Research on Dynamic Prediction Model of Orders in E-Commerce Distribution Center for Intelligent Scheduling

Yu Wang and Zhuan Wang(✉)

School of College of Mechanical Engineering, University of Science and Technology Beijing, Beijing, China
wangzhuan@ustb.edu.cn

Abstract. Based on the intelligent development and order production of e-commerce distribution center, intelligent scheduling of e-commerce distribution center is imperative. A dynamic prediction model based on BP neural network is proposed, which can quantitatively predict the order quantity between scheduling time and the cut-off time of the current wave, so as to support the decision-making of scheduling mode and scheduling batch. Firstly, a time-series structure model of order quantity with minute as time granularity was established. The time step and statistical time period were used to calculate the input parameters of neural network, and the prediction time range was taken as the output parameters of the neural network. The prediction model was obtained through multiple training, and the corresponding network model of the required prediction period was called to achieve the dynamic prediction target; Secondly, GA (Genetic Algorithm) was used to optimize the weights and threshold values of BP neural network; Finally, taking the daily order data of a pharmaceutical e-commerce distribution center as an example, the prediction model was verified. The result shows that the BP neural network model optimized by GA has better prediction effect, and the accuracy of the model meets the requirement.

Keywords: Order predict · BP neural network · Genetic algorithm · Intelligent scheduling

1 Introduction

In recent years, B2C e-commerce logistics has developed rapidly. E-Commerce orders show the characteristics of multiple varieties, small batch and high timeliness requirements. In order to improve efficiency, e-commerce distribution centers divide orders into multiple distribution waves, each wave corresponds to a set of order execution time points, and each wave order must be put into production according to the execution time. Therefore, for the increasing massive order pool, how to effectively organize production scheduling according to the wave, so that all orders can complete the corresponding operations before the order deadline, and improve the sorting efficiency has become the key to order production scheduling. The traditional e-commerce distribution center

Z. Zeng et al. (Eds.): ECIT 2022, AHE 11, pp. 248–260, 2023.
https://doi.org/10.2991/978-94-6463-005-3_25

adopts the manual order mode, which has some problems, such as passive and inefficient single link, lack of standardization of logic and rules and so on. The intelligent order scheduling of e-commerce distribution center depends on the flexible prediction of future order. Therefore, dynamic prediction of future orders is the key of intelligent production scheduling decision-making of e-commerce distribution center.

Order predicting has always been a hot issue concerned by scholars at home and abroad. Many mature time series models and methods, such as moving average, exponential smoothing and ARIMA model, have been widely used [2, 12]. Due to the strong randomness and discreteness of order arrival, the traditional time series prediction model depends on the time series with linear or known statistical distribution, and its prediction result accuracy is often difficult to meet the practical needs [6]. With the development of artificial intelligence, the order predicting model based on neural network is widely used. The basic BP neural network method is gradually used in predicting research, which has certain advantages. For example, Guo [4] tested and compared the same group of sample data through SVR and BP neural network model, and concluded that BP neural network model has higher prediction accuracy. However, BP neural network may have problems such as over fitting, slow convergence and local minimization. Using optimization algorithms such as particle swarm optimization [9], genetic algorithm [10] and bat algorithm [5] to improve neural network is also a practical, effective and widely used prediction method.

At the application level of order prediction, most studies focus on the sales prediction or demand prediction of enterprise orders. For example, Ge [2] used the grey system theory and BP neural network prediction optimized by genetic algorithm to realize the prediction of sales orders. L.H. Wang [11] and others established a short-term sales volume prediction model applied to e-commerce accounting through AdaBoost and BP. Gong [3] proposed an enterprise order predicting model based on RBF neural network by combining minimum orthogonal multiplication algorithm (OLS) and evolutionary particle swarm optimization (EPSO). In order to solve the problem unreasonable personnel allocation in the storage system caused by the uncertainty of order arrival of e-commerce enterprises, Yang [13] proposed a smooth order prediction model based on Monte Carlo index, applied the predicted value of the order to the picking capacity of the distribution center. Leung [7, 8] used genetic algorithm to support batch decision-making of e-commerce distribution center, and applied AR-ANFIS model to predict the hourly arrival volume of E-Commerce orders. Finally, the overall solution was taken as a decision support system, in order to help decision makers determine when to release batch orders.

To sum up, the application research of order predicting for intelligent scheduling and production optimization of distribution center has just started, it mainly predicts the order volume in a fixed period, but the research on the order prediction in the dynamic time period is rare. The research goal of this paper is to establish the order prediction model in the dynamic time period by using the BP neural network method. The prediction model can provide effective data support for intelligent scheduling mode and batch decision-making, which is a direction worthy of exploration and research.

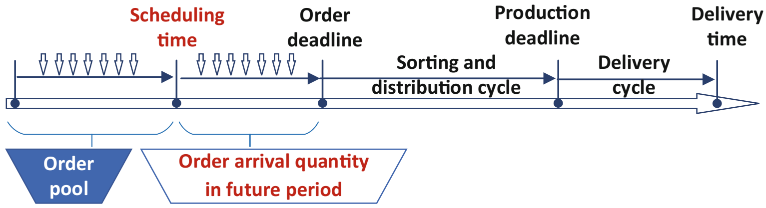

Fig. 1. Wave order execution process

2 Analysis of Order Predicting Problems

2.1 Wave Order Execution Process

The order of the e-commerce distribution center is divided into multiple waves according to the delivery time. Each wave corresponds to one to determine the delivery time, production (sorting and packaging) completion deadline, order deadline and order placement (assigned to the distribution center). The order of each wave should be processed before the production deadline. The time axis of the order execution process is shown in Fig. 1.

2.2 Intelligent Scheduling Principle

In the e-commerce distribution center, the order sorting and distribution efficiency is directly related to the production batch, that is, the larger the batch, the higher the sorting processing efficiency. Therefore, the basic principle of production scheduling decision is to maximize the production scheduling batch as much as possible on the premise of ensuring that the production of the wave order is completed before the completion deadline of the wave production. First, you need to determine the scheduling time point, that is, when to schedule. When the task of the previous batch is about to be completed, you need to start scheduling the next scheduling task. After determining the scheduling time point, the key to scheduling is to determine the scheduling mode and scheduling batch.

The basic original steps of scheduling are as follows: first, by predicting the number of orders placed from the current time to the order deadline of the current wave, combined with the number of unprocessed orders in the current order pool, obtain the total amount to be processed before the order cut-off of the current wave, and then judge which scheduling mode to enter according to the system scheduling mode threshold. There are three main system scheduling modes: first, the current wave scheduling mode, that is, the scheduling task is released from the current wave order. Second, the mixed wave scheduling mode, that is, the scheduling task is released from the orders of the current wave and the next wave. Third, the clear wave scheduling mode, that is, the batch scheduling releases all orders in the current wave order pool. Under the current wave scheduling mode and mixed wave scheduling mode, the scheduling batch of this batch needs to be determined in combination with the system capacity. The order intelligent scheduling process is shown in Fig. 2.

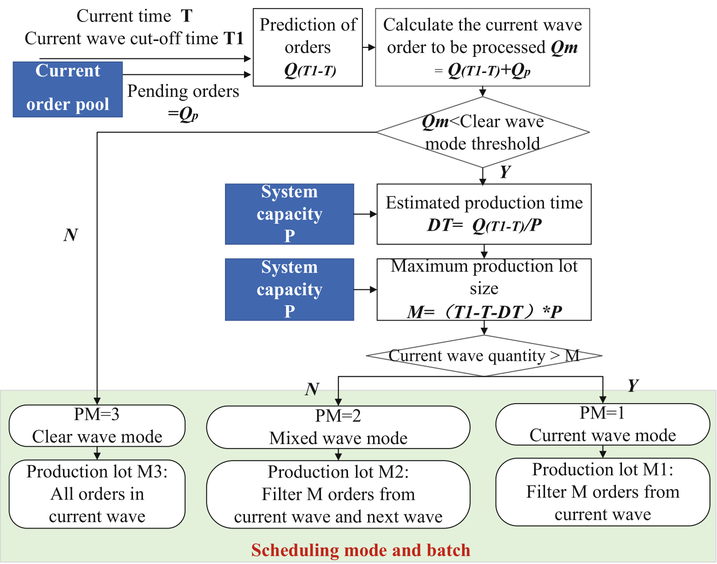

Fig. 2. Order intelligent scheduling process

This paper will take the order quantity in the e-commerce distribution center as the research object, and propose the dynamic prediction problem of order quantity, that is, predict the expected arrival quantity of orders in any period of time in the future for each time point of the distribution center, so as to realize the known production scheduling time point and accurately predict the order quantity from the current time to the order deadline, so as to support the decision-making of production scheduling mode and production scheduling batch in intelligent production scheduling.

2.3 Order Predicting Ideas

In order to support the intelligent scheduling decision, it is necessary to predict the order from the current time to the order deadline. However, the scheduling time point changes dynamically according to the system production and backlog, that is, the predicted time point and the predicted period range fluctuate with the scheduling demand. Different from the traditional single quantity prediction in fixed period, it is difficult to achieve the prediction effect in dynamic period by using more traditional prediction methods. Therefore, it is difficult and necessary to design the order dynamic prediction model of e-commerce distribution center for intelligent production scheduling.

The core of time series prediction is to predict future data based on historical data. Since the arrival volume of orders in different time periods fluctuates strongly, when predicting orders in the future, it is considered to take minutes as the time granularity, and the input and output are defined in the form of statistical values calculated by the time granularity set. Therefore, in order to achieve better prediction accuracy and effect, in addition to the internal parameters of neural network, the prediction step parameters of order time series and the time period for counting order quantity are also particularly important. Based on the above prediction ideas, this paper will select the time series prediction method based on BP neural network to construct the order dynamic prediction model, and carry out example verification through a large number of experiments and evaluation to determine the optimal parameter combination.

3 Order Predicting Model Based on BP Neural Network

BP neural network is a typical error back propagation artificial neural network. It consists of three parts: input layer, hidden layer and output layer. Each layer consists of several neurons. Under the action of activation function and threshold value, each neuron transmits the received information to the next neuron with a certain weight, and adjusts the weight and threshold size in time according to the error of back propagation. The structure of the training model is shown in Fig. 3. BP network has good non-linear approximation ability and clutter information processing ability. E-commerce order arrival has some randomness and uncertainty. It is theoretically feasible to build a prediction model based on BP network. The construction idea of the BP network model is to determine the input and output of the model, select the appropriate number of hidden layers and hidden layer nodes, properly transfer functions and training functions to train the model to achieve the desired prediction results.

Firstly, data mining and preprocessing are carried out on the original order, and the original order is processed into a continuous order time series with minute as the time granularity.

Secondly, Determine the input layer structure of the model. The input layer of the model needs to be confirmed by defining two important parameters. Define the prediction time step variable TimeStep, which means to predict the future order quantity by several time steps in advance, and represents the previous n periods The actual arrival quantity of orders (T, T-1,…, T-N) determines the number of nodes in the network input layer. The predicted input time length TimeSeg indicates how long the order statistics are used, which determines the value of nodes in the input layer.

Finally, the output layer structure of the model is determined. Through the test of multiple groups of sample data, in this problem, the fitting effect of single output neural network is obviously better than that of multi output neural network, so the prediction model in this paper adopts single output layer structure. The output layer structure needs to define the prediction time range parameter Tf and generate the output values of Tf neural networks through cyclic calculation, which respectively represent the number of

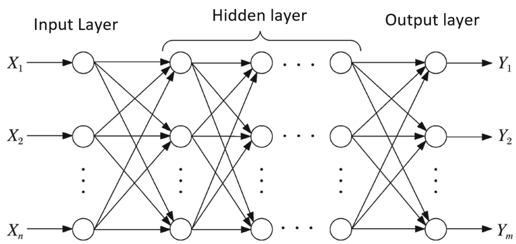

Fig. 3. Structure of BP neural network

orders in the next 1 min, 2 min,…, Tf min. All the generated networks are stored in the array. By calling the corresponding networks, you can predict the order in the required period, so as to achieve the dynamic prediction. The input and output description of a single sample that predicts the order quantity in Tf minutes in the future at time T is shown in Table 1. Q represents the order arrival quantity, and the subscript represents the time range of its statistical order quantity.

According to the above defined sample input and output, after all the data sets are normalized, the training set and test set shall be divided. The main parameters of BP neural network are defined, such as the number of hidden layer nodes, learning rate and so on. Then, train through a large number of order sample data, and store the trained Tf networks in the array. In practical applications, the order is predicted in the future by directly calling the corresponding trained neural network in the array, and the model training process is shown in Fig. 4.

Table 1. Definition of neural network input and output

Parameter type	Parameter name	Parameter value
Input	X1	Q(T-TimeStep*TimeSeg: T-(TimeStep-1)*TimeSeg)
	X2	Q(T-(TimeStep-1)*TimeSeg: T-(TimeStep-2)*TimeSeg)
	…	…
	X(timestep-1)	Q(T-2TimeSeg: T-TimeSeg)
	X(timestep)	Q(T-TimeSeg: T)
Output	Y	Q(T: T + Tf)

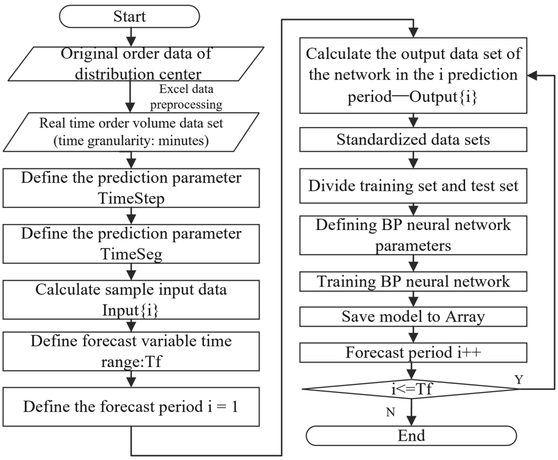

Fig. 4. Model training process

4 Optimization of Model Based on GA

BP neural network is prone to fall into local optimization in the training process, which has certain limitations. Therefore, this paper uses genetic algorithm to optimize the weight and threshold of BP, so as to improve the accuracy and efficiency of BP neural network prediction model.

The specific steps of using GABP algorithm to predict the order quantity are as follows:

Step 1: Normalize the preprocessed order time series historical data.

Step 2: Initialize the population. Each individual in the population is a real number string composed of the weight between the input layer and the hidden layer, the threshold of the hidden layer, the weight between the hidden layer and the output layer, and the threshold of the output layer.

Step 3: Calculate the fitness function. In this paper, the absolute value of the error between the expected output and the predicted output is taken as the individual fitness function f, and its calculation method is as in (1).

$$F = k\left(\sum_{i=1}^{m} |d_i - p_i|\right) \tag{1}$$

In (1), m represents the total number of output nodes of neural network, d_i represents the expected output of the ith node of the neural network, p_i represents the prediction output of the ith node of the neural network, and k is the coefficient.

Step 4: Select. This method uses roulette selection operator, that is, the selection strategy based on fitness proportion to select the chromosomes in each generation of population. Selection probability Q_i is calculated as in (2) and (3).

$$\hat{f} = \frac{\sum_{i=1}^{P} F}{P} \tag{2}$$

$$Q_i = \frac{f_i}{\sum_{i=1}^{P} f_i} (i = 1, 2, \ldots, P) \tag{3}$$

Step 5: Cross. Because the individual adopts real number coding, the crossover operation method adopts real number crossover method, and the kth gene W_k and lth gene W_l the crossover operation at j bit is shown in (4) and (5).

$$W_{kj} = W_{kj}(1 - b) + W_{lj} b \tag{4}$$

$$W_{lj} = W_{lj}(1 - b) + W_{kj} b \tag{5}$$

Step 6: Mutation operation. The jth gene of the ith individual was selected for mutation operation, as shown in (6).

$$W_{ij} = \begin{cases} W_{ij} + \left(W_{ij} - W_{max}\right) f(g) r \geq 0.5 \\ W_{ij} + \left(W_{min} - W_{ij}\right) f(g) r < 0.5 \end{cases} \tag{6}$$

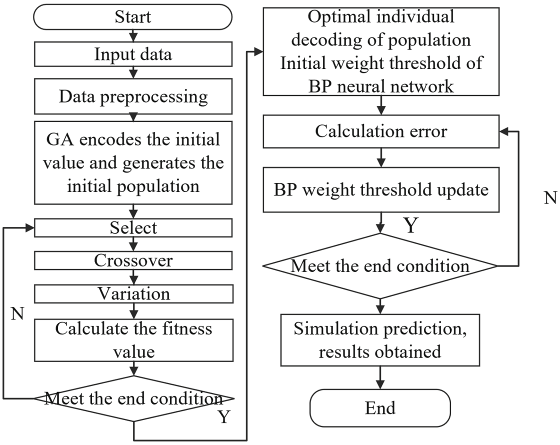

Fig. 5. GABP model training process

$$f(g) = r_2(1 - g/G_{max}) \tag{7}$$

W_{max}, W_{min} are the maximum and minimum of gene W_{ij}, r is a random value in [0,1], r_2 is a random value, g is the current number of iterations, G_{max} is the maximum evolution algebra.

Step 7: Judge the termination condition of the algorithm. Repeat the above steps until the algorithm iteration algebra reaches GenMax. If the termination conditions are met, the global optimal individual will be returned accordingly. Otherwise, the evolutionary algebra will increase and turn to step 2 to continue the optimization.

Finally, the optimal individual obtained by genetic algorithm is decoded and assigned to the connection weight and threshold of BP neural network. With the help of BP neural network algorithm, the BP neural network prediction model is trained to obtain the best solution of order dynamic prediction. GABP optimization steps are shown in Fig. 5.

5 Empirical Research

The order sample data used in this paper is the detailed data of orders of M pharmaceutical e-commerce distribution center from a total of 16 normal days. Its statistical fields include WMS receiving time, outbound date, SKU, and other logistics information. According to the statistics of source data by order line items, there are 427674 orders in total. According to the arrival trend, the order basically presents the time series fluctuation of daily cycle, and the consistency between peak time and trough time is strong. The peak arrival time of daily order is 10:00 a.m. and 22:00 p.m., respectively, and the order arrival volume is less from 1:00 a.m. to 6:00 a.m.

After a lot of experiments, it is determined that the internal parameters of BP neural network are set as 10 hidden layers, 300 training times, 0.00001 target minimum error and 0.05 learning rate. Levenberg Marquardt algorithm is applied. The parameters of the outer loop of the model include Timestep, TimeSeg and Tf, and the value range and parameter description are shown in Table 2.

Table 2. Model parameter description

Parameter type	Parameter name	Value range
External circulation parameters	TimeStep	1,5,10,15,20,25
	TimeSeg	1,10,20,30,60
	Tf	0–60 (Integer)
Neural network parameters	Train set Test set	Train set = total samples * 0.8 Test set = total samples * 0.8
	Hidden layers	10
	Learning rate	0.05
	Minimum error	0.00001
	Training times	300

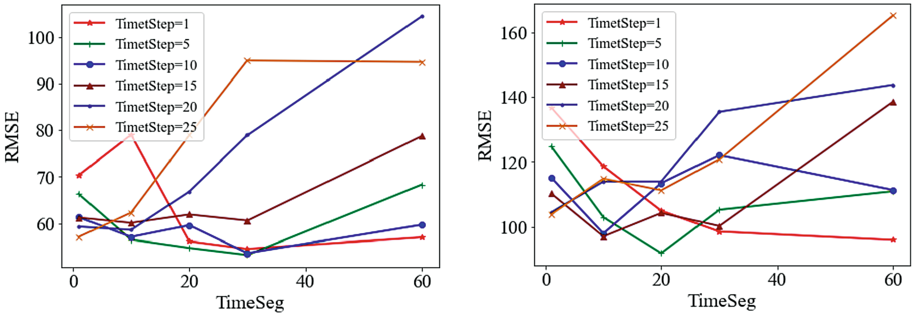

Fig. 6. Comparison diagram of model parameters

5.1 Analysis of Model Results

The model error comparative analysis selects three kinds of typical prediction effect evaluation index, Root mean square error (*RMSE*), Mean absolute percentage error (*MAPE*), coefficient of determination (R^2). *RMSE* and *MAPE* measures the absolute size of the deviation between the real value and the predicted value. The smaller the value is, the better the model effect is. R^2 reflects the fitting effect. The closer its value is to 1, the better the fitting effect of the prediction model is. The closer its value is to 0, the worse the fitting effect of the prediction model is.

Select the typical prediction range Tf as 30 min, that is to predict the next half an hour, train the model 5 times under each parameter combination, and count the average value of prediction effect index, so as to reduce the error caused by single model training inaccuracy. Compare the output results of the model as shown in Fig. 6, TimeStep = 5, TimeSeg = 30, the model result is the best, R^2 is 0.96, RMSE is 53.22. As above, in the case of predicting one hour order quantity, Tf = 60, TimeStep = 5, TimeSeg = 20, the model result is the best, which is 0.97, RMSE is 91.7.

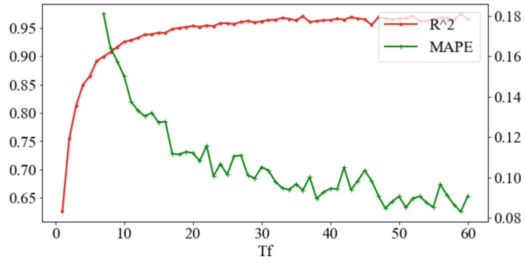

Fig. 7. Comparison chart of prediction and evaluation indexes

According to a large number of experiments, TimeStep = 5 and TimeSeg = 30 are selected as the external circulation parameters of the prediction model to predict the order quantity in any time period within 1 h, and the prediction and evaluation indexes under different Tf parameters are counted, as shown in Fig. 7. According to the model results, since Tf = 30 min, R^2 is stable above 0.96 and MAPE is stable below 10%. It is proved that in the dynamic prediction of the actual order of the e-commerce distribution center, the value accuracy effect of the prediction period of more than 30 min is better.

5.2 Comparison of Results of GABP Model

Based on the above optimal model results, the external circulation parameters of BP neural network prediction model are set as TimeStep = 5, TimeSeg = 20, Tf = 60, that is, the order in the next hour is predicted. A total of 4570 order samples are selected for experimental comparison of prediction models. The prediction effect indexes of BP neural network model are R^2 is 0.97, RMSE is 99.23 and MAPE is 8.70%.

In order to verify the prediction accuracy of the model, this paper selects the weighted average method and one-time exponential smoothing method to compare the prediction results with the BP neural network model. Based on the simple moving average method, the weighted average method is optimized with the weight of the data in the group. In this paper, the number of moving periods is 2, and the weights of the first and second periods are 0.2 and 0.8 respectively, that is, the orders in the next hour are predicted through the weighted moving average of the orders in the first two hours. Exponential smoothing method is a special weighted moving average method. In the comparative experiment of this paper, the smoothing constant is 0.5.

The comparison of various effect indexes of the prediction model is shown in Table 3. BP neural network has obvious advantages over traditional prediction methods in order dynamic prediction, and its prediction accuracy and other evaluation indexes are obviously better than other models, with better prediction ability.

Based on the optimal model results, TimeStep = 5, TimeSeg = 20, Tf = 60 are selected, and GABP model is used to predict and optimize. The parameters are as follows: the evolutionary termination algebra is 50, the population size is 20, the cross probability is 0.7, and the mutation probability is 0.1. The evolutionary iteration is used to obtain the optimal fitness value when running to the 48th generation, and the network performance is the best.

Table 3. Comparison results of traditional models

	R^2	*RMSE*	*MAPE*
BP neural network	0.97	99.23	8.70%
Weighted moving average method	0.68	326.55	33.79%
Exponential smoothing method	0.53	403.98	45.54%

Table 4. Comparison of optimization results

	RMSE	R^2	*MAPE*	Proportion (MAPE < 10%)	Proportion (MAPE < 20%)
BP	99.23	0.97	9.87%	69%	89%
GABP	88.97	0.98	8.70%	73%	90%

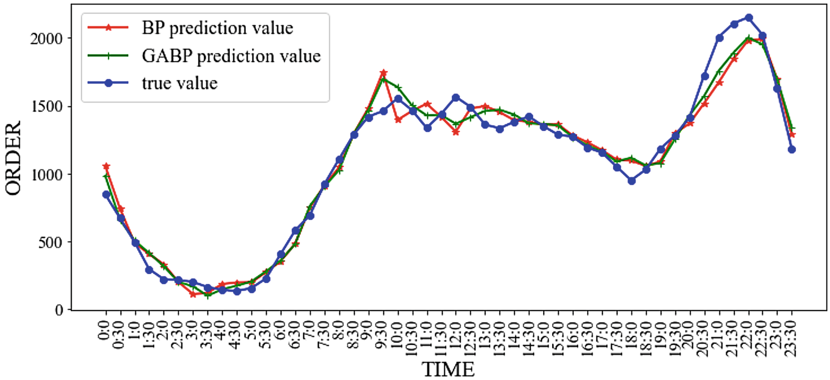

Fig. 8. Fitting diagram of prediction model

In the experiment, besides RMSE and MAPE, the error interval distribution of all test sample sets is also calculated. The error is calculated by dividing the difference between the real value and the predicted value by the real value to further verify the model comparison results. The results of comparison between the GA optimized model and the original model are shown in 0The optimized model is 0.975, RMSE is 88.97, MAPE is 8.70%, and the percentage of samples with prediction error less than 10% is 73%, which is higher than that of BP prediction model by 4%, which verifies the effectiveness of the model optimization (Table 4).

The test set has 4570 sample points in total. In order to visualize the comparison effect of the model, this paper selects the values of the whole point and the half point of the time in the test set sample in one day, and 48 sample points are selected to make the fitting curve, as shown in Fig. 8.

It can be seen that the BP neural network prediction model optimized by genetic algorithm has better fitting effect than the predicted value and real value of BP neural network, which further proves the effectiveness of model optimization.

6 Conclusion

Based on BP neural network, this paper constructs an order dynamic prediction model for intelligent scheduling of e-commerce distribution center, and optimizes the prediction model by genetic algorithm, so that the prediction model has better prediction accuracy. Finally, the accuracy and effectiveness of the model are verified by the example data of M pharmaceutical e-commerce distribution center. The research results of this paper provide important input support for intelligent scheduling mode and batch decision-making of e-commerce distribution center, effectively improve the sorting efficiency of e-commerce distribution center, and provide a reference for further theoretical research on order dynamic prediction in e-commerce distribution center.

References

1. Ge N, Sun LY, Zhao P (2018) Forecast and analysis of sales volume based on ARIMA time series model. J Beijing Union Univ (Nat Sci) 32:27–33
2. Ge YQ, Wang XZ, Wang AM (2012) Research on forecasting of refrigerator orders demand based on improved grey neural network. Comput Simul 29:219–222
3. Gong RR (2011) Research of business order forecast model based on radial basis function neural network using OLS and EPSO algorithms. Comput Eng Appl 47:224–226+243
4. Guo X, Liu C, Xu W (2014) A prediction-based inventory optimization using data mining models. In: Seventh International Joint Conference on Computational Sciences & Optimization, vol. 1, pp 611–615. IEEE Computer Society
5. Han C, Zhang Q (2020) Optimization of supply chain efficiency management based on machine learning and neural network. Neural Comput Appl 33(5):1419–1433. https://doi.org/10.1007/s00521-020-05023-1
6. Hofmann E, Rüsch M (2017) Industry 4.0 and the current status as well as future prospects on logistics. Comput Ind 89:23–34
7. Leung KH, Choy KL, Siu PKY (2018) A B2C E-commerce intelligent system for Re-engineering the E-Order fulfilment process. Expert Syst Appl 91:386–401
8. Leung KH, Choy KL, Ho G (2019) Prediction of B2C e-commerce order arrival using hybrid autoregressive-adaptive neuro-fuzzy inference system (AR-ANFIS) for managing fluctuation of throughput in e-fulfilment centres. Expert Syst Appl 134:304–324
9. Wang CB, Li YB (2013) The order forecasting research based on improved particle swarm optimization and neural network. Mech Eng 5:137–139
10. Sheng WS, Zhao HC, Sun YW (2019) Sales forecasting model based on BP neural network optimized by improved genetic algorithms. Comput Syst Appl 28:200–204
11. Wang LH (2021) Forecast model of short-term sales in e-commerce based on BP-AdaBoost. Comput Syst Appl 30:260–264
12. Wang JL, Kuo JH, Chou SY (2010) A comparison of bullwhip effect in a single-stage supply chain for autocorrelated demands when using Correct, MA, and EWMA methods. Expert Syst Appl 37:4726–4736
13. Yang W, Zhang K, Zhao J (2019) Forecast and decision analysis of exponential smoothing order based on Monte Carlo. Packag Eng 40:155–161