Received: 27 September 2020 | Revised: 18 November 2020 | Accepted: 20 November 2020
DOI: 10.1002/cpe.6151

SPECIAL ISSUE PAPER

WILEY

Practical model with strong interpretability and predictability: An explanatory model for individuals' destination prediction considering personal and crowd travel behavior

Juanjuan Zhao[1] | Jiexia Ye[1] | Minxian Xu[1] | Chengzhong Xu[2]

[1]Shenzhen Institutes of Advanced Technology, China Academic of Science, Shenzhen, China
[2]State Key Lab of IOTSC, Department of Computer Science, University of Macau, Macau SAR, China

Correspondence
Minxian Xu, Shenzhen Institutes of Advanced Technology, China Academic of Science, Shenzhen, China.
Email: mx.xu@siat.ac.cn

Funding information
National Key R&D Program of China, Grant/Award Number: No.2019YFB2102100; National Natural Science Foundation of China, Grant/Award Number: No.61802387, No.61806192; National Natural Science Foundation of Shenzhen, Grant/Award Number: No.JCYJ20190812153212464, No.JCYJ20190812160003719; Science and Technology Development Fund of Macao S.A.R (FDCT), Grant/Award Number: 0015/2019/AKP; Shenzhen Discipline Construction Project for Urban Computing and Data Intelligence; Shenzhen Engineering Research Center for Beidou Positioning Service Improvement Technology, Grant/Award Number: No.XMHT20190101035; SIAT Innovation Program for Excellent Young Researchers

Abstract
Real-time individuals' destination prediction is of great significance for real-time user tracking, service recommendation and other related applications. Traditional technology mainly used statistical methods based on the travel patterns mined from personal history travel data. However, it is not clear how to predict the destinations of individuals with only limited personal historical data. In this paper, taking the public transportation metro systems as example, we design a practical method called practical model with strong interpretability and predictability to predict each passenger's destination. Our main novelties are two aspects: (1) We propose to predict individuals' destination by combining personal and crowd behavior under certain context. (2) An explanatory model combining discrete choice model and neural network model is proposed to predict individuals' stochastic trip's destination, which can be applied to other transportation analysis scenarios about individuals' choice behavior such as travel mode choice or route choice. We validate our method based on extensive experiments, using smart card data collected by automatic fare collection system and weather data in Shenzhen, China. The experimental results demonstrate that our approach can achieve better performance than other baselines in terms of prediction accuracy.

KEYWORDS
crowd behavior, discrete choice model, individual destination prediction, neural network

1 | INTRODUCTION

Recently, a large number of location-based services, for example, navigation service,[1] service recommendation and so on,[2] have gained significant attention. Due to the development of communication technology and ubiquity of human location tracking devices,[3,4] we can collect long-term human travel data. These data provide us opportunity to predict individuals' location. In this paper, taking the public transportation metro systems as example, we predict metro passengers' destinations at individual level based on the collected data by AFC system, through which we can obtain when and where a passenger enters and leaves a metro system. The method could be used to other similar applications, such as vehicular destination prediction on highways,[5] bike sharing users' destination prediction,[6] user mobility in mobile cloud computing[7] and so on.

Traditional technologies are mainly using analytic statistics methods to predict individuals' destinations based on personal travel patterns, such as spatial patterns reflecting co-occurrence of locations;[8-11] spatiotemporal sequence patterns[12,13] reflecting the reasons why a user visits and leaves a location at a certain time, for example, for commuting, studying, and so on. These methods assume that the predicted user has regular commuting pattern and sufficient historical data. However, it is not clear how to predict the destination of a passenger who has limited personal data or whose pattern has no obvious regularity.

The motivation of this paper is based on the intuition that travel activities of a passenger are made up of fixed daily traveling (e.g., traveling for commuting, studying) and flexible traveling (e.g., traveling for entertainment, shopping). The fixed traveling has spatiotemporal regularity, of which the destinations can be uniquely inferred based on the travel patterns extracted from the long-term travel data of passengers. The flexible traveling has spatiotemporal irregularity, of which the destination can be predicted by combined personal and crowd travel behavior under certain contexts (e.g. time, weather, and so on). In addition, in view of that both of the interpretability and prediction accuracy of the prediction model are all significant in the field of transportation for better supporting decision-making. Therefore, we design an explanatory model to predict individuals' destinations based on sparse observations including real-time origin information and historical trips, the major **contributions** are as follows:

- We propose a practical model with strong interpretability and predictability called as PMIP to predict passengers' destinations at individual level. The model is especially applicable for the passengers with limited personal historical data.
- We treat differently among passengers' regular trips and stochastic trips according to personal spatiotemporal travel patterns, and develop a kernel density-based clustering algorithm to extract each passenger's travel patterns for the efficiency and robustness.
- We present a two-view-based hybrid model that predicts individuals' stochastic trips. The model uses a conditional discrete choice model (DCM) with interpretability to handle the direct correlation with history individual and crowd destination choice behavior, and utilizes Neural Network to strengthen the model's predictive ability by using context factors.
- We evaluate our method using realistic smart card data and weather data over 4 months in Shenzhen, China. The experimental results validate the effectiveness of our method compared with other baselines.

The rest of this paper is organized as follows. Section 2 outlines the related work of this paper. Section 3presents the overview of our problem and solution. Section 4 details our proposed solution for passengers' destination prediction. The empirical evaluations of the method based on real-world data are discussed in Section 5, and we conclude the work in Section 6.

2 | RELATED WORK

The purpose of user movement or location prediction is to determine the location of a user or a vehicle by observing the beginning of the trip. By now, there have been some studies addressing this problem. We can categorize existing approaches into two groups by the theoretical principles, statistics-based models and machine learning-based models.

(1) **Statistics-based models:** They are based on the frequent patterns extracted from individuals' historical travel data and use probabilistic methods to predict target user's movement. (i) Based on individuals' spatial patterns, which reflects the co-occurrence of locations, Morzy et al. used an Apriori-based and PrefixSpan algorithm to discover the pattern of moving objects,[8,9] and define some movement rule matching strategies to predict each moving object location. Jeung et al. estimated a moving object's future locations based on its spatial sequence pattern as well as existing motion functions of the object's recent movements.[10] Zheng et al. conducted a personalized location recommendation system by mining the correlation between locations.[11,14]

(ii) Based on users' spatial-temporal sequence patterns. The patterns reflect the co-occurrence of locations considering time information, e.g, Alice tends to go from location A to location B during 6:00 AM –7:00 AM of a day. Li et al.[13] proposed a hierarchical clustering method to find periodic patterns. Zhao et al.[15] proposed to use sequential, overlapped and fixed time length slots to automatically extract the spatiotemporal patterns over the AFC data. (iii) Based on semantic patterns, which refer to the "general" geographical consequence of locations visited by individuals[16] and reflect the semantic reason why individual travel from some locations to other locations. For example, Bruce always goes to restaurant *B* for lunch around 12.30 PM. This is achieved by combining the raw mobility trajectories with related contextual data(e.g., POI). Yan et al. proposed a hybrid Model for analyzing and transforming raw mobility data (GPS) to meaningful semantic abstractions.[17]

Most of these travel patterns-based method have a pre-assumption that sufficient personal history data can be obtained. However, it is not clear how to predict the destinations of individual users with only a few personal historical information. In such cases, it is necessary to combine the crowd travel behavior to provide complementary information. Though we combine the crowd travel behavior and use a linear DCM to predict individuals' destination;[18] however, due to incapability of modeling nonlinearity relationship of DCM and the lack of considering context influences, we further improve the approach[18] in this paper.

(2) **Machine learning-based models:** They use machine learning methods to predict a object movement. In Kaggle-ECML/PKDD, Brebrisson et al. used a multilayer neural network to predict the taxi destination, given the beginnings of trajectories.[19] Lv et al. considered taxi trajectories as two-dimensional images, and adopted multilayer convolutional neural network to predict taxi destination.[20] Endo et al. employed recurrent neural

networks, which represented trajectories as discrete features in grid space and fed the sequences of grids into the RNN model, to predict taxi or user's destination.[21] Rossi[22] proposed a RNN-based approach and encoded the semantics of visited locations by combining geographical information to predict the exact coordinates of taxi's destination. Wang et al. proposed efficient attention-based deep learning framework for sharing bike trip destination prediction.[6]

However, due to the lack of individual's long-term travel data, these machine learning-based methods did not make full use of individual travel information.[21,22] In addition, it is difficult for these models to explain the interpretability, which is significant to support decision-making in transportation field. In this paper, we propose an explanatory model to predict individuals' destinations by combining individual and crowd travel behavior.

3 | OVERVIEW

3.1 | Notations and problem formulation

We first introduce several fundamental concepts:

Definition 1. (*Trip*). A trip of a passenger, denoted as tr, is related to four attributes s_o, s_d, t_o, and t_d, representing the trip's starting station, ending station, starting time and ending time.

Definition 2. *(Trip Sequence).* A trip sequence TS is associated with $TS.id$ uniquely representing a passenger, and a temporally ordered sequence of trips of the passenger denoted as $Tr = \{tr_1, tr_2, \dots, tr_P\}$ collected by AFC system.

Definition 3. *(Travel Pattern).* A travel pattern P of a passenger is used to describe the travel's spatiotemporal regularity, which is associated with four attributes $P.t_b$, $P.t_e$, $P.s_b$ and $P.s_e$. That refers that a passenger often goes from station $P.s_b$ to station $P.s_e$ during a fixed time period $[P.t_b, P.t_e]$ of a day, and the proportion of the active days of such trips is greater than a certain threshold τ (e.g, τ=80%).

Problem definition: Given a metro network consisting of N stations $S = \{s_1, s_2, \dots, s_N\}$ and the historical AFC data of all passengers, and other associated data (e.g., weather information), this paper aims to predict the destination (ending) station of a passenger given the starting station s_o and time t_o.

3.2 | Motivation

Figure 1 presents the distribution of passengers according to the number of active days and the number of trips based on four months of AFC data in Shenzhen, China. We can observe that there are more than half passengers which only have less than six trips as shown in Figure 1(A), and less than four active days as shown in Figure 1(B). For such passengers with only a few historical travel information, we need to combine crowd's travel behavior to predict the destinations. In addition, there have been some studies summarizing that users' travel behavior has temporal dependency (e.g., 24-h periodicity and 7-day periodicity)[23-26] and related with other factors (e.g., weather, special events, and so on).[27] Therefore, we can combine individual travel behavior, crowd travel behavior, and other context information to predict individual's destination.

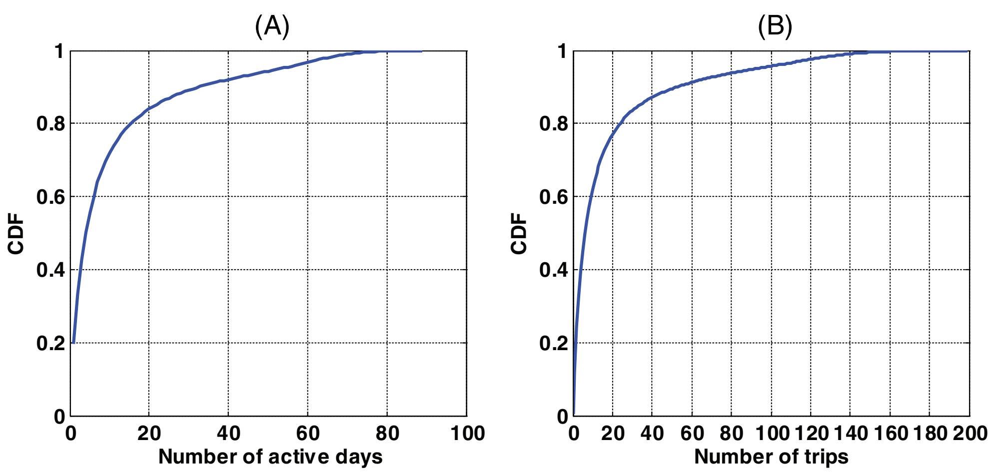


FIGURE 1 Passengers' cumulative distribution based on: (A) the number of active days. (B) The number of trips

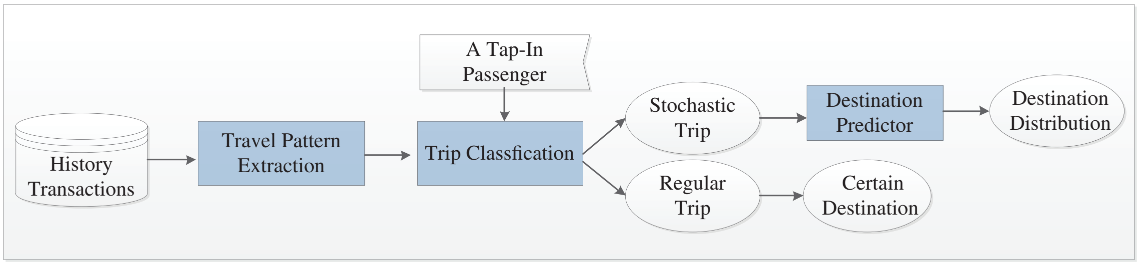


FIGURE 2 Processing flowchart of Practical model with strong interpretability and predictability

3.3 | Analysis framework

Figure 2 presents the processing flowchart of our approach for individuals' travel destination prediction. First, we extract the spatiotemporal travel patterns of individual passengers based on their own historical smart card data. Based on these patterns, given a target trip of a passenger with tap-in information, we classify the trip as regular trip or stochastic trip by finding a matched travel pattern. If the trip belongs to a travel pattern, the trip is regular and the destination can be uniquely determined, else the trip is a stochastic trip and we use a stochastic trip destination predictor to estimate the proportion that the passenger will destine to each station by combining both individual and crowd travel information under certain context information.

In order to predict stochastic trip's destination, we use a DCM with interpretability to capture the definite relationship between the prediction target and personal or crowd travel behavior, and use neural networks to strengthen the model's predictive ability by combining context factors. The reason we do not use a multilayer neural network or deep learning-based method to learn a nonlinear mapping function from observed features to stochastic trip's destination is the drawback of being a black-box. However when studying the demand in transportation applications, it is of utmost importance that we understand what are the key factors to affect passengers' travel behavior, which can help user to make decision.

4 | DESTINATION PREDICTOR

4.1 | Regular trip destination predictor

This section aims to predict the regular trip's destination of a passenger P. We first extract the passenger's travel patterns $M=\{m_1, m_2, \dots, m_{|M|}\}$ referred to Definition 3, then give a trip tr of the passenger who enters a metro system in station $tr.s_o$ at time slot $tr.t_o$, we first classify the trips by matching it with the passenger's travel pattern set M. If there exists a travel pattern m_i satisfying $[m_i.t_b, m_i.t_e] \cap tr.t_o \neq null$, we classify the trip as a regular trip and set $m_i.t_e$ as the destination of the trip tr, otherwise the trip is a stochastic trip, and the destination is predicted by the method in next section. The individual travel pattern is extracted as follows.

Given the history trip sequence $P=\{p_1, p_2, \dots, p_{|P|}\}$ of a passenger, this step aims to extract the passenger's spatiotemporal travel patterns as described in Definition 4. We first use kernel density estimation method to capture the passenger's travel regularity in time dimension, then obtain the regularity in space dimension and construct the travel patterns. Compared with existing methods, this method is more flexible since it can extract each passenger's travel patterns with precise time ranges. The specific process includes two steps as follows.

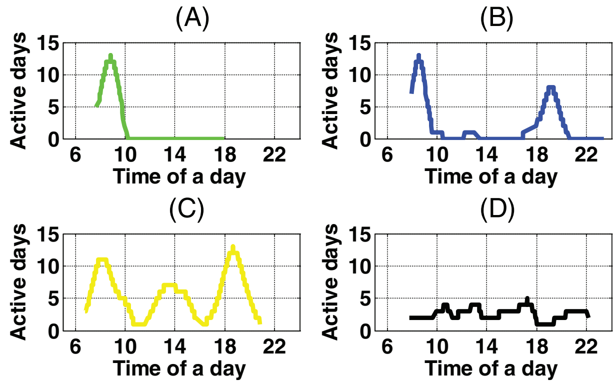
Step 1: This step aims to extract the significant time periods of a day which the passenger mainly travels at. Different passengers may have different significant travel time periods, as shown in Figure 3. There are four specific passengers with different number of significant time periods. We can observe that the passengers in Figure 3 (A), (B), (C), (D) have one, two, and three, and no obvious significant time periods, respectively, according to the number of peaks. Therefore, we need to extract each passenger's temporal travel regularity according to their own trips.

Given the passenger's historical trip set P, we first calculate the trip time distribution in a day.

We choose the common used gaussian kernel function to obtain the local density ρ_i of the passenger's trip time in a day by Equation (1), where $(x_1, x_2, \dots, x_n)$ is the start time slots of all trips in P, n is the number of trips, and h is the bandwidth that determines the size of the local time range affected by the kernel function. The h is determined by the taken time of a trip and set according to the actual scene.

$$f(x;h) = \frac{1}{nh}\sum_{i=1}^{n}\left[\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-x_i)^2}{2h^2}}\right]. \tag{1}$$

FIGURE 3 Trip distribution of some individuals on a day



Then, we choose some points as the center of the significant time periods of a day of the passenger, where the point x satisfies two conditions: (i) with highest local density in the time range $[x-0.5h\,x+0.5h]$. (ii) the total proportion of active days of segments falling in the range is larger than a predefined value λ (e.g ., $\lambda = 5$). Given a time center x_i satisfying the two conditions, the range of significant time period is set to $[x-0.5h\,x+0.5h]$. All significant time periods extracted are represented as $C=\{c_1, \ldots, c_{|C|}\}$.

Step 2: We construct the passenger's travel patterns by extracting the O-D pair during each significant time period of C. Given all trips in a significant time period $c_i \in C$, if there exists an O-D pair (s_o, s_d), and the proportion of the active days of the trips is greater than a certain threshold τ, we create a travel pattern p with $p.s_b = s_o$, $p.s_e = s_d$, and the $p.t_b$ and $p.t_e$ are set to the begin and the end of c_i. Otherwise, we assume that there is no travel pattern in c_i. Based on analysis above, given a passenger, we extract all the travel patterns, dented as $M=\{m_1, m_2, \ldots, m_{|M|}\}$.

Mining individual passenger's latest travel pattern based on passenger's own latest history travel information can help to classify the predicted trip more accurately. However, updating the travel patterns too regularly could increase system costs. Therefore, we use a compromise that updates individual travel pattern once per week. That means if we want to predict the destination of a trip of a passenger at nth week, we use all history travel information of the passenger before the nth week as the training data for KDE.

4.2 | Stochastic trip destination predictor

In this section, we focus on predicting the destination of the passenger's stochastic trip tr. We first depict the input features fed into stochastic trip's destination predictor, then introduce the framework of the predictor.

4.2.1 | Feature extraction

Given the passenger P's stochastic trip tr with starting station $tr.s_o$ and starting time $tr.t_o$ on a day, in order to predict the destination, we use three types of features as follows: individual features, crowd features, and context features.

(1) Individual features F_u: we extract individual features about the passenger P's travel behavior based on the historical travel information from two perspectives: sequential perspective and semantic perspective. The sequential perspective refers that an individual passenger who enters station $tr.s_o$ usually destines to a few stations. Semantic perspective refers that a passenger often stays at a few stations during specific time period. For example, a passenger often stays at the workspace A during the daytime and stays at home B during night. Accordingly, we extract the two types of features.

(a) Semantic destinations distribution F_u^m: the number of the times the passenger stayed around each station during the time interval $[tr.t_o - h\; tr.t_o + h]$ are extracted. The definition of a passenger staying state at a station is as follow. If the passenger's two successive trips meet that the ending station of a ride is the beginning station of the next trip, and the time duration is greater than 2 h, we assume that the passenger stayed around the station during the time interval. In general, the locations a user often stays are different on different days (e.g., weekday and weekend), and that on the days same with $tr.w_d$ has more correlation than that on different days. So we extract them on the two types of history days, respectively, and use $F_u^m = \{f_a, f_a'\}$ to represent the semantic features, where both a and a' are N dimensional vectors and represent the times of the passenger staying at each station at the days same and different with $tr.w_d$, respectively. For example $f_a = \{a_1, a_2, \ldots, a_{|N|}\}$, where a_i represents the number of times that the passenger stayed around station s_i during time period $[tr.t_o - h\; tr.t_o + h]$ on the historical days same with $tr.w_d$.

(b) Sequential destinations distribution F_u^q: the history times that the passenger travelled from station s_o to each other station is extracted. As that for semantic features, we use two N dimensional vectors $F_u^q = \{f_b, f_b'\}$ to represent the times of the passenger destining to each station at the days same and different with $tr.w_d$, respectively.

(2) Crowd features F_g: the individual passenger's historical data can be very sparse. We try to use the crowd destination distribution to provide complementary information. Crowd travel behavior has temporal dependency which refers that the traffic flow at a certain time is usually correlated with various historical observations.[23-26] For example, the destination distribution of crowd from a given origin station at a day is affected by that of past days, and has obvious 7-day periodicity. Therefore, we extract the crowd's destination distribution during the time interval T_i of past $|G|$ days with same week attribute $tr.w_d$ and denoted by $G \times N$ dimensional matrix as $F_g = \{g_{|G|}, g_{|G|-1}, \dots, g_1\}$, where $g_m = \{g_m^1, g_m^2, \dots, g_m^{|N|}\}$, g^n represents the flows of the passenger destining to each station at the past gth week.

(3) Context Features F_c: we further consider other context features, including the day of the week, weekday/weekend, holidays, meteorological information (e.g., rainfall, temperature, air quality). They are all related with traffic tasks. For example, we calculate the similarities of the destination distributions by cosine distance, between rainy day and sunny day, between rainy days, and between sunny days at same time slot of day. They are 0.51, 0.601, 0.634, respectively. That means that the more similar the weather conditions between 2 days are, the more similar the destination distributions between the two days are.

4.2.2 | Prediction model

We build a model for predicting the passenger stochastic trip's destination. The extracted features can be further categorized into two groups. The first group is alternative dependent features, which include individual features(F_u^m, F_u^q) and group features (F_g), where each feature is assigned to a particular alternative, and has direct impact on the alternative. For example, among the individual features $f_a = \{a_1, a_2, \dots, a_{|N|}\}$, a_i belongs to the destination station s_i. The second group are alternative independent features including all context features which are not assigned to any one of these alternatives and have indirectly impact on all destination alternatives.

In order to offer a reasonable explanation for the influence of various features on individual passenger's destination choice, we combine a linear DCM fed in alternative dependent features and non-linear Neural Networks model fed in alternative independent features for individual passenger destination prediction. There are two reasons why we devise two separate components DCM and Neural Networks to address this issues. *First*, from the feature space's perspective, the features used by DCM and those used by neural Networks do not have any overlap, providing different views on passenger's destination choice. *Second*, from the model's perspective, the DCM predicts the destination distribution in terms of the history data about each destination station, which has direct and obvious correlation. So we use a linear model DCM for the interpretability. While the influence of context features is more about a nonlinear interpolation, we use Neural Networks as a supplement to learn the nonlinearity and strengthen the prediction result.

In order to model the interpretable relationship between the alternative-dependent features and all alternatives (destination stations), we choose the most suitable DCM—Conditional Multinomial logit model (CMNL) due to the statistical foundations and capability to represent individual choice behavior realistically. The CMNL model is based on random utility maximization theorem. In our case, utility theorem can be understood as that the probability of the passenger destining to a station is related to the station's utility value, and the larger the utility value is, the higher the probability the station is chosen.

Given the stochastic trip Tr of passenger P and alternative related $d \times N$ dimensional features X (where the nth column represents the features of the nth alternative), the matching utility function of the ith alternative (destination station) is calculated by $U_i = \sum_{j=1}^{d} \beta_j \times x_{j,i}$, where $\beta_1, \dots, \beta_d$ is a set of parameters to be trained. The probability of choosing $s_i \in S$ for multinomial logit is defined as:

$$P(i|Tr) = P(U_i > \max(U_j)) = \frac{\exp^{U_i}}{\sum_{j \in N} \exp^{U_j}}.$$

In order to flexibly add other components as a supplement to CMNL to obtain better prediction result, and conveniently construct optimizers in training when minimizing the loss function, we can write CMNL in modern machine learning libraries. For that we define $\beta = \{\beta_1, \beta_2, \dots, \beta_d\}$ as a kernel with $1 \times d$ size, and define the set of variables X as an image of size $N \times d$, where d is the number of dimensions of alternative dependent features. We get the utility of N alternatives $U = \{U_1, \dots, U_N\}$ by doing a convolution between the kernel and the image. The probabilities can be obtained by using a softmax activation function. The categorical cross-entropy as $H_n(\sigma, y_n) = -\sum_{j \in c_n} y_{i,n} \log[\sigma(V_n)_i]$ are used as the loss function. This method allows us to use conventional deep learning libraries to implement the multinomial logit model, giving us very high flexibility and efficiency in modifying the structure and learning the parameters.

In order to add the influence of context factors on passengers' destination choice behavior to strengthen the model, we take advantage of the neural network approach by adding an external value $E_{m,i} = \psi(X')$ to the corresponding utility function $U_{m,i}$ of the DCM model, where X' is the

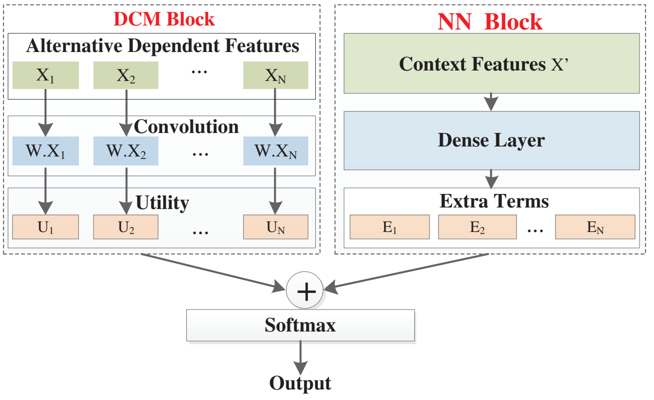


FIGURE 4 The final model combining both the convolution approach of writing Conditional Multinomial logit model and neural network

ensemble of context features and $E_{m,i}$ is a N dimensional vector. ψ is the function defined by multiple neural network dense layers and corresponding activation functions. Correspondingly the utility functions are written as: $U_n = \beta X^T + u_n + \xi_n$.

Overall, the final model combining CMNL fed with alternative dependent features and a complement neural network term fed with context features is shown in Figure 4. On the left hand side of the figure, the multinomial logit model is written in modern machine learning libraries, and the kernel corresponds to the β parameters. That applying a convolution layer to the input features gives us the utility functions as in CMNL. We can flexibly change the model. The design idea of the model can be extended to other CMNL-based applications, where both of the interpretability and prediction accuracy are all significant. On the right hand side, the DNN component is used to enhance each utility function and increase predictive accuracy.

5 | PERFORMANCE EVALUATIONS

This paper focuses on the questions as follows: (1) How does our proposed model perform? (2) how does each component in our model contribute to the predicted result? (3) How is the interpretability of the model? (4) If the extracted features fed into the predictor are really useful?

5.1 | Experimental configurations

5.1.1 | Datasets

In the experiments, we use the two datasets: smart card transition data of metro system and weather data in Shenzhen, China over 4 months from March 1, 2016 to June 30, 2016. There are 118 stations in the metro system. We partition all data into three nonoverlapped sets: training set, validation set, and test set by a ratio of 2:1:1. We use the first 2 months as training set, the third month as validation set, and the rest as test data to guarantee an unbiased result.

(1) ***Smart Card data***: Each passenger needs to tap the smart card for charge when entering origin metro station and leaving destination station. Each record includes card ID, station, transaction time and transaction type (tap-in or tap-out).

(2) ***Weather data***: We use weather data collected by city-level information. Each record consists of district, time, rainfall, temperature, humidity, wind speed, and wind direction.

5.1.2 | Parameter settings

Let us recall that, in the step of individual passenger's spatiotemporal travel pattern extraction, there are two parameters needed to set. The first one is h that is used to control the range of significant time period. In this section, we set h to 2 hours, since the duration of 99% trips in our data is less than 2 hours. Of course it can be set according to actual scenarios. Another one is the threshold τ that is used to judge whether a travel pattern is existed at a given significant time period, in this paper we set $\tau = 80\%$. In addition, for stochastic trip's destination prediction, the destination distribution of the crowd in past N weeks are extracted. In this paper, we set N to 4 which can get the best model performance.

In addition, in stochastic trip destination predictor, our model is implemented with Tensorflow 2.1. The NN Block component uses two hidden layers. The number of corresponding hidden units are tuned within the ranges of [8, 16, 32, 64] on the validation set, and the best performance is 16. The learning rate is tuned within 0.01–0.001.

5.1.3 | Evaluation metrics

In this paper, we evaluate model performance to predict trips' destinations by accuracy (*Acc*). Suppose there are N_{all} trips needed to predict, and N_{cr} trips are correctly predicted. Then the accuracy is defined as $Acc = N_{cr}/N_{all}$.

5.1.4 | Baselines

We compare the performance of our method with some baselines. As mentioned in Section 2, the existing machine learning based models for user's destination prediction do not consider individuals' travel patterns due to the lack of long-term personal historical travel data. Therefore they are not applicable to the scenario of this paper. Therefore, we compare our method with existing statistics methods based on individual travel patterns, and select five baselines as follows:

Emp: Empirical Estimation represents the prediction method based on the naive empirical knowledge where we consider the most frequently visited destinations of each individual passenger.

SP: Spatial pattern-based method,[14] the method is based on individual's spatial patterns that reflect location co-occurrence. Given a passenger's trip with tap-in time and station information, in order to predict the trip's destination, the method first extracts the destination distribution from the passenger's history trips with same tap-in station, then take the destination station with the highest proportion as the predicted value. If such historical trips don't exist, we take destination with the highest proportion based on crowd history trips.

STP: Spatiotemporal pattern-based method,[23] which is similar as the method of SP, except that the extracted travel patterns of each passenger not only reflect the location co-occurrence but also reflect the travel regularity in time dimension.

SMP: Semantic pattern-based method, which first extracts the passenger's staying stations in different time periods of a day, then takes the staying station with the highest proportion as the destination according to the passenger's tap-in time.

Bayes: Bayesian Network is a typical graph-based algorithm,[27] which is a representative of the probability-based models. We feed the extracted features of our method into Bayesian Network to predict individuals' destination.

5.2 | Results and analysis

In this section, we first compare our algorithm with other baselines, then measure the contribution of each component of our model. Then we measure the impact of different extracted features and analyze the parameters in DCM.

5.2.1 | Performance comparison with baselines

We compare our method with various baselines in terms of accuracy metric and highlight the best performance with bold font. Since passengers' travel behavior has different characteristics during different time periods on different days (weekday and weekend), for example, the regular tips in peek periods are much more than those of low hours. We divide a day into two parts according to users' commuting time: considering the time period 7:00 a.m. to 9:00 a.m. and 5:00 p.m. to 7:30 p.m. as peak period, and other time as low period. Accuracy comparison between our method and baselines during different time periods on different days are shown in the following Table 1. As we can see from the table, our model achieves the best performance among them during all time period of all days with an average performance gain of 12%. Moreover, we can observe that all methods in peek period at weekday achieve more than 50% accuracy, and are obvious better than those of other periods. That suggests that the destination choices are relatively stable on metro system and the statistical relationship is better to model the destination prediction problem. In

TABLE 1 Comparison among different methods

Method	*Weekday-Peek*	*Weekend-peek*	*Weekday-low*	*Weekend-low*
Emp	0.501	0.439	0.402	0.381
SP	0.547	0.481	0.443	0.426
STP	0.594	0.524	0.482	0.461
SMP	0.491	0.421	0.392	0.359
Bayes	0.687	0.611	0.548	0.499
PMIP	**0.776**	**0.704**	**0.689**	**0.645**

addition, Bayes performs slightly well than SP, STP, and SMP. That's because it predicts passengers' destination considering more influence factors, including spatial pattern, spatiotemporal pattern, semantic pattern, and other factors.

5.2.2 | Component analysis

PMIP is composed of two independent components, regular trip component (RC) and random trip component. The random trip component consists of two blocks, DCM block with interpretability, and neural networks component (NN) to model the nonlinearity for supplement. In order to evaluate the effect of these components or blocks, we remove them from the original model to observe the performance of the corresponding degraded model. It is noteworthy that there is no overlap between the features fed in the two blocks of the random trip component. To achieve a fair comparison, when removing one block, we feed all features to another block. The results are shown in Table 2, where *PMIP-RC*, *PMIP-DCM*, and *PMIP-NN* refer to the methods where regular trip component is removed from *PMIP*, and DCM is removed from *PMIP*, and neural networks component is removed from *PMIP*, respectively.

The results indicate that the prediction performance decreases no matter which component to remove. In other words, all components make contributions to improve the model performance.

First, we can also observe that when the regular trip component is removed, the model performance becomes worst, which illustrates that regular trip component plays more important role in individual destination prediction. Moreover, as we expected, the raising effect is more obvious in peek hours. This is because the destination choices are less predictable on the weekend when people generally do not follow regular travel patterns. That is very easy to explain by the result in Figure 5, which shows the ratio of destinations that could be inferred and the accuracy rate in different time periods. The figure shows that the regular trips in AM and PM peak hours are much more than other hours. The finding agrees with reality as well, i.e. most users regularly and stably commute back and forth to work during peak hours on week days. Figure 5(B) shows that the regular trip component accurately predicts passenger regular trip's destinations with an average accuracy greater than 95%.

Second, although neural network can work well to model the nonlinearity in many tasks, we can observe in Table 2 that neural network does not perform better than DCM alone. That is because the neural network can not automatically distinguish the features directly related to an alternative and those related to all alternatives, especially, in the cases that the available data is not very sufficient. However, the discrete choice method designed with prior knowledge can address such direct relationship.

TABLE 2 The results of Practical model with strong interpretability and predictability (PMIP) with removing different components

Method	*Weekday-Peek*	*Weekend-Peek*	*Weekday-Low*	*Weekend-Low*
PMIP-RC	0.634	0.564	0.521	0.487
PMIP-DCM	0.676	0.644	0.651	0.612
PMIP-NN	0.692	0.657	0.663	0.631
PMIP	**0.776**	**0.704**	**0.689**	**0.645**

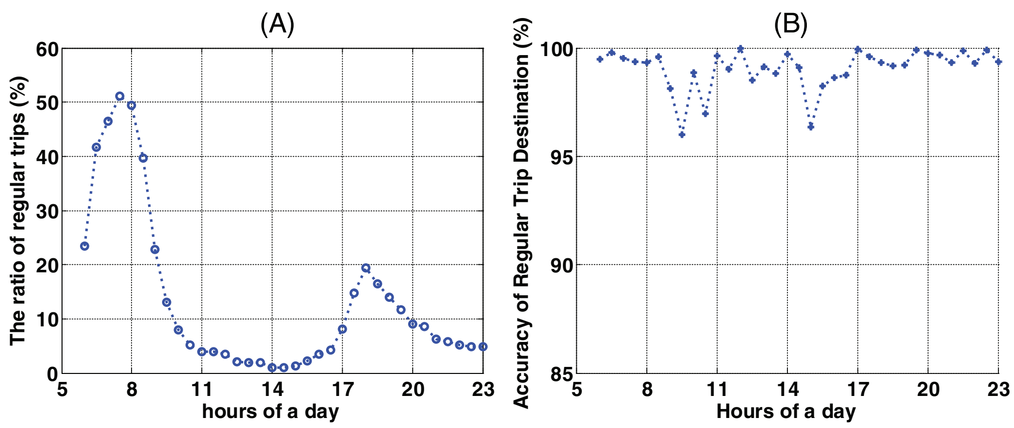


FIGURE 5 Regular trips' destination prediction

Method	*Weekday-Peek*	*Weekend-Peek*	*Weekday-Low*	*Weekend-Low*
No-IF	0.534	0.492	0.445	0.421
No-CF	0.613	0.554	0.526	0.512
NO-MF	0.721	0.649	0.638	0.602
NO-TF	0.669	0.602	0.597	0.638
ALL	**0.776**	**0.704**	**0.689**	**0.645**

TABLE 3 Results of Practical model with strong interpretability and predictability with removing different features

Features	f_a	f'_a	f_b	f'_b	g_1	g_2	g_3	g_4
Parameters	0.2781	0.2647	0.1211	0.1452	0.1045	0.0945	0.0722	0.0772

TABLE 4 Parameters of discrete choice model component

5.2.3 | Choice features and parameter analysis

In this paper, we use three types of features: individual features (IF), crowd features (CF), context features containing meteorological factors (MF), and time factors (TF). In order to verify the features fed into the PMIP are really useful, as shown in Table 3, we remove each of them and feed remaining features to PMIP, and observe that the prediction performance decreases clearly no matter which feature to remove. This indicates that all features make contributions to the model. Moreover, as we can see in Table 3, their effects are different, the individual and crowd features are more important than other factors.

The parameters of β of different features fed into DCM are shown in Table 4 Among these alternative related features in Section 4, we can observe that the increase of any feature corresponding to a destination station makes positive contribution to the probability that the passenger chooses the station. Among these features, the individual features make the biggest contribution, followed by group features and context features, which is consistent with the above analysis. Moreover, the effects of the group features on the past N weeks follow expected rules, where the longer the duration between a past week and target week is, the weaker the impact of the past week on target week is.

6 | CONCLUSIONS AND FUTURE WORK

In this paper, we focus on individuals' travel destination prediction on a metro system with AFC data. In particular, we motivate and design a novel approach called PMIP with two key technical components for regular trip destination predictor and stochastic trip destination predictor. Especially, an explanatory model combining neural network and discrete logit model is proposed to predict individual's stochastic trip's destination, which can be applied to other transportation analysis scenarios about users' choice behavior such as travel mode choice or route choice. More importantly, we conduct and evaluate PMIP based on real-world data in the metro network of Shenzhen, China. The experiments demonstrate that our PMIP model outperforms other baselines in terms of prediction accuracy. In the future work, we will improve the method to apply it to other transportation analysis scenarios about individuals' choice behavior such as travel mode choice or route choice.

ACKNOWLEDGEMENTS

This work is supported by the National Key R&D Program of China (No.2019YFB2102100), National Natural Science Foundation of China (No.61802387, No.61806192), National Natural Science Foundation of Shenzhen (No.JCYJ20190812153212464, No.JCYJ20190812160003719), Shenzhen Engineering Research Center for Beidou Positioning Service Improvement Technology (No.XMHT20190101035), Science and Technology Development Fund of Macao S.A.R (FDCT) under number 0015/2019/AKP, Shenzhen Discipline Construction Project for Urban Computing and Data Intelligence, and SIAT Innovation Program for Excellent Young Researchers.

ORCID

Juanjuan Zhao https://orcid.org/0000-0003-1002-9272

REFERENCES

1. Yan L, Shen H, Chen K. MobiT: distributed and congestion-resilient trajectory-based routing for vehicular delay tolerant networks. *IEEE/ACM Trans Netw*. 2018;26(3):1078-1091.
2. Yan L, Shen H. TOP: optimizing vehicle driving speed with vehicle trajectories for travel time minimization and road congestion avoidance. *ACM Trans Cyber-Phys Syst*. 2019;4(2):1-25.

3. Chen K, Tan G. SatProbe: low-energy and fast indoor/outdoor detection via satellite existence sensing. *IEEE Trans Mob Comput*. 2019;1-1. https://doi.org/10.1109/TMC.2019.2954873. https://ieeexplore.ieee.org/abstract/document/8908759/?casa_token=KYQXJchjXO8AAAAA:NLMfdmKDG4JWHgAxU2esp-6WfRdYOPAvUA6RvdOyK3_AfxXQFRaHj7eBgHvjsLBoMzEbB4jHas4.
4. Chen K, Tan G. BikeGPS: localizing shared bikes in street canyons with low-level GPS cooperation. *ACM Trans Sens Netw*. 2019;15(4):1-28.
5. Yang Y, Xie X, Fang Z, Zhang F, Wang Y, Zhang D. VeMo: enabling transparent vehicular mobility modeling at individual levels with full penetration. *MobiCom*. 2018;11:1-12. https://doi.org/10.1145/3300061.3300130.
6. Wang W, Zhao X, Gong Z, Chen Z, Zhang N, Wei W. An attention-based deep learning framework for trip destination prediction of sharing bike. *IEEE Trans Intell Transp Syst*. 2020;1-10.
7. Xu M, Tian W, Buyya R. A survey on load balancing algorithms for virtual machines placement in cloud computing. *Concurr Comput Pract Exp*. 2017;29(12):e4123. e4123 cpe.4123. https://doi.org/10.1002/cpe.4123. https://ieeexplore.ieee.org/abstract/document/9152145?casa_token=MRnqzy-L2ccAAAAA:avzgI0EfMUIAh8TwzR5fnwDPYZNgRSj7NZCWsG7LZeY8ZWjAiGlZGUidFAqjdrxKapm0JAiWixI.
8. Morzy M. Prediction of moving object location based on frequent trajectories. Paper presented at: Proceedings of the Computer Information Sciences-iscis, International Symposium; November 2006; Istanbul, Turkey.
9. Morzy M. Mining frequent trajectories of moving objects for location prediction. Paper presented at: Proceedings of the International Conference on Machine Learning Data Mining in Pattern Recognition; 2007; Springer, Berlin, Heidelberg.
10. Jeung H, Liu Q, Shen HT, Zhou X. A hybrid prediction model for moving objects. Paper presented at: Proceedings of the IEEE International Conference on Data Engineering, Cancun, Mexico; 2008.
11. Zheng Y, Zhang L, Xie X, Ma W. Mining interesting locations and travel sequences from GPS trajectories. Paper presented at: Proceedings of International conference on World Wide Web 2009, Madrid, Spain; 2009:791-800.
12. Monreale A, Pinelli F, Trasarti R, Giannotti F. WhereNext: a location predictor on trajectory pattern mining. Paper presented at: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Paris, France; 2009.
13. Li Z, Han J, Ming J, et al. MoveMine: mining moving object data for discovery of animal movement patterns. *ACM Trans Intell Syst Technol*. 2011;2(4):1-32.
14. Zheng Y, Zhang L, Xie X, Ma WY. Mining correlation between locations using human location history. Paper presented at: Proceedings of the Workshop on Advances in Geographic Information Systems, Seattle, Washington; 2009.
15. Zhao J, Tian C, Zhang F, Xu C, Feng S. Understanding temporal and spatial travel patterns of individual passengers by mining smart card data. Intelligent Transportation Systems (ITSC). Paper presented at: Proceedings of the 2014 IEEE 17th International Conference, Qingdao, China; 2014:2991-2997.
16. Parent C, Spaccapietra S, Renso C, et al. Semantic trajectories modeling and analysis. *ACM Comput Surv*. 2013;45(4):1-32.
17. Yan Z, Parent C, Spaccapietra S, Chakraborty D. A hybrid model and computing platform for spatio-semantic trajectories. Paper presented at: Proceedings of the Extended Semantic Web Conference, Heraklion, Crete, Greece; 2010.
18. Wang H., Zhao J., Ye K., et al. A destination prediction model for individual passengers in urban rail transit. Paper presented at: Proceedings of the 2020 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS), Shenzhen, China; 2020:1-6.
19. De Brebisson A, Simon E, Auvolat A, Vincent P, Bengio Y. Artificial neural networks applied to taxi destination prediction; 2015. arXiv: Learning.
20. Lv J, Li Q, Sun Q, Wang X. T-CONV: a convolutional neural network for multi-scale taxi trajectory prediction. Paper presented at: Proceedings of the 2018 IEEE International Conference on Big Data and Smart Computing (BigComp), Shanghai, China; 2018:82-89. https://doi.org/10.1109/BigComp.2018.00021.
21. Endo Y, Nishida K, Toda H, Sawada H. Predicting destinations from partial trajectories using recurrent neural network. Paper presented at: Proceedings of the Pacific-asia Conference on Knowledge Discovery & Data Mining, Jeju, Korea; 2017:160-172.
22. Rossi A, Barlacchi G, Bianchini M, Lepri B. Modelling taxi drivers' behaviour for the next destination prediction. *IEEE Trans Intell Transp Syst*. 2019;21(7):2980-2989. https://doi.org/10.1109/TITS.2019.2922002.
23. Zhao J, Qu Q, Zhang F, Xu C, Liu S. Spatio-temporal analysis of passenger travel patterns in massive smart card data. *IEEE Trans Intell Transp Syst*. 2017;18(11):3135-3146.
24. Wu Z, Pan S, Long G, Jiang J, Zhang C. Graph WaveNet for deep spatial-temporal graph modeling. Proceedings of the 28th International Joint Conference on Artificial Intelligence, IJCAI 2019, August 10–16, 2019:1907-1913; Macao, China.
25. Guo S, Lin Y, Feng N, Song C, Wan H. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. Paper presented at: Proceedings of the 33rd AAAI Conference on Artificial Intelligence; January 27 - February 1, 2019:922-929; Honolulu, Hawaii.
26. Chen C, Li K, Teo SG, et al. Gated residual recurrent graph neural networks for traffic prediction. Paper presented at: Proceedings of the 33rdAAAI Conference on Artificial Intelligence, AAAI 2019, The 31st Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The 9th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019; January 27 - February 1, 2019:485-492; Honolulu, Hawaii.
27. Sun J, Zhang J, Li Q, Yi X, Zheng Y. Predicting citywide crowd flows in irregular regions using multi-view graph convolutional networks. *CoRR*. 2019;abs/1903.07789. https://ieeexplore.ieee.org/abstract/document/9139357?casa_token=1Gf0NSIW64kAAAAA:uRnXEhvP_lxalSsyZ1u375bPsNmZl8qXQNTQsARaTUDwUWo8-PggE2FXgA3CltBh41zIAj2Egcw.

How to cite this article: Zhao J, Ye J, Xu M, Xu C. Practical model with strong interpretability and predictability: An explanatory model for individuals' destination prediction considering personal and crowd travel behavior. *Concurrency Computat Pract Exper*. 2020;e6151. https://doi.org/10.1002/cpe.6151