Research on Cargo Volume Prediction and Personnel Scheduling in Logistics Centers

Dongping Sheng, Zhongyuan Ma, Jie Yang, Hongliang Wang, Chun Su*

Changzhou Institute of Technology, Changzhou, China

ABSTRACT

With the rise of e-commerce logistics networks, cargo volume prediction in sorting centers has become increasingly important and a key research topic. The aim of this study is to predict the daily cargo volume and daily cargo volume of 57 sorting centers in the future, in order to optimize the human resource allocation of logistics network sorting centers. LSTM model, ARIMA model, and integer programming model based on genetic algorithm were established according to the different problems. Regarding problem one: In the task of predicting daily and hourly cargo volume for the next 30 days, the ARIMA model is used. The ARIMA model has shown advantages in short-term forecasting and has been applied to predict the hourly cargo volume of each sorting center within 30 days. By combining these two models, the final prediction data can be obtained. Regarding problem two: Draw a directed graph to visually display the data and provide visual support for subsequent analysis. Based on the average transportation volume of sorting centers and changes in transportation routes over the past 90 days, use a mathematical model to calculate the rate of change in cargo volume after changing transportation routes. For sorting centers that have not been affected, sum the rate of change and calculate the average. Finally, use mathematical models to calculate the revised predicted daily and hourly cargo volumes. Regarding problem three: constructing an integer linear programming model based on genetic algorithm, whose decision variables involve the attendance of formal and temporary workers in different shifts of each sorting center. In the model construction, strict constraints were set to ensure the completion of cargo volume processing, prioritize the use of formal workers, and maintain a balance of actual hourly labor efficiency. Use a linear programming solver to accurately calculate the optimal attendance allocation for each shift in each sorting center. Regarding problem four: Taking SC1 sorting center as an example, and based on the prediction results of problem two, a shift attendance plan model for formal and temporary workers was constructed. On this basis, this study aims to minimize the total number of formal and temporary workers as the objective function, and considers the constraint conditions that the attendance rate of each formal worker does not exceed 85% and the continuous attendance days do not exceed 7 days. I wrote code to solve the model and successfully obtained an optimized scheduling plan.

Keywords: LSTM model, ARIMA model, integer programming, genetic algorithm, linear programming

1 INTRODUCTION

The efficient operation of e-commerce logistics networks depends on the synergistic effects of their internal links. In the simplified architecture of the network, the sorting center plays a core role, responsible for classifying packages according to predetermined paths and transporting them to downstream nodes to ensure that packages can be accurately and accurately delivered to end consumers [1], Given the strategic position of sorting centers in logistics networks, improving their management efficiency has a profound impact on optimizing the overall network's performance and operating costs [2].

In the context of e-commerce logistics networks, cargo volume prediction in sorting centers has become a key research topic [3]. Accurately predicting the cargo volume of the sorting center is crucial for subsequent resource allocation and decision-making [4]. Once managers can grasp the cargo volume information of each sorting center in advance for a certain period of time in the future, they can make targeted resource allocation, thereby achieving maximum operational efficiency [5]. Generally speaking, the main objectives of cargo volume prediction include predicting the daily cargo volume of each sorting center based on historical data and logistics network configuration information, as well as using historical hourly cargo volume data to predict the hourly cargo volume of each sorting center [6]. In addition, the cargo volume prediction of the sorting center is closely related to its transportation routes in the logistics network. By conducting in-depth analysis of the flow of goods along various transportation routes, the network connectivity between sorting centers can be revealed, providing strong support for cargo volume prediction [7]. When the logistics route is adjusted, the impact of the adjustment information on the cargo volume prediction of the sorting center should be considered to ensure the accuracy of the prediction results [8].

Based on the cargo volume prediction results of the sorting center, personnel scheduling and scheduling have become important issues that need to be addressed in the future [9]. The human resources of the sorting center are mainly composed of formal workers and temporary workers. Regular workers are usually long-term employees who possess high work efficiency and stable job performance [10]. Temporary workers are recruited based on the actual demand for goods, and their numbers can be flexibly adjusted. However, compared to regular workers, temporary workers have lower work efficiency and higher employment costs. Therefore, in terms of personnel arrangement, it is necessary to fully consider the predicted cargo volume and the characteristics of formal and temporary workers, in order to maximize work efficiency and minimize labor costs [11].

Overall, for the efficient operation of e-commerce logistics networks, the following four problems can be solved:

(1) Work 1: Based on the daily cargo volume data of 57 sorting centers in Annex 1 for the past 4 months and the hourly cargo volume of 57 sorting centers in Annex 2 for the past 30 days, predict the daily and hourly cargo volume for the next 30 days;

(2) Work 2: Analyze the impact of transportation routes on the cargo volume of each sorting center based on the average cargo volume of each transportation route between each sorting center over the past 90 days in Attachment 3. Then, based on the route adjustment, adjust the cargo volume data of the 57 sorting centers. Finally, use the modified data from Attachment 1 and Attachment 2 to predict the daily and hourly cargo volume of 57 sorting centers for the next 30 days;

(3) Work 3: Based on the prediction results of work 2, establish a model to determine the attendance of each shift in each sorting center for the next 30 days, and record the results in Result Table 5. On the premise of meeting the daily demand for handling goods, the total number of person days should be minimized as much as possible, while maintaining a balance of actual hourly human efficiency every day. Specifically, there are 60 formal workers in each

sorting center, and priority should be given to using these formal workers. If additional personnel are needed, temporary workers should be used;

(4) Work 4: Regarding the scheduling problem of SC1 sorting center, we need to build a model based on the prediction results of work 2 to develop a detailed shift attendance plan for 200 formal and temporary workers in the next 30 days. The plan must follow the following guidelines: the attendance rate of formal workers shall not exceed 85%, and continuous attendance shall not exceed 7 days; On the premise of ensuring that the daily cargo volume is processed, the number of people and days should be optimized; Maintain a balance of daily hourly human effects; The attendance rate of formal workers also needs to be balanced.

2 RELATED WORK

2.1 Work one

To solve problem one, which is to predict the daily and hourly cargo volume of 57 sorting centers for the next 30 days, we need to first determine the target cargo volume and related constraints. Considering that the prediction of arrival volume involves different time periods, a suitable model should be selected for solution. The specific steps are as follows: Firstly, visualize the original cargo volume data and analyze the relationship between historical cargo volume and time. Through observation, we found that the cargo volume data may have time series dependence and long-term trends. Therefore, we chose the Long Short Term Memory Network (LSTM) deep learning model to predict the cargo volume of 57 sorting centers in the next 30 days. The LSTM model can capture the long-term dependency relationship between historical cargo volume and time, effectively transmit and express information in long-term time series, and predict future cargo volume. Secondly, for the prediction of hourly cargo volume, data visualization shows a periodic relationship between historical cargo volume and time. In addition, holidays or external factors can lead to shorter time scales for hourly changes in cargo volume. Therefore, the auto regressive moving average model (ARIMA) can be chosen for prediction. The ARIMA model can predict future short-term cargo volume based on seasonal changes and trends in historical cargo volume. By adjusting ARIMA parameters, the accuracy of predicting hourly cargo volume for the next 30 days can be improved.

2.2 Work two

Work 2 provides data on the average cargo volume and future changes in transportation routes between sorting centers over the past 90 days. Based on the content of Annex 4, the sorting center that is transported to the destination can be used as the center point, and the cargo volume can be analyzed to determine the changes in cargo volume after the transportation route is changed. Since the total amount of goods transported remains constant, it can be determined that the missing goods have been allocated to other routes, and the rate of change in the amount of goods on the route can be calculated based on this. Then, for the first work, the predicted data can be multiplied by the corresponding rate of change to use as the corrected prediction data. For those sorting centers that are not affected, the rate of change can be summed and averaged to calculate the corrected prediction result based on the comprehensive rate of change.

2.3 Work three

Work three is an integer programming problem. In this problem, it is necessary to reasonably arrange the attendance of each sorting center for each shift based on the demand for goods in the next 30 days, in order to meet the demand for goods while minimizing the total number of person days and maintaining the actual hourly efficiency as balanced as possible. Firstly, prepare the data obtained from work two and visualize it, which can visually observe the increase or decrease in monthly cargo volume, so that we can arrange the number of employees in the future. Secondly, based on the known conditions extracted from the work, determine the constraint conditions, design the objective function, and ensure that the number of arranged person days is as small as possible and the actual hourly human efficiency is balanced, thus establishing an integer programming model. Finally, perform corresponding calculations to obtain a reasonable allocation plan.

2.4 Work four

This issue requires studying the scheduling problem of specific sorting centers. Given the prediction results of work two, the aim is to develop a detailed shift attendance plan for the next 30 days for 200 employees (including regular and temporary employees) at SC1 sorting center. The plan must strictly adhere to the following guidelines: the attendance limit for formal employees is 85%, and the number of consecutive working days must not exceed seven days; Optimize the number of people and days while ensuring daily cargo handling volume; Maintain a daily balance of hourly productivity; At the same time, ensure that the attendance rate of formal employees is evenly distributed. To achieve the above goals, we adopted integer programming method, established key decision variables, and constructed corresponding constraint systems. By developing a MATLAB solution program, we can obtain a scheduling plan that is both efficient and meets all established standards.

3 MODEL ASSUMPTION AND BUILDING

3.1 Model assumption

In this study, we constructed an analytical framework and advanced the research work based on the following assumptions:

(1) We assume that the analyzed historical cargo dataset does not contain outliers, ensuring that all data is the true output of business activities;

(2) Given the time series characteristics of historical cargo data, we assume that it contains both cyclical and seasonal patterns, which will provide a basis for our trend analysis and prediction;

(3) In the context of considering route adjustments, we assume that the cargo volume corresponding to the braking route will be reasonably allocated based on its proportion in the overall transportation volume;

(4) We assume that employees work for a fixed eight hours per day, excluding any form of rest or other potential changes in working hours.

3.2 Problem one model establishment and solution

3.2.1 Data preprocessing

Firstly, visualize the historical cargo volume of different sorting centers, perform zero interpolation on missing values of hourly cargo volume, and do not remove abnormal cargo volume values as they are considered real data. The visualization of some cargo volume data is as follows:

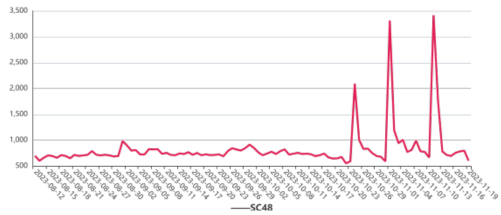

Fig.1: Changes in cargo volume at SC48

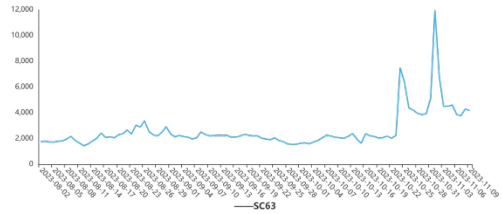

Fig.2: Changes in cargo volume at SC63

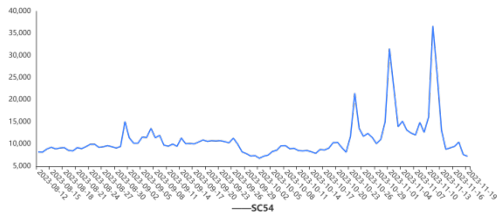

Fig.3: Changes in cargo volume at SC54

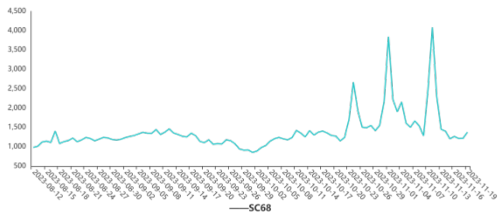

Fig.4: Changes in cargo volume at SC68

As can be seen from the above figure, the trend of cargo volume changes in some sorting centers shows an overall upward trend. The majority of the cargo volume changes in sorting centers are mainly concentrated in the two time periods after October 1st and before November 11th. The cargo volume changes before October 1st are relatively stable. There are also a few sorting centers where the trend of cargo volume changes is not very obvious, such as SC14 and SC16.

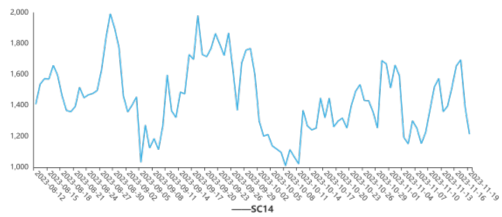

Fig.5: Changes in cargo volume at SC14

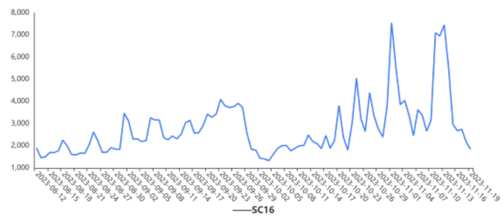

Fig.6: Changes in cargo volume at SC16

Summarize the historical cargo volume of 57 sorting centers and analyze the overall trend of daily cargo volume. The overall trend is shown in Fig.7.

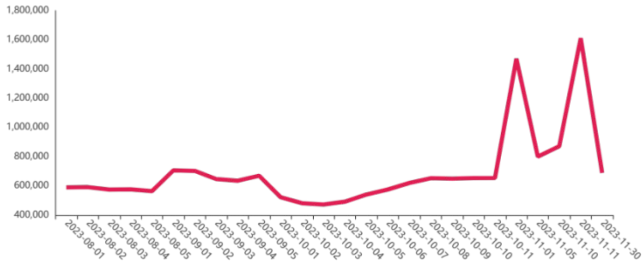

Fig.7: Changes in overall cargo volume

Through visual analysis of the overall trend of cargo volume changes, it can be concluded that the overall trend of the 57 sorting centers is that there is not much change in cargo volume from August to September, but there is a significant upward fluctuation from October 1 to mid-November, during which the cargo volume increases rapidly.

Due to the long-term dependence and instability of the cargo volume data in this work, a Long Short Term Memory Network (LSTM) model is adopted.

3.2.2 Model establishing and solving

The ARIMA model, also known as the auto regressive differential moving average model, mainly consists of three parts: auto regressive model (AR), differential process (I), and moving average model (MA). In practical applications, the ARIMA algorithm can be used to predict future time series data. It usually requires a training process, during which it learns past data patterns in order to better predict future data The parameters of the ARIMA algorithm can be estimated through optimization algorithms to minimize prediction errors. The specific formula is shown in equation (1).

$$N_t=\theta_1 N_{t-1}+\ldots\theta_p N_{t-p}+e_t-\theta_1 e_{t-1}-\cdots-\theta_q e_{t-q} \tag{1}$$

Where, N_t is stationary time series, e_i is white noise sequence, $\varphi_1 \ldots \varphi_p$ is the coefficients of the AR model, $\theta_1 \ldots \theta_q$ is the coefficients of the model, p and q is the order of the model.

According to the ARIMA model process shown in Fig.11, the prerequisite for the application of this model is that the data must have stationarity. Therefore, it is necessary to perform ADF tests on time series data to determine whether the null hypothesis of non-stationarity can be rejected at a significance level ($P < 0.05$). Taking the data from sorting center SC54 as an example, the specific results of ADF tests, including variables, difference order, T-test results, and Akashi Information Criterion (AIC) values, are detailed in Table 1.

Table 1: ADF Inspection Form

Variable	Difference order	t	P	AIC	Critical value		
					1%	5%	10%
Cargo quantity	0	-8.619	0.000***	1303.687	-3.498	-2.891	-2.583
	1	-6.594	0.000***	1307.109	-3.504	-2.894	-2.584
	2	-5.705	0.000***	1311.214	-3.51	-2.896	-2.585

Where, * * *, * * and * represent significance levels of 1%, 5%, and 10%, respectively.

For variable SC54, when the difference is of order 0, the P-value is 0.000, indicating a high degree of significance in the results. Therefore, the non-stationary null hypothesis is rejected statistically, indicating that the time series is stationary. When the difference is of order 1, the P-value is 0.000, which also shows extremely high significance. We rejected the non-stationary null hypothesis and confirmed that the time series is stationary. Continuing until the difference is of order 2, the P-value is still 0.000, showing strong significance.

In summary, the ARIMA model is suitable for use, so the model was established using MATLAB to write code. Using time as the horizontal axis and cargo volume as the vertical axis, the hourly cargo volume of 57 sorting centers for the next 30 days was calculated.

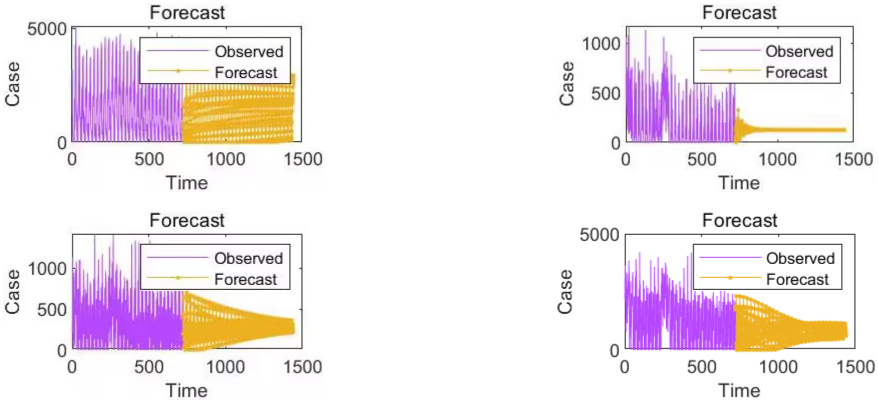

Fig.8: ARIMA Model Prediction Chart

Fig.8 reveals the prediction results of the ARIMA model for future cargo volume, which provides a series of predicted values that exhibit a certain degree of volatility overall and suggest the potential impact of uncertainty factors on future cargo volume. At certain time points, there is a gap between actual observed values and predicted values, which may be due to various unforeseeable events that occur during the operation process, including emergencies, holidays, promotional activities, etc., which may affect actual cargo volume. However, the ARIMA model has shown high accuracy in daily cargo volume prediction.

3.3 Problem two model establishment and solution

3.3.1Data preprocessing

Compare the cargo volume of each starting sorting center that arrives at each sorting center. The comparison can clearly and intuitively show the distribution of cargo volume and facilitate the calculation of weights. The bar charts for partial cargo volume comparison are shown in Fig.9 to Fig.12.

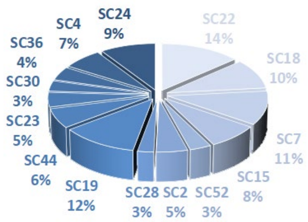

Fig.9: Cargo volume on the branch line reaching the sorting center SC15

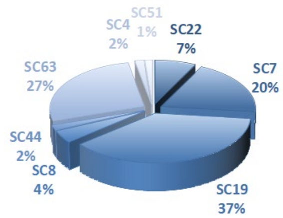

Fig.10: Cargo volume on the branch line reaching the sorting center SC8

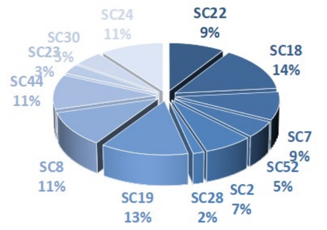

Fig.11: Cargo volume on the branch line reaching sorting center SC5

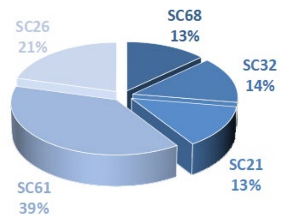

Fig.12: Cargo volume on the branch line reaching sorting center SC10

Fill in the corresponding cargo volume from the origin sorting center to the arrival sorting center, and then calculate the total cargo volume for each arrival sorting center to determine the route change. Based on this, it can be concluded that the total cargo volume after re pairing the route is less than before pairing, and the difference in shipment volume can be calculated to consider the redistribution of cargo volume for the changed route. Finally, the required rate of cargo volume change can be obtained.

In order to solve the rate of change, we need to comprehensively refer to the changes in historical cargo volume and routes. According to the data in Annex 3 and Annex 4, the starting sorting center is not absolutely balanced in terms of quantity when it arrives at the sorting center. There are as many as 17 starting sorting centers when it arrives at the sorting center SC47, while there are only 3 starting sorting centers when it arrives at sorting center 60. Based on the definition given in the work, the cargo volume prediction of the sorting center is related to the transportation route. Adjusting the transportation route can make the cargo volume transport optimal. The weight of each starting sorting center and the total cargo volume can be determined based on the number of arriving sorting centers and the weight of each arriving sorting center. The excess cargo volume that needs to be allocated can be allocated according to the weight of each arriving sorting center. The calculated allocation cargo volume is shown in Table 2.

Table 2: Allocation of cargo volume to sorting centers by weight

Arriving at the sorting center	SC5	SC7	SC8	SC10	SC15	SC25	SC47	SC51	SC57	SC60
Allocation of goods quantity	154	154	154	154	375.5	375.5	230	200	100	100

After calculating the allocated goods quantity into the total quantity, the change rate can be calculated using the formula (allocated goods quantity)/(unassigned goods quantity) * 100% -1=change rate. The calculated change rate is shown in Table 3:

Table 3: Change rate of arrival at sorting center

Arriving at the sorting center	SC5	SC7	SC8	SC10	SC15	SC25	SC47	SC51	SC57	SC60
Allocation of goods quantity	0.0121	0.0129	0.0262	0.0034	0.041	0.003	0.022	0.316	0.233	0.014

3.3.2 Corrected model establishing and solving

For solving the corrected predicted data, the original (predicted data) * (rate of change + 1) = corrected data can be used to calculate the corrected daily and hourly cargo volumes. By establishing the above model, it is possible to predict the daily and hourly cargo volume of each sorting center for the next 30 days, and adjust the cargo volume allocation according to changes in transportation routes, making the management and prediction of sorting centers and cargo volume more accurate

3.4 Problem three model establishment and solution

3.4.1Establishing an integer programming model

This problem involves the effective utilization and scheduling of human resources to ensure daily cargo processing without increasing additional costs and minimizing human resource waste. Therefore, it belongs to the scheduling problem in operations research and is also an optimization problem that requires finding the optimal human resource scheduling solution while satisfying constraints. Based on this, we choose an integer programming model and list the known information based on the stem information:

(1) There are a total of 57 sorting centers, each with 60 formal workers;

(2) Based on the daily demand for goods for the next 30 days, priority should be given to using regular workers before considering using temporary workers;

(3) The maximum efficiency for formal workers in the sorting center is 25 packages per hour, while the maximum efficiency for temporary workers is 20 packages per hour;

(4) There are a total of 6 shifts, each lasting 8 hours, and each person can only take a maximum of 1 shift.

The shift schedule information is as follows: 00:00-08:00, 05:00-13:00, 08:00-16:00, 12:00-20:00, 14:00-22:00, 16:00-24:00. Based on known conditions and requirements, we design decision variables, objective functions, and constraints:

(1) Decision variables

x_{ijk}: Number of formal workers used by sorting center j in the kth shift on the ith day (positive integer);

x_{ijk}: Number of temporary workers used by sorting center j in the kth shift on the ith day (positive integer).

(2) Objective function

The purpose of establishing an objective function is to meet the demand for goods while minimizing the total number of person days and maintaining the actual hourly human efficiency as balanced as possible. This is shown in equation (2).

$$minN = \sum_{d=1}^{30} \sum_{i=1}^{57} \sum_{b=1}^{6} (x_{ijk} + y_{ijk}) \tag{2}$$

(3) Constraints

1) Daily cargo volume processing completed for each sorting center

The amount processed by the employees arranged for each sorting center on a daily basis should be equal to the cargo volume of that sorting center on that day. The expression for constraint 1 is shown in equation (3).

$$8 \times \sum_{k=1}^{6} (25x_{ijk} + 20y_{ijk}) \geq Q_{ij} \tag{3}$$

2) The number of formal workers arranged for work in each sorting center cannot exceed 60

According to the known conditions, there are 60 formal workers in each sorting center, so when arranging, the number of formal workers in each sorting center cannot exceed 60. The expression for constraint 2 is shown in equation (4).

$$\sum_{b=1}^{6} x_{ijk} \leq 60 \tag{4}$$

3) Priority should be given to formal workers, followed by temporary workers

From the known conditions, it can be seen that, due to the maximum human efficiency of formal workers being 25 packages per hour, there are 5 more packages processed per unit time than temporary workers. Considering the maximum use of human resources, formal workers are prioritized. The expression of constraint 3 is shown in equation (5).

$$\sum_{k=1}^{6} x_{ijk} \geq \sum_{k=1}^{6} y_{ijk} \tag{5}$$

4) Try to balance the actual hourly human efficiency as much as possible every day

The actual hourly efficiency of each day should be balanced as much as possible. In order to avoid situations where some periods are excessively busy while others are relatively idle, it is hoped that the workload and output of employees can be more balanced and stable throughout the entire workday. The expression of the constraint conditions is shown in equations (6) and (7):

$$25\sum_{k=1}^{3} x_{ijk} + 20\sum_{k=1}^{3} y_{ijk} \geq \frac{1}{2} Q_{ij} \tag{6}$$

$$20\sum_{k=4}^{6} x_{ijk} + 20\sum_{k=4}^{6} y_{ijk} \geq \frac{1}{2} Q_{ij} \tag{7}$$

4) Non negative integer constraint

x_{ijk} represents the number of formal workers used by sorting center *j* in the k^{th} shift on the i^{th} day should be constrained to a positive integer. The integer limit is implemented in MATLAB by specifying the variable as an integer. The expression for constraint 5 is shown in equation (8).

$$x_{ijk}, y_{ijk} \geq 0 \tag{8}$$

In summary, the following model can be obtained by summarizing decision variables, objective functions, and constraints, as shown in the following equation.

$$minN = \sum_{d=1}^{30} \sum_{i=1}^{57} \sum_{b=1}^{6} (x_{ijk} + y_{ijk}) \tag{9}$$

$$s.t.\begin{cases} 8 \times \sum_{k=1}^{6}(25x_{ijk} + 20y_{ijk}) \geq Q_{ij} \\ \sum_{b=1}^{6} x_{ijk} \leq 60 \\ \sum_{k=1}^{6} x_{ijk} \geq \sum_{k=1}^{6} y_{ijk} \\ 25\sum_{k=1}^{3} x_{ijk} + 20\sum_{k=1}^{3} y_{ijk} \geq \frac{1}{2} Q_{ij} \\ 20\sum_{k=4}^{6} x_{ijk} + 20\sum_{k=4}^{6} y_{ijk} \geq \frac{1}{2} Q_{ij} \end{cases} \tag{10}$$

3.4.2 Model solving based on genetic algorithm

Genetic algorithm (GA) is a global optimization algorithm that simulates the process of biological evolution. By simulating mechanisms such as natural selection, genetics, and mutation, it searches and optimizes solutions to problems. Its uniqueness lies in its ability to effectively deal with complex, nonlinear, multi-modal, and other problems, conduct global searches in the solution space, and find the optimal or approximate optimal solution to the problem The calculation steps of the GA algorithm are as follows:

Step 1: Encoding. Use real number encoding to prevent getting stuck in local minima, using the reciprocal of error E as the fitness value, that is:

$$f_i = \frac{1}{E} \tag{11}$$

Where, f_i represents the fitness of individual $\boldsymbol{i}$.

Step 2: Generate an initial population. Randomly generate M individuals to form a population, and GA continuously evolves based on the initial point

Step 3: Selection operation. The purpose of the selection operation is to find the optimal individual, so that it can be used as the parent to reproduce offspring. The method of individual selection is roulette wheel, which means that individuals with a larger proportion of fitness are more likely to be selected.

$$P_i = \frac{f_i}{\sum_{i=1}^{M} f_i} \tag{12}$$

Where, P_i represents the probability of individual i being selected.

Step 4: Cross operation. Using the arithmetic cross method, the kth chromosome c_k and ith chromosome c_i undergo structural exchange at the intersection position j to generate new chromosomes c'_{kj} and c'_{ij}. The cross-operation method is as follows:

$$\begin{cases} c'_{kj} = c_{kj}(1-d) + c_{ij}d \\ c'_{ij} = c_{ij}(1-d) + c_{kj}d \end{cases} \tag{13}$$

Where, d is a random number between (0,1).

Step 5: Mutation operation. Mutation is the process of mutating the selected individual to form a new individual according to a specific probability. Mutation selects the yth gene of the xth chromosome according to a random probability r for mutation, and obtains the mutated chromosome c'_{xy}. The expression is:

$$c'_{xy} = \begin{cases} c_{xy} + (c_{xy} - c_{max})f(t), r > 0.5 \\ c_{xy} + (c_{min} - c_{xy})f(t), r \le 0.5 \end{cases} \tag{14}$$

Where, c_{max} represents the upper bound of gene c_{xy}; c_{min} represents the lower bound of gene c_{xy}; $f(t) = r(1 - t/G_{max})$, where t is the current iteration number and G_{max} is the maximum evolution number; R is a random number between (0,1).

Replace the mutated new individual into the new population and repeat the operations from equation until the optimal solution appears.

Firstly, establish an integer programming model to determine the objective function and constraint conditions, and then use genetic algorithm to write code to solve the personnel scheduling optimization problem

3.5 Problem four model establishment and solution

(1) Decision variables

Based on work 4, provide more detailed personnel scheduling requirements, as well as requirements for A formal employees to go to sorting center B to sort goods, and C temporary workers to go to sorting center D to sort goods. Here, set formal employees as the 0-1 decision variable x_{ijk}, satisfy the following relationship:

C: Whether formal employee i starts work on the *jth* day, if starting work, it is 1, otherwise it is 0.

y_j: The number of temporary workers to be hired on the jth day.

r_i: The attendance rate of formal employee i, which is the number of days attended within 30 days divided by the total number of days.

(2) Objective function

Optimize the number of people and days while ensuring daily cargo handling volume; Maintain a daily balance of hourly productivity; At the same time, we ensure that the attendance rate of formal employees is evenly distributed. To achieve the above goals, we have determined the following objective function, as shown in equation (15).

$$Min\ N = \sum_{i=1}^{30}\sum_{j=1}^{57}\sum_{k=1}^{6} x_{ijk} + \sum_{j=1}^{30}\sum_{k=1}^{6} y_{jk} \quad (15)$$

(3) Constraints

Cargo volume processing constraint: the primary constraint for scheduling is to complete the cargo volume processing. The sum of the work duration multiplied by the time efficiency should be greater than or equal to the total cargo volume. The expression for constraint one is shown in equation (16).

$$8 \times \sum_{j=1}^{6}(25x_{ij} + 20y_j) \geq Q_i \quad (16)$$

Attendance rate constraint: According to known conditions, the attendance rate of each formal worker cannot exceed 85%. The expression for constraint condition two is shown in equation (17).

$$\sum_{j=1}^{30} x_{ij} \leq 0.85 \times 30 \quad (17)$$

Attendance days constraint: According to the known conditions, each formal worker cannot continuously attend for more than 7 days. The expression for constraint condition three is shown in equation (18).

$$\sum_{j=k}^{k+6} x_{ij} \leq 7 \quad (18)$$

Try to balance the actual hourly human efficiency as much as possible: For the constraint of balancing the actual hourly human efficiency, it should be set to a relatively small value to ensure a relatively balanced allocation of manpower every day. Usually, it can be attempted to set a certain proportion of the total number of people, which is 5% of the total number of people. For example, sorting point SC1 is 18. The expression for constraint condition four is shown in equation (19).

$$|\sum_i x_{ij} - \frac{x_{ij}+\sum_i y_j}{30}| \leq \varepsilon \quad (19)$$

Try to balance the attendance rate of formal workers as much as possible: The selection of the balance constraint for the attendance rate of formal workers depends on the requirements for the balance of attendance rate of formal workers. If you want the attendance rate of formal workers to be as average as possible, you can choose a smaller value, such as 1% of the total attendance rate, for example, the sorting point SC1 is 0.6. The expression of constraint condition five is shown in equation (20).

$$|\frac{1}{30}\sum_{j=1}^{30} x_{ij} - x_{ij}| \leq \delta \quad (20)$$

Non negative integer constraint: x_{ij} represents whether formal worker i starts work on the jth day. If he starts work, it is 1, otherwise it is 0. Therefore, this variable should be a non-negative integer. The expression for constraint six is shown in equation (21).

$$x_{ij} \in \{0,1\},\ y_i \in Z^+ \quad (21)$$

In summary, the following model can be obtained by summarizing decision variables, objective functions, and constraints, as shown in the following equation.

$$Min\ N = \sum_{i=1}^{30}\sum_{j=1}^{57}\sum_{k=1}^{6} x_{ijk} + \sum_{j=1}^{30}\sum_{k=1}^{6} y_{jk} \tag{22}$$

$$s.t.\begin{cases} 8 \times \sum_{j=1}^{6}(25x_{ij} + 20y_j) \geq Q_i \\ \sum_{j=1}^{30} x_{ij} \leq 0.85 \times 30 \\ \sum_{j=k}^{k+6} x_{ij} \leq 7\ \ \forall\ i, \forall\ k = 1,2,\dots,24 \\ |\sum_i x_{ij} - \frac{x_{ij}+\sum_i y_j}{30}| \leq \varepsilon \\ |\frac{1}{30}\sum_{j=1}^{30} x_{ij} - x_{ij}| \leq \delta \\ x_{ij} \in \{0,1\}, y_i \in Z^+ \end{cases} \tag{23}$$

Based on establishing an integer programming model to determine the objective function and constraint conditions, the results could be obtained correspondingly.

4 CONCLUSIONS

The ARIMA auto regressive differential moving average model can handle nonlinear and non-stationary time series, consider the influence of historical data, and have high accuracy in future predictions. For the accuracy requirements of the prediction results, it is more convenient to adjust the parameters. However, the ARIMA model requires a stationary assumption, and if the time series is non-stationary, data preprocessing is required and the long-term prediction effect is not ideal

The integer programming model can provide precise solutions, especially for problems that require integer values, such as calculating the optimal number of formal and temporary workers in this problem. In solving integer programming problems, integer constraints may make the problem non convex or non-polynomial time solvable, making the solution difficult.

Genetic algorithm has scalability in computation and is easy to combine with other algorithms, which can better improve the accuracy of results. The programming implementation of genetic algorithm computation is quite complex. Firstly, the problem needs to be encoded, and after finding the optimal solution, the problem needs to be decoded. And the training process takes a long time when the results are required to be more accurate.

5 ACKNOWLEDGEMENTS

This work is supported by ministry of education industry-university cooperative education project (Grant No.: 231106441092432), the research and practice of integrating "curriculum thought and politics" into the whole process of graduation design of Mechanical engineering major: (Grant. No.: 30120300100-23-yb-jgkt03), research on the integration mechanism of "course-training-competition-creation-production" for innovation and entrepreneurship of mechanical engineering majors in applied local universities (Grant. No.: CXKT202405), Mechanical manufacturing equipment design school-level "gold class" construction project (Grant. No.: 30120324001).

6 DATA SOURCES

Question C of the 14th Annual MathorCup Math Application Challenge Competition 2024(http://mathorcup.org/detail/2438)

REFERENCES

[1] Zhai, J., & Cao, J. (2016). A combined prediction model based on time series ARIMA and BP neural network. *Statistics and Decision Making*, 04: 29-32.

[2] Chen, S. J. (2017). A Logistics sorting system based on multi AGV scheduling. *Shenzhen University*.

[3] Che, G. Q. (2020). Heuristic mixed integer linear programming for flight conflict resolution. *Zhejiang University of Technology*.

[4] Zhang, P. L. (2021). Research on optimization of storage capacity of fast-moving consumer goods based on order prediction. *North China Electric Power University*.

[5] Gong, B. B. (2021). Prediction of natural gas imports in China based on GA-ELM model. *China University of Petroleum*.

[6] Wang, G. N., & Tang, X. P. (2022). Research on rural emergency logistics distribution path based on simulated annealing and floyd optimization algorithm. *Software Engineering*, 25(12): 9-12.

[7] Yang, J., Yang, X. D., & li, L. S. (2022). Optimization of stacker crane path based on genetic simulated annealing algorithm. *Logistics Technology*, 41(08): 119-123.

[8] Yan, N. (2022). Research on volume prediction and route optimization of H company's container dumping transportation. *Guizhou University*.

[9] Xu, B., & Wang, Q. D. (2023). A fault-tolerant method for AUV collaborative localization based on Mahalanobis distance and neural network assistance. *Chinese Command and Control Society*.

[10] Shen, Q. M. (2023). Empirical Comparative Study of ARIMA, LSTM and their combination models. *Suzhou University*.

[11] Xu, B., & Feng, Z. (2023). Study on electronic product assembly line balance based on integer programming and simulation. *Mechanical & Electrical Engineering Technology*, 1-8.