Mixed logit model based on nonlinear random utility functions: A transfer passenger demand prediction method on overnight D-trains

Bing Han (✉ 17120365@bjtu.edu.cn)  
Beijing Jiaotong University

Shuang Ren  
Beijing Jiaotong University

Research Article

Keywords: Newly-added overnight D-trains, Nonlinear random utility functions, Metropolis-Hastings Algorithm, Improved Fireworks-Simulated Annealing Algorithm

 Posted Date: June 8th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-408461/v1

License: ☺️ ⚡️ This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License

Version of Record: A version of this preprint was published at Soft Computing on January 24th, 2022. See the published version at https://doi.org/10.1007/s00500-021-06621-4.
Mixed logit model based on nonlinear random utility functions: A transfer passenger demand prediction method on overnight D-trains

Bing Han · Shuang Ren

a. School of Computer and Information Technology, Beijing Jiaotong University, 100044, Beijing, China

Received: date / Accepted: date

Abstract In recent years, with the development of high-speed railway in China, the operating mileage and passenger transport capacity have increased rapidly. Due to the high density of trains in the daytime, we usually set up skylights at night within 0:00-6:00 on high-speed railway for comprehensive maintenance, which contradict with the operation demand of D-series overnight high-speed trains (overnight D-trains for short). In order to dynamically adjust the operation plan of overnight D-trains with skylights coordinately, it is necessary to predict the passenger demand of newly-added overnight D-trains. Therefore, the purpose of this paper is to predict transfer passenger demand by formulating a mixed logit model based on nonlinear random utility functions for different transport modes. Firstly according to Maximum Simulated Likelihood Method, the likelihood function of this mixed logit model is proposed to maximize the overall utility value of different passenger groups. And then we adopt Metropolis-Hastings Algorithm to iteratively solve the probabilities of discrete random variables in utility functions. After that, we estimate the unknown distributions of elements in parameter vectors and solve the optimal solution of this model by traditional algorithms, basic heuristic algorithms and improved heuristic algorithms including Improved Fireworks-Simulated Annealing Algorithm which is proposed in this paper, respectively. Finally, a real-world instance with related data of Beijing-Shanghai corridor, is implemented to demonstrate the performance and effectiveness of the proposed approaches.

Keywords Newly-added overnight D-trains · Nonlinear random utility functions · Metropolis-Hastings Algorithm · Improved Fireworks-Simulated Annealing Algorithm

1 Introduction

With the continuous improvement of railway network, the connectivity across the regions and travel demand of passenger are growing with an unprecedented speed. However, due to the heavily congested passenger flow in peak hours, sometimes passenger demand still cannot be satisfied even with the maximum departure frequency. To release traffic pressure and solve the transportation problems, such as the mismatch between transportation resource allocation and passenger demand, overnight D-trains have now been constructed and operated between cities (e.g., Beijing-Guangzhou, Shanghai-Shenzhen, Beijing-Shanghai, etc.). Additionally, the skylight time of the railway is closely related to practical operation plan of overnight D-trains, e.g., the carrying capacity, departure quantity and departure frequency, in essence determines the number of passengers allowed to board.

In the beginning of overnight D-trains operation, the reasonable prediction including transfer passenger demand or induced passenger demand, is an important basis for formulating the train operation scheme to maximize benefits. And considering the optimal combination utility value with transfer modes and transport modes of different passenger groups in the prediction methods is also a new problem, which can improve the accuracy of the prediction results with more practical to a great extent. This research intends to explore this issue explicitly.

1.1 Literature review

As a key component of passenger transport organization, the prediction of passenger demand has received significant attention from both researchers and practitioners in the transportation field. In general, the prediction methods can be divided into the following major categories.
The related factors methods are to find out the factors that affect passenger flow, such as competitive relationship among various means of transportation, the development level of social and economic, technical condition, national income, personal attributes and so on, then to establish the function relationship and prediction model between the passenger flow and various influence factors. Some experts generally believe that passenger demand comes from various activities in social production, such as Jonas et al. (2004), Wardman (2006), Ahn et al. (2017), Oh et al. (2017a), Oh et al. (2017b), Gelhausen et al. (2018) and so on. The way they see it is that the development of society and economy has increased all kinds of links to different regions which is the direct force to increase of passenger demand. In other words, passenger demand is a derivative demand which implied in certain activities.

The time series analysis methods are based on the time which is seen as the independent variable, to establish passenger flow prediction models, such as Wang et al. (2012) and Yin et al. (2017). There are also some models which can be combined with time series methods, for example, seasonal model, such as Tian et al. (2011) and Guang et al. (2017); grey model, such as Li et al. (2007) and Huang et al. (2014); neural network model, such as Chen et al. (2001), Dia (2001), Stella et al. (2006), Eleni (2007), Eleni et al. (2010), Wei et al. (2012), Li et al. (2014), Glišović et al. (2016) and Ivanov et al. (2018); and MD model, such as Wang et al. (2017).

Spatio-temporal prediction methods are the latest methods in recent years because the prediction of passenger demand is not only influenced by dynamical of time, but also limited by the scale of transport network, which means it has spatial and temporal dependence. There are also some models shown in Table 1.

In addition, accurate prediction under atypical conditions, such as vehicular crashes, inclement weather, and holidays, is crucial to traffic management, such as Castro et al. (2009b), (2009a), Roos et al. (2018) and so on.

In all prediction methods, the logit discrete choice model is one of the most widely used models to study transfer passenger demand prediction or market shares of some transportation, such as train, airplane and so on. There are some improved logit models shown as follows.

(a) Multinomial logit (MNL) model and mixed multinomial logit (MMNL) model, such as Hess et al. (2005), Jou et al. (2011), Lee et al. (2016) and so on. MNL model is the most commonly used in practice with the advantages of simplicity, reliability and easy implementation. An important property of MNL model is Independence from Irrelevant Alternatives (IIA), which means the ratio of the choice probabilities of any two alternatives is unaffected by the systematic utilities of any other alternatives. However, MNL model also has some inherent theoretic defects and leads to the need for more refined models. MMNL model offers significant advantages over MNL model by allowing for random taste variation across decision makers. But the biggest drawback of MMNL model is the fact that integrals representing choice probabilities do not have a closed-form expression and need to be approximated through simulation.

(b) Nested logit (NL) model, generalized nested logit (GNL) model, cross-nested logit (CNL) model and random parameter nested logit model (RPNL), such as Jung et al. (2014), Cheng et al. (2015) and Yu et al. (2018b). In order to overcome the restriction of MNL model, NL model is first proposed as an extension of MNL model. NL model partitions the choice set into several sub-nests and puts alternatives that are similar and maybe correlate with each other in the same sub-nest, so that the correlation among alternatives within each sub-nest can be captured. However, no correlation across nests can be depicted and when alternatives cannot be partitioned into well separated nests, NL model is not appropriate. CNL model is a direct extension of NL

Table 1 Recent publications about spatio-temporal prediction methods

| Literature | Predition models | Temporal correlation | Spatial correlation |
|------------|------------------|----------------------|--------------------|
| Yu et al. (2018a) | Attention Based Spatial-Temporal Graph Convolutional Networks (ASTGCN) | Spatiotemporal attention mechanism | Graph Convolutional Networks (GCN) |
| Lin et al. (2018) | Graph Convolutional Network with Data-driven Graph Filter (GCN-DDGF) | GCN\textsubscript{reg}-DDGF (Long Short Term Memory, LSTM) | GCN\textsubscript{reg}-DDGF |
| Han et al. (2019) | Spatial-temporal Graph Convolutional Neural Networks for metro (SGCNNmetro) | Time series analysis | Convolution operation of stereogram |
| Guo et al. (2019) | Spatio-Temporal Graph Convolutional Networks (STGCN) | Temporal gated-convolution | Spatial graph-convolution |
| Wang et al. (2019) | Dynamic Spatial-Temporal Graph Convolutional Neural Networks (DSTGCNN) | Temporal convolution | Spatial graph-convolution |
| Geng et al. (2019) | Spatiotemporal Multi-Graph Convolution Network (ST-MGCN) | Contextual Gated Recurrent Neural Network (CGRNN) | Spatial feature (GCN) |
| Zhao et al. (2019) | Temporal Graph Convolutional Network (T-GCN) | Temporal feature (Gated Recurrent Unit, GRU) | Spatial feature (GCN) |
model, where each alternative may belong to more than one nest and the correlation across nests can be estimated.

(c) Mixed logit (MXL) model is suitable for handling random preference problems, such as correlation among alternatives and random coefficients. And it allows the unobserved factors to follow any distributions. See subsubsection 2.2.1 for details.

To the best of our knowledge, the majority of existing literature about researching the logit models based on utility functions, devote to one objective, i.e., maximizing the utility values of passengers. For comparative convenience, we list the detailed characteristics of some closely related references in Table 2, including transport modes, variables in utility functions and formulated models.

| Literature          | Transport modes                  | Variables                                           | Models                                                                 |
|---------------------|----------------------------------|-----------------------------------------------------|------------------------------------------------------------------------|
| Hess et al. (2005)  | Airplane                         | Access time, frequency, fare                        | Mixed logit model: The coefficient values were estimated and the difference between passenger groups were analyzed. |
| Park et al. (2006)  | Airplane, High-speed railway      | Access time, frequency, fare                        | Binary logit model: The coefficient values and the elasticity of each variable were estimated.                        |
| Jou et al. (2011)   | High-speed railway, Hyper-speed transportation Korea train express | In vehicle travel time, out of vehicle travel time, fuel cost Fare, access time, travel time, frequency | Mixed logit model: The importance of variables related to different passengers was analyzed.                        |
| Jung et al. (2014)  | High-speed railway               | Travel cost, travel time, service frequency, safety, duty free shopping availability | Multinomial and nested logit model: The different choices about different transport modes were analyzed.                |
| Lee et al. (2016)   | High-speed railway               | Passenger income, rapidity (travel time), economy (travel expenses), comfort (seat grades) | Mixed logit model: The coefficient values were estimated and the characteristics of business or leisure passengers in choosing transport modes were analyzed. |
| Han et al. (2020)   | Different railway traffic modes   | Travel expenses                                      | Mixed logit model based on improved nonlinear utility functions: The coefficient values and the optimal solutions of the model were solved by two improved heuristic algorithms. |
| This paper          | D-series overnight high-speed train | Travel expenses, Travel time, Arrival time, Passenger income | Mixed logit model based on nonlinear random utility functions: The probabilities of discrete random variables can be solved and the coefficient distributions and the optimal solutions of the model are estimated by an improved heuristic algorithm. |

Nowadays, few studies have been done on the market share of transport modes after newly-added overnight D-trains. Obviously, running overnight D-trains is not only convenient for long distance travel demand of passengers, but also improves the capacity and volume of the railway. So it has a good market demand and passenger demand, and there is also an urgent to accurately predict market share and transfer passenger demand of newly-added overnight D-trains. As mentioned above, there are two unsolved problems in the existing literature as follows.

(a) Many studies have explored transfer passenger demand prediction approaches for different transport modes, i.e., logit models, in which the heart is to establish utility functions in a linear way but without considering that there are nonlinear relationships between the utility values and the influence factors about different income passenger groups choosing different transport modes.

(b) The main focus of recent researches in Table 2 is how to reasonably and comprehensively choose the variables in the models when studying the influence factors of different transport modes. The number of variables involved in each reference is relatively small, so some traditional softwares, such as SPSS or ALOGIT, and traditional methods, such as Newton Method, can usually be used to estimate unknown parameters. Nevertheless, the growing variables make the time complexity large by Newton Method and it may lead to non-convergence of the results. Therefore, Han et al. (2020) adopted some heuristic algorithms to solve the problem and the experiment showed that heuristic algorithms were better than the traditional methods.

Specifically, the detailed contributions of this paper are summarized as follows.

(a) Some literature show that there is not a simple linear function relationship between the utility value of passengers and the influence factors about the travel intention of passengers. So in this paper, we establish different nonlinear utility functions which are mainly including the passenger income function, the ratio function of travel expenses to travel time value and the arrival time function for different types of transport modes. And we consider the travel intention of passengers to choose different combinations of transfer modes and transport modes.

(b) Due to the randomness of passengers choosing different transfer modes and different transport modes, total
travel expenses and total travel time is uncertainty. In order to better describe uncertainty factors in the real situation and make the model we established much closer to our actual decision-making optimization, this paper considers total travel expenses and total travel time as discrete random variables in the objective function and aims to maximize the total utility value of all involved passengers.

(c) In this paper, we adopt some representative basic heuristic algorithms, including Ant Colony Algorithm, Simulated Annealing Algorithm and Fireworks Algorithm, and then propose an improved combinatorial heuristic algorithm, i.e., Improved Fireworks-Simulated Annealing Algorithm, to estimate the unknown distributions of elements in parameter vectors of utility functions and solve the optimal solution of the model, respectively. Finally the comparison results of their running time, optimal solutions and convergence speed show that the improved algorithm we proposed is better than others in all respects, obviously.

The technical route of the specific research in this paper is shown in Figure 1. And the rest of this paper is organized as follows. In Section 2, a mixed logit model based on nonlinear random utility functions is formulated to prediction transfer passenger demand. And to reduce computational complexity, we adopt Maximum Simulated Likelihood Method to construct a corresponding likelihood function. Then, Section 3 develops some algorithms, including Metropolis-Hastings Algorithm and some heuristic algorithms to solve different problems in this paper, respectively. Next, we also design one case to demonstrate the effectiveness of the proposed approaches in Section 4. Finally, some conclusions and further studies are presented in Section 5.

2 Mathematical formulation

To characterize this problem of interest in a mathematical way, this section will explicitly discuss the formulation process for a mathematical model for transfer passenger demand prediction. It can be seen in Figure 1 that modeling technique in this paper is in the part of blue bottom. Firstly, a mixed logit model based on nonlinear random utility functions is established, and then to reduce the complexity of solving this model, we adopt Maximum Simulated Likelihood Method to restructure a corresponding likelihood function for the objective function.

2.1 Problem description

There are three types transport modes are considered and available for passengers when they travel along the route in this study (i.e., air, highway and railway). In order to simplify or idealize some realistic conditions, throughout this paper the following assumptions are made in order to formulate the problem.

Generally, the travel process of passengers from origin city to destination city can be divided into three stages: from a certain point in origin city to airport or train station or bus station, take an airplane or train or intercity bus (with different seat grades), from airport or train station or bus station to a certain point in destination city. Therefore many combination results will be produced after choosing different transfer modes and transport modes. Taking Beijing to Shanghai as an example is shown in Figure 2 which is the basis for setting the proposal probabilities of random variables in subsequent sections.

In this study, two types of discrete random variables are taken into consideration. The first type is expense denoted
Assumptions

(A1) The maximization of utility:

According to Utility Theory proposed by Gao (2011), passengers will evaluate the utility value of different available transport modes and always choose one transport mode which has maximum utility.

(A2) Under the condition of (A1), the complex relations between different passenger income levels and the utility values of choosing different transport modes are as follows.

(a) About air, passenger income is proportional to utility value, i.e., the higher passenger income is, the more utility value will be and vice versa.

(b) About railway, when passenger income is in a certain range, utility value is the highest and when income is lower or higher, utility value will decrease.

(c) About highway, passenger income is inversely proportional to utility value, i.e., the lower passenger income is, the more utility value will be and vice versa.

by \( \hat{e} \). It aims to produce the total travel expenses including transfer cost and price ticket for different passenger groups choosing different transfer modes and transport modes. The second type is corresponding time denoted by \( \hat{t} \), which indicates the total travel time including transfer time, waiting time and in-vehicle time (or running time).

In fact, passenger groups with different income levels have different consumption habits and the model is not developed at the level of the individual but at the level of the income groups. According to the ‘per capita disposable income of urban residents’ of the National Bureau of Statistics in China, we divided all passengers into three passenger groups with different income levels, i.e., low-income passenger group (Low- for short) whose monthly income is less than CNY3000, medium-income passenger group (Medium- for short) whose monthly income is between CNY3000 and CNY8000 and high-income passenger group (High- for short) whose monthly income is larger than CNY8000. Generally speaking, low-income passenger group tends to be more sensitive to total travel expenses, so they probably do not choose transport modes with high-quality, fast and high-price. And medium-income passenger group often seeks a more comfortable transport mode with acceptable price and good service quality. Finally, for high-income passenger groups, the choice of transport modes usually focuses on speed and comfort, and they often choose transport mode which is time-saving, comfortable and better service quality. Therefore, the probabilities of choosing different transfer modes and transport modes are different for passengers with different income levels, the probability distribution regularities of discrete random variables are as follows.

(a) About low-income passenger group, the lower total travel expenses are, the more probabilities will be, and travel time has little effect on them.

(b) About medium-income passenger group, all probabilities are similar and do not change with total travel expenses and total travel time.

(c) About high-income passenger group, the less total travel time is, the more probabilities will be, and travel expenses have few effect on them.

2.2 Mathematical model

Generally speaking, safety, rapidity, economy, convenience and comfort are the main factors affecting passengers’ choices of travel modes. Before this study, authors conducted a Stated Preference (SP) Survey with travel alternatives composed of a combination of attributes and values of each level to gather travel choice data. The survey asked 880 respondents (passengers who have experience in travelling the Beijing-Shanghai route by different transport modes) to choose one influence factor among multiple influence factors. The survey results of passenger travel considerations analysis are shown in Table 3. In addition, there are the travel intention results of different passenger groups about existing transport modes shown in Table 4.

Effective responses from passengers as shown in the table are applied in the construction of utility functions. On the whole, according to the travel characteristics of passengers,
the most important factors affecting travel intention of passengers are safety, which cannot be quantified and has little difference for each transport mode, rapidity, which can be quantified as travel time, and economy, which can be quantified as travel expenses. In addition to these objective factors, the income levels of passengers and arrival time also indirectly determine their travel intention to a large extent.

Therefore, in order to make the model much closer to the actual case, this paper not only considers two objective factors including travel expenses and travel time, but also adds passenger income and arrival time to construct utility functions. See subsubsection 2.2.2 for details.

2.2.1 Mixed logit model

The mixed logit (MXL) model (McFadden, et al. (2000), Hensher, et al. (2003) and Train (2003)) extends the standard logit model and overcomes the limitations in MNL model and NL model by allowing one or more parameters in the model to be randomly distributed. It means that different decision makers may have different preferences and the IIA property in MNL model no longer holds in this model. In this paper, we adopt mixed logit model to solve market shares and predict passenger demand. Firstly the standard logit model mainly includes the following definitions.

(a) Model parameters: A set of individuals (passenger groups \( Q = \{q | q = 1, 2, 3\} \) where \( q = 1 \) is low-income passenger group, \( q = 2 \) is medium-income passenger group and \( q = 3 \) is high-income passenger group), a set of available alternatives (transport modes \( M = \{m | m \in N_r\} \), a set of measured attributes \( X = \{e_m, t_m, s_p\} \) where \( s_q \) is monthly income of passenger group \( q \) of the individuals and their alternatives.

(b) Decision rules: The individual \( q \) will select the alternative \( m \) which maximizes individual utility. Therefore, the utility value is based on comparison and evaluation of each decision maker’s attractiveness to alternatives with different characteristics, which can be expressed as functions of different attributes.

(c) It is not possible to possess complete information about all attributes that determine one decision, so errors can arise for specification or observational reasons. To take into account the unobserved measurement error, the true attribute \( X^* \) or its function \( f(X^*) \) are effectively replaced by \( X + \varepsilon \) or \( f(X^*) + \varepsilon \), respectively. In general, for individuals who have the same set of alternatives and face the same constraints, it can be assumed that the unknown random disturbance term \( \varepsilon \) is independent and obey the identical distribution with the same mean and variance.

Based on the usual mentality of passengers for making choice, we suppose that a passenger group \( q \) always chooses one transport mode \( m \) with the maximum utility value. More precisely in this paper, a utility \( U_{mq} \) can be represented by two kind of components, i.e., observed representative component, \( V_{mq} \) which is a function of observable attributes \( X \), and an unknown random disturbance term, \( \varepsilon_{mq} \) which represents unobserved attributes or measurement errors or observation errors. Therefore the utility value \( U_{mq} = V_{mq} + \varepsilon_{mq} \) is random and associated with a probability, i.e.,

\[
P_{mq} = P\{U_{mq} > U_{lq}, \forall m \neq l \in M, q \in Q\},
\]

or more explicitly, i.e.,

\[
P_{mq} = P\{\varepsilon_{lq} \leq \varepsilon_{mq} + \left(V_{mq} - V_{lq}\right), \forall m \neq l \in M, q \in Q\}.
\]

This means that the probability of a passenger group \( q \) choosing one transport mode \( m \) equals the probability that the utility of choosing \( m \) is greater than that of any other choices. Depending on above joint probability distribution of disturbance term \( \{\varepsilon_{mq}, m \in M, q \in Q\} \), a specific random utility model can be obtained.

According to (c), if each distribution of \( \varepsilon_{mq} \) is assumed to be the Gumbel extreme value distribution and has the following cumulative distribution function, i.e.,

\[
P\{\varepsilon_{mq} \leq \varepsilon\} = \exp\left[-\exp\left(-\frac{\varepsilon - \varepsilon_{mq}}{\sqrt{6} \cdot \delta^2}\right)\right],
\]

with mean zero and variance \( \delta^2 \). So the distribution function about the difference between two random errors \( \varepsilon_{mlq} = \varepsilon_{mq} - \varepsilon_{lq} \) is

\[
F(\varepsilon_{mlq}) = \frac{\exp(\varepsilon_{mlq})}{1 + \exp(\varepsilon_{mlq})}.
\]

---

### Table 3 The weight of travel consideration factors (Unit: %)

| Factors     | Safety | Rapidity | Economy | Convenience | Comfort |
|-------------|--------|----------|---------|-------------|---------|
| Primary     | 51.1   | 30.2     | 8.2     | 4.0         | 1.2     |
| Second      | 4.5    | 34.3     | 33.4    | 11.5        | 2.5     |
| Third       | 5.6    | 12.6     | 32.0    | 22.9        | 23.2    |

### Table 4 The travel intention survey results

| Classification | Proportion (Unit: %) | Travel intention (Unit: person) |
|----------------|----------------------|---------------------------------|
|                |                      | Air | Highway | Railway | Low-income | 17  | 25  | 26  | 113 |
|                |                      |     |         |         | Medium-income | 46  | 57  | 43  | 302 |
|                |                      |     |         |         | High-income      | 37  | 57  | 17  | 240 |
In standard logit model, the probability of passenger group q choosing transport mode m can now be expressed as follows,

\[ p_{mq} = \frac{1}{\sum_{l \in M} \exp(V_{mlq} - V_{lq})} = \frac{\exp V_{mq}}{\sum_{l \in M} \exp V_{lq}}. \]  

(5)

The MXL model considers that the unknown parameter vector to be estimated is not a fixed value, but a random vector, which obeys a certain distribution due to different preferences of passengers with different income levels. Therefore, the probability of passenger group q choosing transport mode m should be the probability expected value of parameter vector \( \theta \) traversing all possible values, and can be expressed as follows,

\[ p_{mq} = \int \frac{\exp V_{mq}}{\sum_{l \in M} \exp V_{lq}} f(\theta | \psi) d\theta. \]  

(6)

The above probability function of MXL model can be regarded as the weighted average of the probability of standard logit model and the weight is determined by the distribution density function \( f(\theta | \psi) \). Among them, the elements in parameter vector \( \theta \) are random variables, which may obey normal distribution, uniform distribution, etc. \( \psi \) are the unknown characteristic parameters to be estimated in density function, such as the mean and variance of normal distribution.

2.2.2 Nonlinear random utility functions

As discussed in previous parts, the authors decide to use not only two major attributes, i.e., 'travel expense' and 'travel time', but also 'passenger income' attribute and 'arrival time' attribute. Next, we explore the function relationship between several attributes and utility value, respectively.

(a) Major attributes: 'travel expense' and 'travel time'

By analyzing the combinations of different transfer modes and transport modes, it can be concluded that when passenger chooses one certain combination, total travel expenses \( e_m \) and total travel time \( t_m \) can be determined accordingly. In other words, they are linked and once the expenses are determined, time will be determined accordingly. Generally, the probability distribution of discrete random variable \( \tilde{X} \) is as follows,

| \( X \) | \( x_1 \) | \( x_2 \) | \( \cdots \) | \( x_i \) | \( x_n \) |
|---|---|---|---|---|---|
| \( P \) | \( p_1 \) | \( p_2 \) | \( \cdots \) | \( p_i \) | \( p_n \) |

and \( E(\tilde{X}) = x_1 \cdot p_1 + x_2 \cdot p_2 + \cdots + x_i \cdot p_i + x_n \cdot p_n \) is called the mean or mathematical expectation and reflects the average level of \( \tilde{X} \). Therefore, in this paper, we adopt the mathematical expectations of two discrete random variables, \( E(\tilde{e}_m) \) and \( E(\tilde{t}_m) \), to calculate the corresponding utility functions.

Because the units of time and expenses are not uniform and belong to different dimensions, it can not be calculated in one function. So we introduce 'time value' attribute to replace 'travel time' in order to unify their dimensions. According to the research of Wardman (1998) about time value, they believe that the marginal utility of time increases with income. The unit (hour) time value of passenger group \( q \) is denoted by \( v_q = \frac{s_q}{30} \), \( q \in Q \) where the denominator is the number of hours converted per month. Generally speaking, the time value of high-income passenger group is higher than that of low-income passenger group and medium-income passenger group.

Obviously, the utility value about rapidity, economy and comfort of transport modes mainly measured by travel time reflected by fatigue recovery time and travel expenses with different seat grades. So the utility of choosing different transport modes is inversely proportional to travel time and directly proportional to travel expenses. According to Shi (2007), the function relationship between them is shown in Figure 3.

![Fig. 3 The relationship between travel time, travel expense and passenger utility](image)

It can be seen from the figure that, on the one hand, the less travel time is, the higher utility value is. But with the increase of time, the increment of utility value brought by saving unit time will decrease. On the other hand, the higher travel expense is, i.e., the higher seat grade is, the higher utility value is. But with the increase of travel expense, the increment of utility value brought by more unit expense will decrease. So in order to describe this relationship more accurately, we establish the following function
where $\phi_m^{(R)}(s_m, t_m; \gamma_m, \alpha_m)$ denotes ratio function of travel expenses to travel time value.

$$\phi_m^{(R)}(s_m, t_m; \gamma_m, \alpha_m) = \ln \left( \frac{E(s_m) + \gamma_m}{E(t_m) \cdot q + \alpha_m} \right).$$  \hspace{1cm} (7)$$

(b) 'Passenger income' attribute

According to assumption (A2), the passenger income functions should be established as curves which can map variables between 0 and 1 to standardize utility functions. When a specific mathematical model is lacking, a sigmoid function is often adopted.

In general, a sigmoid function has the following characteristics. This function and its inverse function are all monotonic. It is a bounded and differentiable mathematical function having a characteristic S-shaped growth curve which is constrained by a pair of horizontal asymptotes as $x \to \pm \infty$, shown in the solid line of Figure 4 and defined by the following formula where $(0, 0.5)$ is a symmetry point, i.e.,

$$S(x) = \frac{1}{1 + e^{-x}} = \frac{e^{x}}{e^{x} + 1}. \hspace{1cm} (8)$$

This function also has a non-negative first derivative at each point. Its derivative function which is bell-shaped can be expressed by itself shown in the dashed line of Figure 4 and defined by the following formula, i.e.,

$$S'(x) = \frac{e^{-x}}{(1 + e^{-x})^2} = S(x) \cdot [1 - S(x)]. \hspace{1cm} (9)$$

Air: \[ \phi_1^{(P)}(s_q, \beta_{11}, \beta_{12}) = \frac{1}{1 + \exp \left( -\frac{s_q + \beta_{11}}{\beta_{12}} \right)}, q \in Q. \] \hspace{1cm} (10)

Highway: \[ \phi_2^{(P)}(s_q, \beta_{21}, \beta_{22}) = 1 - \frac{1}{1 + \exp \left( -\frac{s_q + \beta_{21}}{\beta_{22}} \right)}, q \in Q. \] \hspace{1cm} (11)

Railway: \[ \phi_3^{(P)}(s_q, \beta_{31}, \beta_{32}) = \exp \left[ -\left( \frac{s_q + \beta_{31}}{\beta_{32}} \right)^2 \right], q \in Q. \] \hspace{1cm} (12)

(c) 'Arrival time' attribute

In order to analyze passengers’ preference and utility for arrival time, Lei (2018) conducted a research in the form of intention survey and let passengers choose the ideal arrival time. The results showed that there were two peaks in passengers’ choice of ideal arrival time, as shown in Figure 6 which was called bimodal distribution. The vertical axis represents probability of passengers who choose corresponding arrival time.

So in this paper, we choose probability density function with bimodal shape to describe this function relationship between them according to Jiang (1997), as shown below where $\phi_m^{(A)}(t_m; \gamma_m, \gamma_m2, \gamma_m3, \gamma_m4)$ denotes arrival time function about passenger group $q$ with transport mode $m$ and where $t_m^{(a)}$ is arrival time of transport mode $m$.

$$\phi_m^{(A)}(t_m; \gamma_m1, \gamma_m2, \gamma_m3, \gamma_m4) = \left[ (t_m^{(a)} - \gamma_m2)^2 + \gamma_m1 \right] \cdot \exp \left[ -\left( \frac{t_m^{(a)} - \gamma_m3}{2\gamma_m4} \right)^2 \right].$$ \hspace{1cm} (13)
Therefore based on the above analysis, we establish different nonlinear random utility functions for different transport modes. The utility value of passenger group \( q \) \((q \in Q)\) choosing transport mode \( m \) \((m \in M)\) expressed by the following function, i.e.,

\[
V_{mq} = \omega_1 \cdot \varphi_m^{(R)}(w_1, t_m, \alpha_m, \beta_m) + \omega_2 \cdot \varphi_m^{(P)}(s_q, \gamma_m) + \omega_3 \cdot \varphi_m^{(A)}(t_m, \gamma_m).
\]

Moreover, the coefficient parameter vector

\[
\theta = (\omega_{m1}, \omega_{m2}, \omega_{m3}, \alpha_m, \gamma_m, \beta_m),
\]

where \( m \in M \), \( \omega_m \) denotes the weight about ratio of travel expenses to travel time value, \( \omega_2 \) denotes the weight of passenger income, \( \omega_3 \) denotes the weight of arrival time, \( \alpha \) and \( \beta \) and \( \gamma \) are position and scale parameters in each formula. The elements in parameter vector \( \theta \) obey different distributions.

Then, in next subsection, the above nonlinear random utility function will be substituted into mixed logit model established in this paper. And to reduce the difficulty of solving this complex model, we will adopt Maximum Simulated Likelihood Method to construct a corresponding likelihood function.

### 2.2.3 Maximum Simulated Likelihood Method

The purpose of this study is to maximize the overall utility of all passenger groups in this model. And we need to estimate the unknown distributions of elements in parameter vector. Because the probability function of MXL model can not get the result by analytical method, it needs to use simulation method to calculate the probability and the optimal parameter value. In this paper, Maximum Simulated Likelihood Method is adopted to solve this problem and the specific steps are as follows.

**Step 1.** Calculate the simulation probability \( P_{mq}^* \).

Under the premise of given characteristic parameters \( \psi \), a parameter vector \( \theta \) is randomly selected from the density function and recorded as \( \theta^w \), \( w = 1 \) in the first extraction. Common random sampling methods are pseudo-random sequence method and quasi-random sequence method. According to **Equation 5**, calculate value of \( P_{mq}^*(\theta^w) \). Repeat iteration for \( W \) times and calculate the mean of all \( P_{mq}^*(\theta^w) \) as the simulation probability, i.e., \( \hat{P}_{mq}^* = \sum_{w=1}^{W} P_{mq}^*(\theta^w) / W \).

**Step 2.** Construct maximum likelihood operator.

The likelihood function is concave and can be constructed as follows under a suitable condition. Therefore, we can obtain a unique solution, i.e., the maximum value of the likelihood function.

(a) Construct simulation likelihood function. For a random sample of size \( M \), the simulation likelihood function can be viewed as the product of the choice probabilities associated with \( Q \) subsets of observations, in which the first subset includes \( q=1 \) individuals observed to have chosen alternative \( 1 \sim m_1 \), the next one \( q=2 \) individuals to have chosen alternative \( m_1 + 1 \sim m_1 + m_2 \), etc., all observations being independent, that is

\[
\phi^* = \prod_{m=1}^{M} \bar{P}_{m1} \cdot \prod_{m=m_1+1}^{m_1+m_2} \bar{P}_{m2} \cdots \prod_{m=m_1+m_2+1}^{M} \bar{P}_{mQ}.
\]

(b) Then in this study, we introduce 0-1 variables \( \gamma_{mq} \) where \( m \in M \) and \( q \in Q \), as selection indicators for selection of different passenger groups, i.e.,
3.2 we adopt 

\[ y_{mq} = \begin{cases} 
1 & \text{passenger group } q \text{ choose transport mode } m, \\
0 & \text{otherwise,} 
\end{cases} 
\]

and \( \sum_{m \in M} y_{mq} = 1, q \in Q. \)

Due to that each passenger group can only choose one transport mode and all passenger groups are independent, the expression of \( \phi^* \) can be simplified by \( y_{mq} \) as follows,

\[ \phi^* = \prod_{q \in Q} \prod_{m \in M} \Phi_{mq}^{y_{mq}}. \quad (16) \]

(c) Because \( \ln \phi^* \) is the monotone increasing function of \( \phi^* \), \( \ln \phi^* \) and \( \phi^* \) have the same extreme point. In order to solve this mixed logit model conveniently, we choose corre-

\[
\max \phi = \ln \phi^* = \sum_{q \in Q} \sum_{m \in M} y_{mq} \cdot \ln \Phi_{mq},
\]

where

\[
V_1q = \omega_{q1} \cdot \ln \left[ \frac{E(\epsilon_1) + \alpha_{q1}}{E(\epsilon_1) + \alpha_{q2}} \right] + \omega_{q2} \cdot \frac{1}{1 + \exp \left( -\frac{s_q + \beta_{q1}}{\beta_{q2}} \right)} + \omega_{q3} \cdot [t^{(q)}_{1} - \gamma_{12}] + \gamma_{11} \cdot \exp \left[ -\frac{(t^{(q)}_{1} - \gamma_{13})^2}{2\gamma_{14}^2} \right],
\]

\[
V_2q = \omega_{q2} \cdot \ln \left[ \frac{E(\epsilon_2) + \alpha_{q1}}{E(\epsilon_2) + \alpha_{q2}} \right] + \omega_{q2} \cdot \frac{1}{1 + \exp \left( -\frac{s_q + \beta_{q1}}{\beta_{q2}} \right)} + \omega_{q3} \cdot [t^{(q)}_{2} - \gamma_{22}] + \gamma_{21} \cdot \exp \left[ -\frac{(t^{(q)}_{2} - \gamma_{23})^2}{2\gamma_{24}^2} \right],
\]

and \( V_3q = \omega_{q3} \cdot \ln \left[ \frac{E(\epsilon_3) + \alpha_{q1}}{E(\epsilon_3) + \alpha_{q2}} \right] + \omega_{q2} \cdot \exp \left[ -\frac{(s_q + \beta_{q1})^2}{\beta_{q2}} \right] + \omega_{q3} \cdot [t^{(q)}_{3} - \gamma_{32}] + \gamma_{31} \cdot \exp \left[ -\frac{(t^{(q)}_{3} - \gamma_{33})^2}{2\gamma_{34}^2} \right], \)

s.t. \( f(\theta|\psi) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[ -\frac{(\theta - \mu)^2}{2\sigma^2} \right], \theta = (\omega_{m1}, \omega_{m2}, \omega_{m3}, \alpha_{m1}, \alpha_{m2}, \beta_{m1}, \beta_{m2}, \gamma_{m1}, \gamma_{m2}, \gamma_{m3}, \gamma_{m4}), \psi = (\sigma, \mu), m \in M, \)

\[ y_{mq} \in \{0, 1\}, \sum_{m \in M} y_{mq} = 1, q \in Q. \]

3 Solution approaches

In this section, we aim to develop solution approaches of the model we established, which are shown in the part of grey bottom in Figure 1, to solve different problems in this paper, respectively. Firstly, we approve Metropolis-Hastings (M-H) Algorithm to solve the probabilities of different passenger groups choosing different transfer modes and transport modes, according to proposal probabilities of discrete random variables. And then, we adopt three basic heuristic algorithms (Ant Colony Algorithm, Simulated Annealing Algorithm and Fireworks Algorithm) and two improved heuristic algorithms to solve the unknown distributions of elements in parameter vector to be estimated and obtain satisfactory solutions of objective function in short time, respectively. Finally, we compare the performance of these algorithms according to a real numerical experiment in Section 4.

3.1 Metropolis-Hastings Algorithm

Given the distribution irregularities in Section 2.1, we adopt Bayesian to deal with the probabilities of discrete random variables with different passenger income levels, namely prior probabilities in Bayesian statistics, and the probabilities are mainly determined by expert experiences and historical facts (sample data or observational data). Bayesian is a method to obtain objective posterior distribution on the basis of subjec-
tive prior distribution and the corresponding Bayesian formula is shown as follows, i.e.,
\[ p(\psi_i \mid x) = \frac{p(x \mid \psi_i) \cdot p(\psi_i)}{p(x)} = \frac{\sum_{j=1}^{n} p(x \mid \psi_j) \cdot p(\psi_j)}{p(x)} \]  
\[(18)\]
where \( p(\psi_i) \) denotes prior probability or distribution of \( \psi_i \), \( p(\psi_i \mid x) \) denotes the posterior probability or distribution of \( \psi_i \), \( p(x) \) denotes the marginal distribution of \( x \), and \( p(x \mid \psi_i) \) denotes the likelihood function of \( x \).

Usually, we need to sample from posterior probabilities of Bayesian. Since marginal distribution is a fixed value, posterior distribution is the proportion to product of prior distribution and likelihood function. The likelihood of marginal distribution may be not considered in the sampling of posterior distribution. But in general, posterior probabilities are difficult to solve, mainly because we can’t map to a random number that obeys Uniform Distribution. Therefore, the most feasible methods are Metropolis Algorithm and Metropolis-Hastings (M-H) Algorithm, so as to solve the posterior probabilities. And typically, prior probabilities in Bayesian are same as proposal probabilities in these two algorithms. Usually as the proposal probability \( q(j \mid i) \) doesn’t satisfy Detail Balance Condition \(^1\) in Metropolis Algorithm, we choose M-H Algorithm to solve this problem.

The M-H Algorithm is developed to construct a transition matrix \( P \) of a time-reversible Markov chain \( \{X_t\} \) with a given desired stationary distribution \( \pi \). At the current state \( i \), the next state \( j \) is proposed with a proposal probability \( q(i, j) \), which is a state transition probability of an arbitrary irreducible Markov chain on the state space \( N \), where \( q(i, j) > 0 \) if and only if \( i = j \), and \( q(i, j) = 0 \) for all \( i \neq j \). Let \( Q \cong \{q(i, j)\}_{i,j \in N} \) be a proposal (transition) matrix. The proposed state transition to \( j \) is accepted with an acceptance probability, i.e.,
\[ \alpha(j \mid i) = \min \left\{ 1, \frac{\pi(j) \cdot q(i \mid j)}{\pi(i) \cdot q(j \mid i)} \right\} \in (0, 1]. \]  
\[(19)\]
and rejected with probability \( 1 - \alpha(j \mid i) \) in which state \( i \). And then we construct the Markov chains by multiplied an acceptance probability, i.e.,
\[ \pi(i) \cdot q(j \mid i) \cdot \alpha(j \mid i) = \pi(j) \cdot q(i \mid j) \cdot \alpha(i \mid j). \]  
\[(20)\]
So we summarize the procedure of M-H Algorithm as follows.

---

**Step 1.** The relevant parameters need to be initialized, including proposal probability \( q \) of random variables, acceptance probability \( \alpha \) and maximum iteration times \( iter_{max} \).

**Step 2.** Do for \( i = 1, 2, ..., iter_{max} \).

**Step 2.1.** Generate a current state \( i \) and a new state \( j \) from the proposed distribution.

**Step 2.2.** Calculate the acceptance probability \( \alpha(j \mid i) \).

**Step 2.3.** Do the random sampling of a data point from \([0, 1]\) interval (uniform distribution), i.e., \( u \sim U \in [0, 1] \).

**Step 2.4.** If \( u \) is a convergence to stationary distribution, i.e., \( u < \alpha \) we retain this current state \( i \).

**Step 3.** Calculate posterior probability, which is equal to the number of preserved states divided by total iteration times.

---

As addressed above, the mathematical expectations of discrete random variables, i.e., \( E(e_m) \) and \( E(l_m) \), which can be calculate by the corresponding probabilities using M-H algorithm, are substituted into the objective function to be optimized.

---

**3.2 Heuristic algorithms**

According to Han, et al. (2020), heuristic algorithms not only cost less time than traditional algorithms (e.g., Newton method), but also outperform traditional algorithms in optimal solution. The specific reasons we summarized are as follows.

Firstly, Newton Method has strict requirements for functions, such that the second derivative of the objective function is continuous and differentiable. Secondly, the computational complexity is large because the inverse matrix or the pseudo-inverse matrix of Hessian matrix need to be solved in each step. Moreover, gradient explosion or gradient disappearance occurs easily in the iteration process. Finally, when the initial point is chosen improperly, it often leads to local convergence or non-convergence of the function.

On the contrary, heuristic algorithms can solve many problems of Newton Method. The main characteristics of some heuristic algorithms can be summed up from four aspects, i.e., a positive feedback mechanism, distributed computation, greedy heuristic search and strong robustness. Positive feedback mechanism can account for rapid discovery of good solutions, distributed computation can avoid premature convergence, and the greedy heuristic search can help find acceptable solutions in early stages of the search process.

Based on the principle of bionics, heuristic algorithms mainly abstract some phenomena in nature into algorithms to deal with the corresponding problems. By using them to approach the optimal solution as much as possible, a relative optimal solution is obtained. Usually, heuristic algorithms are not easy to fall into the local optimum but easy to find the global optimal solution. Moreover, the optimal solution of a search problem can be obtained in short time. For NP
problems, a better solution can also be obtained in polynomial time.

In this subsection, five heuristic algorithms including Ant Colony Algorithm (which is an imitative animal algorithm), Simulated Annealing Algorithm (which is an imitative physics algorithm) and Fireworks Algorithm (which is an imitative chemistry algorithm), Imporved Simulated Annealing Algorithm (which is proposed by Han et al. (2020)), Imporved Fireworks-Simulated Annealing Algorithm (which is proposed in this paper) are used to solve the unknown distributions \( f(\theta|\psi) \) of elements in parameter vector to be estimated, and the optimal solution of this model, i.e., \( \phi(\theta) \).

3.2.1 Ant Colony Algorithm

Ant Colony Algorithm (ACA), first proposed by Italian scholar, Marco Dorigo (1992) in his doctoral dissertation, is designed by imitating the cooperative manner of an ant colony and the characteristics of ants foraging behaviors, and then abstracting this manner into mathematical description. In biology, the foraging behaviors of an ant colony have the following characteristics.

(a) While building the paths from their nest to food, ants will deposit and sniff a chemical substance, which can mark the paths and provide ants with the ability to communicate with each other, called pheromones.

(b) Generally speaking, ants essentially move at random, but they always choose one path with higher concentration of pheromones and release a certain amount of pheromones to enhance the concentration of pheromones on this path. Therefore, the higher the concentration of pheromones is, the shorter the distance of corresponding path will be.

(c) With the continuous actions of ant colony, the shorter paths are more frequently visited and become more attractive for the subsequent ants. By contrast, the longer paths are less attractive because the pheromones trail will evaporate with the passing of time. Finally, the shortest way from nest to food is found.

Nowadays, this algorithm is widely used in optimization problems. And the basic idea of applying it is that a feasible solution is expressed by a walking path of an ant and all paths of the whole ant colony constitute the solution space. Finally, the whole ant colony will concentrate on one path which corresponds to the optimal solution. So by analyzing the foraging process of ant colony and aiming at the research content of this paper, we solve estimate the unknown distributions of elements in parameter vector and the optimal solution of model by Ant Colony Algorithm and detailed solution process is summarized by mathematical method as follows.

\begin{itemize}
  \item [Step 1.] The relevant parameters need to be initialized, including ant colony size (the total number of ants) \( ant\_max \), the pheromones volatilization coefficient \( \rho \), total pheromones released by ants for one iteration \( Q \) (constant), a constant of transfer probability \( p_t \), maximum number of iterations \( iter\_max \). Initial pheromones \( \tau(0) = \phi(\theta) \).
  \item [Step 2.] Do for \( i = 1, 2, \cdots, iter\_max \),
    \begin{itemize}
      \item [Step 2.1.] Do for \( t = 1, 2, \cdots, ant\_max \),
        \begin{itemize}
          \item [Step 2.1.1.] Each ant is randomly placed in different positions, and the next position of ant \( t \), namely next feasible solution, is determined according to transfer probability \( p_t \), i.e.,
          \begin{equation}
            p_t = \frac{\max_{\tau(i) \in [1, ant\_max]} \tau(i) - \tau(i)}{\max_{\tau(i) \in [1, ant\_max]} \tau(i)}.
          \end{equation}
        \end{itemize}
      \item [Step 2.1.2.] Do following judgements about the distance between ant \( t \) and the position with the highest concentration of pheromones (i.e., the current maximum of the function):
        \begin{itemize}
          \item [(a)] Local Search: The closer the distance is, the transfer probability \( p_t \) will be smaller, and variable value \( \theta \) will tend to be fine-tuning, i.e., \( \theta(i + 1) = \theta(i) + \frac{\lambda}{r_n} \) where \( r_n \) is a 0-1 random number, and \( \lambda = 1/(t+1) \) is the expected level of \( \theta \) and it gradually decreases as the iteration progresses.
          \item [(b)] Global Search: The farther the distance is, the transfer probability \( p_t \) will be larger, and the algorithm will tend to search for the optimal value in a wider range, i.e., \( \theta(i + 1) = \theta(i) + \frac{\lambda}{r_n} \cdot (ub - lb)/2 \), where \( ub \) is upper bound and \( lb \) is lower bound of \( \theta \).
        \end{itemize}
      \end{itemize}
    \end{itemize}
  \end{itemize}
  \item [Step 3.] Calculate the pheromones of each path, and update the concentration of pheromones by the iteration formula as follows, i.e., \( \tau(i + 1) = (1 - \rho) \cdot \tau(i) + Q \cdot \phi(\theta) \). At the same time, the optimal solution of the current iteration is recorded.
\end{itemize}

3.2.2 Simulated Annealing Algorithm

Simulated Annealing Algorithm (SAA), first proposed by American physicist, Metropolis (1953), is initially inspired by the change rules of internal molecular state and internal energy of solids in the process from high temperature to low temperature. This algorithm takes the temperature of the solid as the control parameter, and with the decrease of temperature, the internal energy of the solid (i.e., the objective function value) decreases gradually until it reaches the global minimum. It is actually a greedy algorithm meanwhile the random factors are introduced in its search process. It accepts a solution worse than the current one with a certain probability, so it is possible to jump out of the local optimal solution and reach the global optimal solution.

Aiming at the research content of this paper, we estimate the unknown distributions of elements in parameter vector and solve the optimal solution of model by Simulated Annealing Algorithm and detailed solution process is summarized as follows.
Step 1. The relevant parameters need to be initialized, including an initial temperature $T_0$, a termination temperature $T_{\text{min}}$, an initial state $\theta_0$, maximum number of disturbance $dis_{\text{max}}$ and maximum number of iterations $\text{iter}_{\text{max}}$. Let current temperature be $T_0$ and current state be $\theta_0$.

Step 2. Do for $i = 1, 2, \cdots, \text{iter}_{\text{max}}$.

Step 2.1. Do for $i = 1, 2, \cdots, dis_{\text{max}}$.

Step 2.1.1. Calculate the internal energy of the current state (objective function value) $\phi(\theta_i)$. Transform the current state $\theta_i$ to a new one $\theta_j$ by exchanging certain elements.

Step 2.1.2. Calculate internal energy of this new state $\phi(\theta_j)$ and increment $\Delta \phi = \phi(\theta_j) - \phi(\theta_i)$. If $\Delta \phi \leq 0$, the new state $\theta_j$ is accepted and let $\theta_0 = \theta_j$; otherwise, the new state is accepted when a random number $\rho (0 < \rho < 1)$ is greater than $P(\phi) = \exp(-\Delta \phi / T_i)$.

Step 2.2. Exit the loop until $T_i < T_{\text{min}}$.

Step 3. Let $T_{i+1} = \lambda \cdot T_i$ where the cooling coefficient $\lambda$ ($0 < \lambda < 1$) is used to control the speed of cooling. The larger the value of $\lambda$ is, the slower temperature will drop.

3.2.3 Fireworks Algorithm

Fireworks Algorithm (FA), first proposed by Chinese professor, Tan (2010) in Peking University, is inspired by the splendid fireworks in the sky. In this algorithm, two explosion (search) processes are employed and mechanisms for keeping diversity of sparks are also well designed. However, the explosion of fireworks are distinguished from good to bad. For a good explosion, the generated sparks are dense and numerous, and vice versa.

For mimicking the process of setting off fireworks, a rough framework of this algorithm is depicted as follows. When a firework is set off, a shower of sparks will fill the local space around a specific point where the firework is set off through the sparks generated in the explosion. When we need to find a point $x$ satisfying $f(x) = y$, we can continually set off ‘fireworks’ in the local space until one ‘spark’ is fairly near the point $x$. For each generation of explosion, we first select $n$ locations, where $n$ fireworks are set off. Then after explosion, the locations of sparks are obtained and evaluated. When the optimal location is found, the algorithm stops. Otherwise, $n$ other locations are selected from the current sparks for the next generation of explosion.

In this paper, we adopt a simplified version of Fireworks Algorithm, called Bare Bones Fireworks Algorithm (BBFWA) proposed by Li (2018). It adopts only three essential operators of Fireworks Algorithm: The explosion operator generates numerous explosion sparks around the locations of the fireworks. The mutation operator mutates the locations of the fireworks to generate mutation sparks. The selection operator selects the fireworks of the new generation from the sparks in the last generation. The mechanism of this algorithm is easy to understand and the time complexity of this algorithm is linear. It can be easily implemented on a variety of platforms, which is convenient for real-world applications. Experimental results on a standard benchmark indicate that its performance is serviceable and stable.

Aiming at the research content of this paper, we estimate the unknown distributions of elements in parameter vector and solve the optimal solution of model by Bare Bones Fireworks Algorithm whose detailed solution process is summarized as follows.

3.2.4 Improved Simulated Annealing Algorithm

The cooling coefficient $\lambda \in (0, 1)$ is an important parameter affecting the convergence of Simulated Annealing Algorithm from the procedure of this algorithm. When this coefficient is too large, the solution accuracy is high but the algorithm runs for a long time. On the other hand, when this coefficient is too small, the search accuracy is low. Therefore, considering this disadvantage, Improved Simulated Annealing Algorithm of cooling coefficient (ISAA-CC) is first proposed by Han et al. (2020) to improve the efficiency of the basic algorithm and the quality of the optimal solution. They improved the cooling strategy in the basic algorithm as follows.

The value of $\lambda$ is set as large as possible at the beginning of the algorithm, so that the algorithm has strong global search ability. With iterations of the algorithm, the value of $\lambda$ decreases, in order that the algorithm can search for the optimal solution better. As a result, a function relationship between the cooling coefficient $\lambda$ and the iteration times as shown in Figure 7 will be established as follows.

$$
\lambda_i = 1 - \frac{1}{1 + \exp \left( \frac{i - a}{b} \right)}, a = \frac{\text{iter}_{\text{max}}}{2}, b = \frac{\text{iter}_{\text{max}}}{2 \ln 9}.
$$
3.2.5 Improved Fireworks-Simulated Annealing Algorithm

Because of the fast convergence speed and running speed of Fireworks Algorithm, the best optimal solution of Improved Simulated Annealing Algorithm, we propose an improved algorithm combining the advantages of both algorithms in this paper, named Improved Fireworks-Simulated Annealing Algorithm (IFW-SAA).

The BBFWA is not globally convergent because there is no stochastic mutation operator. This situation can be fixed easily by setting the reduction coefficient $C_r$ and the amplification coefficient $C_a$ or adopting a suitable mutation operator if necessary. According to Improved Simulated Annealing Algorithm in subsection 3.2.4, we introduce the cooling strategy into Fireworks Algorithm. A function relationship between the reduction coefficient $C_r \in (0, 1)$ and the iteration times is shown as Equation 22. A similar function relationship between the amplification coefficient $C_a \in (1, 2)$ and the iteration times is shown as Equation 23. The values of $C_r$ and $C_a$ are set as large as possible at the beginning of the algorithm, so that the algorithm has strong global search ability. With the iterations of the algorithm, the values decrease, in order that the algorithm can search for the optimal solution better.

$$C_r = 1 - \frac{1}{1 + \exp\left(-\frac{i - a}{b}\right)}, a = \frac{\text{iter\_max}}{2}, b = \frac{\text{iter\_max}}{2\ln9}. \quad (22)$$

$$C_a = 2 - \frac{1}{1 + \exp\left(-\frac{i - a}{b}\right)}, a = \frac{\text{iter\_max}}{2}, b = \frac{\text{iter\_max}}{2\ln9}. \quad (23)$$

In the following Section 4.2, we use these heuristic algorithms to solve the problem, respectively, and then compare the performance of them. The results show that heuristic algorithms proposed in this paper are better than others not only in running time but also in optimal solution.

4 Numerical experiment

In this section, we will give in-depth analysis and predict transfer passenger demand of newly-added overnight D-trains. Due to the representation of Beijing-Shanghai high-speed railway in China and the plan to open its eastern line, it is urgent for us to predict the passenger demand on this line, and to judge the feasibility of operation plan about overnight D-trains. Therefore, this real-world case in Beijing-Shanghai corridor is implemented to illustrate the applications of the proposed model and algorithms, which is shown in the part of yellow bottom in Figure 1.

In this numerical experiment, the set of transport modes is $M = \{m | m = 1, 2, 3, 4, 5\}$. For newly-added transport mode, i.e., D-series overnight high-speed train (overnight D-train for short) where $m = 5$, the most competitive transport modes are airplane where $m = 1$, intercity bus where $m = 2$, T-series express train (T-train for short) where $m = 3$, and G-series high-speed train (G-train for short) where $m = 4$.

The algorithms in this paper are all coded and solved by Python 2.7 Version in Pycharm on Window 7 personal computer with 3.4 Gb processor. And all data used in numerical experiments are shown in Appendix A including full expenses and time of different transport modes in Beijing-Shanghai corridor.

4.1 The computation of Metropolis-Hastings Algorithm

In this paper, according to the different proposal probabilities of discrete random variables, M-H Algorithm is used to solve the probabilities of different passenger groups choosing different transfer modes when travelling in different transport modes. We set that the iteration times and the number of samples are all equal to 1000, and the results are shown in Figure 8. The five radar charts are about airplane, intercity bus, T-train, G-train and newly-added overnight D-train, successively.

Three words or abbreviated letters at each corner denote transfer mode in Beijing, seat grade of transport mode and transfer mode in Shanghai, successively. Among them, ‘AE’ is short for ‘economy class of airplane’, ‘AB’ is short for ‘business class of airplane’, ‘IS’ is short for seat of intercity bus, ‘TS1’ is short for ‘hard seat of T-train’, ‘TS2’ is short for ‘hard sleeper of T-train’, ‘TS3’ is short for ‘soft sleeper of T-train’, ‘GS’ is short for ‘second class seat of G-train’, ‘GF’ is short for ‘first class seat of G-train’, ‘DS’ is short for ‘second class seat of overnight D-train’, ‘DC’ is short for ‘couchette of overnight D-train’.

We can see from Figure 8 that the results obtained by M-H Algorithm do not make much different from the initial proposal distribution, still conform to the description of distribution regularities in Section 2.1 and are consistent with the actual situation.
Mixed logit model based on nonlinear random utility functions: A transfer passenger demand prediction method on overnight D-trains

Five irregular closed-loops of medium-income passengers in radar charts show that the probabilities of different combinations of transport modes and transfer modes to be chosen are approximately equal. Generally speaking, less time corresponds to higher expense about transfer modes, i.e., bus, metro or taxi, and more comfortable seats correspond to higher prices about transport modes, i.e., airplane, intercity bus, T-train, G-train or overnight D-train. Therefore, the corresponding probabilities about each irregular closed-loops of low-income passengers and those of high-income passengers in each radar chart are just the opposite. The fundamental reason for this result is that travel time is the primary factor for high-income passengers, while travel expense is the primary factor for low-income passengers.

4.2 The computation of five heuristic algorithms

Five heuristic algorithms are adopted to solve the mixed logit model based on non-linear random utility functions established in this paper. The distributions of parameter vector to be estimated, i.e., \( f(\theta|\psi) \), and the optimal solution of the model, i.e., \( \phi(\theta) \), are obtained. For the mean errors, the lower boundaries of the search space are \( lb=-1 \) and the upper boundaries of the search space are \( ub=1 \). For the standard deviations, the lower boundaries of the search space are \( lb=0 \) and the upper boundaries of the search space are \( ub=1 \). There are initialized parameters:

(a) Ant Colony Algorithm: \( ant_{max} = 100, \rho = 0.9, Q = 1, p_0 = 0.2 \) and \( iter_{max} = 100 \).

(b) Simulated Annealing Algorithm: \( T_0 = 100, T_{min} = 1e-3, \lambda = 0.9 \) and \( iter_{max} = 100 \).

(c) Fireworks Algorithm: \( C_a = 1.2, C_r = 0.8 \) and \( iter_{max} = 10000 \).

(d) Improved Simulated Annealing Algorithm: \( T_0 = 100, T_{min} = 1 \) and \( iter_{max} = 90 \).

(e) Improved Fireworks-Simulated Annealing Algorithm proposed in this paper: \( iter_{max} = 10000 \).

By editing and running codes of five heuristic algorithms, the estimations of characteristic parameters in parameter vector, including the mean errors (‘Mean’) and standard deviations (‘Std.’ for short) are shown in Table 5. Moreover, the probabilities that different (Low-, medium- and high-) income passenger groups choose different transport modes (including airplane, intercity bus, T-train, G-train and newly-added overnight D-train) and market shares of different transport modes are shown in Table 6.

Fig. 8 The probabilities of different passenger groups choosing different transport modes

There are the contrast of five heuristic algorithms as follows Table 7 and Figure 9 in this numerical experiment. Figure 9 shows the market shares of different passenger groups choosing different transport modes where ‘ACA’ is short for Ant Colony Algorithm, ‘SAA’ is short for Ant Colony Algorithm, ‘BBFWA’ is short for Ant Colony Algorithm, ‘ISAA-CC’ is short for Improved Simulated Annealing Algorithm proposed by Han et al. (2020), ‘IFW-SAA’ is short for Improved Fireworks-Simulated Annealing Algorithm. By comparison from five heuristic algorithms, the market shares of overnight D-train are very close and the maximum difference is not more than 0.0328. The running time in Table 7 is unified into the time spent in about 10000 iterations. It can be concluded from this table that the optimal solutions of five heuristic algorithms are also very close and the maximum difference is not more than 0.123. Moreover, improved
Table 5 The characteristic parameters in distributions

|                      | $\omega_{a1}$ | $\omega_{a2}$ | $\omega_{a3}$ | $\omega_{a4}$ | $\beta_{h1}$ | $\beta_{h2}$ | $\gamma_{a1}$ | $\gamma_{a2}$ | $\gamma_{a3}$ | $\gamma_{a4}$ |
|----------------------|---------------|---------------|---------------|---------------|--------------|--------------|--------------|--------------|--------------|--------------|
| **1. Estimated by Ant Colony Algorithm.** |
| Airplane (m = 1)    | 0.3828        | -0.091        | 1.0           | 0.0728        | 0.9297       | 0.6656       | 0.6368       | 1.0           | 0.9492       | 1.0           | 1.0           |
| Intercity bus (m = 2) | -0.308        | -0.727        | 1.0           | 1.0           | 0.0015       | 1.0           | 1.0           | 1.0           | 0.1308       | 0.4558       | 0.3953        |
| T-train (m = 3)     | -0.068        | -0.083        | 1.0           | 1.0           | 0.00016      | 1.0           | 1.0           | 1.0           | 0.3337       | 0.3958       | 1.0           |
| G-train (m = 4)     | -0.149        | -0.516        | 0.5108        | 0.5026        | 0.7882       | 1.0           | 1.0           | 1.0           | 0.6945       | 0.2048       |
| Overnight-D-train (m = 5) | -0.075       | -0.158        | 0.2434        | 1.0           | 0.7227       | 1.0           | 1.0           | 1.0           | 0.204       | 0.3024       |

| **2. Estimated by Simulated Annealing Algorithm.** |
| Airplane (m = 1)    | 0.3506        | 0.4971        | 0.4639        | 0.4699        | 0.8743       | 0.2251       | 0.7532       | 0.4359       | 0.5242       | 0.0533       | 0.9419        |
| Intercity bus (m = 2) | 0.3294        | 0.7016        | 0.8196        | 0.9126        | 0.7292       | 0.9628       | 0.6885       | 0.5499       | 0.4594       | 0.8805       | 0.2607        |
| T-train (m = 3)     | 0.0911        | 0.4051        | 0.2431        | 0.0259        | 0.0271       | 0.8466       | 0.7078       | 0.1635       | 0.3096       | 0.6289       | 0.0598        |
| G-train (m = 4)     | 0.6344        | 0.4578        | 0.2366        | 1.0           | 0.2149       | 0.3499       | 0.5912       | 0.7649       | 0.7681       | 0.2046       | 0.4066        |
| Overnight-D-train (m = 5) | 0.0991       | 0.8939        | 0.4564        | 0.8787        | 0.2029       | 0.2574       | 0.3998       | 0.3646       | 0.2419       | 0.6775       | 0.6749        |

| **3. Estimated by Improved Simulated Annealing Algorithm.** |
| Airplane (m = 1)    | 0.3189        | 0.2061        | 0.3846        | 0.858         | 0.3311       | 0.1394       | 0.228         | 0.1024       | -0.793       | 0.0929       | 0.2391        |
| Intercity bus (m = 2) | 0.0889        | 0.535         | 0.9139        | -0.825        | -0.3095      | 0.8744       | 0.0201       | -0.264       | -0.799       | -0.856       | 0.0598        |
| T-train (m = 3)     | 0.0891        | 0.3928        | 0.3842        | 0.7944        | 0.8481       | 0.2491       | 0.2367       | 0.8071       | 0.2029       | 0.5793       | 0.4572        |
| G-train (m = 4)     | 0.5143        | 0.6639        | 0.4033        | 0.5991        | 0.7114       | 0.4434       | 0.6903       | 0.6238       | 0.1065       | 0.2885       | 0.6471        |
| Overnight-D-train (m = 5) | -0.049       | -0.094        | -0.408        | -0.222        | -0.0725      | -0.238       | 0.112         | -0.298       | 0.6209       | 0.7199       | -0.526        |

| **4. Estimated by Improved Simulated Annealing Algorithm.** |
| Airplane (m = 1)    | 0.2528        | 0.6475        | 0.4726        | 0.7014        | 0.0415       | 0.9628       | 0.5843       | 0.4304       | 0.2579       | 0.0003       |
| Intercity bus (m = 2) | 0.5218        | 0.1805        | 0.8141        | 0.3913        | 0.3902       | 0.2061       | 0.1569       | 0.5717       | 0.9906       | 0.7378       |
| T-train (m = 3)     | 0.3746        | 0.5265        | 0.1299        | 0.6585        | 0.8458       | 0.9971       | 0.6368       | 0.5717       | 0.9906       | 0.7378       |
| G-train (m = 4)     | 0.5734        | 0.1189        | 0.2945        | 0.6552        | 0.9152       | 0.6849       | 0.9693       | 0.0254       | 0.6743       | 0.8189       |
| Overnight-D-train (m = 5) | 0.4326       | 0.8543        | 0.9408        | 0.5682        | 0.8031       | 0.8172       | 0.6036       | 0.8881       | 0.1996       | 0.5429       | 0.4429        |

| **5. Estimated by Improved Fireworks-Simulated Annealing Algorithm.** |
| Airplane (m = 1)    | 0.0981        | 0.6839        | 0.8978        | 0.2046        | 0.0172       | 0.3747       | 0.8119       | 0.5483       | 0.4304       | 0.2579       | 0.0003       |
| Intercity bus (m = 2) | 0.0994        | 0.9018        | 0.8516        | 0.1244        | 0.9591       | 0.1418       | 0.5957       | 0.4641       | 0.6899       | 0.3841       | 0.2122       |
| T-train (m = 3)     | 0.2510        | 0.8887        | 0.8799        | 0.5691        | 0.3876       | 0.3952       | 0.6312       | 0.7765       | 0.6432       | 0.7046       | 0.4054       |
| G-train (m = 4)     | 0.5290        | 0.4641        | 0.5625        | 0.9621        | 0.6306       | 0.8354       | 0.4844       | 0.6194       | 0.7091       | 0.7358       | 0.1314       |
| Overnight-D-train (m = 5) | 0.8138       | 0.9463        | 0.1675        | 0.8872        | 0.4814       | 0.5056       | 0.4777       | 0.7685       | 0.8087       | 0.2352       | 0.4968       |
Mixed logit model based on nonlinear random utility functions: A transfer passenger demand prediction method on overnight D-trains

Table 6 The market shares

| Probabilities    | Low- | Medium- | High- | Market shares |
|------------------|------|---------|-------|---------------|
| 1. Solved by Ant Colony Algorithm. |
| Airplane         | 0.0669 | 0.0977 | 0.1191 | 0.2837 |
| Intercity bus    | 0.0679 | 0.0607 | 0.0537 | 0.1823 |
| T-train          | 0.0352 | 0.0351 | 0.0349 | 0.1052 |
| G-train          | 0.0697 | 0.1285 | 0.0389 | 0.2371 |
| D-train          | 0.0399 | 0.1136 | 0.0382 | 0.1917 |
| 2. Solved by Simulated Annealing Algorithm. |
| Airplane         | 0.0935 | 0.1185 | 0.1086 | 0.3206 |
| Intercity bus    | 0.0384 | 0.0342 | 0.0322 | 0.1048 |
| T-train          | 0.0245 | 0.0239 | 0.0221 | 0.0705 |
| G-train          | 0.0928 | 0.1563 | 0.0961 | 0.3452 |
| D-train          | 0.0465 | 0.0557 | 0.0567 | 0.1589 |
| 3. Solved by Fireworks Algorithm. |
| Airplane         | 0.0754 | 0.1061 | 0.2077 | 0.3892 |
| Intercity bus    | 0.0345 | 0.0425 | 0.0508 | 0.1278 |
| T-train          | 0.0398 | 0.0544 | 0.0533 | 0.1475 |
| G-train          | 0.0611 | 0.0516 | 0.0601 | 0.1728 |
| D-train          | 0.0549 | 0.0541 | 0.0537 | 0.1627 |
| 4. Solved by Improved Simulated Annealing Algorithm. |
| Airplane         | 0.0751 | 0.0948 | 0.0874 | 0.2573 |
| Intercity bus    | 0.0564 | 0.0705 | 0.1118 | 0.2387 |
| T-train          | 0.0506 | 0.0511 | 0.0446 | 0.1463 |
| G-train          | 0.0667 | 0.0617 | 0.0634 | 0.1918 |
| D-train          | 0.0505 | 0.0531 | 0.0623 | 0.1659 |
| 5. Solved by Improved Fireworks-Simulated Annealing Algorithm. |
| Airplane         | 0.0378 | 0.0452 | 0.1526 | 0.2356 |
| Intercity bus    | 0.0533 | 0.0569 | 0.0539 | 0.1391 |
| T-train          | 0.0498 | 0.0509 | 0.0539 | 0.1546 |
| G-train          | 0.0972 | 0.0941 | 0.0986 | 0.2869 |
| D-train          | 0.0189 | 0.0268 | 0.1381 | 0.1838 |

Table 7 The comparison of five heuristic algorithms in results

| Basic heuristic algorithms | Running time | Optimal solution |
|----------------------------|--------------|------------------|
| ACA | 2819s | -2.7389 |
| SAA | 1593s | -2.6517 |
| BBFVA | 1529s | -2.6742 |
| Improved heuristic algorithms | | |
| ISAA-CC | 1229s | -2.6491 |
| IFW-SAA | 963s | -2.6159 |

In summary, the aim of the model we established is to maximize the log-likelihood function and then in order to maximize utility of all passenger groups. It can be seen from Figure 10 that, as the iteration times increase, both Ant Colony Algorithm and Simulated Annealing Algorithm fluctuate greatly during the search process and have the slowest convergence speed that converge to the optimal solution more than 9000th iterations. Due to the advantage of Fireworks Algorithm, the improved algorithm we proposed in this paper has fastest convergence speed, and it converges to the optimal solution until about 500th iterations. The second is Fireworks Algorithm and it converges to the optimal solution until about 5300rd iterations. Moreover, Improved Simulated Annealing Algorithm converges to the optimal solution until about 7600th iterations.

4.3 Transfer passenger demand prediction

From Figure 9, we can see that the rankings of market shares solved by all algorithms have no big differences. Among them, G-train is one of the most popular transport modes in Beijing-Shanghai corridor at any income levels not only because of its high speed but also due to its appropriate departure time and high running frequency. Meanwhile, the main reasons we concluded in this paper why newly-added overnight D-trains attract passengers are as follows.

(a) Long-distance and high-speed operation at night is not only more comfortable than travelling by trains in the daytime, but also saves the night accommodation costs for passengers.

(b) Overnight D-trains run at a maximum speed per hour of 250 kilometers and are faster than overnight K-trains and overnight T-trains which run at a maximum speed per hour of 160 kilometers.

(c) The operation range of overnight D-trains has been increased from about 1500 kilometers to about 2500 kilometers, so it is more suitable for passengers who choose long-distance and fast railway travel, especially those aim at business and travelling.

Finally, if we know the data of total passenger demand between Beijing and Shanghai in a certain day or a certain year, based on the above numerical results, the transfer passenger demand of new transport modes, such as newly-added overnight D-trains in this paper, can be solved.

5 Conclusions and future researches

The passenger demand prediction is a common method for managing the passenger flow in order to reduce the mismatch between transportation resources allocation and pas-
To research transfer passenger demand of newly-added overnight D-trains and characterize the problem in a mathematical way, a mixed logit model based on nonlinear random utility functions is formulated in this paper. According to Maximum Simulated Likelihood Method, the likelihood function of the model is formulated to maximize the overall utility of passenger groups. In particular, two types of discrete random variables are introduced in utility functions, i.e., total travel expenses and total travel time, and we adopt M-H Algorithm to iteratively solve the probabilities.

While we can use traditional algorithms to solve the model established in this paper, the calculation complexity is high due to relatively more variables and parameters and it becomes a significant problem in the solution process. In view of this fact, the model is solved by five heuristic algorithms, respectively, and these algorithms are compared in running time and optimum solutions. Furthermore by performing a real-world instance in Beijing-Shanghai corridor, we prove that the improved heuristic algorithm we proposed is superior to other algorithms not only in running time but also in optimal solution.

Further researches can focus on the following several aspects. (a) This study only considers transfer passenger demand at Beijing station and Shanghai station. In fact, the passenger demand is also related to stations along the road in reality. Thus, with a series of passenger demand data, how to generate the robust and reliable passenger demand prediction model is a significant topic in the further research. (b) Due to the over-length of this paper, we only design three typical basic heuristic algorithms to solve the problem. So more effective and improved heuristic algorithms can also be studied in our future research.

Acknowledgements This research is supported by National Natural Science Foundation of China (Project 62072025).

Compliance with ethical standards
(a) Conflict of Interest: Author Bing Han and author Shuang Ren declare that they have no conflict of interest.
(b) Ethical approval: This article does not contain any studies with human participants performed by any of the authors.
(c) Informed consent: Informed consent was obtained from all individual participants included in the study.
Authors’ contribution

Bing Han: Literature search and review, model and algorithm establishment, numerical experiment, manuscript writing and editing
Shuang Ren: Guidance and recommendation

References

Ahn Y, Kowada T, Tsukaguchi H, Vandebona U (2017) Estimation of passenger flow for planning and management of railway stations. Transportation Research Procedia, 25:315–330

Castro-Neto M, Jeong Y, Jeong MK, Han LD (2009a) Aadtt prediction using support vector regression with data-dependent parameters. Expert Systems with Applications, 36(2):2979–2986

Castro-Neto M, Jeong YS, Jeong MK, Han LD (2009b) Online-svr for short-term traffic flow prediction under typical and atypical traffic conditions. Expert Systems with Applications, 36(3):6163–6173

Chen H, Grant-Muller S (2001) Use of sequential learning for short-term traffic flow forecasting. Transportation Research Part C: Emerging Technologies, 9(5):319–336

Cheng H, Yang X (2015) Random parameter nested logit model for combined departure time and route choice. International Journal of Transportation Science and Technology, 4(1):93–106

D M, KE T (2000) Mixed mnl models for discrete response. Journal of Applied Econometrics, 15(5):447–470

Dia H (2001) An object-oriented neural network approach to short-term traffic forecasting. European Journal of Operational Research, 131(2):253–261

Dorigo M (1992) Optimization, learning and natural algorithms. PhD thesis, Italy: Department of Electronics, Politecnico di Milano.

Eleni I (2007) Prediction of non-recurrent short-term traffic patterns using genetically optimized probabilistic neural networks. Operational Research, 7(2):171–184

Eleni I, Karlaftis M, Golasia J (2010) Spatio-temporal short-term urban traffic volume forecasting using genetically optimized modular networks. Computer-Aided Civil and Infrastructure Engineering, 22(5):317–325

Gao H Y (2011) Western Economics. China Renmin University Press, Beijing.

Gelhausen MC, Berster P, Wilken D (2018) A new direct demand model of long-term forecasting air passengers and air transport movements at german airports. Journal of Air Transport Management, 71:140–152

Geng X, Li Y, Wang L, Zhang L, Yang Q, Ye J, Liu Y (2019) Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. In: The 33rd AAAI Conference on Artificial Intelligence (AAAI),

Glišović N, Milenković M, Bojić N, švdlenka L, Avramović Z (2016) A hybrid model for forecasting the volume of passenger flows on serbian railways. Operational Research, 16(2):271–285

Guang Z, Yang I, Jian L (2017) Forecast of short-term passenger flow of urban railway stations based on seasonal arima model. In: International Conference on Electrical and Information Technologies for Rail Transportation

Guo S, Lin Y, Feng N, Song C, Han W (2019) Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In: The 33rd AAAI Conference on Artificial Intelligence (AAAI),

Han B, Ren S, Bao J (2020) Mixed logit model based on improved nonlinear utility functions: A market shares solution method of different railway traffic modes. Sustainability, 12(4):1406

Han Y, Wang S, Ren Y, Wang C, P G, Chen G (2019) Predicting station-level short-term passenger flow in a city-wide metro network using spatiotemporal graph convolutional neural networks. ISPRS International Journal of Geo-Information, 30(8):1–25

Hensher D, Greene W (2003) The mixed logit model: the state of practice. Transportation, 30(2):133–176

Hess S, Polak JW (2005) Mixed logit modeling of airport choice in multi-airport regions. Journal of Air Transport Management, 11(2):59–68

Huang ZJ, Feng S (2014) Grey forecasting model in the application of railway passenger flow prediction research. Technology and Economy in Areas of Communications, 16(1):57–60

Ivanov V, Osetrov E (2018) Application of artificial neural networks and singular-spectral analysis in forecasting the daily traffic in the moscow metro. In: European Physical Journal Web of Conferences, Jiang R (1997) One new probability density function with bimodal shape. Journal of Changsha Communications University, 13(1):1–8

Jonas B, Mindaugas M, Vladis I, Skirmantas M (2004) Analysis and forecast of the dynamics of passenger transportation by public land transport. Transport, 19(1):3–8

Jou RC, Hensher DA, Hsu TL (2011) Airport ground access mode choice behavior after the introduction of a new mode: a case study of taoyuan international airport in taiwan. Transportation Research Part E: Logistics and Transportation Review, 47(3):371–381

Jung SY, Yoo KE (2014) Passenger airline choice behavior for domestic short haul travel in south korea. Journal of Air Transport Management, 38:43–47

Lee JK, Yoo KE, Song KH (2016) A study on travelers’ transport mode choice behavior using the mixed logit model: A case study of the seoul-jeju route. Journal of Air Transport Management, 56:131–137
We set Beijing municipal government as origin and Shanghai municipal government as destination, and then the total expenses and time of different transport modes in Beijing–Shanghai corridor (OD) are shown in Figure 2. Amount them, the data of travel expenses (Unit: CNY) and travel time (Unit: hour) for different transport modes are from Qunar.com and 12306.com, and the data of transfer expenses and time for different transfer modes are from Baidu Map. Due to the large number of certain transport modes, the range of airplane arrival time is from 10:00AM to 6:00PM, the range of G-train arrival time is from 10:00AM to 6:00PM, and the range of D-train arrival time is from 7:00AM to 10:00PM the next day.

And then total travel expenses (Unit: CNY) and total travel time (Unit: hour) of different combinations of trans-
Table 8 The data of travel expenses and travel time

1. Airplane (Belongs to air)

| Travel Type | Price ticket | Transfer cost and time | Details |
|-------------|--------------|------------------------|---------|
|             |              | (Airport ↔ Government)  |         |
| Beijing municipal government → Beijing capital airport → Shanghai hongqiao airport → Shanghai municipal government | Economy class 1450 | Bus/Metro (O) 28 | 1h9mins (Metro line 2 → Airport express line) |
|             | Business class 5680 | Taxi (O) 94 | 47mins |
|             |                      | Bus/Metro (D) 5 | 47mins (Metro line 2) |
|             |                      | Taxi (D) 64 | 48mins |

2. Intercity bus (Belongs to highway)

| Travel Type | Price ticket | Transfer cost and time | Details |
|-------------|--------------|------------------------|---------|
|             |              | (Bus station ↔ Government)  |         |
| Beijing municipal government → Beijing station → Shanghai station → Shanghai municipal government | 342 | Bus/Metro (O) 2 | 53mins (Bus No.120) |
|             |              | Taxi (O) 25 | 25mins |
|             |              | Bus/Metro (D) 3 | 31mins (Metro line 1) |
|             |              | Taxi (D) 23 | 17mins |

3. T-train (Belongs to railway)

| Travel Type | Price ticket | Transfer cost and time | Details |
|-------------|--------------|------------------------|---------|
|             |              | (Train station ↔ Government)  |         |
| Beijing municipal government → Beijing station → Shanghai station → Shanghai municipal government | Hard seat 177.5 | Bus/Metro (O) 3 | 24mins (Metro line 2) |
|             | Hard sleeper 304.5 | Taxi (O) 24 | 14mins |
|             | Soft sleeper 476.5 | Bus/Metro (D) 3 | 20mins (Metro line 1) |
|             |                      | Taxi (D) 27 | 13mins |

4. G-train (Belongs to railway)

| Travel Type | Price ticket | Transfer cost and time | Details |
|-------------|--------------|------------------------|---------|
|             |              | (Train station ↔ Government)  |         |
| Beijing municipal government → Beijing south station → Shanghai hongqiao station → Shanghai municipal government | Second class seat 553 | Bus/Metro (O) 4 | 38mins (Metro line 2 → line 4) |
|             | First class seat 933 | Taxi (O) 25 | 25mins |
|             |                      | Bus/Metro (D) 5 | 45mins (Metro line 2) |
|             |                      | Taxi (D) 65 | 50mins |

5. Overnight D-train (Belongs to railway)

| Travel Type | Price ticket | Transfer cost and time | Details |
|-------------|--------------|------------------------|---------|
|             |              | (Train station ↔ Government)  |         |
|             |              | |         |
|             | Second class seat 426 | Bus/Metro (O) 4 | 38mins (Metro line 2 → line 4) |
|             | Couchette 630 | Taxi (O) 25 | 25mins |
|             |                      | Bus/Metro (D) 5 | 45mins (Metro line 2) |
|             |                      | Taxi (D) 65 | 50mins |

port modes and transfer modes, and proposal probabilities (i.e., prior probabilities subjectively based on Section 2.1) of different (Low-, medium- and high-) income passenger groups are shown in Table 9.
Table 9 The total expenses, total time and proposal probabilities

| Combinations (OD)                      | Travel expenses | Travel time | Low- | Medium- | High- |
|---------------------------------------|-----------------|-------------|------|---------|-------|
| **1. Airplane (Belongs to air)**       |                 |             |      |         |       |
| Metro → Economy class → Metro         | 1523            | 6.016       | 2/9  | 1/8     | 1/10  |
| Metro → Economy class → Taxi          | 1582            | 6.033       | 7/36 | 1/8     | 1/20  |
| Taxi → Economy class → Metro          | 1589            | 5.649       | 1/6  | 1/8     | 1/5   |
| Taxi → Economy class → Taxi           | 1648            | 5.666       | 5/36 | 1/8     | 3/20  |
| Metro → Business class → Metro        | 5713            | 6.016       | 1/9  | 1/8     | 1/10  |
| Metro → Business class → Taxi         | 5772            | 6.033       | 1/12 | 1/8     | 1/20  |
| Taxi → Business class → Metro         | 5779            | 5.649       | 1/18 | 1/8     | 1/5   |
| Taxi → Business class → Taxi          | 5838            | 5.666       | 1/36 | 1/8     | 3/20  |
| **2. Intercity bus (Belongs to highway)** |             |             |      |         |       |
| Bus → Bus seat → Metro                | 347             | 19.899      | 2/5  | 1/4     | 1/10  |
| Bus → Bus seat → Taxi                 | 367             | 19.666      | 3/10 | 1/4     | 1/5   |
| Taxi → Bus seat → Metro               | 370             | 19.433      | 1/5  | 1/4     | 3/10  |
| Tax → Bus seat → Taxi                 | 390             | 19.2        | 1/10 | 1/4     | 2/5   |
| **3. T-train (Belongs to railway)**    |                 |             |      |         |       |
| Metro → Hard seat → Metro             | 1835            | 16.65       | 2/13 | 1/12    | 1/30  |
| Metro → Hard seat → Taxi              | 1975            | 16.5334     | 11/78| 1/12    | 1/15  |
| Taxi → Hard seat → Metro              | 2045            | 16.4834     | 5/39 | 1/12    | 1/10  |
| Taxi → Hard seat → Taxi               | 2185            | 16.3667     | 3/26 | 1/12    | 2/15  |
| Metro → Hard sleeper → Metro          | 3105            | 16.65       | 4/39 | 1/12    | 1/30  |
| Metro → Hard sleeper → Taxi           | 3245            | 16.5334     | 7/78 | 1/12    | 1/15  |
| Taxi → Hard sleeper → Metro           | 3315            | 16.4834     | 1/13 | 1/12    | 1/10  |
| Taxi → Hard sleeper → Taxi            | 3455            | 16.3667     | 5/78 | 1/12    | 2/15  |
| Metro → Soft sleeper → Metro          | 4825            | 16.65       | 2/39 | 1/12    | 1/30  |
| Metro → Soft sleeper → Taxi           | 4965            | 16.5334     | 1/26 | 1/12    | 1/15  |
| Taxi → Soft sleeper → Metro           | 5035            | 16.4834     | 1/39 | 1/12    | 1/10  |
| Taxi → Soft sleeper → Taxi            | 5175            | 16.3667     | 1/78 | 1/12    | 2/15  |
| **4. G-train (Belongs to railway)**    |                 |             |      |         |       |
| Metro → Second class seat → Metro     | 562             | 7.417       | 2/9  | 1/8     | 1/10  |
| Metro → Second class seat → Taxi      | 583             | 7.201       | 7/36 | 1/8     | 1/5   |
| Metro → Second class seat → Taxi      | 622             | 7.5         | 1/6  | 1/8     | 1/20  |
| Taxi → Second class seat → Taxi       | 643             | 7.284       | 5/36 | 1/8     | 3/20  |
| Metro → First class seat → Metro      | 942             | 7.417       | 1/9  | 1/8     | 1/10  |
| Taxi → First class seat → Metro       | 963             | 7.201       | 1/12 | 1/8     | 1/5   |
| Metro → First class seat → Taxi       | 1002            | 7.5         | 1/18 | 1/8     | 1/20  |
| Taxi → First class seat → Taxi        | 1023            | 7.284       | 1/36 | 1/8     | 3/20  |
| **5. Overnight D-train (Belongs to railway)** |             |             |      |         |       |
| Metro → Second class seat → Metro     | 435             | 14.12       | 2/9  | 1/8     | 1/10  |
| Metro → Second class seat → Taxi      | 456             | 13.9        | 7/36 | 1/8     | 1/5   |
| Metro → Second class seat → Taxi      | 495             | 14.2        | 1/6  | 1/8     | 1/20  |
| Taxi → Second class seat → Taxi       | 516             | 13.98       | 5/36 | 1/8     | 3/20  |
| Metro → Couchette → Metro             | 639             | 14.12       | 1/9  | 1/8     | 1/10  |
| Metro → Couchette → Taxi              | 660             | 13.9        | 1/12 | 1/8     | 1/5   |
| Metro → Couchette → Taxi              | 699             | 14.2        | 1/18 | 1/8     | 1/20  |
| Taxi → Couchette → Taxi               | 720             | 13.98       | 1/36 | 1/8     | 3/20  |