Advances in Computer, Signals and Systems (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2024.080301
ISSN 2371-8838 Vol. 8 Num. 3

Research on Logistics Network Based on ARIMA Prediction Model and Multi-Objective Optimization

Wei Guo[a,*,#], Jing Zhao[b,#], Dianzhe Yang[c,#], Canglong Zhang[d], Ping Li[e]

College of Intelligence and Information Engineering, Shandong University of Traditional Chinese Medicine, Jinan, China

[a]18364013993@163.com, [b]zhaojing20242024@163.com, [c]15668316916@163.com, [d]15169779017@163.com, [e]15964596939@139.com

[*]Corresponding author

[#]Co-first author

Keywords: ARMA, multi-objective optimization, logistics network, emergency movement, simulated annealing

Abstract: This paper carries out a profound research on the problem of emergency transportation and structure optimization of goods in logistics network. First of all, a prediction model based on the transportation volume of historical routes is established, and data preprocessing is carried out for logistics routes and cargo volume data. Then, the ARIMA model is selected for prediction, and the daily cargo volume of each route is trained and tested during the period from 2023-01-01 to 2023-01-31, and then the ARIMA model is evaluated and adjusted by the white noise test. The ARIMA model is used to predict the future line cargo volume of DC14→DC10, DC20→DC35, DC25→DC62, and the results are analyzed and interpreted, and the prediction results are presented in the paper in the form of a three-line table. Next, on the basis of the previous prediction, with the objectives of minimizing the number of changes in cargo volume before and after the closure of DC5 and making all parcels flow as quickly and normally as possible, we introduce constraints such as the workload being balanced as much as possible, maximum cargo volume constraints, non-negative constraints, total volume constraints, etc., establish a bi-objective non-linear integer planning model, and use simulated annealing algorithm to solve the model.

1. Introduction

With the rapid development of China's e-commerce industry and the continuous expansion of market share, the e-commerce logistics network has also been rapidly developed and improved. However, under the influence of holidays and promotional activities such as "Double Eleven" and "618", the order volume of e-commerce users fluctuates significantly, while emergencies such as the new coronary pneumonia epidemic also have an impact on the logistics network, resulting in the deactivation rate of logistics sites rising. Against this background, national policies have also proposed a series of measures to support the healthy development of the e-commerce industry, including strengthening the construction of the e-commerce logistics system and optimizing logistics services. Therefore, how to use big data technology and intelligent algorithms to predict the cargo volume of each logistics site and route in the e-commerce logistics network, and design a reasonable logistics adjustment program to cope with the impact of factors such as emergencies and holidays on the logistics network will become an urgent problem in the operation and management of the e-commerce logistics network. Based on the above background, this study aims to explore how to realize the integration of logistics resources, structural adjustment, and cost reduction through the establishment of a reasonable e-commerce logistics network planning model, so as to adapt to the development trend of the digital era and the requirements of national policies.

2. Related Works

The design and path optimization of logistics and distribution networks has been an important topic in the field of logistics management. In Duan Jiawei et al.'s study [1], the application of economics based on advanced mathematical modeling thinking was proposed to explore the application of mathematical models in economics, which provided a theoretical basis for the subsequent research. In her study [2], Tingting Zhao focused on the design of logistics and distribution network in Liaodong Bay living area, which provided a practical case for the logistics layout in a specific region. Meanwhile, Wang Yuxue and Wang Ziqiang [3] proposed the method of constructing the mathematical model of underground logistics system network, which provided new ideas for the design of logistics system in special environment. Wen Yi in his paper [4] analyzed the structural characteristics of regional logistics networks and discussed how to optimize these structures to improve efficiency. Cao Yaqun [5], on the other hand, explored the application of mathematical modeling in the teaching of logistics management from the teaching point of view, which provides new teaching ideas for the cultivation of logistics management talents. In addition, Li Qi [6] proposed a regional logistics network construction method based on the gravity model in his research, which can more accurately describe the interactions between nodes in the logistics network. And Xiao Jianmei et al [7] used discrete particle swarm optimization algorithm to solve the logistics distribution vehicle path problem, which provides an effective path planning tool for actual logistics distribution. Le Yixiang et al [8] proposed an improved ant colony algorithm for solving the logistics distribution path optimization problem, which has a strong global search capability in the optimization process. And Li Yongxian et al [9] studied the simulation optimization of vehicle path problem in logistics distribution system, which provides theoretical support for the practical operation of logistics distribution system. Finally, Liu Beilin and Gao Shuang [10] and Jiang Ling and Shen Guilan [11] explored the algorithmic study of the path optimization problem of distribution vehicles, which involved the application of ant colony algorithm and provided multiple feasible solutions to solve the path optimization problem in practical logistics and distribution. These scholars explored the logistics distribution network design and path optimization problems from different perspectives, which provided theoretical and methodological support for improving the efficiency of logistics distribution and reducing costs.

3. Theory and Method

3.1 Data Preprocessing

Firstly, each logistics site and each route are regarded as a node, and the whole logistics network is represented by a graph in graph theory. In this case, the processing capacity of each node and the transportation capacity of each edge are the maximum value of historical cargo volume. By visualizing and analyzing the routes generated by different logistics nodes as well as the cargo volume, we get the average transportation volume of some logistics site nodes from 2021-01-01 to

2022-12-31 as well as the logistics cargo volume and trend of each month as shown in Fig.1.

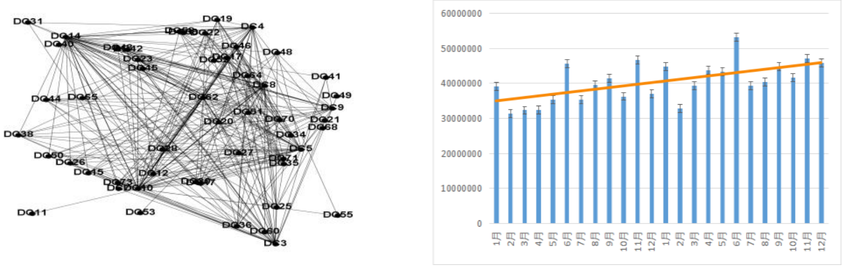

Figure 1: Selected logistics sites and total logistics volume per month

3.2 Time Series Analysis

It is very important to perform data preprocessing before using ARIMA model to forecast the time series. First of all, to determine the stability of the time series, by running the ADF unit root test and other stability tests, the smoothness test can determine whether the time series is stable or not, through the original time series plot can be seen that the sequence fluctuation is large, ACF and PACF plot test. As can be seen from Figure 2, the ACF behaves like trailing, but the data behind does not converge, which indicates that the time series has a monotonic trend, while the PACF is truncated and the coefficients fluctuate around the upper and lower ranges of the zero axis, so it is judged that the original time series is a non-stationary series.

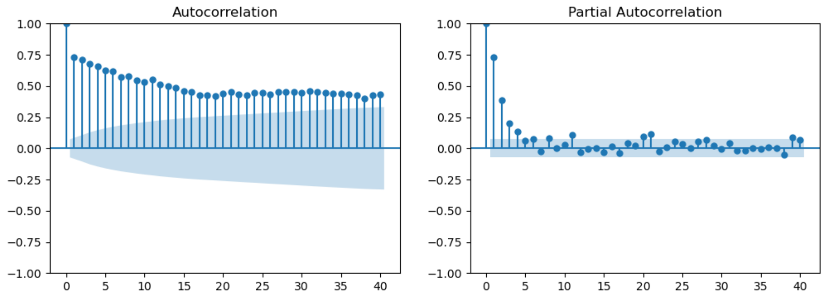

Figure 2: ACF and PACF Raw Data

Due to the instability of the time series, it is necessary to perform a first-order difference process on the original time series, which we do by calling the diff() function of the Pandas library in order to stabilize it. From the first order difference sequence plot it can be seen that the quantity of goods fluctuates up and down around a value and the stability magnitude is increased so that the time series can be analyzed. Again ACF and PACF plot test is performed to observe the truncation and the results are shown in Fig.3.

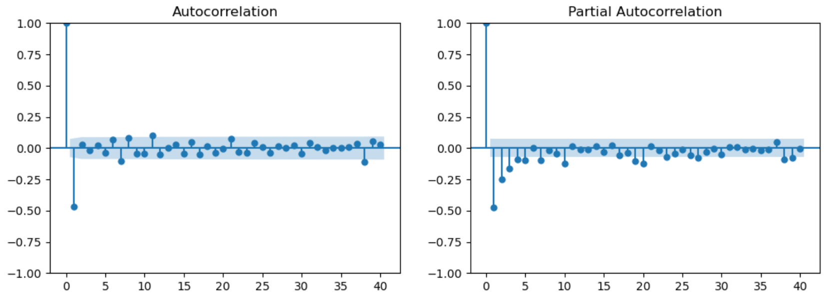

Figure 3: ACF and PACF post-differential data

Differential ACF and PACF are performed on the original data, so the series after first-order differencing is a smooth series, then the value of d is equal to 1, p and q are equal to 2. Line Cargo

Volume after d order differencing and applying autoregressive $AR(p)$ and sliding average model $MA(q)$ can be expressed as

$$(y')t = \alpha_0 + \sum_{p}^{i=1} (y'')_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \beta_i \varepsilon_{t=i} \tag{1}$$

The order difference equation can be expressed as follows.

$$y_t' = \Delta^d y_t = (1-L)^d y_t \tag{2}$$

Substituting (1) into (2), the following equation can be obtained.

$$\left(1 - \sum_{i=1}^{p} \alpha_i L^i\right)(1-L)^d y_t = \alpha_0 + \left(1 + \sum_{i=1}^{q} \beta_i L^i\right)\varepsilon_t \tag{3}$$

where L is the lag operator that satisfies

$$(1-L)y_t = y_t - Ly_t = y_t - y_{t-1} \tag{4}$$

Where p represents the number of lags in the time series data itself used in the prediction model, also called the AR/Auto-Regressive term; and d represents the order in which the time series data is stabilized by differentiation, also called Integrated term; and q is the lag of the prediction error used in the prediction model, also called the MA/Moving Average term; the autocorrelation (ACF) and partial autocorrelation function (PACF) plots are obtained from the plot_acf() and plot pacf() functions of the Pandas library, which can be used to determine the optimal order of the ARIMA model $(p = 2, d = 1, q = 2)$, which means second-order autoregression, first-order differencing, and second-order moving average.

3.3 Scheduling Model

Step 1. Model assumptions

- Assumption1: Observations are independent and identically distributed at each point in time.
- Assumption2: Historical data can be used to predict the future, and the predictions are accurate.
- Assumption3: All decision variables are non-negative.
- Assumption4: The lines shown are the remaining lines after the DC5 logistics site connection.

Step 2. Establishment of the objective function

Objective 1: The number of changes in cargo volume before and after the closure of DC5 is as small as possible.

The total cargo volume before the closure of the DC5 logistics site can be transferred to other logistics routes, and the total cargo volume after the closure of the DC5 logistics site can be reasonably allocated to meet the objective requirements, taking into account the normal value of the cargo volume and the maximum load capacity of each route. When the total cargo volume to be allocated and the number of logistics routes acceptable for allocation is large, the change of cargo volume of these routes $\left(x_{ij}{}^{B} - x_{ij}{}^{L}\right)$, compared to a certain number of cargo volume allocated to a smaller number of routes in the case of the product of the result of the larger number, want to meet the target only need to take the minimum value of the product can be. Therefore, the objective function is set as follows.

$$\prod_{j=1}^{n} \prod_{i=1}^{n} \left(x_{ij}{}^{B} - x_{ij}{}^{L}\right) = \min \tag{5}$$

Objective 2: Make all packages flow normally as much as possible

Need to meet the number of parcels to minimize the number of abnormal flow, here the abnormal flow that is in a number of objective function and constraint function of the limit, in the allocation of cargo volume after the remaining can not participate in the flow of cargo volume, here the total number of cargo volume of the DC5 logistics site is known before the closure of the volume, then need to be the total of changes in the line freight volume than the total number of cargo volume before the closure of the logistics site of the DC5, and will be subtracted from the result of the minimum value of the two , even if the number of goods that can not flow normally is minimized, that is, we can achieve the normal flow of all packages as far as possible. The objective function is set as follows.

$$m-\sum_{j=1}^{n}\sum_{i=1}^{n}\left(x_{ij}{}^{B}-x_{ij}{}^{L}\right)=min \tag{6}$$

Step 3. Determine constraints

Constraint 1: Balance workloads as much as possible

Since each of the remaining lines (here we have removed the lines connected to DC5) have different predictions, and the maximum cargo volume of each line is also different. And the same amount of cargo is allocated to different lines resulting in different workload rates of the line, based on this, we consider setting two constants $\delta1,\delta2$, which will be expressed as the equalization degree, and set different values of the equalization degree by giving different values of $\delta1,\delta2$. Since the workload is to be equalized as much as possible, the given values need to fluctuate within a certain range, the maximum range being $0\sim1$. As it is not easy to calculate using the workload rate function, an improved function is introduced as follows.

$$\delta1\leq\frac{\left(x_{ij}-\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}x_{ij}}{n}\right)}{c_{ij}}\leq\delta2 \tag{7}$$

Workload ratio $=$ (Average value of daily load/maximum value of daily load)* 100%

Constraint 2: Maximum Freight Volume Constraint

Since the freight volume of each route cannot exceed its maximum freight volume, if the maximum freight volume is exceeded under the influence of various factors, the exceeding part of the freight volume to be flowed is called abnormal flow. Sij is called the abnormal flow. The constraint function is as follows.

$$0<x_{ij}{}^{L}<c_{ij} \tag{8}$$

Constraint 3: Non-negative constraint

Since the normal flow of all lines before DC5 shutdown and the flow of each line after DC5 shutdown are greater than zero and integer, the constraint function is

$$x_{ij}^{L},x_{ij}^{B}\in z \tag{9}$$

Constraint 4: Aggregate volume constraint

Since the sum of the changes in the freight volume of each route before and after the closure of DC5 is less than the sum of the freight volume of the routes connected to DC5 before the closure of DC5. Taking into account the influence of the above constraints and optimization objectives, there may be a part of the goods that can not be normal flow of the situation, the constraint function is as follows

$$\sum_{j=1}^{n}\sum_{i=1}^{n}\left(x_{ij}^{L}-x_{ij}^{B}\right)\leq m \tag{10}$$

4. Results and Discussions

4.1 Time Series Model Results

In this model, the method of dividing the training set and the test set is adopted as the fixed time point division method. The logistics data of the three routes DC14→DC10, DC20→DC35, and DC25→DC62 during the period of 2021-01-01 to 2022-10-31 were used as the training set to train the ARIMA model; the training based on the ARIMA model used the logistics data of the three routes during the period of 2022-11-01 to 2022-12-31 as the test set that is used to evaluate the prediction results of the model. Based on the cargo volume data of each logistics site and route from 2021-01-01 to 2022-12-31, with the help of the tested ARIMA model, the logistics routes and cargo volume from 2023-01-01 to 2023-01-31 are predicted, and prediction and planning are carried out for each logistics site and route step by step, and the planning and adjustment of logistics network is carried out based on the prediction results, in order to model the DC14-DC10 training and testing as an example: three lines, DC14→DC10, DC20→DC35, DC25→DC62, can be predicted using the above method, and the results are shown in Fig.4 and Table 1.

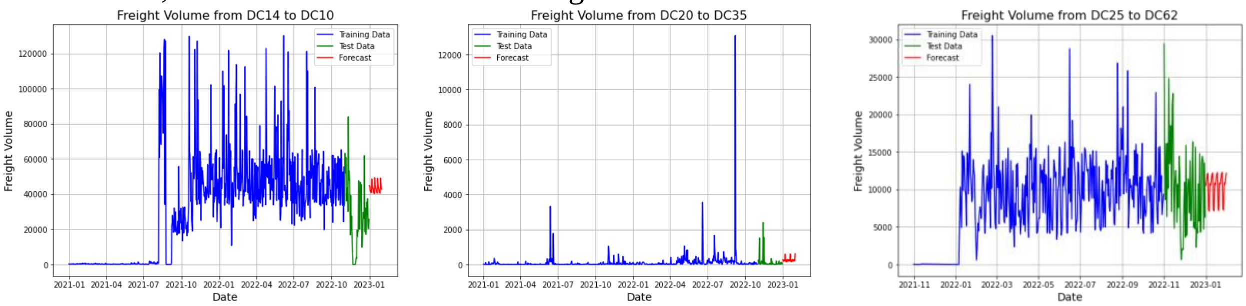

Figure 4: Line Training, Testing and Prediction

Table 1: Line prediction results

Dates	DC14→DC10	DC20→DC35	DC25→DC62
2023/01/01	44788.85	266.55	10539.79
2023/01/02	44548.54	229.42	11149.80
2023/01/03	41781.91	377.00	11946.61
2023/01/04	41871.69	223.05	12134.31
2023/01/05	40770.90	220.50	7283.93
2023/01/06	40990.89	218.83	7101.45
2023/01/07	48579.22	212.42	10784.59
2023/01/08	46132.08	293.09	10576.56
2023/01/09	43104.81	244.20	11327.70
2023/01/10	43024.09	382.48	12025.65
2023/01/11	40834.24	226.93	12200.41
2023/01/12	41687.54	223.74	7336.73

Predicting future logistics flows based on historical and real-time data enables efficient decision-making to cope with unforeseen situations and business needs. If the model predicts that there will be significant fluctuations in the cargo volume of a route, the logistics plan can be adjusted in advance, such as increasing the capacity and strengthening the load capacity of the logistics site, in order to ensure the smooth flow of logistics. At the same time, if the model predicts that a bottleneck may occur at a logistics site or route, corresponding remedial measures can be arranged in advance, such as increasing the number of logistics sites, adjusting the route path, etc., to ensure smooth logistics.

4.2 Scheduling Model Solution

Here, the simulated annealing algorithm in the heuristic algorithm is designed to solve the scheduling model. First, the initial value is constructed, the initial value of temperature T is one of the important factors affecting the global search performance of the simulated annealing algorithm, a high initial temperature will have a high probability of searching the global optimal solution, but therefore a lot of computational time will be spent; on the contrary, computational time can be saved, but the global search performance may be affected. Based on the comprehensive analysis of the amount of data corresponding to the problem and the desired goal to be achieved, T=300 is chosen as the initial value here. The cooling coefficient is set to 0.99, the error range is set to 0.003, the ending temperature is set to 1, the diversion ratio is set to 0.13, the upper limit of the weight is set to 12, and the perturbation ratio of each time is set to 0.4. Then the following results are obtained: the number of lines in which the volume of goods changes before and after the shutdown of DC5 is 3, which are DC5→DC10, DC5→DC15, DC5→DC20; the network load before and after the shutdown is shown in Table 2.

Table 2: Network Load after DC5 Shutdown

Area	Quantity (pieces)	Workload (%)
DC10	14213	82.62
DC15	19385	112.54
DC20	16667	96.70
DC25	21457	124.34
DC30	18128	105.02
DC35	14938	86.52
DC40	18018	104.41
DC45	18297	105.98
DC50	20410	118.13
DC55	18285	105.95
DC60	16294	95.32

In all the adjustments, there were site overloads, there were 7 sites were DC15, DC25, DC30, DC40, DC45, DC50, DC55, with an average load factor of 103.41%.

5. Conclusion

Through the research in this paper, we have deeply explored and analyzed the problem of emergency cargo movement and structural optimization in logistics networks. However, the research in this paper still has some limitations. In building the prediction model, we only considered the historical data without taking into account the influence of possible external factors on the flow of goods, such as the weather, traffic conditions, and so on. Secondly, for solving the bi-objective nonlinear integer programming model, although the simulated annealing algorithm can find a better solution to a certain extent, its efficiency and accuracy still need to be further improved. Future research directions can consider introducing more external factors and real-time data to improve the accuracy and stability of the prediction model; secondly, we can try to use more efficient solution algorithms, such as genetic algorithm and particle swarm algorithm, to optimize the cargo scheduling scheme and improve the overall efficiency and response speed.

References

[1] Duan, J., Li, X., & Cao, S. (2020). Application of Mathematical Modeling Thinking in Economics Based on Advanced Mathematics. Modern Marketing: management Edition, (2), 81.

[2] Zhao, T. (2020). Design of Logistics Distribution Network in Liaodong Bay Living Area. China Storage and Transportation, (3), 119-121.

[3] Wang, Y., & Wang, Z. (2020). Construction of Mathematical Models for Underground Logistics System Networks. Journal of Engineering Mathematics, 37(6), 664-672.
[4] Wen, Y. (2020). Analysis and Optimization of Regional Logistics Network Structure [D]. Changsha: Hunan Normal University.
[5] Cao, Y. (2017). Application of Mathematical Modeling in Teaching of Logistics Management. Journal of Chifeng College: Natural Science Edition, 33(5), 214-215.
[6] Li, Q. (2021). Research on Construction of Regional Logistics Network Based on Gravity Model [D]. Taiyuan: North University of China.
[7] Xiao, J., Huang, Y., Li, J., & Wang, X. (2005). Vehicle Routing Problem in Logistics Distribution Based on Discrete Particle Swarm Optimization. Systems Engineering, (04), 97-100.
[8] Le, Y., Zhou, L., Yue, Q., & Sun, Q. (2006). An Improved Ant Colony Algorithm for Solving Logistics Distribution Path Optimization Problem. Computer Integrated Manufacturing Systems, (06), 905 -910.
[9] Li, Y., Hu, X., & Xiong, Y. (2006). Simulation Optimization of Vehicle Routing Problem in Logistics Distribution System and Its Progress. Management Science, (04), 2-9.
[10] Liu, B., & Gao, S. (2008). Research on Algorithm of Distribution Vehicle Routing Optimization Problem. Business Economics, (12), 31-33+119.
[11] Jiang, L., & Shen, G. (2010). Research on Optimization of Logistics Distribution Vehicle Routing Problem Based on Ant Colony Algorithm. Journal of Capital University of Economics Journal of Capital University of Economics and Business, (01), 71-74.