Journal of Computer Science and Electrical Engineering
ISSN: 2663-1946
DOI: 10.61784/jcseev6n2100

RESEARCH ON THE PREDICTION OF EMERGENCY TRANSPORTATION OF E-COMMERCE LOGISTCIS PARCELS BASED ON ARIMA TIME SERIES

Rong Rong*, Shi Hu, Hang Yang, Xu Han
Zhaotong University, Zhaotong 657000, Yunnan, China.
Corresponding Author: Rong Rong, Email: 371295467@qq.com

Abstract: The time series prediction model can be used to predict the future change trend of the series to provide support for decision-making. For the prediction study of multivariate time series, in order to solve the prediction model of the cargo volume of the line and predict the daily cargo volume of each line during January 2023, if the new logistics site and transportation route are added, how to add more forecast is reasonable, and further explore the robustness of the established network. Using ARIMA (time series prediction model) and MATLAB programming, and combined with the topsis algorithm of entropy weight method, the prediction model of e-commerce emergency parcel transportation based on ARIMA time series is constructed to solve the problem of emergency logistics package.
Keywords: ARIMA, Logistics Transportation, Logistics Forecast, Time Series

1 INTRODUCTION

The emergency transportation of e-commerce logistics is a field that has been increasingly valued along with the development of e-commerce, network and economy in recent years. Electricity logistics network between logistics sites and logistics routes, the holidays and "double a", "618", the influence of promotional activities, electricity user order quantity will occur significant fluctuations, and earthquake emergencies or natural disasters lead to temporary or permanent suspension, the processing of the package will be urgently diverted to other logistics sites, these factors will affect the line transport number of parcels, and the number of various logistics site processing package. If you can predict the number of parcels of each logistics site and line, the manager will be able to arrange transportation, sorting and other plans in advance, so as to reduce operating costs and improve operational efficiency. In particular, when some sites are temporarily or permanently suspended, designing the logistics network adjustment scheme based on the forecast results, the processing capacity of each logistics site and the transportation capacity of the route will greatly reduce the impact of the shutdown of the logistics site on the logistics network and ensure the normal operation of the logistics network. This paper discusses how to update and obtain new data information in real time for e-commerce network logistics parcels in the face of emergency, predict the route and place of emergency transportation, and better solve the problem of emergency transportation.

2 CORRELATION STUDIES

In the e-commerce logistics prediction research, Shi Haiyang[1]In view of the logistics distribution decision problem of B2C e-commerce wave orders with dynamic release and strong heterogeneity, a forward dynamic planning method based on timing prediction is proposed to find the optimal strategy. Wang Yun et al[2]To study the sharing strategy of demand forecast information under the third-party logistics distribution, the problem of demand forecast information sharing of platform providers under the three market power structures of suppliers, platforms and logistics providers is proposed. Yin Xueming et al[3]Use the machine learning method to optimize the network of the export e-commerce logistics. Wang Jianguo[4]Logistics the logistics studied based on the time series. Draw lessons from the research results at home and abroad, this paper puts forward based on ARIMA time series of electricity logistics package emergency dispatching prediction model, through the thorough analysis of historical logistics data, predict the future period of time logistics package flow trend and demand, provide electricity enterprises and logistics enterprises with scientific decision -making basis, better grasp the dynamic change of logistics demand.

Time series prediction is a means analysis method to predict real-time data in the future period of time based on the previous data trend and cyclical development. It is widely used in predicting e-commerce logistics, meteorology, social and economic development, transportation and other aspects. When using this kind of method, Li Zhaoxi, Liu Hongyan, etc[5]In multivariate time series prediction methods incorporating global and sequence features, time series prediction models are able to estimate the future trends of sequence. Liu He et al[6]A yield prediction method based on multivariate time series model, namely multivariate long and short-term memory neural network (LSTM) is proposed. Jiang Qi et al[7]Construct the US dollar exchange rate forecast based on the ARIMA model. Therefore, this paper draws on the application of time series model in other fields, and establishes the prediction model of e-commerce logistics parcel emergency transportation based on ARIMA time series, so as to provide scientific and accurate prediction for logistics demand.

3 MODEL ESTABLISHMENT

3.1 Data Selection and Preprocessing

The cargo volume data transferred between different logistics sites during 2021-01-01 to 2022-12-12-31. The logistics network has 81 logistics sites and 1049 lines. The lines are oriented, such as line DC1 DC2 and line DC2 DC1 are considered to be two or two lines. It is assumed that the processing capacity of each logistics site and the maximum transportation capacity of each route are its historical maximum. Through the real-time data analysis in 2023, how to establish new routes and solve the important assessment of e-commerce logistics network for different logistics sites and routes. If the new logistics sites and transportation routes are added, the prediction of how to add is the network robust.

Before the prediction and solution of the model, the data is analyzed first. It is necessary to predict the number of parcels in each logistics site and line, so that managers can arrange transportation, sorting and other plans in advance, so as to reduce operating costs and improve operational efficiency. In particular, when some sites are temporarily or permanently suspended, designing the logistics network adjustment scheme based on the forecast results, the processing capacity of each logistics site and the transportation capacity of the route will greatly reduce the impact of the shutdown of the logistics site on the logistics network and ensure the normal operation of the logistics network.

3.2 Method Description

Time series prediction is a means analysis method to predict the real-time data in a period of future based on the previous data trend and periodic development. It is widely used in forecasting e-commerce logistics, meteorology, social and economic development, transportation and other aspects. The ARIMA model, the full name differential auto regressive moving average model, is one of the important methods used for time series prediction. It combines the characteristics of auto regressive (AR) and moving average (MA), and transforms the non-stationary time series into the stationary time series through the differential operation. The ARIMA model essentially consists of three parts: AR (p-order auto regressive model) + I (i-order difference) + MA (q-order moving average model). In the use of the time series ARIMA model for prediction, the entropy weight method topsis algorithm and the MATLAB transformation can be used to solve and build the data decision tree to establish the ARIMA model. As shown in the equation (1) and equation (2).

AR part: $y\_t = \phi\_1 y\_\{t-1\} + \phi\_2 y\_\{t-2\} + ... + \phi\_p y\_\{t-p\} + \varepsilon\_t$ （1）

MA part: $\varepsilon\_t = \theta\_1 \varepsilon\_\{t-1\} + \theta\_2 \varepsilon\_\{t-2\} + ... + \theta\_q \varepsilon\_\{t-q\} + \omega\_t$ （2）

ϕ and θ is the parameter of the model, ε is the error term, ω is the white noise. The formula of the model can be expressed as follows equation (3) :

$$\left[\backslash phi(B)(1-B)\hat{}\, dY\_t = \backslash theta(B) \backslash epsilon\_t \right] \quad (3)$$

Where: (\ phi (B)) and (\ theta (B)) are polynomials for the lag operator B, representing the autoregressive and moving average parts, respectively.(P) is the number of autoregressive terms, which determines the order of (\ phi (B)).(D) is the difference order number used to convert non-stationary time series into stationary time series.(Q) is the number of moving average terms that determines the order of (\ theta (B)).(Y_t) is the value of the time series at time point (t).(\ epsilon_t) is the white noise sequence, namely the error term.

① The autoregressive part (AR): (\ phi (B)) describes the data relationship between the current time series and the past values. By fitting the autoregressive terms, the model is able to capture the autocorrelation of the time series.

② Differential part (I): ((1-B) ^ d) is a differential operation used to eliminate trends and seasonality in the time series. The choice of the difference order (d) plays a crucial role in ensuring the smoothness of the time series.

③ The moving average part (MA): (\ theta (B) \ epsilon_t) describes the relationship between the current error term of the time series and its past error term. The moving mean term is able to capture the random fluctuations in the time series.

3.3 Site Importance Considerations

Based on the importance of different logistics sites to the route. The Topsis algorithm of entropy weight method is used to select the forward treatment of the line importance index.

Very large index (benefit index), as equation (4):

$$X_{ij} = \frac{X_j - X\min}{X\max - X\min} \quad (4)$$

Minimization index (cost-type index), as equation (5)::

$$X_{ij} = \frac{X\max - X_j}{X\max - X\min} \quad (5)$$

The importance of evaluating the different sites considers the following factors:

①Total amount of goods delivered by the site e1: The more goods delivered by the site, the higher the ability to deliver the goods, and the higher the importance;
② Total amount of goods accepted by the site e2: The more the total amount of goods received by the site, the site has a strong ability to store goods and the higher its importance;
③ Times of goods sent by the site e3: the more times of goods sent by the site, the frequency of goods sent by the site is high and its importance is high;
④ Number of goods received by the site e4: the more times of goods received by the site, the frequency of goods received by the site is high and its importance is high.
⑤ Average e5 of the site: the higher, the higher the importance.
The calculated results of the Topsis evaluation algorithm are shown in Table 1.

Table 1 Results of the Topsis evaluation algorithm

Index Value	Positive ideal solution distance （D+）	Negative ideal solution distance （D-）	Comprehensive score index	Serial number
DC1	0.99997881	0.00005372	0.00005732	78
DC10	0.4090077	0.78455619	0.65732233	1
DC11	0.99372429	0.01556041	0.01541727	62
DC12	0.93759114	0.10666566	0.10214504	41
DC13	0.99632336	0.01291416	0.01279596	63
DC14	0.41989102	0.76842821	0.64665133	2
DC15	0.92763763	0.23743553	0.20379449	22
DC16	0.99310733	0.01855573	0.0183418	60
DC17	0.87502927	0.31842096	0.26680707	10
DC18	0.9828484	0.05318883	0.05133873	48
DC19	0.88030704	0.31033595	0.26064568	11
DC2	0.999466	0.00191081	0.00190818	71
DC20	0.90628994	0.26124162	0.22375551	16
DC21	0.89833688	0.23882336	0.21001733	19
DC22	0.88738109	0.29777126	0.25125146	13

4 ARIMA MODEL PREDICTION AND ERROR ANALYSIS

4.1 Prediction Results

The analysis method of the mathematical chart two-dimensional model was used, using the MATLAB model to predict the time series table, and the various values were analyzed in the time series table. Write down the analysis and comparison chart of the front information data in the analysis table, predict whether it is more than stable (the fluctuation ratio of before and after the data is limited), bias the time series (autocorrelation analysis), and estimate the p and q values according to the censoring situation. Analysis was combined with the information criterion AIC and BIC values (the lower the better), and the comprehensive analysis was combined with the time series analysis map to obtain the order results of backward prediction.
Time series analysis (ARIMA) is based on the historical period data: the goodness of fit of the model R^2 is 0.621, and the model performs well. The forecast results for the next 10 phases are respectively table 2.

Table 2 Predicted the results

Order (time)	Forecast results
1	25909.51
2	25944.53
3	25979.56
4	26014.58
5	26049.60
6	26084.62
7	26119.64
8	26154.67
9	26189.69
10	26224.71

The prediction results of line DC14-DC10, DC20-DC35, DC25-62 are shown in Figure 1, 2 and 3.

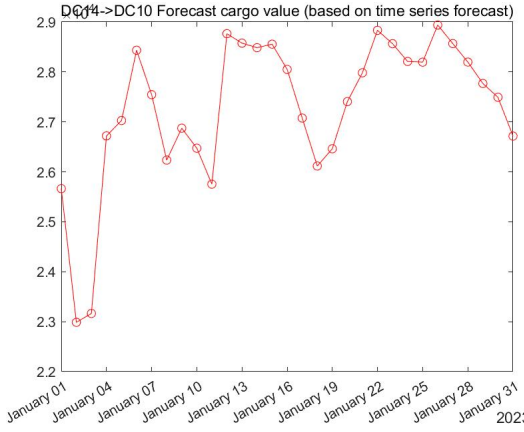

Figure 1 Prediction results of DC14-DC10

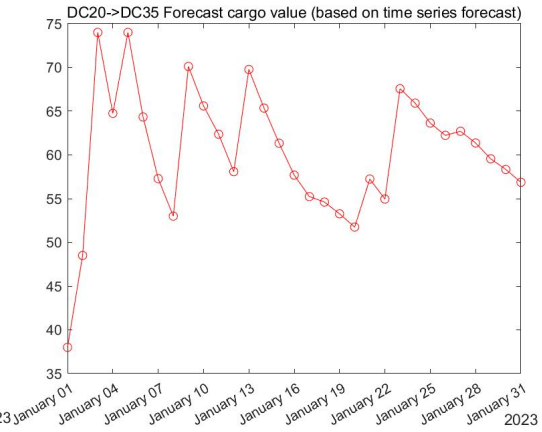

Figure 2 Prediction results of DC20-DC35

In January 2023, the cargo volume forecast and analysis value of DC14 DC10 and DC20 DC35 will be transported online. It can obviously get the forecast value of the goods on the line site in that month, and the fluctuation of the quantity of goods has obvious fluctuation, and the forecast value of the goods out of the line when excluding various assumptions. From the analysis of the data graph in Figure 3, it can be concluded that the cargo volume of the site was significantly low before the beginning of the month, and gradually increased later, with obvious fluctuation. From January 16, the fluctuation of the predicted cargo value was significantly reduced.

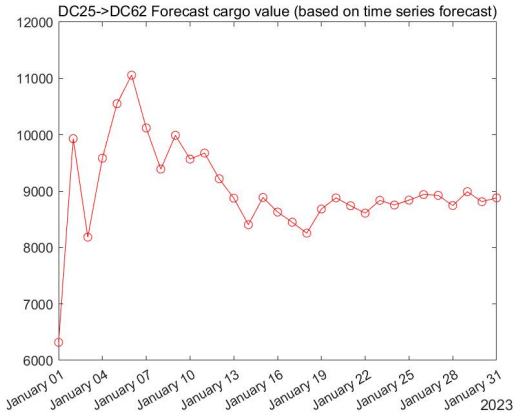

Figure 3 Prediction results of DC25-DC62

4.2 Error Analysis

To find the optimal parameters for the relevance criterion of AIC information, the result of the model is the validation of the ARIMA model as shown in Figure 4. Model parameters are shown in Table 3.

Figure 4 Results plot of model prediction fit and actual values

Table 3 Table of model parameters

ARIMA Model (0,1,1) test table		
Item	symbol	price
sample size	Dfresiduals	723
	N	726
Q statistics	Q6 (P value)	0.017(0.897)
	Q12 (P-value)	1.75(0.941)
	Q18 (P-value)	17.249(0.140)
	Q24 (P-value)	30.989(0.029**)
	Q30 (P-value)	40.377(0.019**)
Information criteria	AIC	16247.18
	BIC	16260.938
goodness of fit	R2	0.621

The results of this test include the number of samples, degrees of freedom, Q statistic and the goodness of fit of the information data model.① The ARIMA model requires that there is no autocorrelation coefficient for the residual of the model (the residual is white noise (P> 0.1)), and the model rechecks the model white noise from the P-value of Q quantity. The information criterion suggests that the two values of AIC and BIC can be used for multiple analytical model comparisons, and naturally the lower the better. R^2 represents the degree of fit of the time series, and the closer to the 1 effect, the better the performance.

5 CONCLUSION

This paper combines the ARIMA model, and discusses the feasibility of predicting the time series model. After parameter optimization and error analysis, the ARIMA model can better reflect the changing trend of the number of parcels transported by lines. This model can grasp the supply and demand relationship of the future market to a certain extent based on the historical data, and provide a scientific basis for enterprises and related departments to make strategic decisions. Thus, when using the huge number of sample data processing model for decision making, on the one hand can use big data analysis technology based on full sample prediction analysis, on the other hand can combine neural network, deep learning, such as artificial intelligence technology, get deeper information from the data, improve cognitive computing ability of data. In addition, with the continuous development of the e-commerce logistics industry, new influencing factors and challenges continue to emerge, which need to constantly update and improve the prediction model. In the future, more advanced time-series prediction methods and technologies, such as deep learning and machine learning, will be further explored to improve the prediction accuracy and efficiency. At the same time, we should pay attention to the new trends and new problems of the industry development, and constantly expand the research field and deepen the research content.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

The Scientific Research Fund of Yunnan Provincial Department of Education in 2024 (2024J1064).

REFERENCES

[1] Shi Haiyang, Sun Lijun, Hu Xiangpei. The decision method of B2C e-commerce orders considering the dynamic release of wave orders. Journal of Management Engineering, 2024, 38 (2): 152-165.
[2] Wang Yun, Zhao Jing, Zhang Fang. Research on the information sharing strategy of demand prediction under the third-party logistics distribution. Journal of Harbin University of Commerce (Natural Science Edition), 2023, 39 (4): 502-512.
[3] Yin Xueming, Wang Changjun. Machine learning-driven export cross-border e-commerce supply chain network optimization. Journal of Donghua University (Natural Science Edition), 2023, 49 (5): 162-170.
[4] Wang Jianguo. Research and application of logistics scheme based on time series. Yunnan Metallurgy, 2023, 52 (5): 178-182.
[5] Li Zhaoxi, Liu Hongyan. Multivariate time series prediction methods incorporating global and sequence features. Journal of Computer Science, 2023, 46 (1): 70-84.
[6] Liu He, Li Yanchun, Du Qinglong, et al. A high water content stage yield prediction method based on a multivariate time series model. Journal of China University of Petroleum (Natural Science Edition), 2023, 47 (5): 103-114.
[7] Jiang Qi, Liu Yongwen. The lar exchange rate forecast based on the ARIMA model. Economic Research Guide, 2022 (20): 69-71.