g., whether it is Monday or not). Note that there are no holidays in the used data. The Category B features include travel frequency and nontravel days. We propose a new Category C feature, the number of potential trips \( p \), which is defined as the difference between the average number of daily trips \( N \) of the passenger in ordinary days and the number of observed trips of the passenger in the
predicted day. Based on Categories A, B, and C features, we employ the L2 regularization model in the trip-making prediction.

For TMF, we predict if a passenger will make a trip. The number of potential trips of a passenger \( p \) is set to the average number of daily trips of the passenger \( N \) in ordinary days. The medians of coefficients in the L2 regularization model are presented in Table 2. Results show that passengers are more willing to travel on weekdays than on weekends. In addition, if a passenger makes a trip in last day, he/she is more active and willing to make a trip in the current day. We also notice that passengers with large number of potential trips \( p \) are more willing to travel.

For TML, we predict if the later trip will occur based on five variables, namely, the order of the later trip \( i \), the number of potential trips \( p \), and the departure time \( t_{i-1} \), the origin \( o_{i-1} \), and the destination \( d_{i-1} \) of last trip. In this study, the departure time \( t_{i-1} \), the origin \( o_{i-1} \), and the destination \( d_{i-1} \) of last trip are converted to a series of binary variables. The departure time of last trip \( t_{i-1} \) is expected to have a great impact in trip-making prediction because the probability of a passenger making the later trip decreases with time. The destination of last trip \( d_{i-1} \) is also important because the last trip of a day usually ends at the same place. Meanwhile, the probability of a passenger making another trip decreases with the number of potential trips \( p \). The coefficients of the variables are unique for each individual and are not presented here.

Our results show that using Categories A + B features can effectively predict if the first trip will occur, and the accuracy of first trip prediction is 88.18% (Table 3, TMF). However, when using the proposed new feature (i.e., the number of potential trips) to replace the Category B features, the prediction accuracy increases to 88.41% (Table 3, TMF). This highlights the effectiveness of the proposed feature in passenger mobility prediction. By combining the Categories A + B + C features, the accuracy of prediction keeps increasing to 88.77%, and the recall increases to 93.96% (Table 3, TMF). The proposed nfc-gram model reaches the highest prediction accuracy (i.e., 88.84%) and the highest recall (i.e., 96.86%). Similarly, the incorporation of the

<table>
<thead>
<tr>
<th>Subproblem</th>
<th>Explanation of the features</th>
<th>Accuracy (%)</th>
<th>Recall (%)</th>
<th>Precision (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMF</td>
<td>A + B</td>
<td>88.18</td>
<td>93.22</td>
<td>92.90</td>
</tr>
<tr>
<td></td>
<td>A + C</td>
<td>88.41</td>
<td>93.84</td>
<td>92.95</td>
</tr>
<tr>
<td></td>
<td>A + B + C</td>
<td>88.77</td>
<td>93.96</td>
<td>92.95</td>
</tr>
<tr>
<td></td>
<td>A + B + C + Π</td>
<td>88.84</td>
<td>96.86</td>
<td>92.63</td>
</tr>
<tr>
<td></td>
<td>( t_{i-1}, o_{i-1}, d_{i-1}, i )</td>
<td>81.55</td>
<td>81.29</td>
<td>87.69</td>
</tr>
<tr>
<td>TML</td>
<td>( t_{i-1}, o_{i-1}, d_{i-1}, i, p )</td>
<td>82.42</td>
<td>84.56</td>
<td>88.25</td>
</tr>
<tr>
<td></td>
<td>( t_{i-1}, o_{i-1}, d_{i-1}, i, p, Π )</td>
<td>82.82</td>
<td>84.65</td>
<td>88.61</td>
</tr>
</tbody>
</table>

Table 4: The average prediction accuracy rates of different models.

<table>
<thead>
<tr>
<th>Subproblem</th>
<th>Model</th>
<th>( t_i ) (%)</th>
<th>( o_i ) (%)</th>
<th>( d_i ) (%)</th>
<th>Average (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAF</td>
<td>MC model</td>
<td>30.5</td>
<td>50.0</td>
<td>43.5</td>
<td>41.3</td>
</tr>
<tr>
<td></td>
<td>n-gram model</td>
<td>43.5</td>
<td>58.6</td>
<td>57.2</td>
<td>53.1</td>
</tr>
<tr>
<td></td>
<td>nfc-gram model</td>
<td>43.5</td>
<td>58.6</td>
<td>57.2</td>
<td>53.1</td>
</tr>
<tr>
<td></td>
<td>nfc-gram model</td>
<td>47.0</td>
<td>63.3</td>
<td>61.4</td>
<td>57.2</td>
</tr>
<tr>
<td>TAL</td>
<td>n-gram model</td>
<td>25.1</td>
<td>45.7</td>
<td>39.2</td>
<td>36.6</td>
</tr>
<tr>
<td></td>
<td>n-gram model</td>
<td>32.6</td>
<td>49.3</td>
<td>45.6</td>
<td>42.5</td>
</tr>
<tr>
<td></td>
<td>nfc-gram model</td>
<td>32.6</td>
<td>49.2</td>
<td>45.3</td>
<td>42.4</td>
</tr>
<tr>
<td></td>
<td>nfc-gram model</td>
<td>46.3</td>
<td>63.2</td>
<td>60.4</td>
<td>56.6</td>
</tr>
</tbody>
</table>
4.2. Trip Attribute Prediction. For trip attribute prediction, we first cluster passengers into different groups based on their temporal mobility patterns. We find that the value of silhouette coefficient reaches the maximum value when the number of groups is set to 3. Hence, the studied passengers are clustered into 3 groups. We develop a hierarchical n-gram model (i.e., nfc-gram model) incorporating a new feature, the trip attraction index, to predict the trip attributes of passengers in each group.

For TAF, only the day of the week is used as the date feature F2 to predict the start time t1, the origin o1, and the destination d1 of the first trip. For TAL, the trip attributes of the later trip are predicted using the trip attributes of last trip (i.e., t1−1, o1−1 and d1−1). We propose a new feature, trip attraction index Φ, to predict trip attributes. The effectiveness of nfc-gram model is validated from three aspects: (1) prediction of individual mobility, (2) prediction of out-passenger flow $f_{out}$ at the crowding station, and (3) prediction of the passenger sources of the crowding station. Several benchmark models, including the MC model, the n-gram model proposed in reference [28], and the nf−gram model (n-gram model incorporating the trip attraction index), are used for comparison. The prediction accuracy rates are listed in Table 4. The MC model cannot capture the individual mobility patterns for both TAF and TAL. The average prediction accuracy rates of the n-gram model is 53.1% for TAF and 42.5% for TAL, implying that n-gram model can better predict passenger mobility attributes than MC model. The nf−gram model achieves the similar prediction accuracy with the n-gram model when predicting trip attributes. The nfc-gram model, which is constructed for each group of passengers specifically, can further improve the average prediction accuracy rates to 57.2% for TAF and 56.6% for TAL.

Out-passenger flow at the crowding station is important for subway management and crowd control. On the event day, 20,823 passengers exited the Window of World station during Tc, and the trip attraction index $\Phi = 1.15$ for nf−gram model and $\Phi = 4.28$ for nfc-gram model according to equations (2) and (3). The out-passenger flow at the Window of World station suddenly increased at 4:00 p.m. and reached its peak at 6:00 p.m. (Figure 5). We find the MC model and the n-gram model are not able to predict the anomalous passenger flow in the large crowding event. By introducing the trip attraction index $\Phi$, the nf−gram model can well reproduce the anomalous passenger flow in crowding events, which suggests that the trip attraction index can be used to improve the performance of trip attribute prediction. In addition, the proposed nfc-gram model can further improve the passenger flow estimation in crowding events (Figure 5).

The trip-making prediction and trip attribute prediction can be conducted as soon as a passenger trip is observed. Hence, we can predict event participants’ trips and corresponding trip attributes before the start of the crowding event. The nfc-gram model can predict 80% of actual passenger flow before the time period of the crowding event $T'$ (Figure 6). This offers sufficient time for deploying crowd control strategies. A total of 2,198 predicted participants arrived at the crowding station in the first trip. Thus, these participants can be predicted at the start of the day (i.e., 6:00 a.m.), therefore providing sufficient time for the implementation of crowd control strategies. For example, we can close streets or place barriers to change the routes of crowds to reduce traffic congestion in streets.

Finally, we predict where the event participants start their trips. Here, only the trips with a destination at the crowding station are analyzed. Table 5 lists the main passenger sources of the event participants from 6:00 p.m. to 7:00 p.m. on the event day. Passengers mainly originated from the subway stations near the crowding station. Table 5 lists that the MC model and the n-gram model could not capture the anomalous OD flow in some crowding events. However, by introducing the trip attraction index, the nf−gram model and nfc-gram model can capture the anomalous passenger.

<table>
<thead>
<tr>
<th>Passenger source</th>
<th>Empirical data</th>
<th>MC model</th>
<th>n-gram model</th>
<th>nf−gram model</th>
<th>nfc-gram model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shenzhen University</td>
<td>873</td>
<td>129</td>
<td>276</td>
<td>504</td>
<td>421</td>
</tr>
<tr>
<td>Qiaocheng North</td>
<td>604</td>
<td>135</td>
<td>522</td>
<td>694</td>
<td>561</td>
</tr>
<tr>
<td>Keyuan</td>
<td>563</td>
<td>198</td>
<td>501</td>
<td>503</td>
<td>673</td>
</tr>
<tr>
<td>Cheggomiao</td>
<td>559</td>
<td>131</td>
<td>481</td>
<td>798</td>
<td>871</td>
</tr>
<tr>
<td>High tech park</td>
<td>545</td>
<td>78</td>
<td>331</td>
<td>428</td>
<td>508</td>
</tr>
<tr>
<td>Huaqiang Road</td>
<td>535</td>
<td>89</td>
<td>434</td>
<td>641</td>
<td>712</td>
</tr>
<tr>
<td>Grand Theater</td>
<td>503</td>
<td>113</td>
<td>605</td>
<td>717</td>
<td>729</td>
</tr>
<tr>
<td>Shopping park</td>
<td>379</td>
<td>71</td>
<td>223</td>
<td>280</td>
<td>255</td>
</tr>
<tr>
<td>Huaqiang North</td>
<td>368</td>
<td>7</td>
<td>233</td>
<td>280</td>
<td>255</td>
</tr>
<tr>
<td>Laojie</td>
<td>361</td>
<td>46</td>
<td>64</td>
<td>198</td>
<td>223</td>
</tr>
</tbody>
</table>

Table 5: The number of event participants from the main passenger sources of crowding station from 6:00 p.m. to 7:00 p.m.
flow in OD pairs, which improves the prediction performance of the model.

5. Conclusions

In summary, we predict the individual mobility of subway passengers in a large crowding event by inducing a new individual mobility feature (i.e., the number of potential trips $p$) and two new cumulative mobility features (i.e., trip generation index $Π$ and trip attraction index $Φ$). Individual mobility prediction is split into the trip-making prediction and the trip attribute prediction, which are respectively solved by the L2 regularization model and the proposed \(nfc\)-gram model. The newly proposed individual and cumulative mobility features, the number of potential trips $p$ and the trip generation index $Π$, can enhance the accuracy of trip-making prediction. The other cumulative mobility feature, that is, trip attraction index $Φ$, captures the anomalous mobility patterns in large crowding events, which improve the prediction accuracy of the trip attributes. The \(nfc\)-gram model is also validated from the aspect of predicting the out-passenger flow at the crowding station and predicting the event participant sources, outperforming several benchmark models.

In this study, we only consider the subway passengers who participate in the crowding events. The taxi passengers and bus passengers involved in crowding events can be further analyzed in the future to improve the accuracy of trip-making prediction and trip attribute prediction. Besides, the trip prediction in the situations when multiple crowding events simultaneously occur may be an interesting future research direction. In addition, passenger demand could be underestimated due to device malfunction or fare evasion [32], which should be further investigated when relevant video surveillance data and subway trip volume data become available.

Data Availability

The subway smart card data and network data used to support the findings of this study have not been made available because of the confidentiality agreement.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The research was supported by the National Natural Science Foundation of China (grant no. 71871224) the 2021 Science and Technology Progress and Innovation Plan of Department of Transportation of Hunan Province (grant no. 212102), and the Hunan Provincial Natural Science Fund for Distinguished Young Scholars.

References


individual mobility for air pollution health studies,” *Environmental Pollution*, vol. 252, pp. 924–930, 2019.


[26] F. Alhasoun, M. Alhazzani, and M. C. González, “City scale next place prediction from sparse data through similar strangers,” in *Proceedings of the KDD ‘17: The 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Halifax NS Canada, August 2017.


