Low-Cost Evaluation Based on Surrogate Model for Multidispatching Sections of High-Speed Railway Scheduling

Xinghao Wang
The State Key Laboratory of Synthetical Automation for Process Industries Northeastern University
Shenyang China
wangxinghao0918@163.com

Dongliang Cui
The State Key Laboratory of Synthetical Automation for Process Industries Northeastern University
Shenyang China
cuidongliang@mail.neu.edu.cn

Ruiguang Liu
The State Key Laboratory of Synthetical Automation for Process Industries Northeastern University
Shenyang China
liuruiguang111@163.com

Pengxin Yang
The State Key Laboratory of Synthetical Automation for Process Industries Northeastern University
Shenyang China
yang_pengxin@163.com

Peng Yue
The State Key Laboratory of Synthetical Automation for Process Industries Northeastern University
Shenyang China
1369187234@qq.com

Xuewu Dai
The State Key Laboratory of Synthetical Automation for Process Industries Northeastern University
Shenyang China
daixuewu@mail.neu.edu.cn

2021 China Automation Congress (CAC) | 978-1-6654-2647-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/CAC53003.2021.9728387

***Abstract*—Under the condition of network operation of high-speed railway, the influence of disruptions on the train scheduling at the current line and related lines is more and more significant. For the multi-region high-speed railway scheduling, the process of network operation and constraint is complex, which leads to a not accurate model. The method of simulation can get an accurate model and evaluate global indicators effectively, but the simulation and optimization process of large-scale high-speed railway network is expensive and time-consuming with the ever-expanding high-speed railway network and intensive High-speed railway traffic. The paper proposes a surrogate model to replace high-speed railway network simulation for low-cost evaluation, which effectively reduces the expense of simulation.**

Keywords—high-speed railway, surrogate model, large-scale, network, expensive simulation, low-cost

I. INTRODUCTION

As the scale of the high-speed railway network continues to expand, the railway network is becoming more complex, with a high traffic density, and the transportation capacity of important hub sections such as Xuzhou-Bengbu is close to saturation. Affected by unexpected events such as rain, snow, and equipment failures, a single point of the road network is delayed. Because the lines are more closely connected to each other and the coupling between the dispatching stations is stronger, the driving conflicts of a single dispatching station will also be spread to other dispatching regions [1]. The delayed train should be made to minimize the delay or even resume the scheduled operation plan, to minimize the impact of the delay on the road network. Nowadays, the dispatching effect is obviously affected by human factors, and the coordination ability is poor and the efficiency is low. The use of computer-assisted dispatchers to optimize the decision-making of the road network is one of the development trends of high-speed rail dispatching technology [2].

Scheduling is usually an NP problem. The larger the scale of the railway network, the exponential growth of the problem scale, the more complex the constraint relationship, the more difficult it is to evaluate the objective function, and the more difficult it is to solve the optimal strategy. For regional scheduling problems with multiple scheduling sections, an overall model can be built for multiple scheduling sections to achieve overall optimization of the road network, but the increase in the scale of the problem makes the problem a large-scale problem, which is difficult to solve. Reference [3] proposed an improved LaGrange relaxation method for train rescheduling, which decomposes complex problems into a series of simpler sub-optimization problems and reduces the number of constraints in the actual optimization process; Reference [4] introduced three common models of integer programming, mixed integer linear programming and alternative graph models to solve the problem of re-scheduling in railway network. When the above-mentioned documents solve the scheduling problem of the railway network, the more complicated the constraints of the road network, the more difficult it is to consider the coordination problem between the various dispatching stations in the road network, and the accuracy of the feasible solution obtained is difficult to guarantee. Reference [5-7] introduced the

This work was supported in part by the National Natural Science Foundation of China under Grant 61790574, U1834211 and 61773111, by the Province Natural Science Foundation of Liao Ning (2020-MS-093, by the Science and Technology R&D Program of the National Railway Corporation under Grant N2019G020, by the Liaoning Revitalization Talents Program under grant XLYC1907132.

use of train dispatching system for train rescheduling. The simulation model is the closest to the real dispatching system. The feasible solution obtained by it has high accuracy, but it takes a long time to obtain the feasible solution and it is expensive to calculate, and poor real-time performance. For the computationally expensive problem of simulation methods, reference [8] introduced a method based on off-line data-driven surrogate model to solve computationally expensive problems, using surrogate model to assist evolutionary algorithm search to find the optimal solution for large-scale problems quickly and accurately. Reference [9] introduces an optimization evaluation method based on surrogate model, which uses surrogate model to replace real simulation system for performance evaluation, thereby greatly shortening the evaluation process of genetic algorithm. The evaluation process based on surrogate model performs global optimization search. There are many methods for constructing surrogate models, such as support vector machine, neural network, Gaussian process, etc. Reference [10] uses Gaussian process to build a surrogate model to solve computationally expensive optimization problems.

We build a surrogate model to replace the simulator to obtain the true response value, so that the railway network optimization process can quickly obtain the optimal goal.

II. Problem Description

A. High-Speed Railway Network

Due to intensive High-speed railway traffic, not only the subsequent trains in this section will be affected by the delay, but the trains in the subsequent dispatching station will be delayed or even blocked when a traffic conflict occurs in a certain section of the railway network. Assuming that the road network contains N_k connected dispatching sections, each dispatching section is provided with a dispatching station corresponding to the train dispatching of the section. According to the position of the dispatching section in the high-speed railway network, the traffic density and the passenger flow density, the corresponding weight coefficients $\{w_1, w_2, \dots \dots, w_{n_k}\}$are assigned to each section. Each dispatch section contains several stations and corresponding inter-station sections. Suppose the dispatch section k contains $J_k - 1$ sections and J_k stations, as $Q_k = \{q_1, q_2, \dots \dots, q_{J_k - 1}\}, S_k = \{s_1, s_2, \dots \dots, s_{J_k}\}$.

Assuming that there are N_t trains running in the railway network, which is denoted as set T, the dispatching section k dispatches I_k trains, and the corresponding train set is $T_k = \{1, 2, \cdots, i, \cdots, I_k\}$, called the train group T_k. When the train in the dispatch section k is delayed, the delay time when the train group T_k reaches the hub station is represented by the delay vector $X_k = \{x_1, x_2, \dots . x_{I_k}\}$, X_k is used to describe the delay time of a dispatch section when the train is handed over. Generally, assuming that the dispatching section k is affected by emergencies, which lead to a long-lasting local traffic interruption, it is difficult to restore punctual driving only through the dispatch of this section and it is expected that the delay time when handing over at the junction station is more than X_k^0, as $X_k \geq X_k^0$, $x_1 \geq x_1^0, x_2 \geq x_2^0, \cdots, x_{I_k} \geq x_{I\,k}^0$. After the train group of dispatching section k is merged with the train group T_m of other dispatching section m in the hub section $m + 1$, a new train group $T_{m+1} = T_k \cup T_m$ is formed, which enters the subsequent dispatching section $m + 2, m + 3, \cdots$ working together. If there are still departure trains or final arrival trains in the hub section $m + 1$, increase or decrease correspondingly on the basis of $T_k \cup T_m$ to form a train group T_{m+1}. Since the train group T_k is delayed by more than D_k^0. If the relevant dispatching section m, $m + 1$, etc. belong to busy lines with dense traffic and small intervals, it is easy to form a conflict relationship with related trains. In order to meet the safe interval, it will affect and squeeze the operation space of other trains resulting in other trains' associated delays on the operation diagram. It is shown that the non-delayed trains also have operation delays, and cannot be operated according to the scheduled operation line resulting in the exchange of the order, which is collectively referred to as delayed propagation.

B. Local Area scheduling

When a train is delayed and deviates from the original basic schedule or daily schedule operation diagram (in this paper, it will not be distinguished, collectively referred to as the basic diagram), the dispatcher will follow the train delay, as well as the line status and equipment conditions, adjust the arrival and departure time of the affected trains at the stations under this section, formulate an adjustment schedule, and reduce delays as much as possible while ensuring the safety of the traffic, so as to restore the running order as much as possible. The process of formulating an adjustment schedule is approximately an optimization decision-making process, and the decision variable is the arrival and departure time of the train at each station. Suppose the planned arrival time of train i at station s_j in dispatch section k is a_{ij}^{k*}, and the planned departure time is d_{ij}^{k*}. Due to the delay, the actual arrival and departure time may be different from the planned time. The actual departure time is d_{ij}^k, the actual arrival time is a_{ij}^k , then $a_{ij}^k - a_{ij}^{k*}$ represents the amount of early or late arrival time, negative value is early, and positive value is late. The delay of the departure time is similarly defined. The goal of formulating an adjustment schedule is usually to reduce delays and restore punctuality. Therefore, the commonly used dispatch objective function is the total delay(TD) of all trains at all stations in the dispatch region, which can be expressed as follows:

$$f_k = \sum_{j=1}^{J_k} \sum_{i=1}^{I_k} (|a_{ij}^k - a_{ij}^{k*}| + |d_{ij}^k - d_{ij}^{k*}|) \qquad (1)$$

This TD objective function reflects the dispatch target of keeping the rescheduled train timetable as close as possible to the original timetable.

Another commonly used objective function for a dispatch region is the minimum travelling time (MTT), which is defined as:

$$c_k = \sum_{i=1}^{I_k} (d_{iJ_k}^k - a_{i1}^k) \qquad (2)$$

where implies how much delay has been increased the minimum actual travelling time. Accordingly, the planned travelling time is:

$$c_k^* = \sum_{i=1}^{I_k} (d_{iJ_k}^{k*} - a_{iJ_k}^{k*}) \qquad (3)$$

The Increment Delay (ID), which is defined as the increment amount of delay in the dispatch region when a train leaves the region. It can be defined as :

$$H_k = \sum_{i=1}^{I_k}((d_{iJ_k}^k - d_{iJ_k}^{k*}) - (a_{i1}^k - a_{i1}^{k*})) \quad (4)$$

Obviously, the increment delay in the section has the same practical significance as the travelling time of the section. The shorter the travelling time, the stronger the passing capacity of this section.

Usually, the scheduling section where the initial delay occurs should be the main purpose of restoring punctual operation, and the objective function f_k is adopted and the objective function c_k should be adopted to improve the passing capacity for the busy section.

We adopt the local area scheduling objective function (1).

The constraints as follows:

a) Minimum interval running time:

$$a_{ij}^k - d_{i,j-1}^k \geq \min\left(a_{ij}^{k*} - d_{i,j-1}^{k*}\right), i \in T_k \quad (5)$$

b) Minimum stop time:

$$d_{ij}^k - a_{ij}^k \geq \min\left(d_{ij}^{k*} - a_{ij}^{k*}\right), i \in T_k \quad (6)$$

c) departure time:

$$d_{ij}^k \geq d_{ij}^{k*} \quad (7)$$

d) Departure interval:

$$d_{ij}^k - d_{i-1,j}^k \geq min(d_{ij}^k - d_{i-1,j}^k), i \in T_k \quad (8)$$

C. Multi-region High-Speed Railway Scheduling

Scheduling under the conditions of the railway network, restoring the punctual operation of the train group T_k in the section k where a traffic conflict occurs should be the main purpose. The train group T_m of other dispatching section m in the hub section $m+1$, a new train group $T_{m+1} = T_k \cup T_m$ is formed, the number of trains increase, which mainly aimed at improving passing ability. Therefore, the scheduling objective function as follows:

$$F = \sum_k w_k f_k + \sum_m w_m c_m \quad (9)$$

where f_k is the objective function for region k as the total delay (TD) and c_m is the objective function for the subsequent dispatching region m after the region k. w_k is the weight for each scheduling region k. Constraints are the same as local area scheduling.

In this paper, we select $k = 1, m = n_k - 1$. Due to the collaborative scheduling between multiple scheduling sections, f_1 is easy to be evaluated, but c_m is a large-scale problem which is difficult to be an accurate model, and the simulation method is used to solve the problem, however, the simulation calculation is expensive, so the surrogate model is used for optimization to solve the c_m of expensive calculation.

III. Surrogate Model and Active Learning

The Gaussian process (GP) surrogate model is selected in this paper. The surrogate model built by the GP model can predict uncertainty well. For a complex high-speed railway network with many trains and stations, the number of decision variables are huge and the objective function are highly nonlinear. As a result, a huge amount of the delay vector $X = \{x_1, x_2, \dots . x_{I_k}\}$ is needed to establish the surrogate model to reach a sufficiently accuracy level. Training such as model leads to expensive calculations.

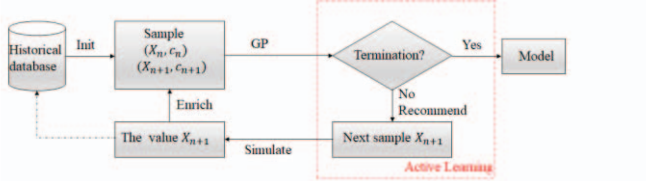

Fig.1. The process of surrogate model

Therefore, the active learning strategy is added to optimize the surrogate model. The surrogate model can be obtained with less sample data and can accurately replace the simulation for f_n evaluation. An element of the active learning strategy is to obtain the uncertainty of the model, so the GP can be more well suited to active learning strategies. Figure 1 shows the process of adding an active learning strategy to build an accurate surrogate model.

A. Gaussian Process Model With Active Learning

Let the sample of Gaussian process is denoted by (X_n, c_n), where X_n is the delay vector $\{x_1, x_2, \dots . x_{I_k}\}$ and c_n is the increment delay (4) when the train group T_{m+1} enters the subsequent dispatching section $m+2, m+3, \cdots$. The input set and output set constitute a set of random variables, the random process they form is the Gaussian process, as follows:

$$c(X) \sim G\left(m(X), \Gamma(X_r, X_t)\right) \quad (10)$$

where $m(X) = E(c(X))$ is the mean function of $c(X)$, $\Gamma(X_r, X_t)$ is the covariance function, as follows:

$$\Gamma(X_r, X_t) = \sigma^2 \exp\left[-\frac{1}{2}(X_r - X_t)^T M (X_r - X_t)\right] \quad (11)$$

where $M = diag[\lambda^{-2}]$ and $\theta = (\lambda, \sigma)$, which is solved by maximum likelihood method.

The probability density function of the Gaussian process, as follows:

$$p(X) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(X - m(X))^2}{2\sigma^2}\right) \quad (12)$$

According to the active learning(AL) strategy recommendation, select the sample X_i whose current prediction value $c(X)$ can be lower than the current observation minimum $c_{min}(X)$ to the greatest extent, who can be represented by the expected value $\mathrm{E}[I(X)]$ of the random variable $I(X)$, as follows:

$$I(X) = \begin{cases} c_{min}(X) - c(X), & c(X) < c_{min}(X) \\ 0, & c(X) \geq c_{min}(X) \end{cases} \quad (13)$$

$$\mathrm{E}[I(X)] = \int I(X) p(\, c(X)) d(\, c(X)) \quad (14)$$

Simplify as follows:

$$\mathrm{E}[I(X)] = (\, c_{min}(X) - m(L))\Phi\left(\frac{c_{min}(X) - c(X))}{\sigma}\right) + \sigma\phi\left(\frac{c_{min}(X) - c(X)}{\sigma}\right) \quad (15)$$

where Φ and ϕ are the cumulative distribution function and probability density function of the normal distribution.

B. *Algorithm Implementation*

The construction process of the surrogate model with the active learning strategy is shown in Figure 2. Continuously recommend sample X_{n+1} through active learning and use simulation to give recommended samples L_{new} increase of delay time $c(X_{n+1})$. Using simulation to continuously give recommended samples $(X_{n+1}, c(X_{n+1}))$ to enrich the sample data to update the surrogate model, which solves the demand of more samples. Using fewer samples, which means less computational cost, builds the surrogate model which can accurately give the value c and participate in railway network optimization, speeds up railway network optimization.

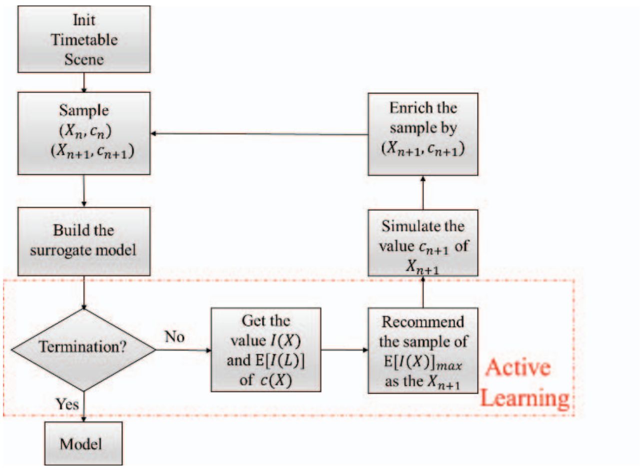

Fig.2. The surrogate model with the active learning strategy

IV. Experiment

In order to verify whether the surrogate model of the active learning strategy in this paper can accurately replace the high-speed railway network simulation and participate in the optimization and adjustment of trains in the network, 16 high-speed railway trains in Shenyang North-Changchun, as shown in Figure 3, are selected as follow-up dispatchers The section is the scenario where the surrogate model replaces the simulation evaluation of the simulator, and the evaluation index is increment delay of the scheduling section.

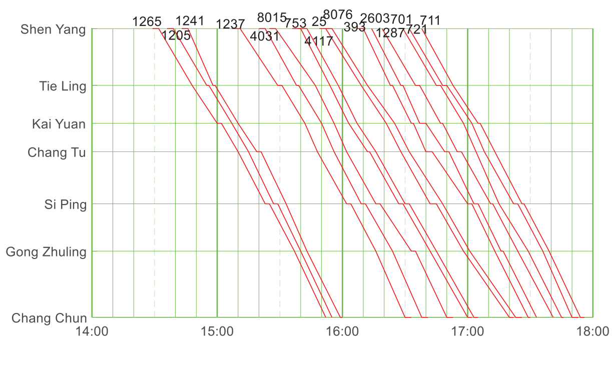

Fig.3. Train plan operation diagram

A. *Scene 1*

When two of the trains entered the area were delayed, the train 1237 was in the late range of Shenyang North Station (5, 30), and the train 4031 was in the late range of Shenyang North Station (5, 30). To verify the accuracy of the model, the two delayed real target value of all the train delays is obtained by the simulator, a total of 676 sample data, the map of the real target value is shown in Figure 4, the minimum value of the delay increment is 0.

Extract 25 initial delay vectors evenly at equal intervals, and update the model 3 times according to the active learning strategy. Finally, the more accurate prediction results of the surrogate model are obtained. The projection map is shown in Figure 5.

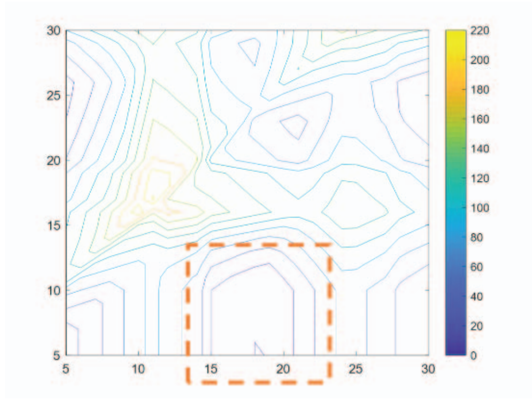

Fig.4. The map of the real target value

According to Figures 4 and 5, the overall effect of the final model is better. The dotted line in the figure is the area where the optimal delay is brightened. The results are basically the same, which can accurately replace the simulator in railway network optimization adjustments.

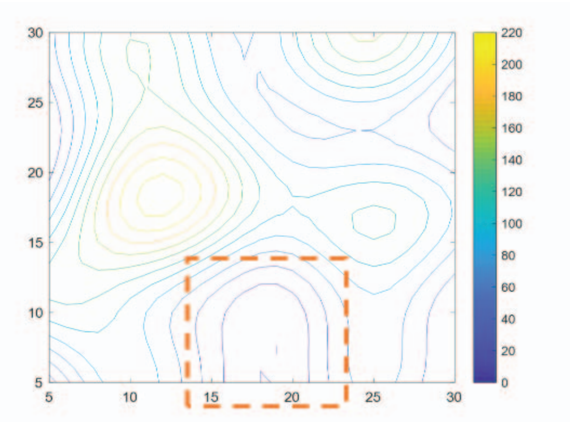

Fig.5. The projection map of prediction results

B. *Scene 2*

When four of the trains entered the area were delayed, train 1265 was in the late range of Shenyang North Station (50,70), train 1241 was in the late range of Shenyang North Station (45,65), and train 1205 was in the late range of Shenyang North Station (30,50), train 1237 is in the late range of Shenyang North Station (15,35). To verify the accuracy of the model, the four delayed real target value of all the train delays is obtained by the simulator, a total of 198841 sample data, the minimum value of the delay increment is 18. 5929 sample data are uniformly sampled, and the model is updated 36 times through active learning. The final output of the surrogate model and the real simulator output are shown in Figure 6. It can be seen that the surrogate model predicts the results near all minimums unanimously.

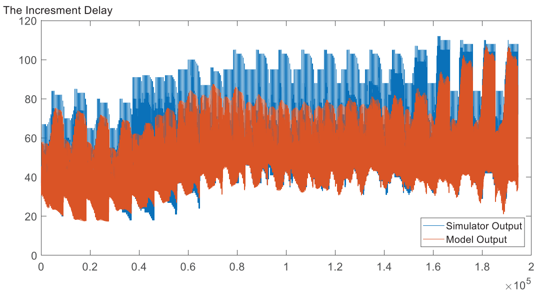

Fig.6. The comparison diagram

The changes in the root mean square error (RMSE) of the model during the iteration process are shown in Figure 7, and the change in the minimum value of the delay increment predicted by the model output is shown in Figure 8.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(F_i(L_i) - L_i)^2} \quad (16)$$

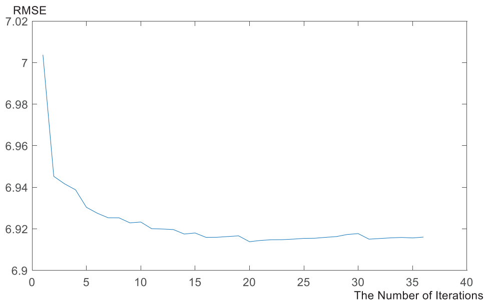

Fig.7. RMSE of the model during the process

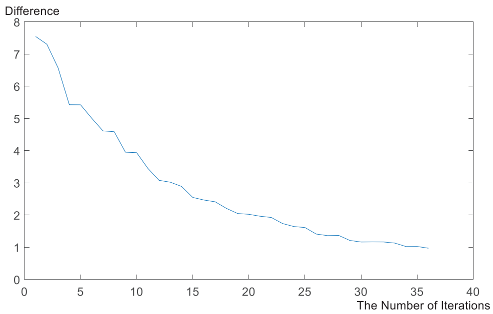

Fig.8. The change of difference by actual and forecast of the increment delay

With the active learning of the model, the overall error of the model prediction data is shrinking. At the same time, the minimum delay increment predicted by the model is gradually approaching the true minimum delay increment. The final minimum delay increment predicted by the model is 17.1, but the position of the minimum delay increment predicted by the model is consistent with the real data, which can ensure that we get the optimal delay vector in the optimization process.

When we use the Gauss Process model without active learning(AL), we can obtain the same consequence compared with the active learning model, though, the number of samples and the time to build the model will be very different. The comparison results are shown in the following table I. The distance represents difference between the optimal solution position of the prediction model and the actual optimal solution. From the table, we can see that it takes more time without AL if we want to get the same consequence.

TABLE I. MODEL COMPARISON

Model Type	*Index*		
	The number of samples	***Time***	***Distance***
With AL	5964	119.28 hours	0
Without AL	8549	170.98 hours	0

Therefore, the surrogate model with active learning can have higher prediction accuracy near the optimal position and can reduce the cost of simulation calculations significantly.

CONCLUSION

We propose a surrogate model of an active learning strategy to replace the simulation evaluation of the simulator, which effectively reduces the cost of simulation, and lays the foundation for the next adjustment of train operation in the railway network. The active learning strategy adopted in this paper makes the prediction results around the best points more accurate to ensure the adjustment of train operation in the railway network, but the accuracy of all prediction data cannot be guaranteed. Therefore, the active learning strategy still needs further improvement.

REFERENCES

[1] PENG Qiyuan, WEN Chao, LUO Jian. Dispatching Scheme for HighSpeed Passenger Lines in China[J]. Journal of Southwest Jiaotong University, 2006, 41(5):541-548.

[2] Cai X , Goh C J . A fast heuristic for the train scheduling problem[J]. Computers & Operations Research, 1994, 21(5):499-510.

[3] XIA Ming, ZHOU Lei- shan, SUN Qi, WANG Yong- ming. Study on the Train Operation Adjustment on Double- track Line Based on Ant Colony Algorithm[J]. logistics technology, 2008, 27(006):61-64.

[4] Meng L , Zhou X . Simultaneous train rerouting and rescheduling on an N-track network: A model reformulation with network-based cumulative flow variables[J]. Transportation Research Part B: Methodological, 2014.

[5] Fang W , Yang S , Yao X . A Survey on Problem Models and Solution Approaches to Rescheduling in Railway Networks[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(6):2997-3016.

[6] Masoud S , Abdolreza S , Amin J . A Multi-Objective Decision Support System for Real-Time train Rescheduling[J]. IEEE Intelligent Transportation Systems Magazine, 2018:1-1.

[7] Chang C S , Thia B S . Online rescheduling of mass rapid transit systems: Fuzzy expert system approach[J]. IEE Proceedings - Electric Power Applications, 1996, 143(4):307-316.

[8] Yang C , Ding J , Jin Y , et al. Offline Data-Driven Multiobjective Optimization: Knowledge Transfer Between Surrogates and Generation of Final Solutions[J]. IEEE Transactions on Evolutionary Computation, 2020, 24(3):409-423.

[9] Wang Ling, Ji Lijun, Zheng Dazhong. Simulation optimization based on surrogate model and genetic algorithm[J]. Control and Decision, 2004, 19(006):626-630.

[10] Liu B , Zhang Q , Gielen G G E . A Gaussian Process Surrogate Model Assisted Evolutionary Algorithm for Medium Scale Expensive Optimization Problems[J]. IEEE Transactions on Evolutionary Computation, 2014, 18(2):180-192.