Machine Learning based integrated pedestrian facilities planning and staff assignment problem in transfer stations

Bisheng He [a,b], Hongxiang Zhang [a], Keyu Wen [c,d], Gongyuan Lu [a,b,1]
[a] School of Transportation and Logistics, Southwest Jiaotong University P.O. Box 610031, Chengdu, China
[b] National United Engineering Laboratory of Integrated and Intelligent Transportation P.O. Box 610031, Chengdu, China
[c] China Railway Economic and Planning Research Institute P.O. Box 100038, Beijing, China
[d] School of Economics and Management, Southwest Jiaotong University P.O. Box 610031, Chengdu, China
[1] lugongyuan@swjtu.edu.cn, Phone: +86 (0) 138 8060 9100

Abstract
Optimizing the pedestrian facilities plan in transfer stations is the problem of adjusting the facilities on the layout of pedestrian flow route and the number of machines in service to service to meet the level of services requirements. In the practice, the operation of pedestrian facilities plan is always associated with the staff assignment. Hence, we develop a machine learning based integrated pedestrian facilities planning and staff assignment optimization model in transfer stations to schedule the pedestrian facilities plan and the staff assignment together. It aims to minimize the staff assignment cost and the deviation of working time of each employee of the station. The minimizing of the deviation gains the fairness of the assignment plan. The facilities plan is enforced by the level-of-services requirement in three performance indicators including transfer capacity, transfer average time and level-of-service. The performance indicators of facilities plans are evaluated by a simulation-based machine learning method. Based on simulation results, the random forest method fits a quantitative relationship among performance indicators of the facilities plans with operation scenario attributes and facilities plan attributes. The experiments on the case study of Xipu station show the integrated model can return pedestrian facilities plans which meet the level of service requirements and assign employees fairly of each period in a day and minimize the labor cost. The solutions of pedestrian facilities plan and staff assignment plan for possible operation scenarios in future are also suggested to station manager by our integrated method.

Keywords
Transfer stations, Facilities plan, Staff assignment, Simulation, Random forest

1 Introduction

Pedestrian Facilities Planning (PFP) is about adjusting the layout of pedestrian flow route and the number of machines in service to meet the passenger movement demand in the station. Given the increasing passenger demand in transfer station in China, the station management face high pressure. Especially, the passenger demand varies from hour to hour in a day. A fixed pedestrian facilities plan usually fails to satisfy passengers' transfer demand, resulting in the bottleneck of the rail transit network. The adjustment of pedestrian facilities in different time periods in one day are applied by station managers' experience currently. Hence, the optimization of PFP is an urgent concern(Hu et al. (2015)).

Staff assignment (SA) of the employees in the station is the most important part of station management. Considering the situation in China, lots of passengers are unfamiliar with the automatic ticket machine and automatic gate for ticket checking, and they are unable to use them correctly and quickly. Some passengers may waste time on walking through passages without conductor's guidance in the station because of low sensitivity to the guide signs. And it is necessary to maintain order by staff if congest happens. Therefore, the staff are assigned to the facilities to assist passengers in passing correctly and quickly, guaranteeing the efficient operation of station. In practice, the employees are associated with pedestrian facilities plan. For example, one employee can handle four gates at the same time, if the pedestrian facilities plan with five gates in service is chosen with the increasement of passenger demand, two employees would be assigned. Hence, the staff assignment plan should be modified to correspond to the adjustment pedestrian facilities plans in time periods.

In all, PFP and SA must be managed simultaneously in the daily operation to meet the high-density transfer stations. There are three challenges in the management of station when applying the integration of PFP and SA into practice.

(1) The performance the pedestrian facilities plan is hard to quantify. A lot of researches are focused on the evaluation of pedestrian flow performance and passenger assignment in the station. The pedestrian route choice model which is developed from the route choice model in road transportation is a widely used to evaluate the pedestrian behavior(Lam et al. (1999) and Hänseler (2016)). However, the pedestrian routes in transfer station are different from the routes in road transportation, they are always overlapped. It is hard for pedestrians to find the optimal paths in most time. Then, the accuracy of pedestrian route choice model need to be improved by other measures. Berbey et al. (2012) and Xu et al. (2013) addressed a probabilistic model and a fuzzy logic approach to modeling passenger behavior on the platform. In other researches, the route choice data collected by Bluetooth and WiFi technologies are investigated. Shlayan et al. (2016) used Bluetooth and WiFi technologies to obtain origin-destination(OD) demands and pedestrian movement path in public transportation terminals. Based on the route choice and waiting time data collected by Bluetooth on two platforms, Heuvel et al. (2015) estimated the impact of vertical infrastructure like escalators and stairs on choosing pedestrian flow routes in train station. Through Bluetooth, WiFi and infrared technology, Heuvel et al. (2016) expanded the research to include station hall and non-train passengers. However, collecting data of pedestrian flow route is a difficult and time-consuming process, and the exact pedestrian flow route are hardly acquired by these technologies due to the date collecting devices only located in specific points. Hence, these technologies need to be improved to acquire more reliable data in a efficient way. Pedestrian simulation has been recognized as a powerful tool under limited data collection and has been implemented to return the performance indicators in the station for the strong computation capability (Hoy et al. (2016), Pu (2017)). However, the pedestrian facilities are fixed as the given conditions for the analysis. Some research focuses on some specific facilities in the station. Hu et al. (2015) addressed a width design of the urban rail

transit stations circulation facilities problem by a simulation-based optimization. Only the width of circulation facilities was analysand. The relationship between the whole pedestrian facilities plan and the performance of the pedestrian is not analyzed. The machine learning method has been used in railway operation with the capacity assessment (Lai et al. (2014)), train delay estimation(Kecman and Goverde (2015)),and track maintenance(Ghofrani et al. (2018)). The machine learning could be a promising way to provide the nonlinear quantitative relationship(Ghofrani et al. (2018)).

(2) The optimization operation of pedestrian facilities is interrelated with the train timetable. The development of comprehensive transportation brings the multi-mode transfer into one station, and leads to more transfer connections, designed by the transportation company. A lot of researchers focus on the optimization and the train service in the station(Wong et al. (2008), Ibarra-Rojas and Rios-Solis (2012), Dollevoet et al. (2014) and Corman et al. (2017)), but the performance indicators about level of services are ignored in their researches. D'Acierno et at. (2017) combined an analytical dwell time model and dthe railway simulation under the crowding level at platforms to support timetabling development of metro, which aims to guarantee an appropriate robustness of rail operation. Minimizing the transfer time or travelling time are always the objective function in their models. Given the tightened transfer time, the station would be into a congested situation with high density of passengers in different time periods, which may lead to negative response and interactions among pedestrians(Bandini et al. (2014) and Pel et al. (2014)). Hence, the cost and difficulty of station management may increase because the passengers are more likely to miss their trains or services and need a longer waiting time for the following services(Tirachini et al. (2013)).

(3) The staff associated with the pedestrian facilities plan should assigned with the practical constraints. Although the staff assignment is a classic assignment problem(Ernst et al. (2004)), similar to the crew scheduling problem(Huisman (2007)), the practical requirements may make the problem more the number of employees is limited, and the labor constraints, like the maximum working time, the fairness of the work plan should be considered.

We develop a machine learning based integrated pedestrian facilities plan and staff assignment method in transfer stations. The purpose of this paper is to use the facilities plan and to assign the staff of the station dynamically to ensure the adaptability of the station to different passenger demands and to minimize the labor cost in one-day operation. We need to state that our research focuses on the transfer between the metro system and railway system in China. This paper contributes to face the three above methodological issues on the integration of PFP and SA.

For point 1, a machine learning method by fitting the train data returned by an agent-based simulation model is developed to evaluate the performance of the transfer management. We use AnyLogic software which is based on the social force model to simulate the passengers behavior in station. Instead of delimiting the pedestrian routes by researchers, origin-destination(OD) is the motive force for pedestrian and the routes are formed automatically. The interaction among pedestrian also be considered in simulation model because of social force model. Then, the simulation model output the performance indicators, including the transfer capacity, the average transfer time and level-of-service of the space in the station by input the train operation attributes and passenger attributes. Then, as the simulation based on the social force model, the simulation output data easily result in a randomness. To acquire a reliable performance indicators, a machine learning, random forest in the paper, fits the relationship between the pedestrian facilities plan and transfer management performance indicators.

For the point 2 and point 3, we develop a mathematical model with the considering of the level of services requirement in each time period. Based on the quantitative performance indicators acquired by the random forest, these requirements which is proposed by the station management could select the the facilities plans to face the challenge of the tightened transfer time. Then the mathematical model can allocate the selected pedestrian facilities plan and assign the station staff, which aims to minimize the staff assignment cost which is the main operation cost of station. The minimizing of the deviation gains the fairness of the assignment plan. Then, the constraints of limitation on the total working time per day, the working time range and fairness of workload are considered.

The next sections of the paper are organized as follows. Section 2 describes the pedestrian facilities planning problem, staff assignment problem and the integration problem. Section 3 presents the mathematical model proposed in this paper for the integration of PFPP and SAP problems, and explains the facilities plan performance quantization method by the random forest and simulation. Section 4 shows the experiments results by the simulation with random forest, and provides computational results on the proposed methods. Section 5 summarizes our contributions to the literature and outlines directions for further research on the integrated problem.

2 Problem Description

This section introduces the pedestrian facilities planning problem and staff assignment problem and how to integrate them.

2.1 Pedestrian facilities planning problem

Pedestrian facilities plan contains pedestrian flow route layout and number of machines in service. It can be adjusted to satisfy different passenger demand.

Two main elements of pedestrian facilities plan

We introduce the two main elements of pedestrian facilities plan.

1)Layout of pedestrian flow route

Pedestrian flow route is the path for passengers walking in the transfer station. By arranging the equipment like railings and barriers, the layout of pedestrian flow route is designed properly to make sure that there is no cross interference or convection among each other. Otherwise, a poor layout of pedestrian flow route usually leads to collisions, conflicts and dwell on the platform occur with passenger movement, which will increase the dwell time in the stations. Hence, an appropriate layout of pedestrian flow route should be selected to meet the operational requirements.

2)Number of machines in service

Except for the layout of pedestrian flow route, the number of machines in service in pedestrian facilities plans plays an essential role in the movement of transfer passenger. Concerning the queuing theory, the number of machines in service could result in the service and pass time in the machine and the neighbor area. As the transfer between the metro system and railway system in China is very complicated, the machines in the transfer station include automatic ticket machines, automatic gates for ticket checking, etc. The passing of each machine will influence each other. Therefore, the station manager can change the number of machines in service by turning off idle machines or turning on more machines to guarantee the level of services requirement.

Evaluation of pedestrian facilities plan

The pedestrian facilities plan is selected by the station manager by its performance under the certain operation scenario. Based on the analysis in (Hänseler (2016)), we propose three indicators to evaluate the performance of pedestrian facilities plan for the transfer station: 1)Transfer capacity. It is the number of passengers transferring from the one system to another in the given time period. These indicators could be used in the oversaturate situation; 2)The average transfer time. It is the average of the time from arrival train to departure train belong to another system for each transfer passenger. It could obtain the operation of a station to meet the tighten timetable requirement; 3) level-of-service. It refers to the pedestrian density of space for each passenger, and usually measure in six levels from A to F. When level-of-service of space factor is low, the train stop time may not be sufficient for passengers to get on or get off the train, and safety accidents may also happen. The station manager can determine to choose which pedestrian facilities plan to improve the level of station management by the performance of these indicators.

Adjustment of pedestrian facilities plan

The adjustment of pedestrian facilities plan can improve the level of station management and save operation cost. Different pedestrian facilities plans are applicable for different operation scenarios.

The operation scenario contains two parts. The first part is the attributes of the passenger. Passenger demand, which affects the number and density of passengers directly, is the main factor in the operation scenario. Passenger characteristics, including whether to carry luggage, whether to buy tickets, etc., may affect the time of passengers receiving all kinds of service. The second part is train operation headway. Train operation headway, which means the change of the train frequency for multi-mode station in the operation scenario. It can affect the arrival and departure density of passengers in the station.

As the operation scenario changes in every time period in the daily operation, the performance of pedestrian facilities plan varies with the change of the three evaluation indicators. On the other side, different operation scenario will set the delicated operation requirements for operation. Corresponding to the performance indicators, the station operation provides three requirements, the minimum capacity, the maximum average transfer time and minimum level-of-service. If the performance of the three evaluation indicators shows that the current pedestrian facilities plan cannot meet the level of services requirement in current operation scenario, the pedestrian facilities plan must be adjusted. Considering the using cost and service life of facilities, under the premise of guaranteeing the level of services requirement, the pedestrian facilities plan with fewer machines in service can be chosen in low demand period. What's more, fewer machines in service means that fewer conductors are required, which can reduce the cost for the station. And it must be stated that changing pedestrian flow route between different pedestrian facilities plans, which can be completed in one minute, cost so little that it can be ignored because of using movable and retractable railings.

2.2 Staff assignments problem

Staff assignment plan is about how many employees should be assigned to and where to assign as well as how long will they work. According to the performance of evaluation indicators in different operation scenarios, the pedestrian facilities plan can be determined. Then, it is necessary to develop an appropriate staff assignment plan for pedestrian facilities plan and to minimize the labor cost at the same time, because labor cost is the main expense in the operation of rail transit stations in China. The pedestrian facilities plan determines the number of employees required. The number of machines in service is different when pedestrian facilities plan changes along with a different period, which leads to a different requirement of employees in one day because the number of machines that one employee can handle is limited.

There are two kinds of employees in the staff of rail transit station in China: regular employees and secondment employees. The beginning and the end of regular employees' working time are fixed. The secondment employees can be assigned to anywhere at any time if the requirement of employees is more than regular employees. However, the cost of assigning one secondment employee is three times more than the regular employees. Besides, the limit of employees' fatigue, including the total working time of each employee, working time range and fairness of workload should be considered. Normally, the total working time of each employee should be less than 8 hours, and at least 1-hour rest is required.

2.3 Integrated pedestrian facilities planning and staff assignment problem

As mentioned above, the simulation model and random forest are aim to obtain performance indicators for different operation scenarios, but the categoric input and output of them are not determined. Therefore, an integrated model is developed to combine the pedestrian facilities plan and staff assignment plan, and it can define the input and output clearly of simulation model and random forest. The purpose of this model is to reduce the operation cost of the station, especially the staff cost, and to improve the transfer passenger satisfaction by according to the evaluation indicators mentioned above.

We firstly develop an agent-based simulation model in Anylogic software to obtain all the evaluation indicators performances of all possible combinations between operation scenarios and pedestrian facilities plans. The simulation model can provide plenty of indicators performances results of each combination to prevent extreme situation of results that affect the final result. However, due to the randomness and large quantities of the simulation results and the nonlinear relationship between operation scenario and pedestrian facilities plan, it is very difficult for station manager to judge which simulation results can be used and choose appropriate pedestrian facilities plan for certain operation scenario. Therefore, we fit the nonlinear relationship between them by using Random Forest, and we obtain the correspondence between the performance of evaluation indicators and the combinations of operation scenarios and pedestrian facilities plans. Then, we use given operation scenario for the operation day to select out all the pedestrian facilities plans which satisfy the performance requirement provided by the station manager. Finally, the number of employees required for each pedestrian facilities plan and the cost of hiring an employee are provided. We use a mathematical model which can assign staff to minimize the operation cost and determine the integral pedestrian facilities plan from all selected out plans to satisfy the performance requirement simultaneously for the station. The whole solution process is shown in Figure 1.

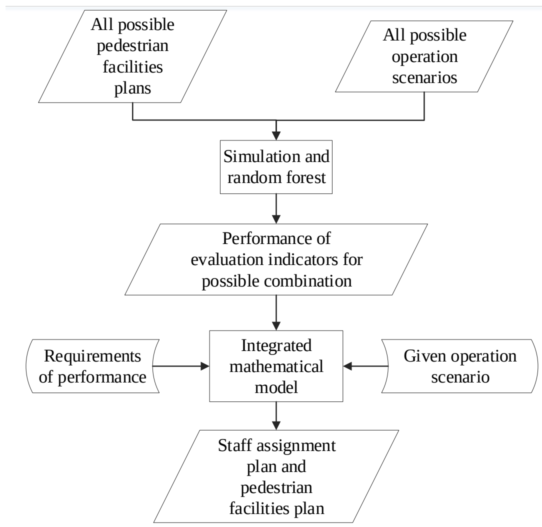


Figure 1: Flow chart of the our method.

3 Model Formulation

In this section, we address our Integrated Pedestrian Facilities Planning and Staff Assignment Problem (IPFPSA) in the transfer station. Firstly, the notation is presented. Then, we will give the mathematic model in detail. Next, the random forest connected the pedestrian facilities plan with the performance indicators in the station in IPFPSA will be introduced.

3.1 Notation

The sets, parameters and decision variables used in this paper are described in Tables 1 and 2 , respectively.

3.2 Modeling assumptions

The Modeling assumptions in our research is list as follows:

(1)Since the operation and adjustment of facilities plan cost are not considered in the daily operation, we do not minimize them in our model.

(2) A facilities plan utilized in station must be associated with a number of employees.

(3)A number of secondment employees could apply to the facilities plan when necessary. The labor cost of secondment employees is much higher than the regular employees.

(4)The number of transfer passenger is given. The station manager could acquire the passenger demand by other prediction methods.

Table 1: Definition of sets and parameters

Symbol	Definition
S	Set of staff, index by s , i.e., $s \in S$
S_{se}	Set of secondment staff in some time periods, index by s , i.e., $s \in S_{se}$
S_{re}	Set of regular staff, index by s , i.e., $s \in S_{re}$
F	Set of pedestrian facilities plan, index by f , i.e., $f \in F$
T	Set of time periods, index by t , i.e., $t \in T$
O_s	The limitation of working time for staff s
$t_{s,b}, t_{s,e}$	The begin time period and end period for staff s
$L_{f,t}$	The level-of-service indicator when facility plan f performance in time period t
$\tau_{f,t}$	The maximum average transfer time indicator when facility plan f performance in time period t
$Cap_{f,t}$	The minimum transfer capacity indicator when facility plan facility plan f performance in time period t
$L_{min,t}$	The minimum level-of-service requirement in time period t
$\tau_{avg,t}$	The transfer time requirement in time period t
$Cap_{min,t}$	The transfer capacity requirement in time period t
N_f	The needed number of staff when using facility plan f

Table 2: Decision variables

Symbol	Definition
$x_{s,f,t}$	0-1 staff assigning variable, equal to 1 if employee s assigned to facilities plan f in time period t, 0 otherwise
$y_{f,t}$	0-1 facilities planning variable, equal to 1 if facilities plan f selected in time period t to meet the station operation requirements, 0 otherwise
z_s	The actual working time for employee s
w	The average working time for all the staff $s \in S$
D_s^+, D_s^-	Deviate time to the average working time for all the staff

3.3 Mathematical Model

The aim of pedestrian facilities planning is to meet the station operation requirements, while the staff assigning focuses on operating the facilities plan in a low cost. Moreover, fairness is included to improve the satisfaction of station staff. Moreover, our constraints include four parts: 1) the station operation requirements for the facilities plan; 2) the staff assignment rules; 3) the computing of the deviation of the working time for each employee to the total staff average; 4) the relationship of facilities plan and staff assignment. IPFPSA is formulated as follows:

$$min \sum_{t \in T} \sum_{f \in f} \sum_{s \in S} c(x_{s,f,t}) x_{s,f,t} + \sum_{s \in S} D_s^+ + \sum_{s \in S} D_h^- . \tag{1}$$

$$c(x_{s,f,t}) = \begin{cases} c_{s,t,re}, s \in S_{re} \\ c_{s,t,se}, s \in S_{se} \end{cases} \tag{2}$$

$$\sum_{f \in F} y_{f,t} = 1 \, \forall t \in T. \tag{3}$$

$$\sum_{f \in F} L_{f,t} y_{f,t} \geq L_{min,t} \quad \forall t \in T \tag{4}$$

$$\sum_{f \in F} Cap_{f,t} y_{f,t} \geq Cap_{min,t} \quad \forall t \in T. \tag{5}$$

$$\sum_{f \in F} \tau_{f,t} y_{f,t} \leq \tau_{avg,t} \quad \forall t \in T. \tag{6}$$

$$\sum_{f \in F} x_{s,f,t} \leq 1 \quad \forall s \in S, t \in \{T | t_{s,b} \leq t \leq t_{s,e}\}. \tag{7}$$

$$\sum_{f \in F} x_{s,f,t} = 0 \quad \forall s \in S, t \in \{T | t \leq t_{s,b}, t \geq t_{s,e}\}. \tag{8}$$

$$\sum_{t \in T} \sum_{f \in F} x_{s,f,t} \leq O_s \quad \forall f \in F, t \in T. \tag{9}$$

$$z_s = \sum_{s \in S} x_{s,f,t} \quad \forall s \in S. \tag{10}$$

$$W = \sum_{s \in S} z_s \setminus n_s. \tag{11}$$

$$z_s = W + D_s^+ - D_s^- \quad \forall s \in S. \tag{12}$$

$$x_{s,t} \leq M y_{f,t} \quad \forall t \in T \tag{13}$$

$$\sum_{s \in S} x_{s,f,t} = N_f y_{f,t} \quad \forall f \in H, t \in T. \tag{14}$$

$$x_{s,f,t} \in 0, 1 \quad \forall s \in S, f \in F, t \in T. \tag{15}$$

$$y_{f,t} \in 0, 1 \quad \forall f \in F, t \in T. \tag{16}$$

$$z_s \in \mathbb{Z} \quad \forall s \in S. \tag{17}$$

$$w \in \mathbb{R} \tag{18}$$

$$D_s^+, D_s^- \in \mathbb{R} \quad \forall s \in S. \tag{19}$$

The objective function (1) minimizes the total staff assigning cost. To obtain the fairness of the staff, the deviations of the average working time are computed. Constraints (2) deal with the situation that the secondment employees working cost is different from the regular employees. Indeed, it is much higher, which is computed by the practical operation experience. Constraints (2) enforce that only one facility plan s is selected in each time period t. Constraints (4), (5), and (6) ensure that in each time period t , the selected facilities plan s must meet the service level requirement $L_{min,t}$, the transfer capacity requirement $Cap_{min,t}$. and the transfer average time requirement $\tau_{avg,t}$, respectively. Constraints (7) and (8) address that each employee s only assigned to one facility plan s in their working time period t, while employee s could not work outside the working time period t. For an employee s, Constraint (9) is defined to ensure that the total working time in one day is less than O_s. The big-M in constraints (13) are used to couple the usage of an employee s with facilities plan f. Constraints (13) imply that the usage of an employee s in time period t will be enforced to be 0 if $y_{f,t}$ is equal to 0, i.e. employee s could not assign to a facilities plan s which does not implement; otherwise, $x_{s,f,t}$ is less than or equal to the value of big-M, i.e. it could be used. Constraints (14) specify that if $y_{r,t}$ is equal to 1, N_f employees are applied to the facilities plan r to operate it. The domain of variables in the model is defined by expressions (16)- (19). The staff assigning and facilities planning are defined as binary variables. The actual working time, the average working time and the deviation time to the average working time are defined as integer variables. As we only focus on one-day's planning, the IPFPSA could be solved by some commercial software. The essential part for the model is how to get the performance of the facilities plan in various scenarios. We will introduce quantization method next.

3.4 Facilities plan performance quantization method

As we stated before, the process of transfer between two modes leads to a nonlinear relationship between the facilities plan and the performance. Regression methods based on machine learning are the common ways to model that. This kind of method has been applied in railway system to predict the railway capacity, train delay, etc. For the machine learning, how to get the learning data is the most challenge task. We develop a simulation system to provide a number of results which connect the performance indicators and input data.

Simulation for transfer station

We choose Anylogic to build the simulation model of passenger transfer progress. Anylogic combines professional discrete event, system dynamics, and agent-based modeling in one platform(Anylogic (2019)). In the software, Rail Library is used to build the train operation simulation model, in which each train are agents with their own states and properties. Pedestrian Library, based on the social force model, is used to build a pedestrian moving simulation model. Pedestrians can be preassigned with individual characteristics in models. In order to integrate train and pedestrians, this paper builds a transfer simulation model, which can simulate the train operation in the station and the complete progress of transfer in one model. Therefore, the change of train operation headway and its influence on operation scenario can be presented in the simulation model. The operation scenario attributes, pedestrian facilities plan attributes and performance indicators are listed in Table 3. Facilities plan attributes have been introduced in Section 2, we will introduce simulation attributes and performance indicators in detail.

Operation scenario attributes

In multi-mode transfer station connecting railway and metro, both of the train operation headway of railway and metro affect the operation scenarios. Whether to buy tickets or not

Table 3: The input and output of the simulation model

Operation scenario attributes	**Facilities plan attributes**	**Facilities Performance indicators**
Headway of railway(HR) $H_{r,t}$,	Layout of pedestrian flow route(LPFR)R_y,	
Headway of metro(HM) $H_{m,t}$,	Automatic gate for railway tickets checking(AGRTC)$N_{rtc,y}$,	Transfer capacity $Cap_{y,t}$,
Metro inbound demand(MID) $D_{in,t}$,	Automatic ticket machines for metro(ATMM)$N_{m,y}$,	Average transfer time $\tau_{y,t}$,
Transfer demand(TD) $D_{tr,t}$,	Automatic gate for metro tickets checking(AGMTC)$N_{mtc,y}$	Service level $L_{y,t}$
Ratio of buying tickets(RBT) RA_t		

is the main passenger characteristic in this transfer progress. Passenger demands include transfer demand and inbound demand. Therefore, inbound demand from the metro, transfer demand of each arrival train, the ratio of buying tickets in transfer passengers, train operation headway of railway and train operation headway of metro are the five variables of operation scenarios.

Facilities plan attributes
As stated before, the layout of pedestrian flow route and number of machines in service are the facilities plan attributes. In multi-mode transfer station, the machines are specified in automatic gate for railway tickets checking, automatic ticket machines for metro and automatic gate for metro tickets checking.

Facilities performance indicators
The performance of evaluation indicators is returned as the results of the simulation model under the given operation scenario and pedestrian facilities plan. The method to obtain the performance of evaluation indicators shown as follows.

1 Transfer capacity

Transfer capacity is the number of passengers given by a simulation method within the time period. In the simulation model, the unit time is set to 10 minutes, so this indicator can calculate the number of passengers transferring successfully in 10 minutes after the train arrived.

2 Transfer time

The average transfer time is chosen to be the performance indicator to evaluate all transfer passengers' transfer time. In multi-mode transfer station, the progress of transfer is divided into 4 parts as shown in expression (20)and the average transfer time τ_{avg} can also be computed.

$$\tau_{avg} = \sum_{i \in N} (\tau_{i,walk1}, \tau_{i,serv} + \tau_{i,walk2} + \tau_{i,wait}) \setminus |N|. \quad (20)$$

Where $\tau_{i,walk1}$ is the time of getting off and walking of ith transfer passenger, $\tau_{i,serv}$ is the time of buying tickets and passing automatic gate for ticket checking of of ith transfer passenger, $\tau_{i,walk2}$ is the walking time after checking ticket of ith transfer passenger and $\tau_{i,wait}$ is the time of waiting to get on the train to transfer of ith transfer passenger.

3 Level-of-service

Considering China's rail transit transfer station, we choose the A-F level-of-service grading standard in (HCM,Highway Capacity Manual (2000)). This paper counts the percentage of time in the A-F level-of-service of the station in per hour, and normalizes the six values to obtain a value of (0,1), which means the level-of-service is better if the value is larger.

Each operation scenario and each pedestrian facilities plan can be combined with a possible input of simulation model, which leads to plenty of results of evaluation indicators. Due to the randomness of the simulation and the nonlinear relationship between parameters and results, the results need to be processed further by machine learning.

Random forest for quantization method

Random forest ensembles a set of decision tree prediction results to gain better predictive results for both classification and regression problems. It is an efficient machine learning model which was used widely for many real-world applications(Shafique and Hato (2017), Kecman and Goverde (2015)). Moreover, the random forest could provide the importance scores of input attributes which is useful for the analysis of the integration model.

The random forest needs lots of learning scenarios before fitting it. Hence, the simulation method is applied to provide a number of scenarios. We choose the random forest to connect the simulation attributes, pedestrian facilities plan attributes and performance indicators in Table 3 with expression (21).

$$\begin{cases} Cap_{y,t} = f_{cap}(H_{r,t}, H_{m,t}, D_{in,t}, D_{tr,t}, RA_t, R_y, N_{rtc,y}, N_{m,y}, N_{mtc,y}) \\ \tau_{y,t} = f_{\tau}(H_{r,t}, H_{m,t}, D_{in,t}, D_{tr,t}, RA_t, R_y, N_{rtc,y}, N_{m,y}, N_{mtc,y}) \\ L_{y,t} = f_L(H_{r,t}, H_{m,t}, D_{in,t}, D_{tr,t}, RA_t, R_y, N_{rtc,y}, N_{m,y}, N_{mtc,y}) \end{cases} \tag{21}$$

After the evolution of performance indicators, the commercial software, like the gurobi or IBM Cplex, could be used to solve IPFPSA, to get the final solution.

4 Numerical experiments

This section present the numerical experiment and solution analysis. The case study of Xipu station is used to model the transfer management. The experiments on the models proposed in this paper have been performed on a laptop computer with i7-6700HQ @ 2.6 GHz CPU and 8.0 GB RAM. The IPFPSA is solved with Gurobi 7.5, the simulation is developed with Anylogic 8.3 , the random forest is trained on Python 3.6 with scikitlearn package on Windows 10.

4.1 The case study of Xipu transfer station

The Xipu station is an appropriate station for the case study as shown in Figure 2. It is a multi-mode station as the intermediate station of the intercity railway system and terminal station of the metro system. Passengers can transfer between railway and metro on the same platform, where the platform 1 provides the transfer from the railway to metro and platform 2 provides the transfer system from metro to railway. Moreover, different ticket systems of railway and metro means more complicated procedure and longer walking distance for transfer passengers. When passengers transferring from metro to railway, they have to wait for check-in to board, which can relieve the pressure of transfer management. Therefore, we focus on the transfer from the railway to the metro on platform 1 with complicated facilities requiring more conductors of the station staff, and more continuous transfer progress.

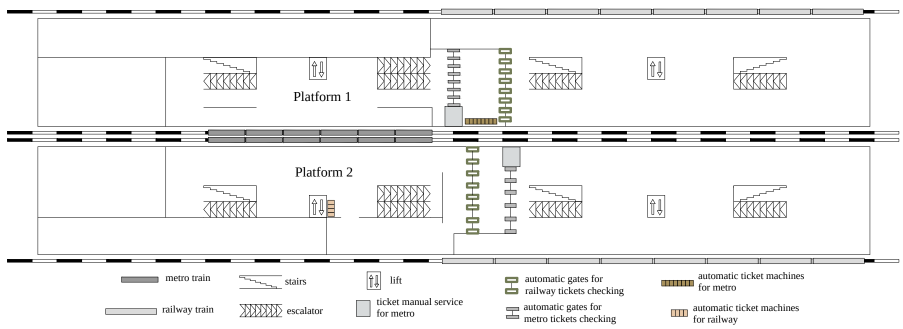


Figure 2: The facilities layout on the platforms in Xipu station

In a limited area on platform 1 as shown in Figure 3, after getting off the arrival train, transfer passengers must complete the transit through checking-out of railway system, ticket purchasing (if not have) and checking-in of metro system by 7 automatic gates for railway tickets checking, 7 automatic gates for metro tickets checking, 8 automatic ticket machines and 1 manual ticket service for metro. With so many machines available for service, the number of chosen machines for service varies in a wide range, increasing the complicate for developing pedestrian facilities plan and staff assignment plan. possible layout of pedestrian flow routes aiming to satisfy different passenger demands can be chosen after checking in the metro system as shown in Figure 3. Since most facilities belong to the metro system, the pedestrian facilities plan and staff assignment plan are mainly designed for metro station management.

4.2 Simulation experiments

simulation model

The transfer simulation model shown in Figure 4 is built in Anylogic and an display of simulation results is shown in Figure 5. The simulation model consists of train simulation and pedestrian simulation:

(1)Train operation: The three blocks of *Delay* control the stop of train and passengers getting on and off, which integrate the train operation and pedestrian.

(2)Passengers getting off and walking: The block *judge1* controls whether the passenger transfer or not.

(3)Passengers checking tickets: The block *judge2* controls whether the passenger needs to buy metro tickets or not. The number of machines in service is the parameter of block *CheckOutCRH*, *BuyTicket* and *CheckInMetro*.

(4)Passengers walking and waiting to get on: The block *Wait*, *GetOn* and *judge3* control the part of waiting to get on the metro and receiving the guidance of conductors to simulate the reality in China.

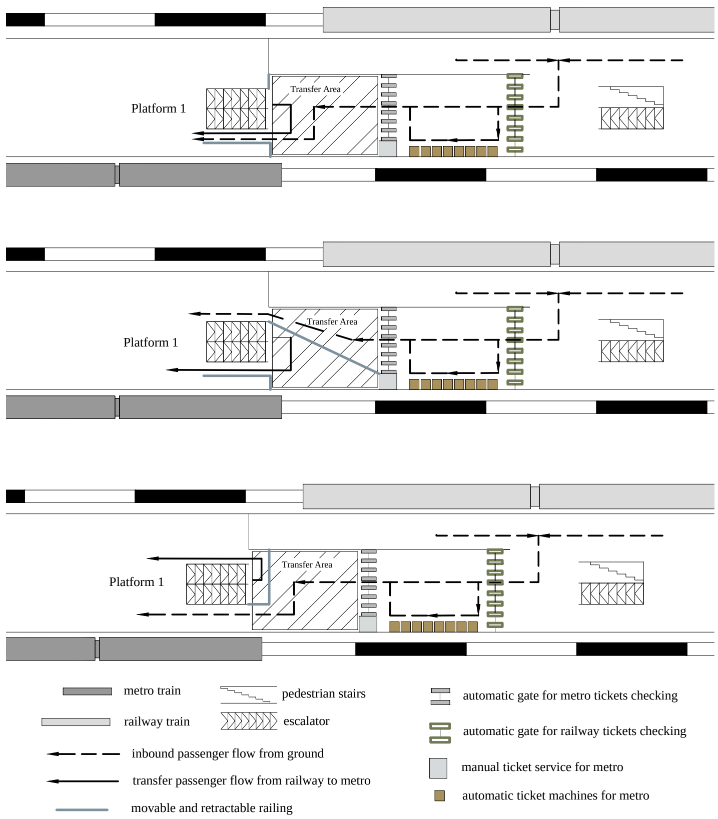


Figure 3: The facilities, three possible layout of pedestrian flow route and progress of transfer on platform 1

(5)Inbound passengers: The demand for inbound passengers also affects operation scenarios.

Parameter settings and experiment

A large number of simulation experiments are performed to obtain the performance of evaluation indicators for all possible operation scenarios with all possible pedestrian facilities plans. As shown in Figure 3, three types of layout of pedestrian flow route are designed. Then, 13 different numbers of machines in service and with its requirement of employees are designed as shown in Table 4. Since the manual ticket service for metro is fixed at 1, it will not be considered in requirement of employees for the facilities plans .

Due to 3 types of layout of pedestrian flow route and 13 different numbers of machines in service, there are 39 pedestrian facilities plans enumerated.

For operation scenarios, we developed over 400 possible operation scenarios in total by changing the attributes for diverse operation scenarios. Each operation scenario will con-

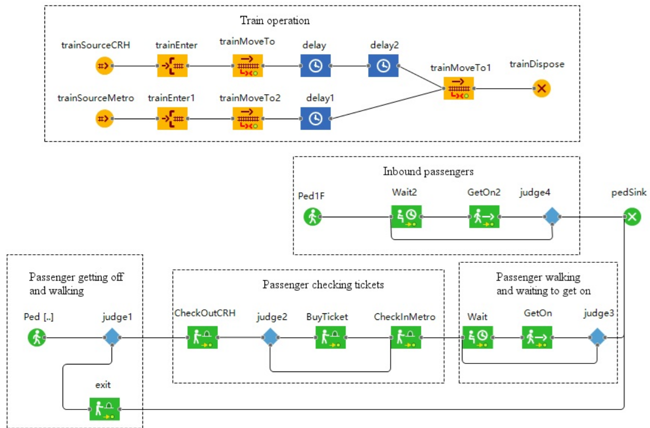


Figure 4: The transfer simulation model in Anylogic.

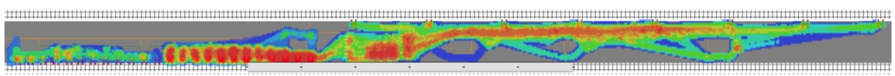


Figure 5: An display of simulation results in Anylogic.

Table 4: Different facilities plans of number of machines in service and requirement of staff

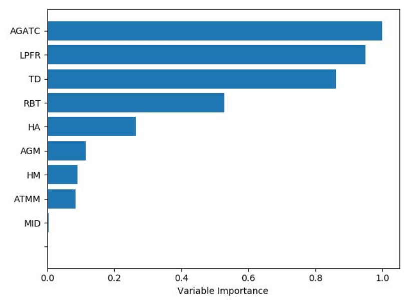
Facilities plan No.	AGRTC	ATMM	AGMTC	Requirement of employees
1	7	8	7	6
2	7	8	6	6
3	7	7	6	6
4	6	7	6	6
5	6	7	5	5
6	6	6	5	5
7	7	6	5	5
8	8	6	4	4
9	9	5	4	4
10	4	5	4	4
11	4	5	3	3
12	4	5	3	3
13	3	6	3	3

tinue at least 1 hour in one day. Considering 39 pedestrian facilities plans to be chosen for 400 operation scenarios, there are more than 15,000 combinations of operation scenarios and pedestrian facilities plans required to be simulated to obtain the performance of evaluation indicators. Considering the randomness of simulation, each combination was simulated for five times. In total, more than 75,000 results of performance of evaluation indicators are provided to random forest.

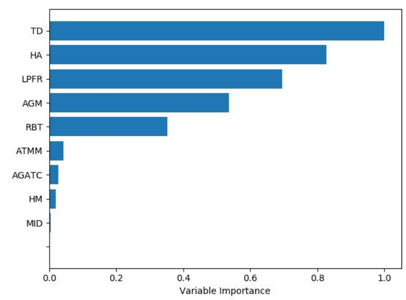
4.3 Random forest results

To improve the accuracy of machine learning, *GridSearchCV* in scikitlearn package is applied for hyperparameter tuning on *max_depth*, *max_features*, *max_leaf_nodes*, etc. Cross validation implement to avoid overfitting. Finally, the scores of three performance indicators are larger than 9.5.The score of the fitting on transfer capacity is 0.955606873045, on average transfer time is 0.97330181704 and on level-of-service is 0.972644965479. It shows the accuracy of random forest is acceptable for the fit.

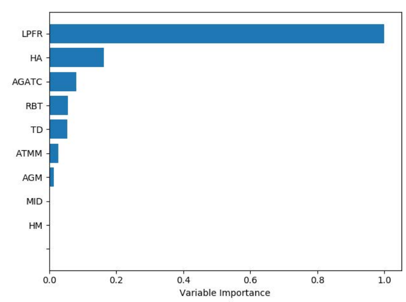
Based on the fitting results, the importance scores are shown in Figure 6. It shows that besides the facilities plan attributes, transfer demand and the ratio of buying tickets get the higher scores. Hence, we analyze the influence of them in the next.



(a) importance scores on fitting transfer capacity



(b) importance scores on fitting transfer time



(c) importance scores on fitting level-of-service

Figure 6: importance scores of the random forest

4.4 The integrated solution

The choosing of employment plan and different diagrams of pedestrian facilities plan and staff assignment plan for each day are results of the integrated mathematical model. With the potential operation scenarios in the future, suggestions for staff adjustment are provided to the station manager.

The cost and work time limit of staff

As stated before, the staff consist of regular employees and secondment employees, whose costs to assign and working time limits are different from each other. 2 teams of regular employees and 1 team of secondment employees can be assigned in Xipu station. The cost, working time range, and upper bound of working time of staff are shown in Table 5. We design 3 employment plans as shown in Table 6 for Xipu station. Each day is divided into 16 periods because the operation time of one day is from 6:00 to 22:00.

Table 5: Cost and work time limit of employees

Staff	Beginning of working time	End working time	Cost	Upper bound
Regular team 1	6:00	1400	250 per day	8
Regular team 2	14:00	22:00	250 per day	8
Secondment team	6:00	22:00	100 per hour	12

Table 6: Number of employees for each employment plan

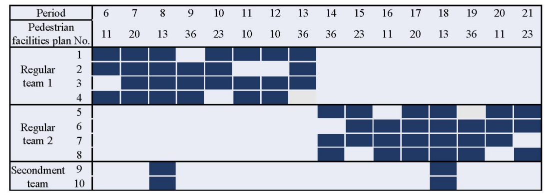
Plan No	Number of employees in Regular team 1	Number of employees in Regular team 2	Number of employees in secondment team
1	3	3	3
2	4	4	2
3	5	5	1

The different solution for four days in a week

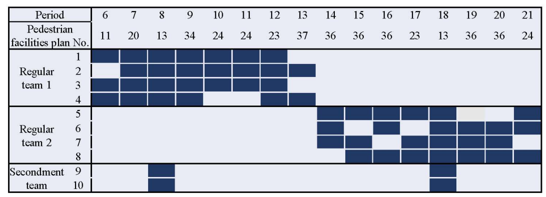
With the given requirement of performance of evaluation indicators, pedestrian facilities plans and staff assignment plans can be obtained by integrated model. Different days in a week need different pedestrian facilities plans and staff assignment plans because of different operation scenarios. A optimal solutions for Monday, Wednesday, Friday and Sunday under employment plan 2 are shown in Figure 7.

Fairness of working time of regular employees is guaranteed in all days according to Figure 7. No difference of working time is more than 1 hour between any two regular employees. We rarely use secondment employees in order to save cost since the secondment employees cost relatively higher. What's more, the integrated mathematical model provides the certain pedestrian facilities plan which can satisfy the given operation scenario and requirement of the performance of evaluation indicators for each period.

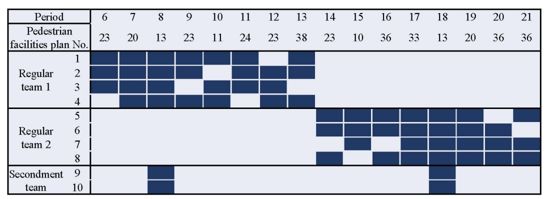
The evaluation of each plan is studied. With the given condition of operation scenarios and requirements of performance, the solution of average working time of regular employees, the total working time of secondment employees and average labor cost of Monday, Wednesday, Friday and Sunday in a week as shown in Table 7.In the table, A is short for Average working time of regular employees per day, B is short for total working time of secondment employees per day, and C is short for Labor cost per day. Average working time of regular employees aims to evaluate the work intensity of the regular. Total working time of secondment employees is aim to evaluate the rationality of the employment plan. Average labor cost aims to help station manager choose employment plan.



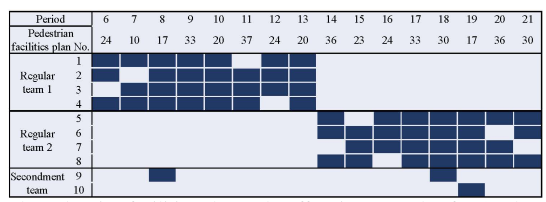
(a) Pedestrian facilities plan and staff assignment plan for Monday.



(b) Pedestrian facilities plan and staff assignment plan for Wednesday.



(c) Pedestrian facilities plan and staff assignment plan for Friday.



(d) Pedestrian facilities plan and staff assignment plan for Sunday.

Figure 7: Solutions of four days in week under employment plan 2

Table 7: The results of different employment plan

	Plan 1			Plan 2			Plan 3		
	A/hour	B/hour	C/money	A/hour	B/hour	C/money	A/hour	B/hour	C/money
Monday	8	8	2300	6.5	4	2400	5.4	2	2700
Wednesday	8	8	2300	6.5	4	2400	5.4	2	2700
Friday	8	9	2400	6.625	4	2400	5.5	2	2700
Sunday	8	10	2500	6.875	3	2300	5.8	0	2500
Average	8	8.75	2375	6.625	3.75	2375	5.525	1.5	2650

By observing Table 7, if the employment plan 1 is chosen, it is impossible for each regular employee to work all the time from beginning to ending continuously. Besides, the average of the total working time of secondment employees is over than 8 means that the cost of assigning secondment staff is high and the station manager can add a regular employee. If the employment plan 3 is chosen, the labor cost is too high for station manager and the average working time of regular employees per day is low which means the regular employees have too more rest time and the cost of employing a regular employee is not fully utilized. Therefore, employment plan 2 is the most appropriate plan. Because the regular employees have a reasonable rest time, the working time of secondment employees is not too long, and the labor cost is the lowest.

Different ratios of buying tickets and transfer demand

With the promotion of e-tickets and QR-code, transfer passenger may no longer need to buy tickets, and due to the development of rail transit, the number of transfer passenger will increase. These will lead new operation scenarios in the station. Therefore, different ratios including 35%, 45% and 55% of buying metro tickets and increasing transfer demand are performed. The current transfer demand is considered low and the other types of operation scenarios with middle and high transfer demands are designed. The results of three different employment plans in Wednesday are reported in Table 8. And A,B and C have the same meaning in Table 7 As reported in Table 8, with the increasing of transfer demand, the

Table 8: The results of different employment plans in the potential operation scenarios

		Low transfer demand			Middle transfer demand			High transfer demand		
	Ratio	A/hour	B/hour	C/money	A/hour	B/hour	C/money	A/hour	B/hour	C/money
Plan 1	0.55	8	8	2300	8	11	2600	8	13	2800
	0.45	8	13	2800	8	16	3100	8	21	3600
	0.35	8	12	2700	8	14	2900	8	18	3300
Plan 2	0.55	6.5	4	2400	6.625	6	2600	6.75	7	2700
	0.45	7.125	4	2400	7.25	6	2600	7.5	9	2900
	0.35	7.25	2	2200	7.25	4	2400	7.5	6	2600
Plan 3	0.55	5.4	2	2700	5.7	2	2700	5.9	2	2700
	0.45	5.9	2	2700	6.2	2	2700	6.6	3	2800
	0.35	6.0	0	2500	6.2	0	2500	6.6	0	2500

working time of employees and labor cost of station increase regardless of whichever employment plan is chosen. However, the operation scenarios whose ratio of buying tickets is 0.45 requiring the highest working time of employees and labor cost. The other 27 results of Monday, Friday and Sunday are similar to this. The reason for it is that when the ratio of buying tickets is 0.55, more employees are required for automatic tickets machines for metro; when the ratio of buying tickets is 0.35, fewer employees are required for automatic gates; while the ratio of buying tickets is 0.45, both automatic tickets machines and automatic gates require more employees. With the high transfer demand and the ratio of buying tickets as 0.45 and 0.35, if employment plan 2 is chosen, the average working time of regular employees is 7.5, which means that some regular employees have to work continuously for 8 hours. The employment plan is not appropriate for the same reason as stated before. The results show that if employment plan 3 is chosen, the working time of employees and labor cost can be both reduced. Each regular employee can have rest time and the secondment employees even can be idle when the ratio of buying tickets is 0.35. Therefore, if transfer demand is high and the ratio of buying tickets is reduced below 0.45, the employment plan 3 is the better choice.

5 Conclusions

This paper proposes a novel mix-integer linear problem formulation to deal with the integrated optimization of pedestrian facilities planning and staff assignment. The staff assigning and facilities planning are defined as binary variables to get the actual operation plan. To obtain the fairness, the actual working time, the average working time and the deviation time to the average working time are computed. The station operation requirements for the facilities plan select the qualified pedestrian facilities plan. The staff assignment rule addressed to obtain the meet the practical constraints. To acquire the performance indicators, an agent-based simulation model on Anylogic is developed to provide a huge of train data for the machine learning. Moreover, the random forest, a machine learning method, performs well to fit the non-linear relationships among the operation attributes, facilities plan attributes and the performance indicators on transfer capacity, average transfer time and level-of-service. The experiment results show the integrated model can return pedestrian facilities plans which meet the level of service requirements and assign employees fairly of each period in a day and minimize the labor cost for Xipu station. Moreover, the solutions of pedestrian facilities plan and staff assignment plan for possible operation scenarios in future show the labor cost for different employment plans. It could help the station manager to select the reasonable employment plan.

Future research efforts can be dedicated to investigating other working rules in practical operation on the staff assignment, and the staff rostering scheduling in a week. What's more, more machine learning methods, like the SVM (Support Vector Machine), ANN(Artificial Neural Network) and RNN(Recurrent Neural Network), could be tested to improve the accuracy.

Acknowledgments
This work is supported by National Natural Science Foundation of China (No. 61603317), National Key Research and Development Program of China (No. 2017YFB1200700-1), and Special Fund for Basic Scientific Research of Central Universities (No. 2682017CX023). The generous support of our colleagues Jun Zhao is greatly appreciated.

References

The AnyLogic Company, 2019. *AnyLogic Help*, Retrieved from https://help.anylogic.com/index.jsp.

Bandini, S., Mondini, M., Vizzari, G., 2014. "Modelling negative interactions among pedestrians in high density situations", *Transportation Research Part C: Emerging Technologies*, vol.40, pp. 251-270.

Berbey, A., Galan, R., Sanz, J.D., Caballero, R., 2012. "A fuzzy logic approach to modelling the passengers' flow and dwelling time", *WIT Transactions on The Built Environment*, vol.128, pp. 359-369.

Corman, F., D'Ariano, A.,Marra, A.D.,Pacciarelli, D.,Samà, M., 2017. "Integrating train scheduling and delay management in real-time railway traffic control", *Transportation Research Part E: Logistics and Transportation Review*, vol.105, pp. 213–239.

D'Acierno, L., Botte, M., Placido, A., Caropreso, C., Montella, B., 2017. "Methodology for Determining Dwell Times Consistent with Passenger Flows in the Case of Metro Service.", *Urban Rail Transit*, vol.3(2), pp. 73–89

Dollevoet, T., Corman, F., D'Ariano, A., Huisman, D., 2014. "A mathematical model for periodic scheduling problems", *Flexible Services and Manufacturing Journal*, vol.26, pp. 490–515.

Ernst, A.T., He, Q.,Jiang, H.,Krishnamoorthy, M.,Sier, D., 2004. "Staff scheduling and rostering: A review of applications, methods and models", *European journal of operational research*, vol.153, pp. 3–27.

Ghofrani, F., He, Q.,Goverde, R.M.,Liu, X., 2018. "Recent applications of big data analytics in railway transportation systems: A survey", *Transportation Research Part C: Emerging Technologies*, vol.90, pp. 226–246.

Hänseler, F.S., 2016. *Modeling and estimation of pedestrian flows in train stations*, INTER,EPFL,Lausanne.

Transportation Research Board, 2000. *Highway Capacity Manual*, Transportation Research Board,Washington, DC.

Heuvel, J.V.D., Voskamp, A., Daamen, W., Hoogendoorn, S.P., 2015. "Using Bluetooth to estimate the impact of congestion on pedestrian route choice at train stations", *Traffic and Granular Flow '13*, pp. 73–82.

Heuvel, J.V.D., Ton, D., Hermansen, K., 2016. "Advances in Measuring Pedestrians at Dutch Train Stations Using Bluetooth, WiFi and Infrared Technology", *Traffic and Granular Flow '15*, pp. 11–18.

Hoy, G., Morrow, E., Shalaby, A., 2016. "Use of Agent-Based Crowd Simulation to Investigate the Performance of Large-Scale Intermodal Facilities", *Transportation Research Record: Journal of the Transportation Research Board*, vol.2540, pp. 20–29.

Hu, L., Jiang, Y., Zhu, J.,Chen, Y., 2017. "A PH/PH (n)/C/C state-dependent queuing model for metro station corridor width design", *European Journal of Operational Research*, vol.240, pp. 109–126.

Huisman, D., 2007. "A column generation approach for the rail crew re-scheduling problem", *European Journal of Operational Research*, vol.180, pp. 163–173.

Ibarra-Rojas, O.J., Rios-Solis, Y.A., 2012. "Synchronization of bus timetabling", *Transportation Research Part B: Methodological*, vol. 46, pp. 599-614.

Kecman, P., Goverde, R.M.P., 2015. "Predictive modelling of running and dwell times in railway traffic", *Public Transport*, vol.7, pp. 295–319.

King, D., Srikukenthiran, S., Shalaby, A, 2014. "Using Simulation to Analyze Crowd Congestion and Mitigation at Canadian Subway Interchanges", *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2417, pp. 27–36.

Lai, Y.-C., Huang, Y.-A., Chu, H.-Y., 2014. "Estimation of rail capacity using regression and neural network", *Neural Computing and Applications*, vol.25, pp. 2067–2077.

Lam, W. H. K., Cheung, C. Y., Lam, C. F., 1999. "A study of crowding effects at the Hong Kong light rail transit stations", *Transportation Research Part A: Policy and Practice*, vol.33, pp.401–415.

Pel, J. A., Bel, N. H., Pieters, M., 2014 "Including passengers' response to crowding in the Dutch national train passenger assignment model", *Transportation Research Part A: Policy and Practice*, vol.66, pp. 111-126.

Pu, Y., 2017. *Capacity Analysis of the Union Station Rail Corridor using Integrated Rail and Pedestrian Simulation*, Department of Civil Engineering,University of Toronto,Toronto.

Shafique, M. A., Hato, E, 2017. “Classification of travel data with multiple sensor information using random forest”, *Transportation research procedia*, vol.22, pp. 144–153.

Shlayan, N., Kurkcu, A., Ozbay, K., 2016. “Exploring Pedestrian Bluetooth and WiFi Detection at Public Transportation Terminals”, *2016 IEEE 19th International Conference on Intelligent Transportation System*, pp. 229-234.

Tirachini, A., Hensher, D. A., Rose, J. M., 2013. “Crowding in public transport systems: Effects on users, operation and implications for the estimation of demand”, *Transportation Research Part A: Policy and Practice*, vol.53, pp. 36–52.

Wong, R. C., Yuen, T. W., Fung, K. W.,Leung, J. M., 2008. “Optimizing timetable synchronization for rail mass transit”, *Transportation Science*, vol.42, pp. 57–69.

Xu, X., Liu, J., Li, H., Zhou, Y., 2013. “Probabilistic model for remain passenger queues at subway station platform”, *J.Cent.South Univ*, vol.20, pp. 837-844.