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

DOI: 10.23977/acss.2024.080315
ISSN 2371-8838 Vol. 8 Num. 3

Analysis of Network Movement Optimization Model Based on Time Series Forecasting and Multi-Objective Integer Optimization

Weiyi Zhang[1,*]

[1]*School of Mathematics and Physics, Southwest University of Science and Technology, Mianyang, China*

[*]*Corresponding author: zwy1063208702@outlook.com*

Keywords: ARIMA, BP Neural Network, Genetic Algorithm

Abstract: This study is devoted to the problem of predicting and optimizing the cargo volume of routes in e-commerce logistics networks. By establishing ARIMA time series and BP neural network prediction models, combined with weighted summation method to accurately predict three lines. A multi-objective integer optimization model is proposed and solved using a genetic algorithm to achieve the adjustment and optimization of daily route capacity in the case of logistics site closure. This study provides an effective network transfer optimization scheme for large logistics companies, which is expected to improve logistics efficiency, reduce costs, meet the challenges of emergencies, and promote the healthy development of the e-commerce logistics industry.

1. Introduction

With the rapid rise of the e-commerce industry, logistics, as an important part of it, faces the challenge of improving efficiency and reducing costs [1]. The e-commerce logistics network is complex, consisting of multiple logistics sites and transportation routes, and its operation is affected by factors such as holidays, promotional activities, and unexpected events [2]. Against this background, this study focuses on a large logistics company and aims to develop a line-volume forecasting model to cope with possible logistics site closures.

In this paper, we delve into the prediction and optimization of the network shifting problem, using Python, MATLAB and other tools to construct a time series prediction model and a multi-objective integer optimization model, and combining genetic algorithms to achieve the optimization goal. To address the research problem, we successfully built ARIMA time series and BP neural network forecasting models to predict the cargo volume of three important routes through weighted summation. Further, a sub-objective function is proposed based on the constraints, and a genetic algorithm is used to solve the multi-objective integer optimization problem to achieve the adjustment and optimization of daily route capacity. The results of the study show that 93.3% of the line capacity operates normally, 26,995 lines experience changes in cargo volume, and 518,074 pieces of cargo fail to flow normally. This study provides useful practical guidance for improving logistics efficiency and coping with unexpected situations.

2. Line cargo volume forecasting model

2.1 Principles of ARIMA and BP neural networks

ARIMA model (Autoregressive Integrated Moving Average model), i.e., Autoregressive Integrated Moving Average Model, also known as Integrated Moving Average Autoregressive Model, is one of the methods for time series forecasting analysis. ARIMA is built on the basis of stable time series, so the stability of the time series is an important prerequisite for modeling [3]. If the time series is unstable, some operations can be performed to stabilize the time series (e.g., logarithmic, differential), and then the ARIMA model is used for forecasting, to obtain the predicted results of the stabilized time series, and then the inverse of the previous stabilized operations (exponential, differential inverse operation) is performed on the predicted results, to get the predicted results of the original data. The procedure is as follows [4].

ARIMA (p, d, q) model:

$$(y')_t = \alpha_0 + \sum_{p}^{i=1} \alpha_i (y')_{t-i} + \epsilon_t + \sum_{i=1}^{q} \beta_i \epsilon_{t-i} \tag{1}$$

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

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

Where p is the number of lags in the time series data itself used in the prediction model, also known as the AR/Auto-Regressive term. d is the order of the time-series data to reach stable discretization, also called Integrated term. q is the number of lags in the prediction error used in the prediction model, also known as the MA/Moving Average term.

The advantage of ARIMA is that the model is very simple and requires only endogenous variables and no other exogenous variables. However, the disadvantages are significant.

It requires the time series data to be stationary, or stable after differencing. If the data is not stable, it is impossible to capture patterns. For example, stock data cannot be predicted by ARIMA because it is unstable and often fluctuates due to policy and news.

ARIMA can essentially capture both linear and nonlinear relationships, but when nonlinear relationships exist in the data, the performance of the ARIMA model may be limited.

BP Neural Network is a common feed-forward artificial neural network and a neural network trained based on the backpropagation algorithm [5]. It is a supervised learning method, widely used in pattern recognition, classification, regression and other tasks. It consists of multiple layers of neurons, including input, hidden and output layers. Each layer contains multiple neurons, which are connected by connections with weights and activated by an activation function.

As mentioned before, ARIMA and BP neural network models have their own advantages and disadvantages, but due to the advantages of processing linear and nonlinear models, respectively, there is a complementary advantage between them, therefore, the combination of the two for cargo forecasting may receive better results.

Assuming that the prediction value of the ARIMA model is $\hat{y}_{\text{ARIMA}}(t)$ and the prediction value of the BP neural network model is $\hat{y}_{BP}$, this paper proposes to take the following steps to construct the combined prediction model.

(1) Using ARIMA model to predict the whole, let the prediction result be y_{ARIMA}, the error is,

$$e_{ARIMA}(t) = \frac{y(t) - \hat{y}_{ARIMA}(t)}{n} \tag{4}$$

Where $y(t)$ is the actual observation and $\hat{y}_{ARIMA}(t)$ is the prediction of ARIMA model.

(2) The BP neural network model is used to predict the whole, and the prediction result is y_{BP}, and the error is,

$$e_{BP}(t) = \frac{y(t) - \hat{y}_{BP}(t)}{n} \tag{5}$$

Where $y(t)$ is the actual observation and $\hat{y}_{BP}(t)$ is the prediction of BP neural network model. $e_{ARIMA}(t)$ and $e_{BP}(t)$ denote the errors of the ARIMA model and the BP neural network model at time t, respectively.

(3) The results of ARIMA and BP neural network prediction are weighted according to the weight of the error, and the prediction results are obtained as follows.

$$\hat{y} = \frac{e_{ARIMA}(t)}{e_{\mathrm{ARIMA}}(t) + e_{BP}(t)} \times \hat{y}_{BP} + \frac{e_{BP}(t)}{e_{\mathrm{ARIMA}}(t) + e_{BP}(t)} \times \hat{y}_{BP} \tag{6}$$

2.2 Predicted results

According to the above principle, the experiment is carried out as follows.

ARIMA model requires the series to satisfy the stability, check the ADF test result, according to analyze the t-value, analyze whether it can override the assumption that the series is not stable ($P<0.05$). Table.1 shows the results of the ADF test. The model requires the series to be stable time series data. The original hypothesis that the series is not stationary can be disproved by the table.

Check the data before and after differentiation to determine whether it is smooth (not much fluctuation up and down), and at the same time bias the time series (autocorrelation analysis) and estimate its p and q values according to the truncation.

Table 1: ARIMA+BP prediction results

ADF checklist							
Variable	Differential order	t	P	AIC	Threshold value		
					1%	5%	10%
CD14-CD10	1	-3.107	0.026**	15810.238	-3.44	-2.866	-2.569
	1	-9.15	0.000***	15794.263	-3.44	-2.866	-2.569
	2	-10.528	0.000***	15840.845	-3.44	-2.866	-2.569

Note: ***, **, * represent 1%, 5%, 10% level of significance respectively.

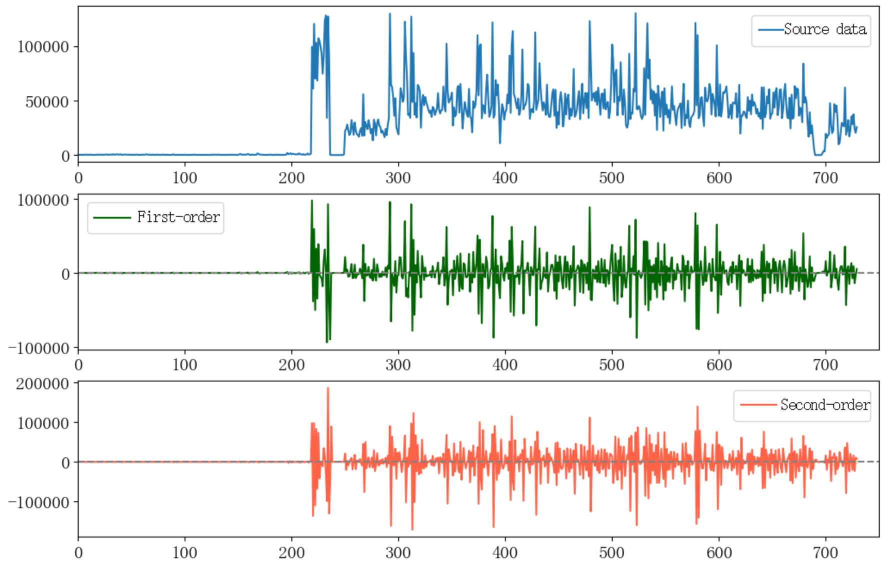

Figure 1: 1st order and 2nd order difference sequence plot

Figure 1 illustrates the autocorrelation diagram (ACF) including the coefficients, upper confidence limit and lower confidence limit.

The horizontal axis represents the number of delays, and the vertical axis represents the autocorrelation coefficient.

If both the autocorrelation and partial autocorrelation plots are trailing, the most significant order (smallest value) in the PACF and ACF plots can be combined as the p and q values.

Truncation is when the ACF or PACF is constant at zero (or fluctuates randomly around zero) after a certain order within the confidence interval.

Trailing is that within the confidence interval, ACF or PACF always has non-zero values, and does not show a constant value equal to zero (or fluctuates randomly around 0) after a certain order.

The error value $e_{BP}(t)$ of the BP neural network is shown in Figure 2.

The prediction results are shown in Figure 3 and Table. 2.

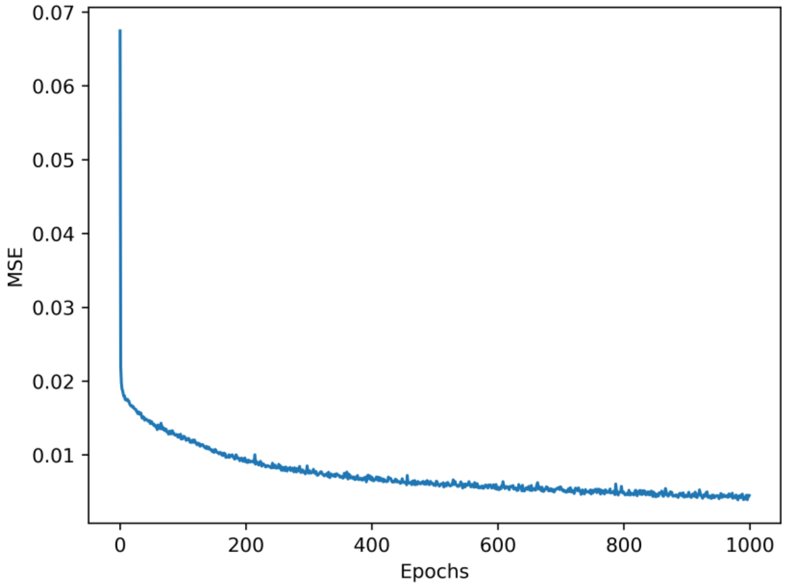

Figure 2: BP error values

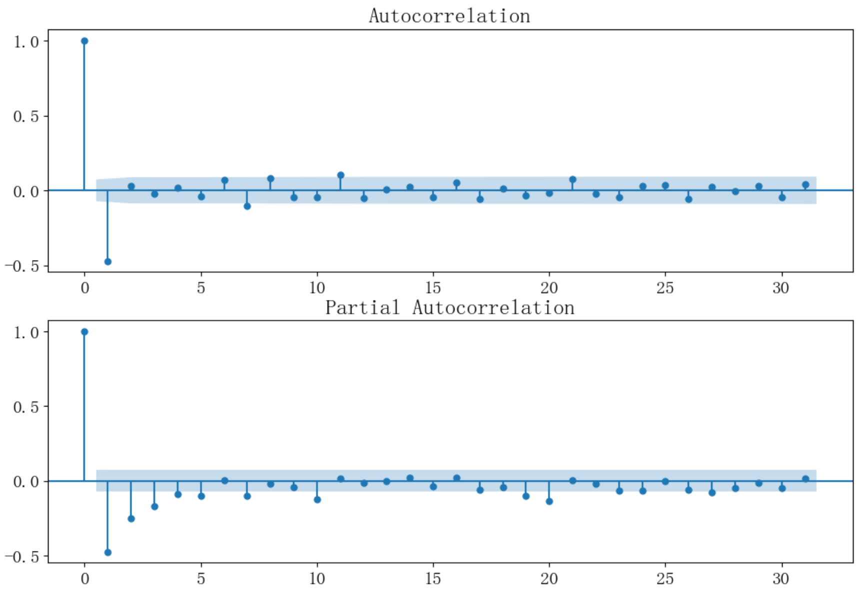

Figure 3: ACF and PACF after first-order differencing

Table 2: ARIMA+BP prediction result

Date	DC14-DC10	DC20-DC35	DC25-DC62
January 1, 2023	27011	80	11452
January 2, 2023	27087	83	12202
January 3, 2023	27226	87	12649
January 4, 2023	27340	88	11299
January 5, 2023	27442	89	9774
January 6, 2023	27536	90	9084
January 7, 2023	27624	90	9710

3. Line cargo adjustment model

3.1 Model building

Assuming that the DC5 node will be shut down on January 1, 2023, and based on the predicted data, a mathematical model is built to solve the problem of: (1) redistributing the volume of shipments in such a way that all packages flow as normally as possible.

(1) All packages will flow as normally as possible after reallocation.

(2) Minimize the number of lines with changes in volume before and after the closure of DC5.

(3) Keeping the daily accumulated total of parcels on each line as small as possible.

(4) Keep the workload of each line as balanced as possible.

Based on these four points, we establish a multi-objective integer optimization model, and take the above constraints as sub-objectives, quantify them and obtain our objective function by weighted sum [6].

$$\min \mathrm{f} = \omega_1 \times \left(\sum_{i=1}^{n}\sum_{j=1}^{31} \mathrm{sign}\left|y_{ij} - x_{ij}\right|\right) + \omega_2 \times \left(\sum_{i=1}^{n}\sum_{j=1}^{31}\left(y_{ij} - \frac{1}{n}\left(\sum_{i=1}^{n} y_{ij}\right)\right)^2\right) + \omega_3 \times \max\left(0, z_j - \sum_{i=1}^{n} m_i\right) \tag{7}$$

The first term of the objective function describes the number of routes that are transformed after adjusting the capacity, and the second term describes the variance of the adjusted capacity of each route, in order to describe the workload of each route is balanced as much as possible. The third term, $z_j - \sum m_i$, describes the number of shipments that did not flow properly on route i on day j. The presence or absence of this term indicates whether or not the shipments flowed properly.

$\omega_1, \omega_2, \omega_3$ denote the weights of the three constraints, we believe that the first thing to do is to ensure the normal operation of the package goods, so the weight of the third item is greater than the first two items, so we finally determine the weights are: 0.3, 0.3, 0.4.

The constraints are that the adjusted capacity is greater than the original capacity and less than the historical maximum capacity and is an integer.

The final model is.

$$\min \mathrm{f} = 0.3 \times \left(\sum_{i=1}^{n}\sum_{j=1}^{31} \mathrm{sign}\left|y_{ij} - x_{ij}\right|\right) + 0.3 \times \left(\sum_{i=1}^{n}\sum_{j=1}^{31}\left(y_{ij} - \frac{1}{n}\left(\sum_{i=1}^{n} y_{ij}\right)\right)^2\right) + 0.4 \times \max\left(0, z_j - \sum_{i=1}^{n} m_i\right) \tag{8}$$

It is constrained as follows.

$$
\text{s.t.} \begin{cases} 0 \leq i \leq n \\ 0 \leq j \leq 31 \\ x_{ij} \leq y_{ij} \leq m_i \\ i, j, x_{ij}, y_{ij}, m_i \in \mathbb{Z}^+ \end{cases} \tag{9}
$$

3.2 Model solving

The heuristic algorithm, Genetic Algorithm, is used to solve the optimization problem. We call the genetic algorithm GA in MATLAB to solve the problem, where we set the population size to 50 and the maximum number of iterations to 1,000,000, etc. The optimal solution y_{ij} is obtained as a matrix.

Genetic algorithm (GA) is a global adaptive probabilistic search algorithm based on the principles of natural selection and genetic genetics, which draws on the natural selection mechanism of biological evolution and the genetic mechanism of genetic recombination and mutation in biological reproduction and evolution [7]. The genetic algorithm is used here mainly to avoid the situation that multiple AGVs go to the same storage node at the same time [2][3][4].

The total daily cargo volume of y_{ij} is obtained from the solution, which is compared with the sum of the historical maximum values of each route, if the former is larger than the latter, it means that the cargo volume of this day cannot flow normally, and the specific difference is the total cargo volume that cannot flow.

3.2.1 Cargo failed to flow normally

After calculation, we get in January 1-31, there are January 9 and January 11 two days of the total amount of goods cannot be normal flow, accounting for 6.7%, the normal flow accounted for 93.3%.

The load situation of the network is shown in Figure 4.

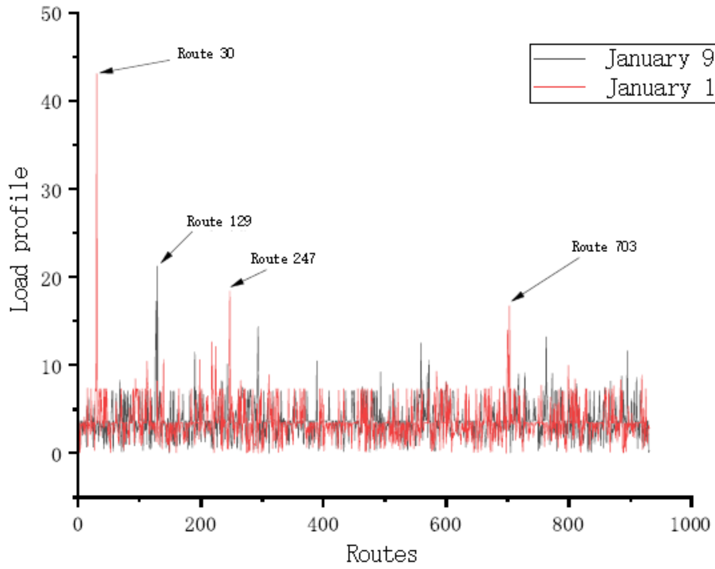

Figure 4: Network load

It can be seen that lines 30, 247, and 703 were more heavily loaded on January 11, and line 129 was more heavily loaded on January 9.

3.2.2 Cargo flow normally

Except for January 9th and 11th, the goods could circulate normally during the other days, so the number of changes in the goods due to the shutdown of DC5 is shown in Table. 3.

Table 3: Normal circulation

1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.10	1.11	1.12	1.13	1.14	1.15	1.16	1.1
931	931	931	930	931	931	931	931	931	NUM	931	931	930	930	931	931
1.17	1.18	1.19	1.20	1.21	1.22	1.23	1.24	1.25	1.26	1.27	1.28	1.29	1.30	1.31	1.17
931	931	931	931	931	931	931	931	931	931	931	931	930	931	931	931

Among them, there is one line unchanged on 4, 14, 15, and 29, and the others are all 931.

The details for January 1, January 15 and January 31 are shown here in Figure 5.

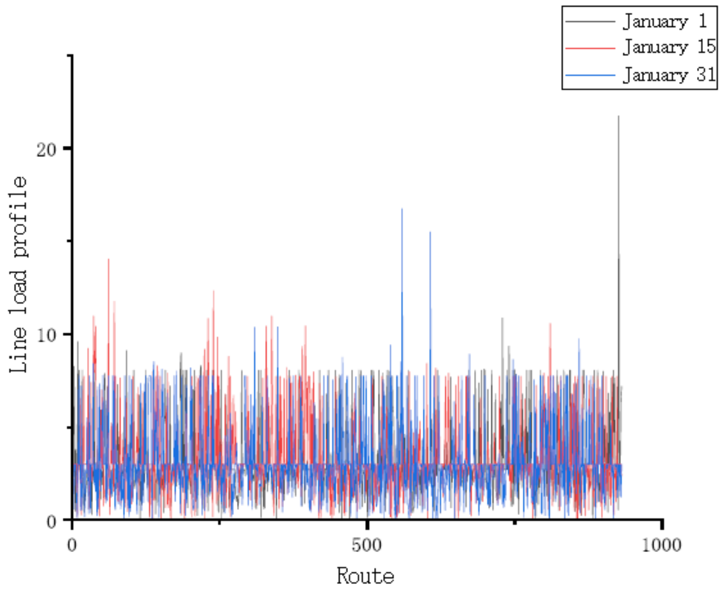

Figure 5: Daily line loadings

Figure 5 shows that most of the lines have a balanced load, and only some lines are overloaded.

4. Conclusions

Through this study, we successfully established a time series prediction model and a multi-objective integer optimization model and applied genetic algorithms to effectively solve the network scheduling problem. For line volume prediction and optimal scheduling, we combined ARIMA and BP neural network models to achieve accurate prediction of the volume of three key lines. For optimal scheduling, we proposed a multi-objective function based on constraints and obtained the best adjustment scheme for daily line capacity by genetic algorithm.

The results of the study show that 93.3% of the line capacity was able to operate normally, 26,995 lines experienced changes in cargo volume, and 518,074 pieces of cargo were not able to flow normally. These results provide practical guidance and decision support for improving logistics efficiency and responding to emergencies. In the future, we will further optimize the model to improve the prediction accuracy and optimization effect, in order to respond to the rapidly changing needs of the e-commerce logistics industry and continue to promote the development and innovation of the logistics industry.

References

[1] Schöder D, Ding F, Campos J K. The impact of e-commerce development on urban logistics sustainability[J]. Open

Journal of Social Sciences, 2016, 4(3): 1-6.

[2] Nie Y, Chen X, Wu X, et al. Emergency Transportation and Structural Prediction and Optimization of Goods Based on ARIMA and Multi-objective E-commerce Logistics Network Model[C]//2023 International Conference on Applied Intelligence and Sustainable Computing (ICAISC). IEEE, 2023: 1-8.

[3] Dimri T, Ahmad S, Sharif M. Time series analysis of climate variables using seasonal ARIMA approach[J]. Journal of Earth System Science, 2020, 129: 1-16.

[4] Stage F, Statements U A P. The ARIMA Procedure [J]. SAS Institute Inc., Cary, NC, USA, 2014.

[5] Cilimkovic M. Neural networks and back propagation algorithm [J]. Institute of Technology Blanchardstown, Blanchardstown Road North Dublin, 2015, 15(1).

[6] Li K , Wang J , Yang L ,et al.Optimization of Medium Voltage Distribution Network Based on Multi-scenario Survivability Analysis[J].Automation of Electric Power Systems, 2014, 38(1):34-37+81.DOI: 10.7500/ AEPS20130218005.

[7] Haldurai L, Madhubala T, Rajalakshmi R. A study on genetic algorithm and its applications[J]. International Journal of computer sciences and Engineering, 2016, 4(10): 139.