Research of the Application of Cargo Volume Forecasting and Simulated Annealing Scheduling Optimization Based on Multi-model Integration in Logistics Centers

Xintao Xu[1, *], Kuiyan Men[2], Yingqiao Li[2], Chuansen Zhang[1] and Hang Su[2]

[1] Department of Civil and Architectural Engineering, Qingdao University of Technology (Linyi), Linyi, China

[2] Department of Management Engineering, Qingdao University of Technology (Linyi), Linyi, China

*Corresponding author:

Abstract. SARIMA model and the XGBoost model are both classical time series forecasting tools, while the simulated annealing algorithm (SA) is an excellent multi-objective planning method. By combining these algorithms, this study aims to forecast the cargo volume of a logistics center and arrange the scheduling of the logistics center accordingly. First, the SARIMA forecasting, and XGBoost hybrid correction model can more accurately predict the cargo volume of this logistics center for the next 30 days. Subsequently, the study will analyze and determine the decision variables and constraints in scheduling planning. Finally, the planning problem will be solved by SA, which enables the minimum number of personnel to meet cargo volume requirements, scheduling constraints, and other constraints. The results of this study have significant practical value, providing a replicable forecasting and planning framework for the dynamically changing market environment, which effectively helps enterprises to anticipate market changes and optimize resource allocation to enhance market competitiveness.

Keywords: the SARIMA model; XGBoost model; Simulated Annealing algorithm (SA); optimize resource allocation.

1. Introduction

In modern logistics management, accurate cargo volume forecasting is a key factor in improving operational efficiency and resource allocation. To meet this challenge, this study firstly adopts SARIMA and XGBoost hybrid correction model with a view to achieve more accurate cargo volume forecasting. After that, multi-objective scheduling planning is performed with the help of SA. This study not only provides a scientific decision support tool for logistics centers, but also this innovative approach will provide an effective solution for the market resilience and resource allocation of enterprises.

2. Data collection and pre-processing

2.1 Data collection

To predict the volume of shipments for the next 30 days and specific hours, this paper collects data on the volume of shipments at a sorting center from August 2023 to December 2023 and sorts the raw data by date and date hour using Python software. To make it easy to handle time series data.

2.2 Filling of missing data

In the process of sorting for time ordering, some of the data were missing. Based on finding out the missing values, we fill the missing values for different dates by Lagrange difference filling method.

First, the Lagrange interpolation polynomial is constructed, based on the known data points, the Lagrange polynomial $L(x)$.

$$L(x) = \sum_{i=0}^{n} y_i l_i(x) \tag{1}$$

Next, compute the basic Lagrange polynomials:

$$l_i(x) = \prod_{j=0,j\neq i}^{n} \frac{x-x_j}{x_i-x_j} \tag{2}$$

Finally, the missing value locations are identified, and the missing values are calculated to obtain the complete time series data.

Through the pivot table analysis found that there is a significant surge in the futures volume during holidays, double eleven, etc., in order not to affect the prediction results, this paper chooses to remove the trend of the time series by one time difference method. After the difference of the data in the ADF test P value from 0.312 to 0.042 can be considered as a smooth series, as shown in Fig .1.

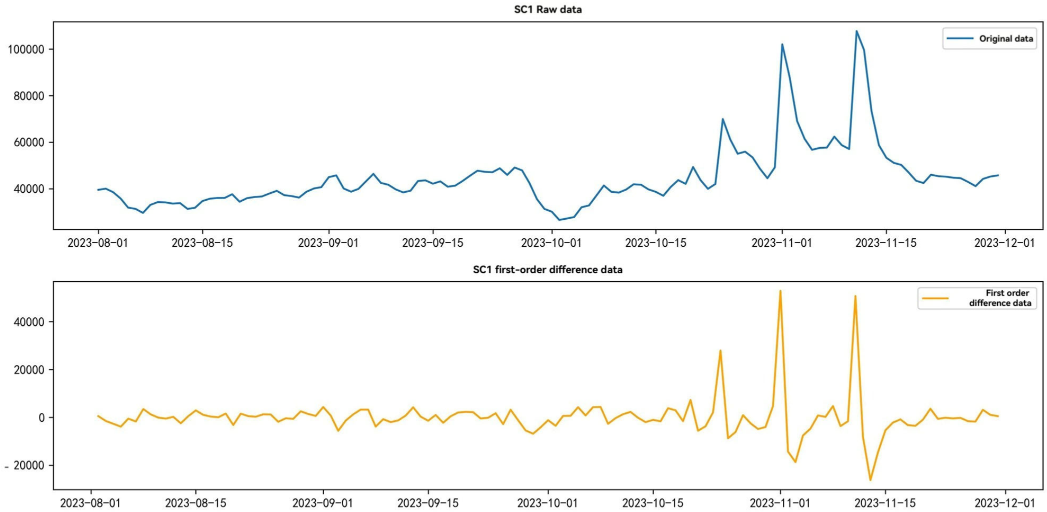

Fig. 1 First-order difference processing of data

3. Cargo volume prediction model based on a mixture of SARIMA and XGBoost

3.1 SARIMA model construction

Considering the characteristics of the dataset in this paper, the single use of time series forecasting model will lead to too fast convergence of the predicted values, so this paper introduces the XGBoost model and establishes a hybrid forecasting model of SARIMA and XGBoost.

SARIMA (Seasonal Autoregressive Integrated Moving Average) time series model is a powerful tool for analyzing and forecasting time series data with seasonal trends. The model contains several parameters that together determine the structure and performance of the model [1-4]. p (autoregressive order): indicates how many past values (i.e., lagged values) the current value of the time series is weighted against to form a forecast. These weights are determined by the autoregressive coefficients.d (difference order): Difference is used to remove trend and seasonality from the time series to make the data smoother. The difference order d indicates how many times the difference operation is performed on the data.q (moving average order): the moving average component is used to model the errors in the time series, assuming that the current error is related to errors from previous periods. q indicates how many past error terms are used in the moving average model.

First, hourly forecasts of cargo volumes are made using the SARIMA model in time series analysis:

$$\left(1-\sum_{j=1}^{p}\phi_j L^j\right)(1-L^s)^d v_{i,h}=\left(1+\sum_{j=1}^{q}\theta_j L^j\right)\dot{\boldsymbol{O}}_{i,h} \tag{3}$$

Where p and q are the orders of autoregression and moving average, respectively, and s, d are the orders and periods of seasonal differentials.

According to the $p.d.q$ parameters set in the SARIMA model of this paper, this paper obtains the daily and hourly cargo volume data of this sorting center for the next 30 days and generates the following prediction graphs as shown in Fig .2.

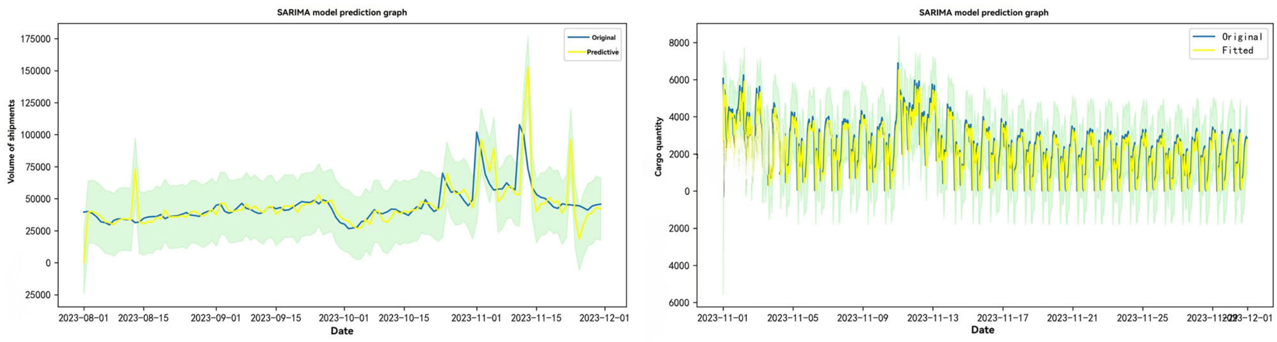

Fig. 2 Forecasted cargo volume per hour and per day

3.2 XGBoost Hybrid Correction

To test the adaptability of the model, this paper fitted the original data with the predicted data and found that the single use of the SARIMA model would lead to the rapid convergence of the predicted future values, which may be more inconsistent with the actual situation. Therefore, to make up for the shortcomings of the SARIMA model prediction, this paper also introduces the XGboost model to predict the amount of goods more accurately per day and per hour in each sorting center for the next 30 days by means of a weighted mixed model [5,6].

XGBoost is an optimized machine learning algorithm that builds multiple basis functions and combines them to form a synthetic algorithm that fits the data well, is capable of handling large-scale datasets and complex models, and also performs well in preventing overfitting and improving generalization. In this paper, based on the establishment of the SARIMA model, the XGBoost model is used to mix and supplement. After the XGBoost model processing, the prediction results obtained, as shown in Fig .3:

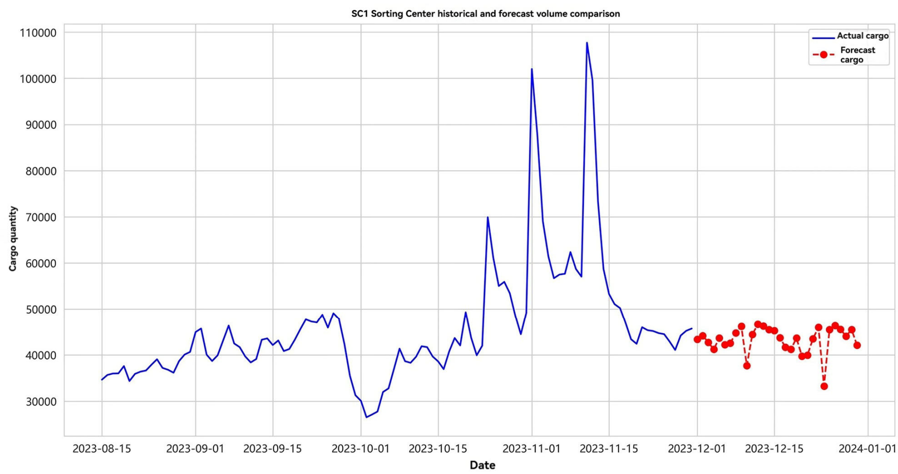

Fig. 3 XGBoost model prediction results

To make up for the shortcomings of a single model and improve the prediction accuracy or stability, so as to obtain more accurate results, this paper carries out a weighted hybrid calculation by assigning different weights to the data of XGBoost and SARIMA models. The details are as follows:

$$Hybrid_t = w_{XGB} \times XGB_t + w_{SARIMA} \times SARIMA_t \tag{4}$$

Where, $Hybrid_t$ is the predicted value of hybrid model at moment $t.XGB_t$ denotes the predicted value of XGBoost model at time t. $SARIMA_t$ denotes the predicted value of SARIMA model at time t. w_{SARIMA} denotes the weight of SARIMA model.

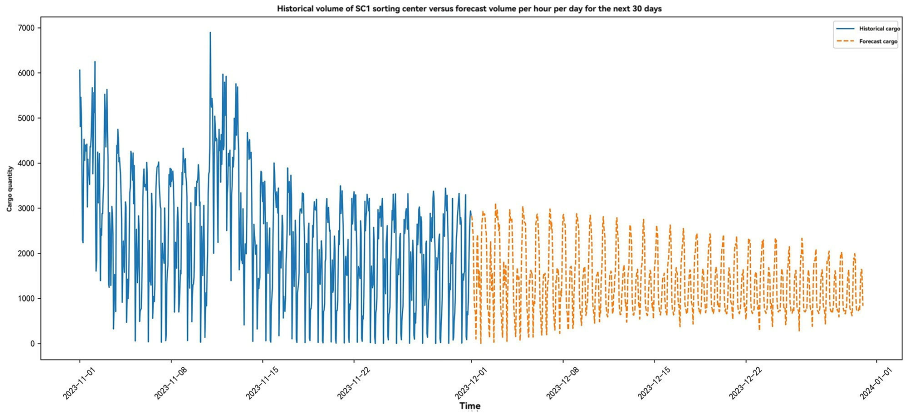

Fig. 4 Hybrid model prediction results

As can be seen in Fig .4, historical volumes show some volatility and have significant peaks in certain time periods. The predicted volume, on the other hand, is predicted based on the historical volume, which is similar to the trend of the historical volume on the whole but differs in specific values. Meanwhile, the volatility of the predicted volume is relatively small, which is since we reduce noise and randomness in building the model.

At the same time, we first take the shipment volume of each sorting center for three months as the data source to establish the resulting time series prediction, which can be seen from the above table, the prediction results show a certain volatility, but the overall trend is upward. This may mean that this time series data may continue to grow in the coming period. In addition, the blue curve shows large fluctuations in some time periods, which reflects the sensitivity of the model to short-term events or seasonal changes. These fluctuations may be caused by some temporary factors, such as promotional activities, holidays, etc., and its prediction results are in line with the actual situation and the prediction effect is better. Choose to call the Pandas library to read the residual data file, and then use the matplotlib library to visualize the data, which indicates that the predicted value of the model is basically in line with the actual value, proving that the robustness of the model is better.

4. Scheduling Planning Model Based on Simulated annealing algorithm(SA)

4.1 Analysis of constraints

In scheduling optimization of a sorting center, the core objective of the optimization is to minimize the cost of headcount while ensuring that the volume processing is done efficiently. Assuming that there are 200 regular workers in the sorting center, the following constraints clearly exist: 1) Ensure that the daily volume can be processed efficiently; 2) Minimize the total number of man-days required to make human resources utilization more efficient; 3) Pay attention to the balance of the actual hourly manpower efficiency per day; 4) The attendance rate of each regular worker does not exceed 85% and the number of consecutive days of attendance does not exceed 7 days [5-8].

4.2 Defining Decision Variables

Set up a binary variable , where $i = 1,2,\ldots,200$ denotes the ith it regular worker, $j = 1,2,\ldots,6$ denotes six shifts, and $t = 1,2,\ldots,30$ denotes the next 30 days. If the ith regular worker attends the ith shift on day t, then $x_{ijt} = 1$; otherwise $x_{ijt} = 0$. Set up the integer variable y_{tj}, which denotes the number of temporary workers required for the jth shift on day t.

$$min \sum_{t=1}^{30} \sum_{j=1}^{6} \left(\sum_{i=1}^{200} x_{ijt} + y_j \right) D_{tj} \tag{5}$$

To equalize the actual hourly ergonomics of each day as much as possible, a penalty term can be introduced to measure the sum of the squares of the deviations of the actual hourly ergonomics of each shift from the average hourly ergonomics of each day. The objective can be expressed as follows: H_{tj} denotes the actual hourly productivity of the jth shift on day t and avg_H denotes the average hourly productivity of all shifts:

$$min \sum_{t=1}^{30} \sum_{j=1}^{6} \left(H_{tj} - avg_H \right)^2 \tag{6}$$

To equalize the attendance of regular workers as much as possible, a penalty term is also introduced, which measures the sum of the squares of the deviations of each regular worker's attendance from the average attendance. Let the attendance rate of the ith worker be O_i and avg_O denotes the average attendance rate of all workers. The objective can be expressed as follows:

$$min \sum_{i=1}^{200} \left(\frac{\sum_{t=1}^{30} \sum_{j=1}^{6} x_{ijt}}{30}_i - avg_0 \right)^2 \tag{7}$$

Combining the above three objectives, a weighted composite objective function is constructed:

$$min\, Z = w_1 \sum_{t=1}^{30} \sum_{j=1}^{6} \left(\sum_{i=1}^{200} x_{ijt} + y_{tj} \right) D_{tj} + w_2 \sum_{t=1}^{30} \sum_{j=1}^{6} \left(H_{tj} - avg_H \right)^2 + w_3 \sum_{i=1}^{200} \left(\frac{\sum_{t=1}^{30} \sum_{j=1}^{6} x_{ijt}}{30} - avg_O \right)^2 \tag{8}$$

4.3 Constructing planning constraints

No more than 85% attendance per regular worker:

$$\frac{\sum_{t=1}^{30} \sum_{j=1}^{6} x_{ijt}}{30} \leq 0.85 \quad \forall i = 1,2,\ldots,200 \tag{9}$$

No more than seven consecutive days of attendance per regular worker. This is achieved by introducing additional logical constraints or by using special mathematical models (e.g., recurrent integer programming), which are simplified here:

$$\text{Ensure absence of 7 consecutive days or more} = 1 \quad \forall i, j \tag{10}$$

The manpower (the sum of regular and temporary workers) allocated to each shift each day should be able to meet the cargo handling requirements of that shift on that day:

$$\sum_{i=1}^{200} x_{ijt} + y_{tj} \geq D_{tj} \quad \forall t, j \tag{11}$$

Variable value range:

$$\begin{cases} x_{ijt} \in \{0,1\} \forall i, j, t \\ y_{tj} \in Z \forall t, j \end{cases} \tag{12}$$

4.4 Model solving and analysis

Simulated Annealing (SA) algorithm is a heuristic global optimization method inspired by the annealing process of solid materials [7-8]. In physics, annealing is the process of heating a metal to a high temperature so that its internal particles are in a highly disordered state, and then cooling it slowly so that the particles are gradually ordered and eventually reach a stable structure with lower energy. SA draws on this natural phenomenon and applies it to solve complex optimization problems, especially those with many feasible solutions and nonlinear, nonconvex objective functions [9-12]. Its specific processing flow is shown in Fig .5.

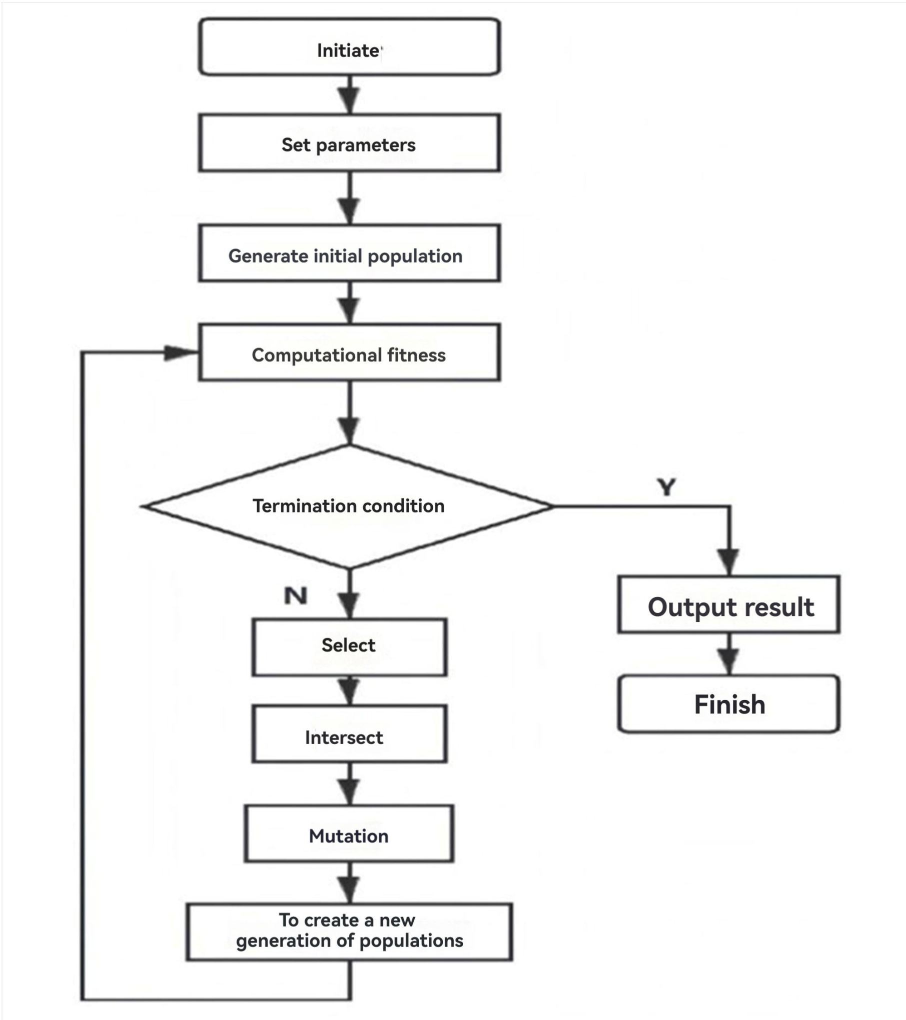

Fig. 5 The sensitivity of the model

4.4.1 Iterative process

Starting from the current solution, a random perturbation (e.g., operations such as swapping, flipping, adding noise, etc.) produces a candidate solution that is different from the current solution. These perturbations are designed to explore neighboring regions in the solution space. The difference between the objective function (or energy) values of the new solution and the current solution is compared and denoted as $\triangle E$. The probability of accepting the new solution is computed according to the Metropolis acceptance criterion:

$$P(accept) = exp(-\frac{\Delta E}{T}) \tag{13}$$

Where $T$is the current temperature. If the new solution is better than the current solution ($\triangle E < 0$), the new solution is accepted; if the new solution is inferior ($\triangle E > 0$), the decision of acceptance is based on this probability. If accepted, the new solution is taken as the current solution; otherwise, the current solution is kept unchanged. Regardless of acceptance, the optimal solution record is updated, and if the new solution is better than the known optimal solution, the optimal record is updated with the new solution.

4.4.2 Cooling process

We are given the initial temperature T_0 and set the initial temperature as 1. The cooling coefficient is selected a , and after each iterative process, the cooling temperature T_1 is obtained, where the cooling equation is $T_l = aT_{l-1}$. After several transfers at temperature T_l, a new cooling temperature is obtained, i.e., $T_{i+1} < T_i$. and the cooling process is repeated at the new temperature, constantly searching for a new solution and alternating with the slow reduction of temperature. The optimal result of the problem is finally obtained.

4.4.3 Ending process

End condition. We choose the termination temperature as $T_{end} = 10^{-50}$ (when the temperature drops to T_{end}, the simulated annealing process is judged to be finished, and then the output solution is the global optimal solution.

The initial temperature T_0 and the termination temperature T_{end} are selected, and the annealing process selects the appropriate cooling coefficients a and let the error acceptance range be 0.005. The solution is solved by using Python. Through SA, we obtain the optimized scheduling table, as shown in Table 1:

Table 1. Scheduling optimization prediction results

sorting center	date	shift	regular worker	temporary worker
SC1	2023/12/1	00:00-08:00	12	212
SC1	2023/12/1	05:00-13:00	17	192
SC1	2023/12/1	08:00-16:00	27	199
SC1	2023/12/1	12:00-20:00	25	196
SC1	2023/12/1	14:00-22:00	16	203
SC1	2023/12/1	16:00-24:00	24	214
SC1	2023/12/2	00:00-08:00	13	199
SC1	2023/12/2	05:00-13:00	16	158
SC1	2023/12/2	08:00-16:00	10	162
SC1	2023/12/2	12:00-20:00	10	153
SC1	2023/12/2	14:00-22:00	48	158

5. Summary

In summary, this study successfully constructed an efficient and accurate cargo volume prediction framework by combining the SARIMA model and the XGBoost model. Based on the predicted cargo volume, the simulated annealing algorithm (SA) is used to realize multi-objective scheduling planning, which enables the minimum number of personnel to meet cargo volume requirements, scheduling constraints, and other constraints, achieving the effect of rationalizing resource allocation. The results show that the integrated model can significantly improve the operational efficiency and market competitiveness of logistics enterprises and provides a useful reference for similar forecasting and planning problems in other industries.

References

[1] Song Ying-Hua,Xu Ya-An,Zhang Yuan-Jin. Research on seasonal PM2.5 prediction based on SARIMA-SVM model[J/OL]. Computer Engineering:1-11[2024-04-13]. https://doi.org/10.19678/j.issn.1000-3428.0068372.

[2] Xing Zhi-Wei, Li Xue-Zhe, Luo Qian, et al. Airport cargo volume prediction based on SARIMA and RBF neural network[J]. Journal of Civil Aviation University of China,2016,34(05):51-55.

[3] Wang Ying, Han Bao-Ming, Zhang Qi, et al. Beijing subway inbound passenger flow prediction based on SARIMA model[J]. Transportation Systems Engineering and Information,2015,15(06):205-211.DOI:10.16097/j.cnki.1009-6744.2015.06.031.

[4] Zheng-G. Container cargo volume prediction based on gray model[J]. China Navigation,2014,37(02):118-121.

[5] Li Zhan-Shan, Liu Zhao-Gang. Feature selection algorithm based on XGBoost[J]. Journal of Communication,2019,40(10):101-108.

[6] Ye Qian-Yi. Research on physical retail sales forecasting based on Xgboost method[D]. Nanchang University,2016.

[7] Chen Xiao-Hong, Hú Wen-Hua, Cao Yu,et al. Application of hierarchical multi-objective linear programming model based on trapezoidal fuzzy number in multi-attribute uncertain decision-making problems[J]. Journal of Management Engineering, 2012, 26(4):7. DOI:10.3969/j.issn.1004-6062.2012.04.026.

[8] Luo Pei-Wen, Xiong Xiao-Yu, Liu Bing-Lin. Forest carbon sequestration analysis and multi-objective planning management strategy based on machine learning and growth prediction[J]. Science and Technology Wind,2023,(19):156-159.DOI:10.19392/j.cnki.1671-7341.202319052.

[9] Zhao D. Research on site selection of highway emergency logistics center based on network robustness and multi-objective optimization model [D]. Shijiazhuang Railway University,2024.DOI:10.27334/d.cnki.gstdy.2024.000022.

[10] Tao Sha, Liu Yang, Zhou Jing. Integrated optimization of project scheduling and staffing considering quality transfer[J]. Journal of Systems Management,2024,33(02):356-367.

[11] Wu Yao-Hua, Wang Sheng-Li, Wang Chang-Xiang. Application of queuing theory in logistics planning[J]. Logistics Science and Technology,2004, (05):53-55.

[12] Yi H. Research on enterprise logistics cost control based on competitive strategy [D]. Beijing Jiaotong University,2012.