A Hybrid Total Logistics Forecasting Method Combined with ARIMA and BPNN

Jianjun Wang1,*, Yaqian Xing1, Shuo Zhang1, Yan Zhou1, Li Li2, Tiantian Nian3

1School of Economics and Management, North China Electric Power University, Beijing, China
2School of Economics and Management, Beijing Information Science & Technology University, Haidian district, Beijing, China
3Jiangsu Pharmaceutical Co., Ltd, Yibin City, Jiangsu Province, China
*Corresponding author’s email: wangjianjunhd@126.com

Abstract. In recent years, with the continuous improvement of China's economic level, Chinese people's living standards are also improving, so that the vigorous development of the e-commerce industry and the number of the total amount of social logistics have increased year by year, which has brought great impetus to the development of the logistics industry. Based on the background, it requires a scientific and reasonable forecasting result of the total amount of logistics for the logistics industry investment. This paper employs an ARIMA and BPNN combined method to forecast the total logistics situation during the "14th Five-Year Plan" period, the results show that the combination forecasting method can take into account the advantages of each single forecast model, and its forecast effect is more ideal.

Keywords. Total logistics forecasting, ARIMA, BPNN, Combined forecasting method

1. Introduction
In recent years, with the stable development of China's economy and the gradual improvement of people's living standards, the domestic e-commerce industry is also increasing rapidly, which makes the social logistics package transportation volume soar, and ranks into the new logistics era of 1 billion levels. All these have given the logistics centers of e-commerce, express distribution centers, and other logistics points severe challenges. There are many kinds of logistics packages and the specifications are different. It is difficult for the traditional logistics system to carry out centralized and batch unified processing. Coupled with the continuous emergence of the huge e-sales transportation problem difficulty, especially in the period of "618", "double 11" and "double 12" of e-commerce festival in China, this problem is more prominent. These changes are not only a sharp increase in the number of packages handled by the logistics industry but also new requirements for high efficiency, low cost, and high flexibility logistics solutions. In this context, it is necessary to reform the logistics system to intelligence and make continuous investment in the logistics system to meet the needs of the logistics business. The total logistics forecast is the basis of decision-making for investment in the logistics system. Scientific and reasonable total logistics forecasts can provide a decision-making basis and reference for the accurate investment of the logistics system. It is very necessary to study the prediction of social logistics total amount.
Many scholars have studied logistics-related forecasting problems. For example, Hu built a BPNN to forecast regional logistics demand [1], Luo Wei built a seasonal sequence forecast model for logistics companies to conduct logistics forecasts [2], Yu used the improved GA model to predict the logistics demand. The prediction results showed that GA can effectively optimize the parameters of the BP network [3]. Zhang built a gray forecast model to predict the logistics demand of Beijing [4], the above forecast models have achieved good prediction results. However, some studies have pointed out that in fact, the total logistics volume is affected by many factors, and accurate prediction of the total logistics volume is a more complex and difficult problem [5]. Therefore, the combination forecasting method for logistics forecasting will effectively improve forecasting accuracy [6]. This paper constructs a combination forecasting model that combines the ARIMA model which has good performance for time-series forecasting problems and the BPNN model which has good nonlinear fitting ability. The combination forecasting model can combine the advantages of each model, so as to achieve a better prediction accuracy.

2. Hybrid total logistics forecasting model combined with the ARIMA and BP

2.1 The ARIMA forecasting model
The annual historical data of total social logistics is typical time-series data. Therefore, in the total social logistics forecasting problems, we should first consider the time series benchmark model ARIMA. ARIMA model is called a differential integrated moving average autoregressive model, also known as an integrated moving average autoregressive model, which is one of the time series forecasting analysis methods. In the ARIMA \((p, d, q)\) model, AR stands for "autoregressive", \(p\) represents the number of autoregressive items; MA stands for "moving average", \(q\) represents the number of moving average items, and \(d\) represents the number of different times (also known as order) to make the data become stationary series. ARIMA model has three basic forms, which are the Autoregressive model (AR), Moving-Average model (MA), and hybrid model (ARIMA: autoregressive moving average).

1) Autoregressive model AR \((p)\) can be expressed as:
If the time series \(\{y_t\}\) meets:

\[
y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t
\]

Where \(\{\epsilon_t\}\) is a sequence of random variables that obey an independent distribution, and meets:

\[
E(\epsilon_t) = 0, \text{Var}(\epsilon_t) = \sigma^2 > 0
\]

Then the time series \(\{y_t\}\) obeys the p-order autoregressive model, which can also be written as \(\phi(B)y_t = y_{t-p}\).

The judgment of the stationary condition is based on the lagged operator polynomial, the formula is \(\phi(B) = 1 - \phi_1 B + \cdots + \phi_p B^p\), if its roots are all outside the unit circle, that is, the roots of \(\phi(B) = 0\) are not less than or equal to 1.

2) The description of the moving average model MA\((q)\) can be expressed as:
If the time series \(\{y_t\}\) is \(y_t = \epsilon_t + \cdots + \theta_q \epsilon_{t-q}\), or can be expressed as \(y_t = \theta(B)\epsilon_t\), then can be described as obeying the q-order moving average model, and it is stable under any conditions.

3) ARIMA \((p, q)\) model:
If the time series \(\{y_t\}\) meet:
\[ y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q} \]  \hspace{1cm} (3)

Then the time series \{y_t\} can be described as following the (p, q)-order autoregressive moving average model. Or denoted as \( \phi(B)y_t = \theta(B)\varepsilon_t \).

When in the special case of \( q=0 \), the ARIMA model is equivalent to AR \((p)\), and when in the special case of \( p=0 \), the ARIMA model is equivalent to MA\((q)\).

According to the above principle, we use the ARIMA forecasting toolbox in SPSS to forecast the total amount of social logistics. First of all, before ARIMA \((p, d, q)\) model, we observe whether the data has seasonal components, so we should make a sequence diagram for observation. Because of the previous steps, the ARIMA model requires that the sequence is stable, so the data should be differentiated and stationarity analysis should be performed to obtain the d value. Observe the autocorrelation graph (ACF) and partial autocorrelation graph (PACF) of the observation sequence to determine the order of ARIMA \((p, d, q)\). The specific determination criteria are shown in Table 1.

| Model         | ACF                      | PACF                      |
|---------------|--------------------------|---------------------------|
| AR\((p)\)     | Attenuation approaches zero | Censored after \(p\)-order |  
| MA\((q)\)     | Censored after \(p\)-order | Attenuation approaches zero |  
| ARIMA\((p, d, q)\) | The attenuation tends to zero after \(q\)-order | The attenuation tends to zero after \(p\)-order |

After the order is determined, the standard for error testing is the significance value of the Ljung-box statistic. This is the test value of the residual error of the random test model. It can indicate whether the specified model is correct. If the significance is less than 0.05, it means that the error of the residual error is not random, which means that there is a structure in the observation time series that cannot be explained by the model. The larger the square of R, the better the fit.

2.2 BPNN forecasting model

The ANN model is the most popular artificial intelligence-based model in forecasting. The advantage of the ANN model is as follows: it can easy to get the nonlinear mapping between the input variables and the output variables; it needn’t too much prior knowledge for the forecasting as so on. In ANN models, the BP training algorithm is widely used, and this type of ANN is named BPNN. The BP training algorithm is an iterative gradient descent algorithm designed to minimize the mean square error between the outputs and the desire values, it uses the gradient-descendent direction to training the weight between the layers. The process is made up of two directions through the layers, one is the forward, and the other is backward. The typical structure of the BPNN is shown in Figure 1.

![BPNN structure of load forecasting](image)

**Figure 1.** The BPNN structure of load forecasting

In BPNN, the input layer is the history load data, such as \( y_t, y_{t-1}, \ldots, y_{t-n} \), the output layer is the forecasting load target \( y_{t+1} \).
the connection weights and node bias will be adjusted iteratively by a process of minimizing the forecasting errors by gradient descent algorithm, the final computational equation of BPNN is [7]

\[ y_t = b_0 + \sum_{j=1}^{q} w_j f(b_j + \sum_{i=1}^{p} w_{ij} y_{t-i}) + \varepsilon_t \]  

(4)

Where \( b_j \) is a bias on the jth unit, \( w_{ij} \) is the connection weight between layers, \( f() \) is the transfer function of the hidden layer, \( p \) is the number of input nodes and \( q \) is the number of hidden layer nodes.

2.3 Hybrid forecasting model

Because ARIMA model can effectively predict the time series, can better capture the linear correlation information of time series itself, and BP neural network can carry out non-linear fitting with arbitrary precision, therefore, this paper combines the two models in the principle of algorithm, which can better reflect the advantages of each model and supplement the disadvantages of each model. Its algorithm framework is shown in Figure 2. It is shown as follows:

![Figure 2. The ARIMA cooperate with Artificial Neural Network approach frame](image)

3. Case study

3.1 ARIMA prediction results

According to the above principles, ARIMA is used to predict the total amount of social logistics in China. It can be obtained from SPSS that ARIMA (2,2,2) is the best fitting. The fitting coefficient is 0.996, which is significant at the significance level of 96.9%. It can be seen from Figure 3 that the average error of the whole forecast model is 4.58%, and the fitting situation is good.

Using the ARIMA model mentioned above, we can get the forecast value of the total amount of social logistics in China from 2019 to 2025, and the results are shown in Table 2.

| Year | Predictive value |
|------|------------------|
| 2019 | 315.3            |
| 2020 | 348.9            |
| 2021 | 384.8            |
| 2022 | 423.4            |
| 2023 | 464.7            |
| 2024 | 508.8            |
| 2025 | 555.9            |

3.2 BPNN prediction results

Based on the above-mentioned BPNN principle, the neural network toolbox in MATLAB is used to predict the total amount of social logistics. The parameters set include the maximum number of training epochs \( \text{iter}=1000 \), the learning rate \( \text{lr}=0.02 \). The \text{str2double} function is used to convert the string type data into numerical data. The system uses \text{newff} function to create a BP neural network according to the obtained parameters. After many experiments, the input variable is set as the total value of the national social logistics of the year and the previous year, and the double hidden layer is set. Each hidden layer node number is 10. After calculation, the neural network prediction result graph as shown in Figure 3 can be obtained.
From the above sample of historical data, it can be seen that the error of BPNN is 1.84%, which is relatively accurate. According to the BPNN model, we can get the forecast value of the total social logistics volume of China from 2019 to 2025, as shown in Table 3.

Table 3. BPNN prediction value of total national social logistics from 2019 to 2025

| Year | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 | 2025 |
|------|------|------|------|------|------|------|------|
| Predictive value | 291 | 327.4 | 335.6 | 366 | 375.6 | 394.2 | 403.8 |

Figure 3. Prediction results of ARIMA and BPNN

3.3 Combined forecast results
According to the results of the two forecast models, the total amount of social logistics will show a steady upward trend in the next seven years. The average error of the two forecasting methods is less than 5%, showing good prediction ability. The prediction effect of BPNN is better than that of ARIMA method, but BPNN belongs to neural network prediction method, and there may be over fitting phenomenon. The ARIMA method is a classic method of time series. Both models have their own advantages. Therefore, this paper chooses to give the same weight to the initial prediction results of the two methods and carry out composite weighted prediction. Both of them are weighted 0.5, and the new national total social logistics forecast results are shown in Table 4.

Table 4. Prediction value of total national social logistics in China from 2019 to 2025

| Year | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 | 2025 |
|------|------|------|------|------|------|------|------|
| Predictive value | 303.15 | 338.15 | 360.2 | 394.7 | 420.15 | 451.5 | 479.85 |

4. Conclusion
In this paper, the total amount of social logistics is forecasted and analyzed. In the forecasting method, the traditional time series forecasting method ARIMA and the more commonly used BP neural network are used. The former can well capture the relevant information of the time series, while the latter can better fit the nonlinear relationship. The combination of the two methods can improve the prediction accuracy of the total logistics volume, and provide a better basis and reference for logistics related investment. From the forecast results, the annual growth of total social logistics reflects the increase of material flow and logistics commodity value year by year, which is a benign development for the logistics industry and society, and the importance of the logistics industry is becoming more and more prominent.

References
[1] Hu P, Mei T, Liu X L and Fan W L, 2016 A method for forecasting regional logistics demand based on BP neural network (American Applied Sciences Research Institute, USA.
Proceedings of 2016 3rd International Conference on Information and Communication Technology for Education (ICTE 2016)) pp 231-234.

[2] Luo W, and Fang X P, 2016 Research and application of logistics enterprise freight demand forecasting based on seasonal time series vol 5 no 1 (Management Science and Engineering).

[3] Yu G Y. 2016 Study on logistics demand forecasting based on improved adaptive genetic algorithm and bp neural network vol 4 (Hong Kong New Century Culture Press Co., Ltd. International Journal of Intelligent Information and Management Science: Hong Kong New Century Culture Press Co., Ltd.).

[4] Zhang H, Zhu J, Zhou L. 2015 Research on logistics demand forecasting and transportation structure of Beijing based on grey prediction model vol 3 no 3 (Science Journal of Applied Mathematics and Statistics) 2015.

[5] He F B, Chang J. 2014 Combined forecasting of regional logistics demand optimized by genetic algorithm vol 4 no 2 Grey Systems: Theory and Application.

[6] Ghodrati B, and Kumar U, 2005 Operating environment-based spare parts forecasting and logistics: a case study vol 8 no 2 International Journal of Logistics Research and Applications pp 95-105.

[7] Tian X, Lu X and Deng X, 2009 A TEI@I-based integrated framework for port logistics forecasting (2009 International Conference on Business Intelligence and Financial Engineering, Beijing) pp 293-296.