

Vegetable Product Pricing Replenishment Strategy Based on WOA

Zhiyuan Wang^{*,#}, Jiachang Liang[#], Jiaxing Zhang[#]

Department of Data Science and Big Data Technologies, Fuyang Normal University, Fuyang, China, 236037

* Corresponding Author Email: 13635542713@163.com

[#]These authors contributed equally.

Abstract. In fresh food superstores, the freshness date of general vegetable commodities tends to be relatively short, and the quality of vegetable commodities will deteriorate with the delay of the sales time. Therefore, how to allow supermarkets to ensure that the daily replenishment of the amount of market demand for various types of vegetable products while trying to meet the premise of maximizing their profits has become an urgent problem, this study aims to propose an innovative vegetable product prediction model based on big data, which adopts the SARIMA model to accurately predict the daily replenishment amount of each type of vegetable product in supermarkets. Meanwhile, based on the replenishment amount, the WOA model takes the addition coefficient and sales of different vegetable products as the dependent variable, and the final profit as the independent variable. The corresponding pricing strategy is formulated to maximize the profitability of the supermarket. Through solving the model, it is found that the maximum profit of the supermarket's vegetable products in the next week is 3,649.65 yuan.

Keywords: Whale Optimization Algorithm, SARIMA Model, Big Data.

1. Introduction

In the fresh food superstore, general vegetable commodities compared to other commodities, their freshness date is often shorter, and the quality of vegetable commodities will deteriorate with the delay of the sales time [1]. Therefore, the superstore usually replenishes its stock every day according to the historical sales volume and demand of each vegetable category and different individual products, to ensure that the daily replenishment volume can maximize its profit under the premise of trying to satisfy the market's demand for each type of vegetable commodity. The supermarket has cauliflower, foliage, aquatic roots and tubers, peppers, eggplants, and edible mushrooms. We need to analyze the relationship between total sales volume and cost-plus pricing for each vegetable category based on the available data so that we can develop the total daily replenishment volume for the coming week and the related pricing strategy, and then maximize the profit of the superstore [2].

The sales volume of vegetable items has obvious seasonal characteristics. Referring to previous studies, when dealing with seasonal time series, seasonal patterns can be extracted by calculating seasonal factors after preprocessing the original time series to reduce noise interference [3]. For the pricing of vegetable commodities, the optimal solution is generally based on the existing data, and then nonlinear programming is used to find the optimal solution.

Methods for dealing with seasonality After preprocessing the original time series to reduce noise interference, the results can be derived by calculating seasonal factors to extract seasonal patterns[4]. To reduce the interference of non-seasonal factors on the solution results, and at the same time have a high prediction accuracy, we use the SARIMA model to solve for the replenishment of fresh vegetables in the coming week. For the pricing strategy, the advantage of nonlinear programming is that it can solve nonlinear problems, but the disadvantage is that the solution process may be more complicated and may not be able to find the global optimal solution. The advantage of the whale intelligent optimization algorithm is that it has a high convergence speed and the ability to jump out of the local optimal solution [5]. Therefore, we use the whale intelligent optimization algorithm to solve the pricing of vegetable individual products [6].

2. The fundamentals of SARIMA model and Whale Optimization Algorithm

2.1. The basic funamental of SARIMA model

In some time, series, there are obvious periodic changes. This cycle is due to seasonal variations (including quarterly, monthly, weekly, etc.) or some other inherent factors. This type of series is called a seasonal series. For example, a series of temperature values for a region (observations taken at hourly intervals) contains annual variations as well as diurnal variations. In the economic field, seasonal series can be seen everywhere. Such as quarterly time series, monthly time series, weekly time series, etc. It is not enough to deal with seasonal time series using only the methods introduced above. One model that describes this type of series is the seasonal ARIMA model, denoted by SARIMA [7]. Earlier literature also called it the multiplicative seasonal model.

Let the seasonal series (monthly, quarterly, weekly, and so on) have a period of s ; that is, observations spaced at intervals s are similar. Firstly, the seasonal difference method is used to eliminate the periodic variation. The seasonal difference operator is defined as

$$\Delta_s = 1 - B^s \quad (1)$$

If the seasonal time series is denoted by y , the first seasonal difference is denoted by

$$\Delta_s \cdot X_t = (1 - L^s) X_t = X_t - X_{t-s} \quad (2)$$

For non-stationary seasonal time series, sometimes needs D times of seasonal difference before it can be transformed into stationary series. On this basis, the model of p -order autoregressive q -order moving average seasonal time series with period S can be established as

$$A_p(B^s) \Delta_s^D X_t = B_q(B^s) \varepsilon_t \quad (3)$$

For the above model, it is equivalent to assume that ε_t is stationary and non-autocorrelated.

When ε_t is nonstationary and there is an ARMA component, then ε_t can be described as

$$\Phi_P(B) \Delta^d \varepsilon_t = \Theta(B) v_t \quad (4)$$

where v_t is the white noise process, p and q denote the maximum number of stages of the non-seasonal autoregressive and moving average operators, respectively, and d denotes the number of first-order (non-seasonal) differences ε_t . Substituting Equation (4) into Equation (3) into the moving average seasonal time series model can be obtained as

$$\Phi_P(B) A_p(B^s) (\Delta^d \Delta_s^D X_t) = \Theta(B) B_q(B^s) v_t \quad (5)$$

Where the subscripts P , Q , p , q denotes the maximum lag order of seasonal and non-seasonal autoregressive and moving average operators, respectively, and d , D denote the number of non-seasonal and seasonal differences, respectively. The above equation is called the order $(p, d, q) * (P, D, Q)_s$ seasonal time series model or the product seasonal model [8].

The seasonal order of the product's seasonal model, i.e., the period length S , can be determined by analyzing the autocorrelation and partial correlogram. If these plots do not exhibit a linear decay trend but instead show a significant peak with a large absolute value and demonstrate oscillating changes at integer multiples of the periodicity, it can be inferred that the time series can be effectively described using the SARIMA model.

2.2. The basic fundamental of Whale Optimization Algorithm

The whale Optimization Algorithm realizes the purpose of optimization search by simulating the predation behavior of whales, such as searching for prey, surrounding prey, and attacking prey with bubble nets [9].

As is shown in Figure 1, whale populations can find out the location of prey and surround the prey during hunting. Before solving the problem, the position of the prey in the space is unknown to the whale population, so in the whale optimization algorithm, it is assumed that the position of the best whale individual in the current population is the position of the prey, and other whales in the population are surrounded by the position of the optimal whale [10].



Figure 1 Schematic of bubble feeding by humpback whales.

Expressed by the mathematical model is.

$$D = |C \cdot X^*(t) - X| \quad (6)$$

$$X(t+1) = X^*(t) - A \cdot D \quad (7)$$

Where t is the current iteration number; X^* denotes the position of the best whale in the current population; X represents the current position of the whale; A and C are vectors, i.e

$$A = 2ar - a \quad (8)$$

$$C = 2r \quad (9)$$

Where a is the convergence factor, with the whale population predation iteration, the value of a decreases linearly from 2 to 0. r is a vector of random numbers with values in the interval $[0,1]$.

To mathematically describe the bubble net predation behavior of whale populations, the whale optimization algorithm designs two different mechanisms. The first is the narrowing surround mechanism, which is realized by decreasing the value of A in Equation (7), where the fluctuation range of $|A|$ will also be reduced by the decrease of a . The second is the spiral update position mechanism, between the whale and the prey, using a spiral equation to mimic the spiral-shaped motion associated with the whale, i.e

$$X(t+1) = D^1 \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t) \quad (10)$$

Where $D^1 = |X^*(t) - X(t)|$ represents the distance between the optimal whale individual and the current whale individual in the TTH iteration. b is a constant; l is a random number in the interval $[-1,1]$.

The whale swims around the prey in a reduced circle following a spiral-shaped path [11]. To obtain a model that simulates this behavior, it is assumed that there is a 50% probability in the optimization process to randomly choose between the shrinking surround mechanism and the spiral updating position mechanism to update the position of individual whales, whose mathematical model is.

$$X(t+1) = \begin{cases} X^*(t) - A \cdot D & , p < 0.5 \\ D^1 \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t) & , p \geq 0.5 \end{cases} \quad (11)$$

Where p is a random number in the interval $[0,1]$.

In the process of prey searching, the evolution of matrix A can be utilized for global exploration as the iterative procedure progresses. During the phase of locating prey, since the exact location is

unknown to the whale population, cooperation among whales becomes essential to acquire this information [9]. Whales employ random individual positions within their population as navigation targets for food-finding purposes, and this mathematical model is described as

$$D = |C \cdot X_{rand} - X| \quad (12)$$

$$X(t+1) = X_{rand} - A \cdot D \quad (13)$$

Where X_{rand} represents the position of a randomly selected individual whale in the current whale population.

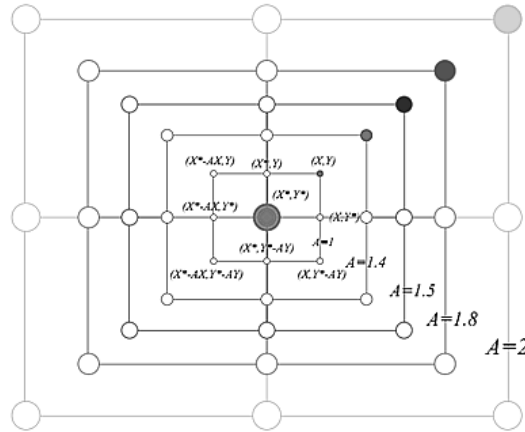


Figure 2 Global Exploration Mechanism for Whale Optimization Algorithm

The whale optimization algorithm is initiated by randomly selecting a set of initial positions for the whale populations. As is shown in Figure 2, during each iteration, an individual whale updates its position based on either randomly selected position information or the position information of the individual whale with the highest fitness value obtained thus far in that iteration [12]. As parameter A linearly decreases from 2 to 0, both the global exploration phase and local exploitation phase of the algorithm transform. When $|A| > 1$, a random whale from the population is chosen; when $|A| < 1$, the whale with the highest current fitness value is selected to update the position of the current individual. By specifying a value for p , it becomes possible for the whale optimization algorithm to switch between utilizing either the shrinking surround mechanism or the spiral updating position mechanism. Finally, termination conditions can be defined to determine when to stop executing the whale optimization algorithm. Theoretically, the whale optimization algorithm can be regarded as a global optimizer because it includes a global exploration stage and a local development stage. In addition, the hypercube model of the whale optimization algorithm defines a search space near the optimal solution, allowing other search entities to take advantage of the current optimal location within that region. Through the adaptive change of the search matrix A , the whale optimization algorithm makes a smooth transition between the global exploration stage and the local development stage. By reducing $|A|$, part of the iteration is assigned to the global exploration phase, and the rest is dedicated to the local development segment.

3. Results

3.1. The prediction of replenishment volume in the next week based on the SARIMA model

We collected the historical sales data for various varieties of vegetable products in a supermarket from July 1, 2020, to June 30, 2023. For the forecast of the sales volume of 6 kinds of vegetable products in the next week, we adopted PyCharm software to solve the model and forecast the sales data within 7 days. (Source: mcm.edu.cn) The first 70% of the historical data of each vegetable product was taken as the training set and the last 30% as the test set, and the time series diagram of 6 different vegetable types was obtained, as shown in Figure 3 to Figure 8.

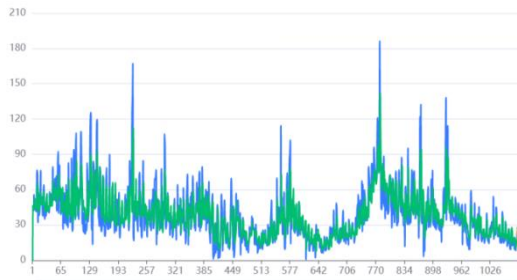


Figure 3 Time series diagram of cauliflower class



Figure 4 Mosaic time series diagram

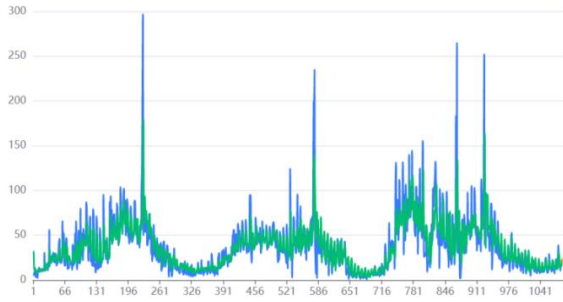


Figure 5 Time series diagram of pepper class

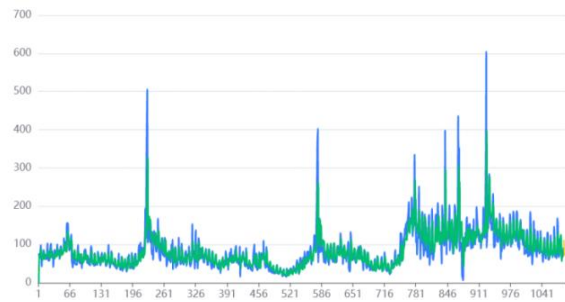


Figure 6 Nightshade time series diagram.

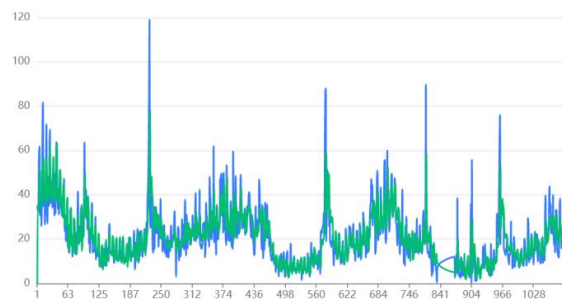


Figure 7 Time series diagram of edible fungi

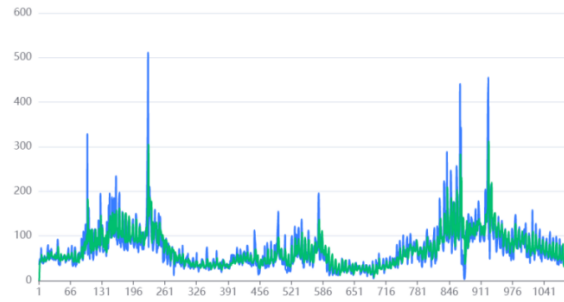


Figure 8 Time series diagram of aquatic rhizomes.

Through Figure 3 to Figure 8, we conducted statistical analysis on the predicted historical sales volume of 6 different vegetables, and the results were shown in Table 1.

In the process of transportation and sales of goods, there will be certain losses, which will cause the quantity of goods that can be sold will be less than the quantity of replenishment, expressed in the formula as follows.

$$M_1(1 - \beta) = M_2 \quad (14)$$

Table 1 Daily sales forecast table for different vegetable categories from July 1, 2023, to July 7, 2023

	Cauliflower vegetables	Variegated vegetables	Chili vegetables	Nightshade vegetables	Edible mushroom vegetables	Aquatic root vegetables
Data	Prediction results	Prediction results	Prediction results	Prediction results	Prediction results	Prediction results
2023/7/1	26.94	180.94	111.78	29.85	64.50	24.92
2023/7/2	23.15	168.93	99.95	30.32	57.69	22.12
2023/7/3	17.06	126.29	72.16	20.89	42.10	16.97
2023/7/4	16.95	139.32	71.55	18.74	42.08	17.20
2023/7/5	16.65	135.86	71.60	19.02	46.71	18.49
2023/7/6	18.07	138.90	77.04	18.93	45.28	20.40
2023/7/7	18.70	150.62	86.11	21.75	50.68	21.16

We calculated the consumption rate of a single product and the total sales volume of a single product in different vegetable categories, calculated the total amount of a single product before depletion, and finally obtained the consumption rate of different categories, as shown in Table 2.

Table 2 Schematic table of consumption rate of different categories

	Cauliflower vegetables	Variegated vegetables	Chili vegetables	Nightshade vegetables	Edible mushroom vegetables	Aquatic root vegetables
Consumption rate	10.97%	16.50%	8.10%	6.37%	7.01%	8.57%

Through data processing, the consumption of different kinds of vegetables in the process of purchase can be obtained. The total amount of replenishment is calculated by equation (14), and the final replenishment amount in the coming week is shown in the table below.

Table 3 Forecast table of daily repetition quantity for different vegetable categories from July 1, 2023, to July 7, 2023

	Cauliflower vegetables	Variegated vegetables	Chili vegetables	Nightshade vegetables	Edible mushroom vegetables	Aquatic root vegetables
Data	Replenishment volume	Replenishment volume	Replenishment volume	Replenishment volume	Replenishment volume	Replenishment volume
2023/7/1	30.26	216.69	121.64	31.88	69.36	27.25
2023/7/2	26.00	202.31	108.76	32.38	62.04	24.19
2023/7/3	19.17	151.24	78.52	22.31	45.27	18.56
2023/7/4	19.04	166.85	77.86	20.02	45.25	18.81
2023/7/5	18.70	162.71	77.91	20.31	50.23	20.22
2023/7/6	20.30	166.35	83.83	20.22	48.70	22.32
2023/7/7	21.01	180.39	93.71	23.23	54.50	23.15

Table 3 shows that we use the SARIMA model to predict the sales volume of 6 kinds of vegetable products in the next week, and the replenishment volume of 6 kinds of vegetable products in the next week is obtained through the transformation of the loss rate of different vegetable varieties in the transportation process. We reserve two decimal places for the forecast result.

3.2. Optimal solution for commodity pricing based on WOA

Through the analysis of the problem, we perform nonlinear polynomial fitting, fit the relationship between the total sales volume of each vegetable category and the cost-plus pricing into a nonlinear function, and establish the relationship between sales volume and pricing. We want to maximize the profit of the supermarket, that is, maximize the profit sum of different categories of goods, and the objective function is:

$$\max y = \sum_{i=1}^6 (X_i r \alpha_i - X_i / (1 - \beta_i) r) \quad (15)$$

The sales volume of different categories in the next week is less than the replying volume in one week (unit: kg), i.e

$$\begin{cases} 0 \leq X_1 \leq 154.47 \\ 0 \leq X_2 \leq 1246.54 \\ 0 \leq X_3 \leq 642.23 \\ 0 \leq X_4 \leq 170.35 \\ 0 \leq X_5 \leq 375.35 \\ 0 \leq X_6 \leq 154.52 \end{cases} \quad (16)$$

For the cost addition coefficient, the requirements should not be too large or too small. The historical data are collated, and the maximum and minimum value of the cost addition coefficient are extracted as constraints, i.e

$$0.023 \leq \alpha_i \leq 3.13 \quad (17)$$

The relationship between the total sales volume of each vegetable category and the cost-plus pricing is fitted to a nonlinear function, and the constraint conditions of sales volume and pricing are established, i.e

$$\begin{cases} X_1 = -4.741 \cdot 10^5 \alpha_1^3 + 7.751 \cdot 10^5 \alpha_1^2 - 3.753 \cdot 10^5 \alpha_1 + 5.118 \cdot 10^4 \\ X_2 = -1.177 \alpha_2^3 + 40.12 \alpha_2^2 - 508.8 \alpha_2 + 2.388 \cdot 10^3 \\ X_3 = -1.357 \cdot 10^3 \alpha_3^3 + 9.875 \cdot 10^3 \alpha_3^2 - 1.655 \cdot 10^4 \alpha_3 + 8.457 \cdot 10^3 \\ X_4 = 4.271 \cdot 10^3 \alpha_4^3 - 1.735 \cdot 10^4 \alpha_4^2 + 1.196 \cdot 10^4 \alpha_4 + 1.84 \cdot 10^3 \\ X_5 = 3.858 \alpha_5^3 - 134.1 \alpha_5^2 + 66.55 \alpha_5 + 1.14 \cdot 10^3 \\ X_6 = -2.643 \cdot 10^4 \alpha_6^3 + 6.339 \cdot 10^4 \alpha_6^2 - 4.353 \cdot 10^4 \alpha_6 + 8.717 \cdot 10^3 \end{cases} \quad (18)$$

Matlab software was used to write the whale intelligent optimization algorithm, the maximum number of populations was set to 100, the maximum number of iterations was set to 1000, and the results are shown in Figure 9.

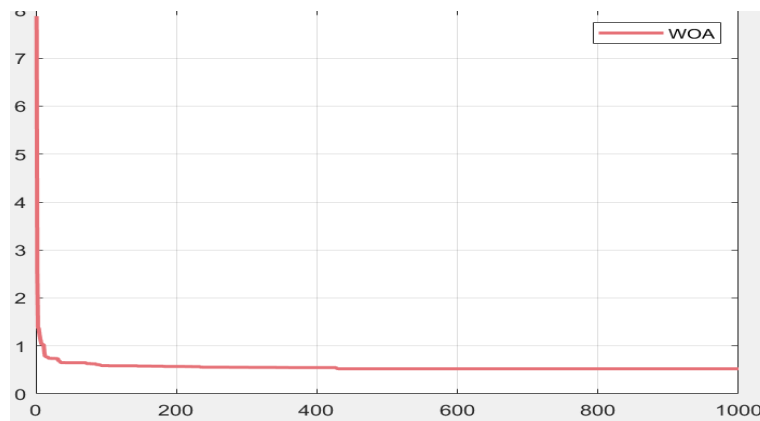


Figure 9 WOA operational function diagram

The maximum profit within a week is 3649.65 yuan. The cost-addition factor of each vegetable category and the corresponding sales volume are shown in Table 4.

Table 4 Cost plus coefficient of each vegetable category and corresponding sales schematic table

	Cauliflower vegetables	Variegated vegetables	Chili vegetables	Nightshade vegetables	Edible mushroom vegetables	Aquatic root vegetables
Addition coefficient	1.57	0.41	1.11	1.06	0.18	0.72
Sales volume (unit: kg)	140.35	1080.23	610.14	142.23	360.12	145.35

Through the data in Table 4, we can know the weighting coefficient of different vegetable products and the sales volume of maximum profit in one week, and we can use the weighting coefficient to predict the pricing of six vegetable products in the next week.

4. Conclusions

In this paper, the prediction of weekly replenishment quantity of vegetable products in fresh supermarkets and the strategy of profit maximization of vegetable products in a week are studied. We

collected relevant historical sales data and used the SARIMA model to predict the replenishment volume of six kinds of vegetable products in the supermarket in the next week. Meanwhile, in order to maximize the profits of the supermarket in the next week, we established relevant equations through the WOA model and finally obtained the profit maximization strategy.

Through the SARIMA model, we predicted the replenishment amount of 6 kinds of vegetable products in the coming week based on the historical data of the supermarket. Among them, the total replenishment amount of cauliflower vegetables in one week is 154.47kg, the total replenishment amount of Variegated vegetables in one week is 1246.54kg, the total replenishment amount of Chili vegetables in one week is 642.23kg, the total replenishment amount of Nightshade vegetables in one week is 170.35kg, the total replenishment amount of Aquatic root vegetables in one week is 375.35kg, and the total replenishment amount of Edible mushroom vegetables in one week is 154.52kg. By predicting the replenishment of 6 kinds of vegetable products, we set the maximum limit range for the price and sales of vegetables in the supermarket in the next week.

Through the pricing prediction of vegetable products by WOA function, we can get that the supermarket profit in the coming week is 3649.65 yuan, among which the price analysis of various vegetable commodities is as follows:

For cauliflower vegetables, we give the cost additional factor, and the final price is the wholesale price of different cauliflower products multiplied by 2.57. For variegated vegetables, we obtain a cost additional factor of 1.41, and finally, we price variegated vegetables by multiplying their wholesale price by 1.41. For chili vegetables, we give a cost additional factor of 2.11, and the final price is the purchase cost price multiplied by 2.11. For nightshade vegetables, we give the cost additional factor of 2.06, and the pricing of different nightshade products is the wholesale price multiplied by 2.06. For edible mushroom vegetables, the final price is the wholesale price of different edible mushroom products multiplied by its cost additional factor 1.18. For water root vegetables, the final price is the wholesale price of water root for different products multiplied by the additional cost factor of 1.72.

Through the study of this paper, we can process the historical data of supermarkets, and use the final forecast results as a reference for fresh supermarkets, to provide reliable data comparison for the purchase of supermarkets in the future and help them reduce costs and improve benefits. At the same time, the results of this study can not only provide a reference for supermarkets to optimize replenishment and pricing and provide intelligent solutions for the supply chain management of vegetable commodities.

References

- [1] Liu L, Zhao L, Ren X. Optimal preservation technology investment and pricing policy for fresh food[J]. *Computers & Industrial Engineering*, 2019, 135: 746-756.
- [2] Wang Z X, He L Y, Zhao Y F. Forecasting the seasonal natural gas consumption in the US using a gray model with dummy variables [J]. *Applied Soft Computing*, 2021, 113: 108002.
- [3] Pham Q V, Mirjalili S, Kumar N, et al. Whale optimization algorithm with applications to resource allocation in wireless networks [J]. *IEEE Transactions on Vehicular Technology*, 2020, 69 (4): 4285-4297.
- [4] Lejarza F, Pistikopoulos I, Baldea M. A scalable real-time solution strategy for supply chain management of fresh produce: A Mexico-to-United States cross border study [J]. *International Journal of Production Economics*, 2021, 240: 108212.
- [5] Ai X, Yue Y, Xu H, et al. Optimizing multi-supplier multi-item joint replenishment problem for non-instantaneous deteriorating items with quantity discounts [J]. *Plos one*, 2021, 16 (2): e0246035.
- [6] He B, Gan X, Yuan K. Entry of online presale of fresh produce: a competitive analysis [J]. *European Journal of Operational Research*, 2019, 272 (1): 339-351.
- [7] He K, Ji L, Wu C W D, et al. Using SARIMA–CNN–LSTM approach to forecasting daily tourism demand [J]. *Journal of Hospitality and Tourism Management*, 2021, 49: 25-33.
- [8] Benvenuto D, Giovanetti M, Vassallo L, et al. Application of the ARIMA model on the COVID-2019 epidemic dataset [J]. *Data in brief*, 2020, 29: 105340.

- [9] Gharehchopogh F S, Gholizadeh H. A comprehensive survey: Whale Optimization Algorithm and its applications [J]. *Swarm and Evolutionary Computation*, 2019, 48: 1-24.
- [10] Rana N, Latiff M S A, Abdulhamid S M, et al. Whale optimization algorithm: a systematic review of contemporary applications, modifications and developments [J]. *Neural Computing and Applications*, 2020, 32: 16245-16277.
- [11] Nasiri J, Khiyabani F M. A whale optimization algorithm (WOA) approach for clustering [J]. *Cogent Mathematics & Statistics*, 2018, 5 (1): 1483565.
- [12] Mirjalili S, Lewis A. The whale optimization algorithm [J]. *Advances in engineering software*, 2016, 95: 51-67.