

JUNSHENG HUANG, Ph.D. candidate¹
(Corresponding author)
E-mail: 19114037@bjtu.edu.cn
TONG ZHANG, Ph.D.²
E-mail: 16114195@bjtu.edu.cn
RUNBIN WEI, Ph.D. candidate¹
E-mail: 19114056@bjtu.edu.cn

Traffic Planning
Original Scientific Paper
Submitted: 14 Oct. 2020
Accepted: 15 Jan. 2021
DOI: 10.7307/ptt.v33i5.3736

¹ Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport
Beijing Jiaotong University
No.3 ShangYuanCun, Haidian District,
Beijing 100044, China

² China Waterborne Transport Research Institute
Ministry of Transport of the People's Republic of China
No.8 West TuCheng Road, Haidian District,
Beijing 100088, China

URBAN RAILWAY TRANSIT TIMETABLE OPTIMISATION BASED ON PASSENGER-AND-TRAINS MATCHING – A CASE STUDY OF BEIJING METRO LINE

ABSTRACT

Due to the congested scenarios of the urban railway system during peak hours, passengers are often left behind on the platform. This paper firstly brings a proposal to capture passengers matching different trains. Secondly, to reduce passengers' total waiting time, timetable optimisation is put forward based on passengers matching different trains. This is a two-stage model. In the first stage, the aim is to obtain a match between passengers and different trains from the Automatic Fare Collection (AFC) data as well as timetable parameters. In the second stage, the objective is to reduce passengers' total waiting time, whereby the decision variables are headway and dwelling time. Due to the complexity of our proposed model, an MCMC-GASA (Markov Chain Monte Carlo-Genetic Algorithm Simulated Annealing) hybrid method is designed to solve it. A real-world case of Line 1 in Beijing metro is employed to verify the proposed two-stage model and algorithms. The results show that several improvements have been brought by the newly designed timetable. The number of unique matching passengers increased by 37.7%, and passengers' total waiting time decreased by 15.5%.

KEYWORDS

urban railway system; train matching; timetable optimisation; AFC data; machine learning.

1. INTRODUCTION

Recent years have witnessed a tremendous increase in the urban railway transit ridership. In China, by the end of 2017, 17 billion ridership had been carried by urban railway transit [1] (Mao 2018). As for the Beijing Metro, 4.53 billion passengers were transported in 2019. The maximum load factor on Line 1 even reached 100% also in 2019. Thus, during the peak hours in the major metropolitan stations, heavy ridership remains a challenge to the daily regular operational schedule. Besides, a growing number of passengers will return to work after the COVID-19 crisis in 2020. How to deal with heavy ridership in metro stations to guarantee public hygiene is the concern of many scholars. One of the practical approaches is travel reservation [2] (Han et al. 2020). Travel reservation in metro stations refers to the reservation of passenger's departure time before travelling, which belongs to travel demand management (TDM). Another approach is the fare-reward scheme [3] (Yang et al. 2018). Passengers will be rewarded a free trip if the travelling period is not during peak hours. Regardless of the travel reservation or the fare-reward scheme, one primary issue is passenger-and-trains matching. This concept refers to providing the probabilities

of passengers matching different trains. Moreover, based on the passenger-and-trains matching, timetable optimisation could become more accurate. Thus, the investigation of timetable optimisation combining passenger-and-trains matching merits further discussion.

Besides passenger-and-trains matching, a well-designed timetable would increase passengers' satisfaction. In addition, a poorly designed timetable would lead to some unexpected situations, e.g., just missing the connecting train [4] (Guo et al. 2016), longer waiting time [5] (Barrena et al. 2014), etc. With this concern, it is essential to elaborately design an efficient, dynamic timetable for the congested urban railway transit to satisfy passenger demands [6, 7] (Robenek et al. 2018, Wang et al. 2015). Specifically, many efforts from literature might help us to better understand what the efficient and dynamic timetable would bring us. An efficient timetable would reduce passengers' travel costs in terms of passengers' waiting time [8] (Zhu et al. 2017), and a dynamic timetable would be more efficient in balancing passenger demands and transit services [9] (Fu et al. 2003). Therefore, at present, timetable optimisation is mainly based on passenger demands. Indeed, passenger demands can be inherent in the Automatic Fare Collection (AFC) data. In general, the AFC data [10] (Jiang et al. 2016) provides detailed individual information including departure station, destination station, entry time and exit time, etc. With the help of useful AFC information, macroscopic and microcosmic statistical indicators would be easily obtained. From the perspective of macroscopic statistical indicator, it includes passenger volume [11] (Shi et al. 2018) at peak hour, etc. Concerning the microcosmic statistical indicator, it consists of passengers' various behaviours [12, 13] (Sun et al. 2012, Zhou et al. 2015), including transfer behaviour, train matching behaviour, walking behaviour, etc. How to infer these implicit macroscopic and microcosmic indicators from the AFC data remains a major challenge. In this line of research, emphasis is placed on inferring passengers matching different trains from the massive AFC data. Similarly, Kusakabe et al. [14] (2010) proposed the "train choice behaviour" but it did not, however, reflect each passenger's train choice behaviour accurately.

In general, the objective of timetable optimisation is heterogeneous. Yang et al. [15] (2017) proposed a bi-objective model to minimise passengers' total travel time and net energy consumption. Binder et al. [16] (2017) delivered a tri-objective model that considers passenger satisfaction, operational costs, and deviation from the original timetable. Parbo et al. [17] (2014) considered transferring time minimisation in the system. Sels et al. [18] (2016) minimised passengers' travel time to optimise the timetable. However, a limited number of research studies have been devoted to timetable optimisation based on each passenger's train matching behaviour obtained by machine learning (ML). In this study, a model to minimise passengers' total waiting time considering each passenger train matching is designed in the context of the peak hour. Additionally, due to the indeterminate train matching of each passenger, the formula of passengers' total waiting time should be well re-designed considering the probabilities of passengers matching different trains. Unlike the formula proposed by Newell [19] (1971), in this study, the probabilities of passengers matching different trains are added in the passengers' total waiting time.

In summary, the main contribution of this paper is developing a two-stage model to deal with the passenger-and-trains matching problem and timetable optimisation. In detail, (1) in the first stage, a likelihood function of the probabilities is proposed to deal with passenger-and-trains matching problem. (2) In the second stage, passengers' total waiting time is considered as the objective to optimise the timetable. (3) This research fills the gap of timetable optimisation based on passenger-and-trains matching behaviour.

The remainder of this paper is organised as follows. The overall problem statement is presented in Chapter 1. Chapter 2 contains a pre-processing for the AFC data. Chapter 3 introduces the models and algorithms utilised for solving the two-stage model. Chapter 4 is utilised to verify the model and the proposed algorithm. The final chapter concludes with a summary of the findings.

2. PROBLEM DESCRIPTION

Due to the overcrowding scenarios at the station, some passengers who miss the first train would be categorised as left-behind passengers. A stampede of left-behind passengers would struggle

to board the next trains. In essence, this problem belongs to the train matching category, which is the microcosmic issue from the passengers’ perspective. Similarly, the passenger flow assignment problem in the metro network is to achieve the macroscopic matching between the passenger volume and the network capacity. However, complexity is a common issue encountered in the traditional passenger flow assignment model [20] (Sun et al. 2015), especially in the context of peak hours. Theoretically, the concept of passenger-and-trains matching is proposed in this paper, motivated by the phenomenon of passengers left behind [21] (Zhu et al. 2018). The probabilities of passengers matching different trains is gathered from the AFC data records and timetable parameters.

To the best of our knowledge, each passenger’s journey time on train, dwelling time at each station, and headway are fundamental factors to affect the probabilities of passengers matching different trains. In general, passenger journey time consists of passenger waiting time, passenger walking time, and in-vehicle consuming time. A simplified diagram (Figure 1), which does not consider passenger transfer behaviour, is applicable to demonstrate the probabilities of passengers matching different trains.

Schematically, passenger and train movements can be described as particular curves temporally and spatially. Indeed, each passenger’s departure time $t_{i,u}^{in}$ and arrival time $t_{i,v}^{out}$ are known from the AFC system. After entering the entry gate,

passengers should walk to the origin platform under their different speeds. Passenger access and egress walking time are approximated by the truncated normal distribution, where it reflects the fluctuations in passenger speeds. Due to different circumstances on the platform, the set of alternative trains which the passenger will probably pick up are denoted in black colour in Figure 1. Herein, the train that the passenger picks up is no longer deterministic. Correspondingly, $P_1^{i,u}$ and $P_2^{i,u}$ are the probabilities of this passenger matching the two black trains, respectively. In this example, this passenger would not pick up the white trains, because the departure or arrival time of these trains goes beyond the passenger’s journey time limit. Thus, the probability of this passenger matching the white trains would be set as zero. The sum of $P_1^{i,u}$ and $P_2^{i,u}$ would be equal to one. Generally, the advantages of the proposed passenger-and-trains matching lie in its ability to give better interpretability in accordance with the practical experience.

3. DATA PREPARATION

3.1 Passenger types

In public transport, a great amount of useful information is implicit in the data resources. In this paper, we only discuss passenger-and-trains matching where their origins and destinations are on the same line. In other words, passenger routes should be modelled explicitly if passengers have transfer

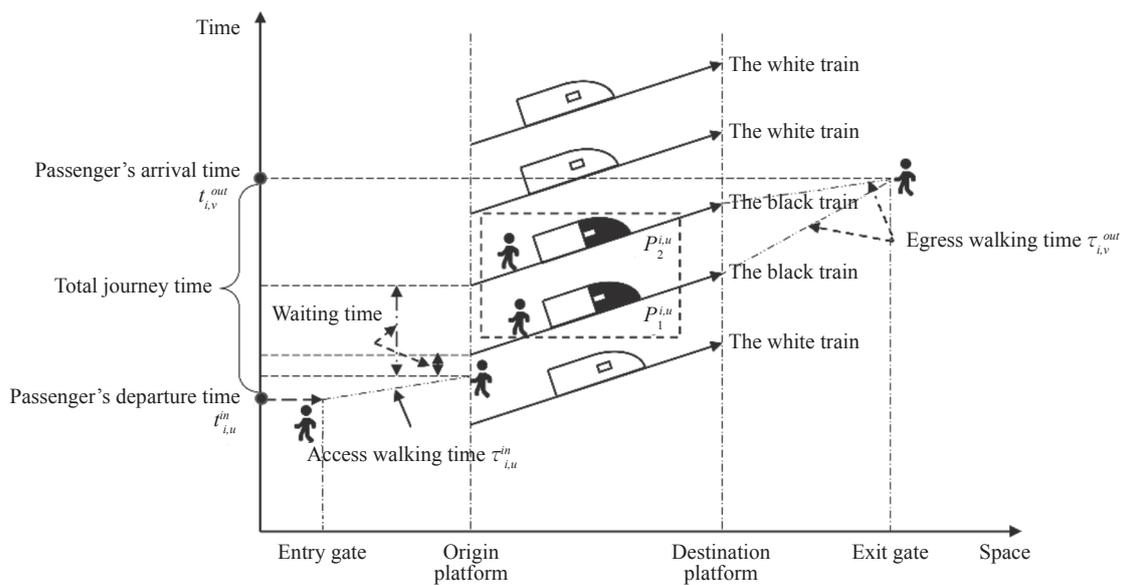


Figure 1 – Possible passenger-and-trains matching

behaviour. In general, three types of passengers should be considered according to their origins and destinations.

Type 1: passengers who transfer from Line 1 are represented by the dot line in *Figure 2*. For this type of passengers, their origins are on Line 1, and in contrast, their destinations are on Line 2. Passenger-and-trains matching at the original station on Line 1 is considered.

Type 2: passengers who transfer to Line 1 are represented by the dash line in *Figure 2*. The destinations of this type of passengers are on Line 1, but their origins, on the other hand, are on Line 4. Passenger-and-trains matching at the transfer station on Line 1 is considered.

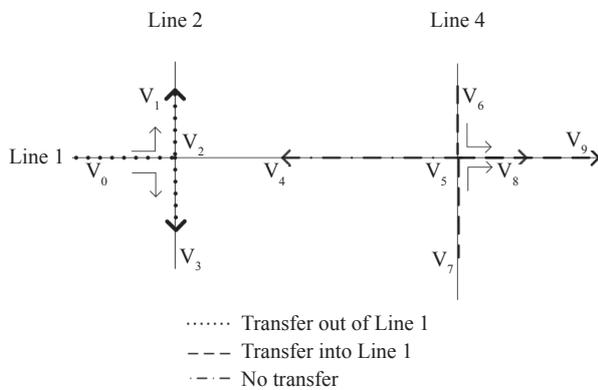


Figure 2 – Passengers' different routes through Line 1

Type 3: passengers who do not have transfer behaviour are represented by the chain line in *Figure 2*. Regarding this type of passengers, their origins and destinations are on the same line, which is why passenger-and-trains matching at the original station is considered.

Without loss of generality, if neither the passengers' origins nor destinations are typically on Line 1, and they would choose Line 1 to reach their destinations, this type of passengers can be categorised as a special combination of Type 1 and Type 2.

3.2 Determining routes by adopting BFS

According to the three types of passengers, the premise of calculating the probabilities of passengers matching different trains is to obtain the passenger's route if the passenger has transfer behaviour. In our paper, the passenger's route is assumed to strictly follow a predetermined route, which is obtained by the Breadth-First Search (BFS). In other words, the proposed BFS is not followed by the passenger and neither are the effects of normal deviations by the deterministic journey time. Then, the passenger's determined route would be decomposed into several parts which belong to different lines. For descriptive convenience, passenger's journey time on each decomposed route is approximately calculated by the proportion for which the length of the decomposed route

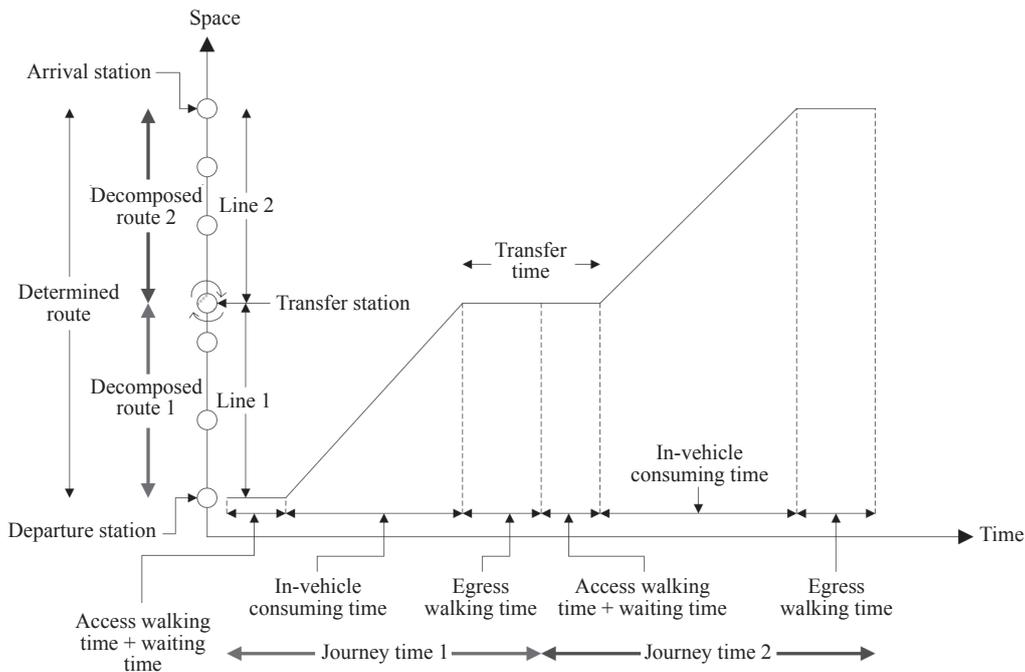


Figure 3 – The components of each decomposed journey time

accounts in the entire route. Therefore, two aspects should be considered. Firstly, in our paper, passenger's journey time on each decomposed route is independent. Secondly, passenger's journey time on each decomposed route also comprises passenger's waiting time, passenger's access walking time, passenger's egress walking time, and in-vehicle consuming time. In other words, passenger's transfer time at the transfer station is roughly separated into three parts (Figure 3), including passenger's egress walking time, passenger's access walking time, and passenger's waiting time.

In general, the BFS is a basic methodological approach that solves graph traversal. The metro network is a complex graph that involves nodes and edges. Specifically, the node represents the station, and the edge represents the traverse between two adjacent stations. Thus, exploring the route between any two stations is the extension of the graph traversal.

In the BFS, firstly, the scheduling discipline of the queue is set as first-in-first-out (FIFO). Secondly, three sets utilised in the BFS are defined to collect different nodes. Namely, W is the set whose nodes are still not accessed; B is the set whose nodes were already accessed; G is the set whose nodes are going to be accessed. Therefore, the algorithm flowchart of the BFS is depicted in Figure 4.

4. METHODOLOGICAL FRAMEWORK

In this study, the passenger's route obtained by the BFS is the first step to address the probabilities of passengers matching different trains. Moreover, the passenger's journey time on each decomposed route is also deterministic. Thus, according to the passenger's journey time and timetable parameters at the current stage, the probabilities of passengers matching different trains would be determined by the following proposed models and algorithms. Note that the timetable parameter is one of the basic factors that affect the probabilities of passengers matching different trains. Then, if timetable parameters are changed, the probabilities of passengers would be changed as well, which seems to be a domino effect. Therefore, it is significant to increase the probability of passengers matching the first train to reduce passenger waiting time by adjusting timetable parameters. Obviously, from the perspective of the urban railway management, it is also useful to adjust timetable parameters which include dwelling time of the train at each station and headway between two consecutive trains. With this concern, a two-stage model is proposed to deal with the relationship between the probabilities of passengers matching different trains and timetable optimisation. Specifically, in stage one, the aim is to obtain the probabilities of passengers matching different

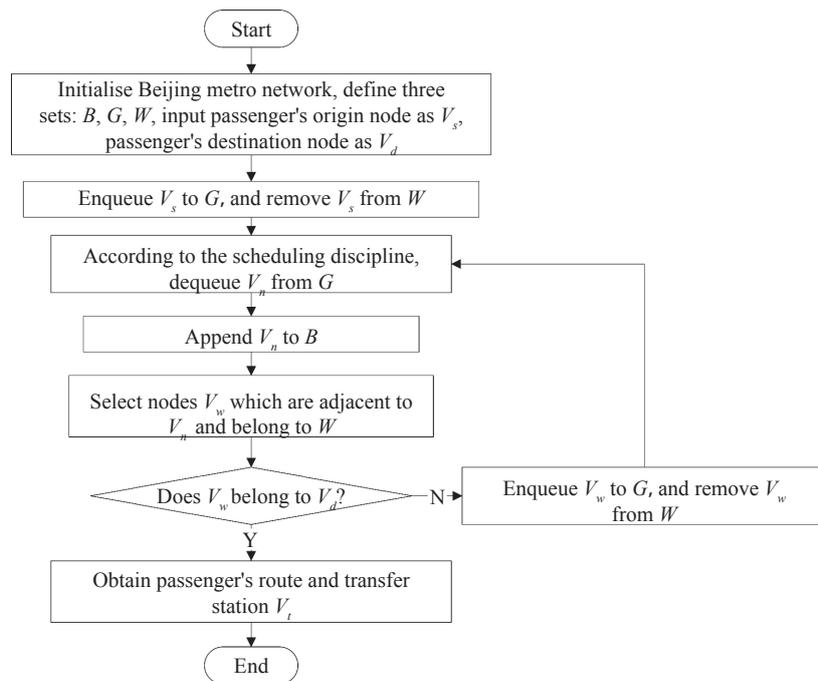


Figure 4 – The flowchart of BFS

trains. In the stage two, the objective is set to reduce passengers' total waiting time by adjusting the timetable.

4.1 Assumptions

Before formulating the two-stage model, several assumptions are provided.

- 1) In our proposed framework, the minimum dataset should include the AFC data and timetable information.
- 2) Passenger route is strictly followed by a predetermined route, which is obtained by the BFS.
- 3) Each passenger's decomposed route is independent.
- 4) Passenger's access and egress walking time on each decomposed route follows the truncated normal distribution.
- 5) The number of alternative trains that the passenger will probably pick up is no more than four.
- 6) Trains are punctual according to the scheduled timetable.
- 7) Trains follow the even schedule with a constant headway.
- 8) The number of trains operating within a timetable is sufficient.

With regards to the assumption 2, two aspects should be considered. Firstly, the passenger's path choice is another interesting research orientation, which provides the demand forecasting in peak hours. Secondly, the probabilities of which train the passenger will pick up should be based on a predetermined route. Thus, for simplicity, it is exceedingly necessary to model passenger route according to the BFS. For assumption 3, the purpose is to split the passenger's transfer time into three parts, including passenger's egress walking time, passenger's access walking time, and passenger's waiting time, to unify each decomposed route. Therefore, the probabilities of passengers matching different trains on each decomposed route would be determined according to the proposed model and algorithm. Regarding assumption 4, this assumption has been proven by some recent studies [22] (Zhang et al. 2015). For assumption 5, undoubtedly, the more alternative trains are set, the more accurate the result would be. Given that computation efficiency is related to the number of alternative trains, it should be noted that the number of alternative trains should not be too large. For assumption 6, the influence of the train's delay on passengers is not considered. For assump-

tion 8, the headway satisfies the minimum headway request and it means that the number of trains operating within a timetable is also sufficient.

4.2 Notations

In this study, the following notations are defined to formulate the two-stage model.

Sets/Matrix:

- J – Set of running trains;
- U – Set of stations, and one practical station is separated into two bi-direction virtual stations, i.e., the number of all station is set as $2N$;
- $\eta^{u,v}(t)$ – Trip demand matrix, i.e., the number of passengers who depart at station u at a time t heading to station v ;
- P_u – The sum of passengers who depart at the station u during the study period $[t_0, t_0+T]$;

Parameters:

- t_0 – Start time of the research period;
- T – Lasting time of the research period, i.e., the study period is $[t_0, t_0+T]$;
- $AT_{j,u}$ – The number of passengers who match train j at the station u according to the corresponding probability;
- $BT_{j,v}$ – The number of passengers who alight from train j at the station v using corresponding probability;
- $x^{i,u,v}(t)$ – 0-1 binary parameter, i.e., if passenger i departs from the station u and his/her destination station is at the station v , the value is one, and vice versa;
- $t_{i,u}^{in}$ – Passage time of passenger i at an entry gate of the station u ;
- $t_{i,v}^{out}$ – Passage time of passenger i at an exit gate of the station v ;
- $\tau_{i,u}^{in}$ – Access walking time of passenger i at the station u ;
- $\tau_{i,v}^{out}$ – Egress walking time of passenger i at the station v ;
- $t_{i,u}^a$ – The arrival time of passenger i on the platform at the station u ;
- $TJ_{j,u}$ – The arrival time of train j at the station u ;
- $TF_{j,u}$ – The departure time of train j from the station u ;
- \overline{m}_i – The latest alternative train for passenger i to match;
- \underline{m}_i – The earliest alternative train for passenger i to match;
- $TR_{j,u}$ – Running time of train j in the traverse between the station u and the station $u+1$;

- $C_{j,u}$ – The number of passengers on the train j when departing from the station u ;
 - C – Prescribed passenger loading capacity of a train;
 - ξ – Allowed full-loaded coefficient;
 - T_G – The required time of a train during a u-turning operation at the terminal;
 - h_{\min} – The minimum headway between two consecutive trains at the station;
 - $TD_{j,u}^u$ – The maximum dwelling time of train j at the station u ;
 - $D_{j,u}^l$ – The minimum dwelling time of train j at the station u ;
 - $W_{i,u}$ – Waiting time of passenger i at the station u ;
 - W – Passengers' total waiting time;
 - $L_i^u(\mathbf{Z}_i^u)$ – The probability of passenger i tapping-out during his/her journey time at the departure station u ;
 - $G(P)$ – The likelihood function of the probabilities of passengers matching different trains;
- Decision variables:
- $TD_{j,u}$ – Dwelling time of train j at the station u ;
 - h – Headway between two consecutive trains at the same station;
 - $P_n^{i,u}$ – Probability of passenger i matching the n th train at the station u , i.e., the maximum of n is set as four, $\sum_{n=1}^4 P_n^{i,u} = 1$, and if n is one, it means that the probability of passenger i matching the first train at the station u is $P_1^{i,u}$.

4.3 Passenger-and-trains matching problem

The probability of passengers matching different trains utilised in stage two should be calculated in stage one. In our research, a complete passenger journey time contains four parts, including passenger's access walking time, passenger's waiting time, in-vehicle consuming time, and passenger's egress walking time. In particular, whether the passenger picks up the train j or not is based on two conditions as follows:

- 1) The arrival time of a passenger on the platform is derived by the passage time of a passenger at the entry gate from the AFC system plus the passenger's access walking time. Moreover, the obtained arrival time of a passenger on the platform should not exceed the departure time of the earliest train j at the departure station.

- 2) If the passenger's egress walking time is subtracted from the passage time of passenger at the exit gate from the AFC system, the result should be larger than the arrival time of the train j at the destination station.

If and only if the above two conditions are satisfied, the probability of passenger matching the train j would be obtained. In other words, the process of calculating the probability of the passenger matching the train j should typically involve three independent parts, namely, the probability of $t_{i,u}^a$ ahead of the departure time of the earliest train j , the probability of passenger matching the train j , and the probability of passenger exiting from the destination station successfully by train j .

(i) Arrival time of passenger i on the platform should not exceed the departure time of the earliest train \underline{m}_i at his/her departure station:

$$\tau_{i,u}^{in} \cdot TN(\bar{\mu}, \bar{\sigma}^2, a, b) \tag{1}$$

$$\tau_{i,v}^{out} \cdot TN(\bar{\mu}, \bar{\sigma}^2, a, b) \tag{2}$$

$$t_{i,u}^a = t_{i,u}^{in} + \tau_{i,u}^{in} \tag{3}$$

$$\psi(t|\bar{\mu}, \bar{\sigma}, a, b) = \begin{cases} 0 & \text{if } t \leq a \\ \frac{\phi(\bar{\mu}, \bar{\sigma}, t)}{\Phi(\bar{\mu}, \bar{\sigma}, b) - \Phi(\bar{\mu}, \bar{\sigma}, a)} & \text{if } a < t < b \\ 0 & \text{if } t \geq b \end{cases} \tag{4}$$

$$\begin{aligned} &P(TF_{\underline{m}_i-1,u} \leq t_{i,u}^a \leq TF_{\underline{m}_i,u}) \\ &= \int_{TF_{\underline{m}_i-1,u}}^{TF_{\underline{m}_i,u}} \psi(t|\bar{\mu}, \bar{\sigma}, a + t_{i,u}^{in}, b + t_{i,u}^{in}) dt \end{aligned} \tag{5}$$

In our research, passenger's access walking time and egress walking time follow the truncated normal distribution (Equation 4), where $\bar{\mu}$ and $\bar{\sigma}$ are respectively the mean and the standard deviation of the normal distribution. a and b are respectively the lower bound and the upper bound of the truncated normal distribution. $\psi()$ is the probability density function (PDF) of the truncated normal distribution. $\phi()$ is the PDF of the normal distribution. $\Phi()$ is the cumulative density function (CDF) of the normal distribution.

(ii) Probability of passenger i matching the train:

$$\begin{aligned} &P(\text{board train } k | TF_{\underline{m}_i-1,u} \leq t_{i,u}^a \leq TF_{\underline{m}_i,u}) \\ &= P_{k-\underline{m}_i+1}^{i,u}(\underline{m}_i \leq k \leq \bar{m}_i) \end{aligned} \tag{6}$$

where the value of $P_{k-\underline{m}_i+1}^{i,u}$ is unknown, and its posterior distribution would be determined in our research.

(iii) Probability of passenger i exiting the destination station successfully by train k :

$$P(TJ_{k,v} + \tau_{i,v}^{out} \leq t_{i,v}^{out} | \text{board train } k) = \int_{TJ_{k,v}}^{\tau_{i,v}^{out}} \psi(t | \bar{\mu}, \bar{\sigma}, a + TJ_{k,v}, b + TJ_{k,v}) dt \quad (7)$$

(iv) Probability of passenger i tapping-out during his/her journey time at the departure station u :

$$P(TJ_{k,v} + \tau_{i,v}^{out} \leq t_{i,v}^{out}, \text{board train } k, TF_{m,-1,u} \leq t_{i,u}^a \leq TF_{m,u}) = P(TF_{m,-1,u} \leq t_{i,u}^a \leq TF_{m,u}) \cdot P(\text{board train } k | TF_{m,-1,u} \leq t_{i,u}^a \leq TF_{m,u}) \cdot P(TJ_{k,v} + \tau_{i,v}^{out} \leq t_{i,v}^{out} | \text{board train } k) = \int_{TF_{m,-1,u}}^{TF_{m,u}} \psi(t | \bar{\mu}, \bar{\sigma}, a + t_{i,u}^a, b + t_{i,u}^a) dt \cdot P_{k-m_i+1}^{i,u} \int_{TJ_{k,v}}^{\tau_{i,v}^{out}} \psi(t | \bar{\mu}, \bar{\sigma}, a + TJ_{k,v}, b + TJ_{k,v}) dt \quad (8)$$

$$L_i^u(\mathbf{Z}_i^u) = \sum_{k=m_i}^{\bar{m}_i} P(TJ_{k,v} + \tau_{i,v}^{out} \leq t_{i,v}^{out}, \text{board train } k, TF_{m,-1,u} \leq t_{i,u}^a \leq TF_{m,u}) = \int_{TF_{m,-1,u}}^{TF_{m,u}} \psi(t | \bar{\mu}, \bar{\sigma}, a + t_{i,u}^a, b + t_{i,u}^a) dt \cdot \sum_{k=m_i}^{\bar{m}_i} P_{k-m_i+1}^{i,u} \int_{TJ_{k,v}}^{\tau_{i,v}^{out}} \psi(t | \bar{\mu}, \bar{\sigma}, a + TJ_{k,v}, b + TJ_{k,v}) dt \quad (9)$$

Due to the independence of (i), (ii), and (iii), this probability (Equation 8) is derived by Equations 5–7. In Equation 9, it should be noted that $\mathbf{Z}_i^u = [P_1^{i,u}, P_2^{i,u}, P_3^{i,u}, P_4^{i,u}]^T$ is the vector that contains the probabilities of passenger i matching train at the departure station u .

(v) The likelihood function of the probabilities of passengers matching different trains:

$$G(\mathbf{Z}_i^u | L_i^u(\mathbf{Z}_i^u)) = \frac{G(\mathbf{Z}_i^u, L_i^u(\mathbf{Z}_i^u))}{G(L_i^u(\mathbf{Z}_i^u))} = \frac{G(L_i^u(\mathbf{Z}_i^u) | \mathbf{Z}_i^u) G(\mathbf{Z}_i^u)}{G(L_i^u(\mathbf{Z}_i^u))} \quad (10)$$

$$G(\mathbf{Z}_i^u | L_i^u(\mathbf{Z}_i^u)) \propto G(L_i^u(\mathbf{Z}_i^u) | \mathbf{Z}_i^u) G(\mathbf{Z}_i^u) \propto G(L_i^u(\mathbf{Z}_i^u) | \mathbf{Z}_i^u) \quad (11)$$

Due to the unknown PDF expression of the probability of passenger matching different trains, the Bayesian estimation is proposed to deal with this issue. According to the Bayesian framework, the prior distribution of \mathbf{Z}_i^u is assumed, and after training with the real AFC data, the posterior distribution of \mathbf{Z}_i^u would be learned. Since almost no further information would be known about the probabilities in advance, then in our research, it is assumed that the prior distribution of \mathbf{Z}_i^u follows the broad distribution, such as the uniform distribution. Therefore, the value of $G(\mathbf{Z}_i^u)$ would be equal to one. Then, Expression 11 could be obtained.

$$\max G(P) = \sum_{u=1}^{2N} \prod_{t=70}^{t_0+T} \prod_{v=u+1}^{2N} \prod_{i=1}^{\eta^{m(v)}} G(\mathbf{Z}_i^u | L_i^u(\mathbf{Z}_i^u)) = \sum_{u=1}^{2N} \prod_{t=70}^{t_0+T} \prod_{v=u+1}^{2N} \prod_{i=1}^{\eta^{m(v)}} \int_{TF_{m,-1,u}}^{TF_{m,u}} \psi(t | \bar{\mu}, \bar{\sigma}, a + t_{i,u}^a, b + t_{i,u}^a) dt \sum_{k=m_i}^{\bar{m}_i} P_{k-m_i+1}^{i,u} \int_{TJ_{k,v}}^{\tau_{i,v}^{out}} \psi(t | \bar{\mu}, \bar{\sigma}, a + TJ_{k,v}, b + TJ_{k,v}) dt \quad (12)$$

Herein, to normalise Expression 11, a constant value G_{norm} is utilised in Equation 12. Undoubtedly, it is conceivable that maximising the likelihood of Function 12 would achieve the replication of reality in the metro system. Then, the probabilities of each passenger matching different trains would be determined by the following algorithm.

4.4 Timetable optimisation

In stage two, the aim is to reduce passengers' total waiting time by adjusting headway and dwelling time at each station. Firstly, the probabilities of passengers matching different trains should be added to the objective. Secondly, microcosmic and macroscopic constraints should both be taken into accounts in our model. For the microcosmic constraint, it indicates how many passengers would be loaded without exceeding the upper bound of the train capacity. To better delineate passenger-and-trains matching, the unit of time is in seconds. For the macroscopic constraint, it illustrates

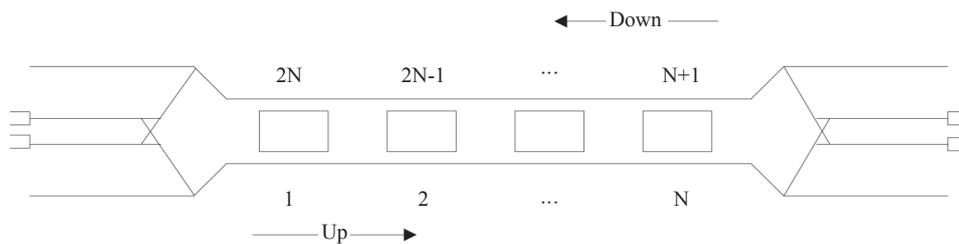


Figure 5 – A simple representation of the urban railway transit line

that train movements are strictly followed by the scheduled timetable, and correspondingly, the stop-pattern is not considered in our research.

Microcosmic constraints from the passenger's perspective

$$AT_{j,u} = \sum_{t=t_0}^{t_0+T} \sum_{v=u+1}^{2N} \sum_{i=1}^{\eta^{u,v}(t)} \sum_{j=\underline{m}_i}^{\bar{m}_i} P_{j-\underline{m}_i+1}^{i,u} \quad (13)$$

$$BT_{j,v} = \sum_{t=t_0}^{t_0+T} \sum_{u=1}^{v-1} \sum_{i=1}^{\eta^{u,v}(t)} \sum_{j=\underline{m}_i}^{\bar{m}_i} P_{j-\underline{m}_i+1}^{i,u} x^{i,u,v}(t) \quad (14)$$

$$C_{j,u} = C_{j,u-1} + AT_{j,u} - BT_{j,u} \quad (15)$$

$$C_{j,u} \leq C\xi \quad (16)$$

Passengers matching different trains are no longer deterministic. In addition, Equation 13 gives the formulation that the number of passengers who match the train j should be explicitly cumulated by the probability of passengers matching the train j . In Equation 14, 0-1 binary variable $x^{i,u,v}(t)$ is determined by the AFC data record. To summarise, the number of passengers on the train j can be calculated as Equation 15. In general, Constraint 16 ensures a finite capacity of a running train to agree with the practical experience.

Macroscopic constraints from an operation perspective

$$TJ_{j,u} = TF_{j,u-1} + TR_{j,u-1} \quad (17)$$

$$TF_{j,u} = TJ_{j,u} + TD_{j,u} \quad (18)$$

$$TF_{j,u} - TF_{j-1,u} = h \geq h_{\min} \quad (19)$$

$$TJ_{j,N+1} - TJ_{j,N} = T_G \quad (20)$$

$$TJ_{j,1} - TJ_{j,2N} = T_G \quad (21)$$

$$TD_{j,u}^l \leq TD_{j,u} \leq TD_{j,u}^u \quad (22)$$

The stop-skip pattern is not considered in our research. Thus, any two consecutive trains should satisfy Equations 17–21. Specifically, for Equation 20, it ensures a u-turning operation at the end terminal station; however, for Equation 21, it ensures the u-turning operation at the start terminal station. Generally, dwelling time at each station is the decision variable, and Constraint 22 gives a boundary condition.

The objective of the stage-two model

$$W_{i,u} = \sum_{j=\underline{m}_i}^{\bar{m}_i} (TJ_{j,u} - t_{i,u}^a) P_{j-\underline{m}_i+1}^{i,u} \quad (23)$$

$$\min W = \sum_{t=t_0}^{t_0+T} \sum_{v=u+1}^{2N} \sum_{i=1}^{\eta^{u,v}(t)} \sum_{j=\underline{m}_i}^{\bar{m}_i} (TJ_{j,u} - t_{i,u}^a) P_{j-\underline{m}_i+1}^{i,u} \quad (24)$$

With the intent of reducing passengers' total waiting time, the objective in stage two is designed to minimise passengers' total waiting time. Due to the indeterminate passengers train matching, the probabilities of passengers matching different trains should be added in Objective 24, and the dwelling time of each station and headway are the decision variables.

4.5 Solution algorithm by implementing MCMC-GASA

In stage two, timetable optimisation considering capacity constraint falls in the category of the NP-hard class [23] (Niu et al. 2013). In other words, this kind of problem is difficult to be solved by commercial optimisation solvers within a reasonable amount of time. Thus, motivated by the artificial intelligent technique, a genetic algorithm (GA) combining simulated annealing (SA) is implemented to optimise the timetable in our research. In general, the advantage of the GA [24] (Yin et al. 2019) lies in its ability to solve the intractable problem efficiently and provide good extensivity. With respect to SA [25] (Kang et al. 2016), it turns out that it does well in searching for the optimal solution. Thus, GASA is utilised to find the suboptimal solution for the stage-two model.

Regarding stage one, the challenge is to determine the posterior distribution to obtain the probabilities of passengers matching different trains. Theoretically, Markov Chain Monte Carlo (MCMC) [26] (Xu et al. 2018) has been widely used in addressing this kind of problem. The step of MCMC is to generate the candidate samples according to prior distribution and to generate the next candidate samples by the random walk process, then to accept the new candidate samples or to reject them based on the selection rules.

Therefore, to solve stage two successfully and efficiently, MCMC-GASA is implemented to deal with this model.

MCMC utilised for stage one

For MCMC, the Metropolis-Hastings (MH) technique is typically utilised for sampling. In general, in a recent study [27] (Lee et al. 2015), certain advantages of the MH sampling from the aspects of its efficiency and its feasibility have been

shown. With regard to the efficiency, unlike the basic rejection rule, the MH algorithm would improve the efficiency of the convergence. Concerning its feasibility, the constant value G_{norm} utilised in normalisation (Equation 24) is not required to reduce the complexity of the calculation. Thus, it is suitable to use the MCMC method to address this issue in stage one. The detailed steps of MCMC are shown as follows:

- Step 1: (Initialisation) Input timetable parameters, AFC data, and maximum iteration number M , for timetable parameters, including $h, TJ_{j,u}, TF_{j,u}$, for AFC data, including $t_{i,u}^{in}, t_{i,v}^{out}$,
- Step 2: (MH sampling) At the iteration t , generate the new candidate probabilities of passengers matching different trains $Z^* = (Z_1^1 \dots Z_1^{2N}, Z_2^{2N} \dots Z_{P_{2N}}^{2N})$, according to the random walk process;
- Step 3.1: (Metropolis acceptance rule) Calculate the acceptance probability of $Z^* : \alpha(Z^* | Z^{t-1}) = \min \left\{ 1, \frac{G(P^*)}{G(P^{t-1})} \right\}$;
- Step 3.2: Generate a random number u_0 from the uniform (0,1) distribution;
- Step 3.3: Determine whether Z^* should be accepted or not; if $u_0 < \alpha(Z^* | Z^{t-1})$, then $Z^t = Z^*$; otherwise $Z^t = Z^{t-1}$;
- Step 4: (Terminate or not) If $t < M$, then $t = t + 1$, and return to Step 2; otherwise, go to Step 5;
- Step 5: Obtain the probabilities of passengers matching different trains.

GASA incorporating MCMC for stage two

In general, the results obtained by the MCMC for stage one are added to the objective in stage two. Then, the main idea of timetable optimisation for stage two is to adjust the headway and train dwelling time at each station by the GASA algorithm. Therefore, it is evident that any iterations of the GASA algorithm for stage two should involve a complete calculating process of the MCMC for stage one. For descriptive convenience, it is essential to specify some core techniques in GASA.

In the framework of GASA for solving stage two, the *chromosome*, *crossover* operation, *mutation* operation, and *Metropolis rule* for iteration should be emphasised. Firstly, a *chromosome* involves the decision variables, which are called “genes.” Specifically, the following vector forms the genes of a chromosome:

$$(TD_{j,1}, TD_{j,2}, TD_{j,3}, TD_{j,4}, \dots, TD_{j,N}, \dots, TD_{j,N+1}, \dots, TD_{j,N+2}, \dots, TD_{j,2N}, h, j \in J)$$

Secondly, for the *crossover* operation, the replacing method is employed. Herein, the crossover operation indicates that a gene value of the chromosome is replaced by the gene value in the same position of another chromosome.

Thirdly, to coordinate the crossover operation, the *mutation* operation is adopted to enlarge the range of feasible solutions. Particularly, gene values in the chromosome are replaced by a random number to meet the boundary *Constraints 19 and 22*.

With respect to the *Metropolis rule* for iteration, this is the core technique for the SA. In that sense, for any iteration, there are two objective values, namely, the new objective value W_{new} and the old objective value W_{old} . Accordingly, it should be noted which objective value should be selected in this iteration. Thus, if the difference between W_{new} and W_{old} is less than zero, then the new objective value W_{new} is accepted with the probability $\rho = 1$. However, if W_{new} is larger than W_{old} , then the new objective value would be accepted with a small probability $\rho = \exp(-(W_{new} - W_{old})/t_c)$, where t_c is the current annealing temperature.

To summarise, the detailed steps of the MC-MC-GASA algorithm are described as follows:

- Step 1: Initialisation
 - 1.1. Set the initial parameters: initial temperature T_{ini} , lowest temperature T_{end} , cooling efficiency λ , population N_{pop} , generation k .
 - 1.2. According to the timetable at the current stage, extract the genes, and generate an initial parent chromosome.
 - 1.3. Copy the parent chromosome in 1.2, perform the *crossover* operation to these two parent chromosomes, and generate two child chromosomes.
 - 1.4. Repeat 1.3 until the population reaches N_{pop} , mark the initial generation as $k=1$, calculate the probabilities of passengers matching different trains according to *Figure 6* for each chromosome, obtain passengers’ total waiting time W_i for each chromosome i .
- Step 2: Creating new chromosomes by *crossover* operation and selecting by *Metropolis rule*.
 - 2.1. If current temperature $t_c \leq T_{end}$, then go to Step 5; otherwise, go to 2.2.
 - 2.2. For any chromosome i , conduct the *crossover* operation to obtain a new chromosome j , calculate the probabilities of passengers matching different trains according to *Figure 6* for chromosome i and j , obtain passengers’ total waiting time W_i and W_j respectively.

2.3. Calculate the difference between W_i and W_j , then, according to the *Metropolis rule*, determine whether the new chromosome is accepted or not.

2.4. Generate a new population $P(k)$ according to 2.2 and 2.3, then go to Step 3.

Step 3: Conduct the *mutation* operation to the population

3.1. Perform the *mutation* operation for 30% chromosomes in the population N_{pop} .

3.2. Replace the old genes with new genes by random number strictly following the boundary *Constraints 19 and 22*.

3.3. Update the new population $P(k)$ according to 3.1 and 3.2, then set $k=k+1$, then go to Step 4.

Step 4: Terminate or not

4.1. If current temperature $t_c \leq T_{end}$, then go to Step 5; otherwise set the current temperature as $t_c = t_c \cdot \lambda$, then return to Step 2.

Step 5: Obtain the optimal solution

5.1. Select the best chromosome from the population as the output result.

5. CASE STUDY OF LINE 1 IN BEIJING METRO

In this section, the case study of Line 1 in Beijing Metro is selected to demonstrate the proposed models and algorithms during the morning peak hour (7:00–8:00) on a weekday. The snapshot of Line 1 is shown in *Figure 6*.

Specifically, Line 1 consists of ten transfer stations. To classify passengers into different types (*Figure 2*), it is necessary to determine the route for each passenger by using the BFS method. Some results are presented in *Table 1*.

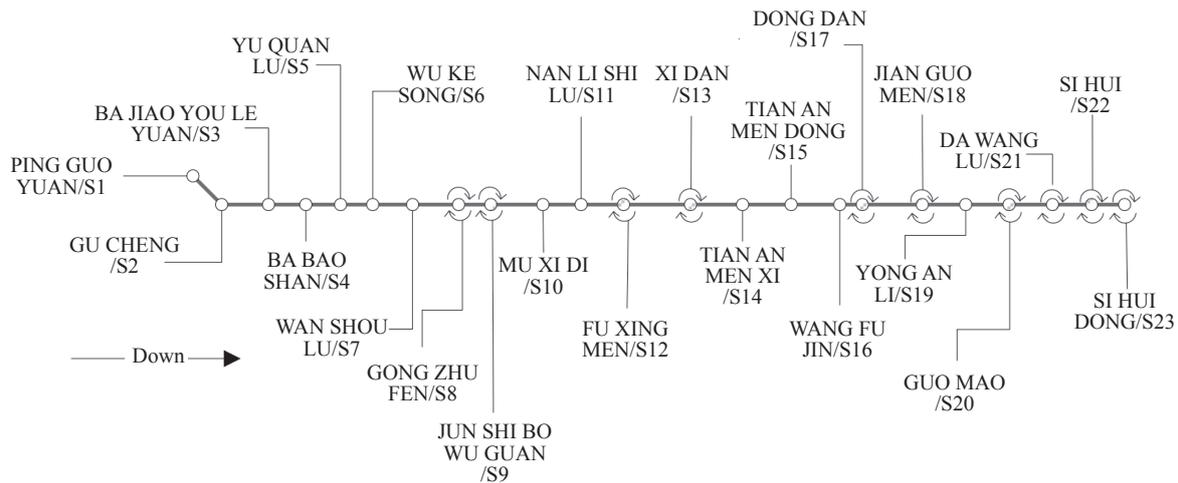


Figure 6 – The snapshot of Line 1 in Beijing Metro

Table 1 – Some results obtained by BFS

AFC ID number	Departure station	Arrival station	Transfer station	Passenger type
1251**21	PING GUO YUAN on Line 1	TAI YANG GONG on Line 10	Out at JUN SHI BO WU GUAN	Type 1
6137**13	PING GUO YUAN on Line 1	SAN YUAN QIAO on Line 10	Out at JUN SHI BO WU GUAN	Type 1
1045**20	PING GUO YUAN on Line 1	SU ZHOU JIE on Line 10	Out at GONG ZHU FEN	Type 1
0405**53	FU CHENG MEN on Line 2	XI DAN on Line 1	Into at FU XING MEN	Type 2
1331**46	DENG SHI KOU on Line 5	SHUANG JING on Line 10	Into at DONG DAN, out at GUO MAO	Type 1 and Type 2
...				

5.1 Results analysis

The experiments were conducted on a personal computer with an Intel Core i5-6200U and 12GB RAM. With respect to GASA, the initial temperature, lowest temperature, cooling efficiency, population, and generation are 100, 0, 0.9, and 55, respectively. Regarding the basic parameters of the train, the number of the vehicle seats is set as 240, so the prescribed train capacity is 1440 persons/train, then the allowed full-loaded coefficient of a train is 150%. Furthermore, the minimum headway is 120 seconds. For MCMC, the maximum iteration is 1000. In addition, the lower bound and the upper bound of the truncated normal distribution, the mean and the standard deviation of the normal distribution are set to 0.3min, 2min, 1.5, and 1, respectively.

AFC data indicate a passenger’s unlink trip, which means that the passenger’s route is unknown. In China, AFC data provides information including the passenger’s ID number, departure station, passage time at an entry gate, destination station, passage time at an exit gate, line number of the departure station, and line number of the destination station. Furthermore, based on the AFC data and the current timetable parameters, the probabilities of passengers matching different trains would be calculated by using the MCMC technique.

Before timetable optimisation, a simple example which includes 50 passengers is shown in *Figure 7*. Obviously, the area of the first train is the largest, which indicates that most passengers would match the first train successfully. It turns out that during the morning peak hours, most passengers would not be willing to wait in case of being late for work or school. However, due to the finite capacity of a running train, a phenomenon in which passengers are left behind typically occurs. Thus, the probabilities of passengers matching the second, third, or fourth train would also be inferred. From *Figure 7*, the probability of passengers matching the second train would be typically larger than the counterpart matching the third train. Correspondingly, the probability of passengers matching the third train would also typically be larger than that matching the fourth train. Essentially, these results are coherent with passengers’ daily habits. Furthermore, a larger experiment during the period 7:00–8:00 which includes 40000 passengers was also conducted. The average probability of passengers matching the first

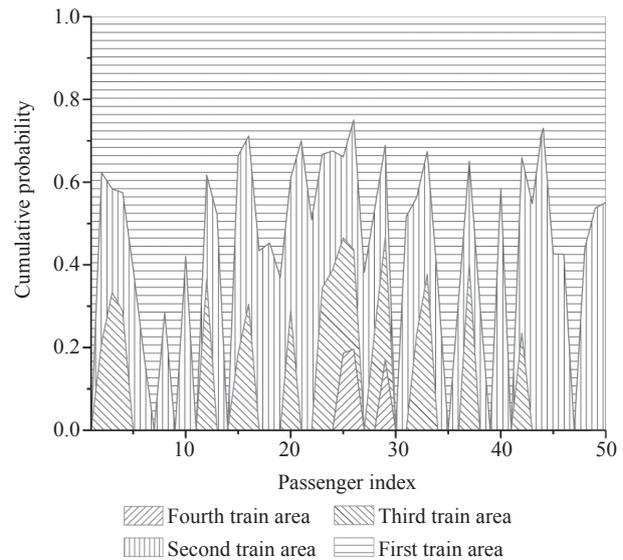


Figure 7 – Results obtained by MCMC

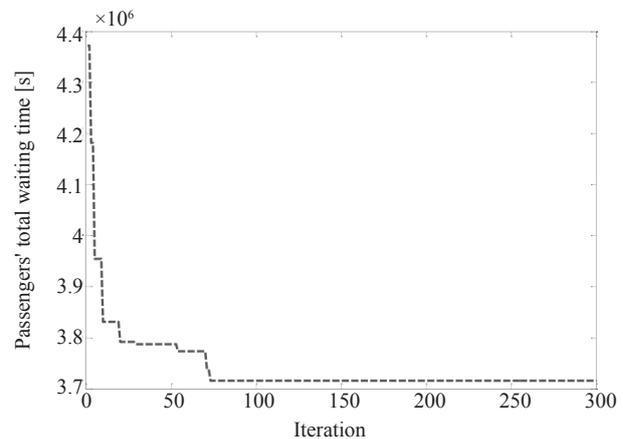


Figure 8 – The optimisation process of MCMC-GASA

train, the second train, the third train, and the fourth train is 0.63, 0.24, 0.09, and 0.04, respectively. The results correspond strictly to those in *Figure 7*.

After timetable optimisation, in *Figure 8*, passengers’ total waiting time decreased by 15.5%, more precisely, from 1218 hours (4387651 seconds) to 1029 hours (3704584 seconds). Secondly, from the perspective of the probability of passengers matching the first train, a passenger whose AFC ID is 6185**19 is taken as an example to demonstrate the efficiency of the new timetable. This passenger entered the station at 8:46:07, and exited at 8:56:09. In *Figure 9*, after calculations by using the MCMC, the train number which matches the passenger’s journey time is two, although the maximum alternative trains are four. Then, in *Figure 9*, the probability of this passenger matching the first train increased by 5%, specifically from 0.5 to 0.525. Thirdly, the

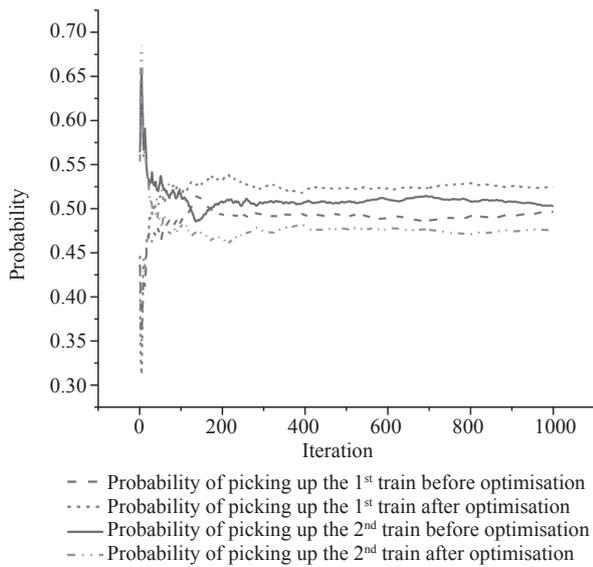


Figure 9 – The process of MCMC for a passenger’s matching different trains before and after timetable optimisation

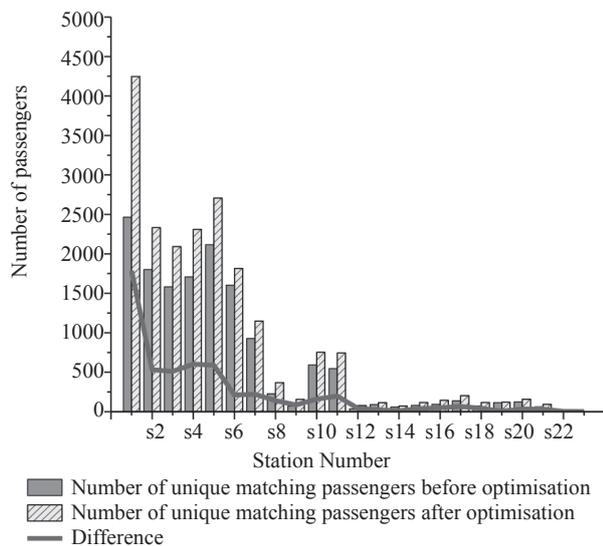


Figure 10 – The number of unique matching passengers before and after timetable optimisation

passenger whose probability of matching the first train is equal to one is defined as the unique matching passenger. From Figure 10, the number of unique matching passengers after timetable optimisation increased remarkably compared to before. To sum up, the reason why passengers benefit a lot from the new timetable is that the objective of the new timetable aims to reduce passengers’ total waiting time. Furthermore, passengers’ total waiting time is based on passengers matching different trains. In other words, shorter passengers’ total waiting time corresponds with higher probability of passengers matching the first train. Another interesting phenomenon is that the number of unique matching passengers at

non-transfer stations is larger than those at transfer stations. Indeed, this phenomenon corresponds with a recent study [28] (Zhang et al. 2018).

Based on the analysis, it can be concluded that the new timetable is more sufficient than the alternatives from the passenger’s perspective. More precisely, several improvements are listed in Table 2. Some timetable parameters before and after optimisation are provided in Table 3.

5.2 Sensitivity analysis

Given that different allowed full-loaded coefficients of the running train would affect the objective value of the stage-two model, several additional numerical simulations were conducted on the MATLAB platform. In Figure 11, passengers’ total waiting time would be shorter when the allowed full-loaded coefficient is higher. In particular, the passengers’

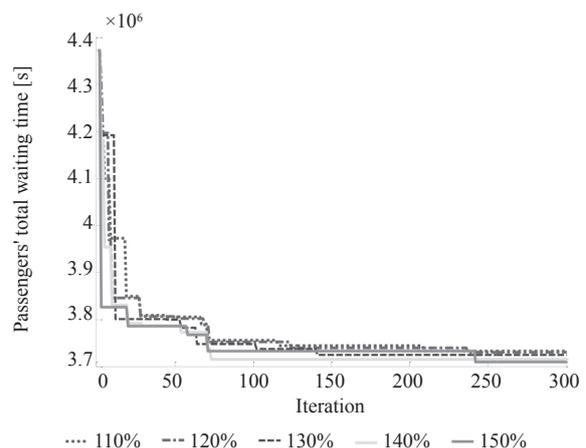


Figure 11 – The comparison of optimisation results under different full-loaded coefficients of a train

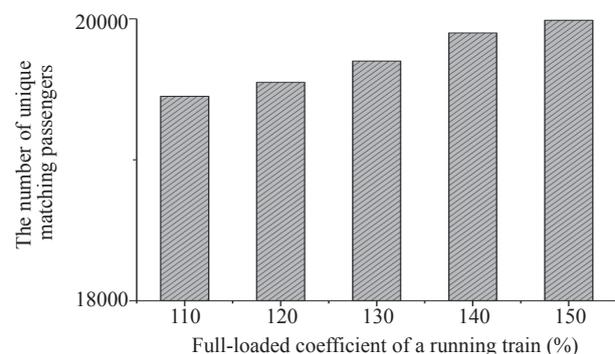


Figure 12 – The number of unique matching passengers under different full-loaded coefficients of a train

total waiting time is associated with the probability of passengers matching different trains. Figure 12

Table 2 – The comparisons between the real timetable and the new timetable

Timetable	Real timetable	The best found by MCMC-GASA
Passengers' total waiting time	1218 hours (0%)	1029 hours (-15.5%)
Unique matching passengers	14514 (0%)	19990 (+37.7%)
Headway	180 seconds	121 seconds
CPU	-	317 minutes

Table 3 – Timetable parameters [s]

Station	Original dwelling time	Optimised dwelling time (Up)	Optimised dwelling time (Down)	Running time between two stations
S1	30	26	25	-
S2	30	24	27	220
S3	30	39	34	150
S4	30	34	32	150
S5	40	46	43	120
S6	40	43	41	140
S7	45	48	45	130
S8	45	45	44	100
S9	40	34	36	100
S10	30	29	30	100
S11	30	31	32	120
S12	50	56	55	80
S13	30	24	26	120
S14	28	28	28	90
S15	30	35	34	90
S16	30	27	27	90
S17	45	45	45	80
S18	45	49	50	110
S19	30	24	26	120
S20	45	47	50	70
S21	30	20	23	120
S22	30	30	29	150
S23	30	31	31	150

indicates how the full-loaded coefficient affects the number of unique matching passengers. When the full-loaded coefficient of a running train increases, the number of unique matching passengers increases as well. For demonstrative purposes, the following equation is utilised to explain the mechanism underlying the phenomena in which passengers' total waiting time decreases when the allowed full-loaded coefficient increases.

Case 1 (unique matching passenger):

$$W_{i,u}^1 = TF_{j,u} - t_{i,u}^a$$

Case 2 (regular matching passenger):

$$\begin{aligned} W_{i,u}^2 &= (TF_{j,u} - t_{i,u}^a)P_1^{i,u} + (TD_{j+1,u} + h + TF_{j,u} - t_{i,u}^a)P_2^{i,u} \\ &+ (TD_{j+2,u} + TD_{j+1,u} + 2h + TF_{j,u} - t_{i,u}^a)P_3^{i,u} \\ &+ (TD_{j+3,u} + TD_{j+2,u} + TD_{j+1,u} + 3h + TF_{j,u} - t_{i,u}^a)P_4^{i,u} \\ &= (TF_{j,u} - t_{i,u}^a) + (TD_{j+1,u} + h)(1 - P_1^{i,u}) \\ &+ (TD_{j+2,u} + h)(1 - P_1^{i,u} - P_2^{i,u}) + (TD_{j+3,u} + h)P_4^{i,u} \end{aligned}$$

Undoubtedly, $W_{i,u}^2$ in case 2 is larger than $W_{i,u}^1$ in case 1. It is indicated that the waiting time of a unique matching passenger is shorter than that of a regular matching passenger. Hence, with a higher full-loaded

coefficient of a running train, the number of unique matching passengers would increase simultaneously. Thus, passengers' total waiting time would be shorter as well.

6. CONCLUSION

To summarise, the mathematical model is formulated to characterise each passenger matching different trains according to the AFC data record. Then, passengers' total waiting time based on each passenger matching different trains is delivered as the objective of the timetable optimisation. Accordingly, two elaborately designed algorithms are proposed to estimate each passenger matching different trains and to optimise timetable, respectively. Moreover, a case study of Line 1 in Beijing metro is utilised to verify the proposed models and algorithms. The results show that the new timetable considering passengers matching different trains would be better and more efficient than the alternative timetable. Particularly, passengers' total waiting time decreased from 1218 hours to 1029 hours, and the number of unique matching passengers increased from 14514 persons to 19900 persons.

Specifically, the train loading and timetable parameters are utilised to build the model of passenger-and-trains matching, and the timetable is optimised according to passenger's matching different trains. However, as mentioned in assumption 2, the passenger's route is obtained by the BFS. In future research, random nature of decisions on the choice of route by passengers will be considered. Timetable parameters on different lines are different, and it indicates that the passengers' different routes would affect the probability of passengers matching different trains. Therefore, the probabilities of route choice for a passenger should be considered for affecting passenger's matching different trains in the future.

As for assumption 4, the advantage of the truncated normal distribution lies in successfully characterising the phenomenon in which the passenger's walking time is not too long nor too short. However, the gender and the age of a passenger would affect the passenger's walking time in the station as well. Thus, passenger's walking time obtained by the truncated normal distribution would not be accurate enough. With this concern, signalling data of the mobile phone will be utilised to verify a passenger's walking time in the future.

The proposed MCMC-GASA is not efficient enough to find the optimal solution. Hence, a more efficient algorithm should be designed to optimise the timetable in the future. Furthermore, the application of MCMC-GASA is limited by the minimum input dataset, and the results accuracy by MCMC-GASA is also affected by the results obtained by the BFS. Therefore, a general integrated algorithm combining the BFS and MCMC-GASA should be constructed in the future.

Essentially, passengers matching different trains in the intermediate station of a complete route would be categorised as the transfer train matching behaviour. Thus, the proposed models would also be applicable for network optimisation. In future work, the external factors and unpredicted delays by network will also be included, and our models will be extended to a network case.

ACKNOWLEDGEMENT

This work was supported by the Fundamental Research Funds for the Central Universities 2020YJS083.

黄俊生, 博士研究生¹ (通讯作者)

电子邮箱: 19114037@bjtu.edu.cn

张桐, 博士²

电子邮箱: 16114195@bjtu.edu.cn

魏润斌, 博士研究生¹

电子邮箱: 19114056@bjtu.edu.cn

¹ 北京交通大学综合交通运输大数据应用技术交通运输行业重点实验室,

中国北京市海淀区上园村3号, 100044

² 交通运输部水运科学研究院物流中心,

中国北京市海淀区西土城路8号, 100088

基于乘客与地铁车次匹配的城市轨道交通时刻表优化: 以北京地铁线为例

摘要

由于高峰期地铁车站内的拥挤现象十分常见, 因此站台上的乘客容易发生滞留。本文首先提出一个可以获取乘客匹配车次的方法, 其次, 为了降低乘客总等待时间, 本文提出基于乘客匹配车次的两阶段时刻表优化模型。第一阶段模型的目标是根据AFC数据和时刻表参数获取乘客匹配不同车次的概率, 第二阶段模型的决策变量包含列车发车间隔和列车在各站的停站时间, 目标函数是使乘客总等待时间最小。鉴于本文提出的两阶段模型的复杂性, 本文提出马尔可夫链蒙特卡洛模拟-遗传模拟退火的混合算法, 并运用北京地铁一号线实际案例予以实证。结果显示, 优化后的时刻表更具优越性, 其中匹配唯一车次的乘客人数提高了37.7%, 乘客总等待时间下降了15.5%。

关键词

城市轨道交通; 车次匹配; 时刻表
优化; AFC数据; 机器学习

REFERENCES

- [1] Mao BH. Public transport capability is an important indicator of national strength in transport. *Journal of Beijing Jiaotong University (Social Science Edition)*. 2018;17: 1-8. Chinese
- [2] Han Y, Zhang T, Wang M. Holiday travel behavior analysis and empirical study with Integrated Travel Reservation Information usage. *Transportation Research Part A: Policy Practice*. 2020;134: 130-151. DOI: 10.1016/j.tra.2020.02.005
- [3] Yang H, Tang Y. Managing rail transit peak-hour congestion with a fare-reward scheme. *Transportations Research Part B: Methodological*. 2018;110: 122-136. DOI: 10.1016/j.trb.2018.02.005
- [4] Guo X, et al. Timetable coordination of first trains in urban railway network: A case study of Beijing. *Applied Mathematical Modelling*. 2016;40: 8048-8066. DOI: 10.1016/j.apm.2016.04.004
- [5] Barrena E, Canca D, Coelho LC, Laporte G. Single-line rail rapid transit timetabling under dynamic passenger demand. *Transportation Research Part B: Methodological*. 2014;70: 134-150. DOI: 10.1016/j.trb.2014.08.013
- [6] Robenek T, et al. Train timetable design under elastic passenger demand. *Transportation Research Part B: Methodological*. 2018;111: 19-38. DOI: 10.1016/j.trb.2018.03.002
- [7] Wang YH, et al. Passenger-demands-oriented train scheduling for an urban rail transit network. *Transportation Research Part C: Emerging Technologies*. 2015;60: 1-23. DOI: 10.1016/j.trc.2015.07.012
- [8] Zhu YT, Mao BH, Bai Y, Chen SK. A bi-level model for single-line rail timetable design with consideration of demand and capacity. *Transportation Research Part C: Emerging Technologies*. 2017;85: 211-233. DOI: 10.1016/j.trc.2017.09.002
- [9] Fu L, Liu Q, Calamai P. Real-time optimization model for dynamic timetabling of transit operations. *Transportation Research Record*. 2003;1857: 48-55. DOI: 10.3141/1857-06
- [10] Jiang ZB, Hsu CH, Zhang DQ, Zou XL. Evaluating rail transit timetable using big passengers' data. *Journal of Computer and System Science*. 2016;82(1): 144-155. DOI: 10.1016/j.jcss.2015.08.004
- [11] Shi JG, Yang LX, Yang J, Gao ZY. Service-oriented train timetabling with collaborative passenger flow control on an oversaturated metro line: An integer linear optimization approach. *Transportation Research Part B: Methodological*. 2018;110: 26-59. DOI: 10.1016/j.trb.2018.02.003
- [12] Sun YS, Xu RH. Rail transit travel time reliability and estimation of passenger route choice behavior. *Transportation Research Record*. 2012;2275: 58-67. DOI: 10.3141/2275-07
- [13] Zhou F, Shi JG, Xu RH. Estimation method of path-selecting proportion for urban rail transit based on AFC data. *Mathematical Problems in Engineering*. 2015; Article ID 350397. 9 p. DOI: 10.1155/2015/350397
- [14] Kusakabe T, Iryo T, Asakura Y. Estimation method for railway passengers' train choice behavior with smart card transaction data. *Transportation*. 2010;37: 731-749. DOI: 10.1007/s11116-010-9290-0
- [15] Yang X, Chen A, Ning B, Tang T. Bi-objective programming approach for solving the metro timetable optimization problem with dwell time uncertainty. *Transportation Research Part E: Logistics and Transportation Review*. 2017;97: 22-37. DOI: 10.1016/j.tre.2016.10.012
- [16] Binder S, Maknoon Y, Bierlaire M. The multi-objective railway timetable rescheduling problem. *Transportation Research Part C: Emerging Technologies*. 2017;78: 78-94. DOI: 10.1016/j.trc.2017.02.001
- [17] Parbo J, Nielsen OA, Prato CG. User perspectives in public transport timetable optimisation. *Transportation Research Part C: Emerging Technologies*. 2014;48: 269-284. DOI: 10.1016/j.trc.2014.09.005
- [18] Sels P, Dewilde T, Cattrysse D, Vansteenwegen P. Reducing the passenger travel time in practice by the automated construction of a robust railway timetable. *Transportation Research Part B: Methodological*. 2016;84: 124-156. DOI: 10.1016/j.trb.2015.12.007
- [19] Newell GF. Dispatching policies for a transportation route. *Transportation Science*. 1971;5(1): 91-105. DOI: 10.1287/trsc.5.1.91
- [20] Sun LJ, et al. An integrated Bayesian approach for passenger flow assignment in metro networks. *Transportation Research Part C: Emerging Technologies*. 2015;52: 116-131. DOI: 10.1016/j.trc.2015.01.001
- [21] Zhu YW, Koutsopoulos HN, Wilson NHM. Inferring left behind passengers in congested metro systems from automated data. *Transportation Research Part C: Emerging Technologies*. 2018;94: 323-337. DOI: 10.1016/j.trc.2017.10.002
- [22] Zhang YS, Yao EJ. Splitting Travel Time Based on AFC Data: Estimating Walking, Waiting, Transfer, and In-Vehicle Travel Times in Metro System. *Discrete Dynamics in Nature and Society*. 2015; Article ID 539756. 11 p. DOI: 10.1155/2015/539756
- [23] Niu HM, Zhou XS. Optimizing urban rail timetable under time-dependent demand and oversaturated conditions. *Transportation Research Part C: Emerging Technologies*. 2013;36: 212-230. DOI: 10.1016/j.trc.2013.08.016
- [24] Yin HD, et al. Optimizing the release of passenger flow guidance information in urban rail transit network via agent-based simulation. *Applied Mathematical Modelling*. 2019;72: 337-355. DOI: 10.1016/j.apm.2019.02.003
- [25] Kang LJ, Zhu XN. A simulated annealing algorithm for first train transfer problem in urban railway networks. *Applied Mathematical Modelling*. 2016;40: 419-435. DOI: 10.1016/j.apm.2015.05.008
- [26] Xu XY, Xie LP, Li HY, Qin LQ. Learning the route choice behavior of subway passengers from AFC data. *Expert Systems with Applications*. 2018;95: 324-332. DOI: 10.1016/j.eswa.2017.11.043
- [27] Lee M, Soh K. Inferring the route-use patterns of

metro passengers based only on travel-time data within a Bayesian framework using a reversible-jump Markov chain Monte Carlo (MCMC) simulation. *Transportation Research Part B: Methodological*. 2015;81: 1-17. DOI: 10.1016/j.trb.2015.08.008

[28] Zhang TY, Li DW, Qiao Y. Comprehensive optimization of urban rail transit timetable by minimizing total travel times under time-dependent passenger demand and congested conditions. *Applied Mathematical Modelling*. 2018;58: 421-446. DOI: 10.1016/j.apm.2018.02.013