

Data-Driven Fresh Food Supply Chain Management and Market Competitiveness Research

Qiliang Ren ^{*,#}, Yifan Kang [#], Wei Zhang [#]

International College, Hebei University, Baoding, China, 071000

* Corresponding Author Email: 15030521013@163.com

[#]These authors contributed equally.

Abstract. Vegetable commodities have the characteristics of short freshness period, variety, and special purchase time. Therefore, it is particularly important to predict and manage the purchase and shipment of vegetable commodities. Good management can significantly improve the profits of supermarkets and save costs. For the relationship between vegetable categories, we studied the seasonal distribution of sales volume of each category of vegetables based on Spearman correlation analysis, and then studied the cyclical pattern of vegetables based on the ARIMA algorithm model with a sliding window of seven days. We also predicted the sales trend of vegetables and predicted the cost data based on the LSTM model, and predicted the vegetable prices and inventory for the next seven days based on the particle swarm optimization algorithm, obtaining the optimal results.

Keywords: ARIMA seasonality analysis, Entropy-Weighted Topsis method, Planning modeling.

1. Introduction

In fresh food superstores, due to the perishable nature of vegetable items, they often become unsaleable the following day. Supermarkets are required to replenish their stock daily based on historical sales data; however, merchants frequently restock without specifying demand. Therefore, the pricing of vegetables plays a crucial role in replenishment and pricing decisions. Constrained by limited sales space, merchants must devise sales and pricing strategies based on historical sales data and customer demand.

The data of this research comes from: www.mcm.edu.cn. Considering that vegetable categories are highly seasonal, we guessed that their data may be cyclical, so we decided to use the ARIMA model for further analysis. LSTM time series forecasting model, based on the entropy weight Topsis method we first pre-process the obtained data, and deal with missing values and outliers of the data, from the two different time dimensions, namely, month and quarter. Starting from the two different time dimensions of month and quarter, we fit the data and analyze the change rule. The LSTM time series prediction model was used to predict the cost data. Construct a comprehensive evaluation model to evaluate the vegetables that affect the importance of single-item vegetables, and solve for the top 33 single-item category vegetables. The planning model was solved using a genetic algorithm^[1] to derive the required replenishment and pricing strategy for individual items. Finally, the hierarchical analysis and entropy weight method were used to analyze the carry-over objective function solution.

2. Vegetable category relationships

2.1. Data analysis and pre-processing

It can be inferred that the vegetable goods carried by the superstore are categorized into six categories. From the demand side, there is often a certain correlation between the sales volume of vegetable goods and time. Here we integrate the daily sales data of the six categories of vegetable commodities and summarize and analyze the sales volume data every month. We can derive the aggregated trend for a total of 36 months from June 2020 to July 2023, as shown in Figure 1.

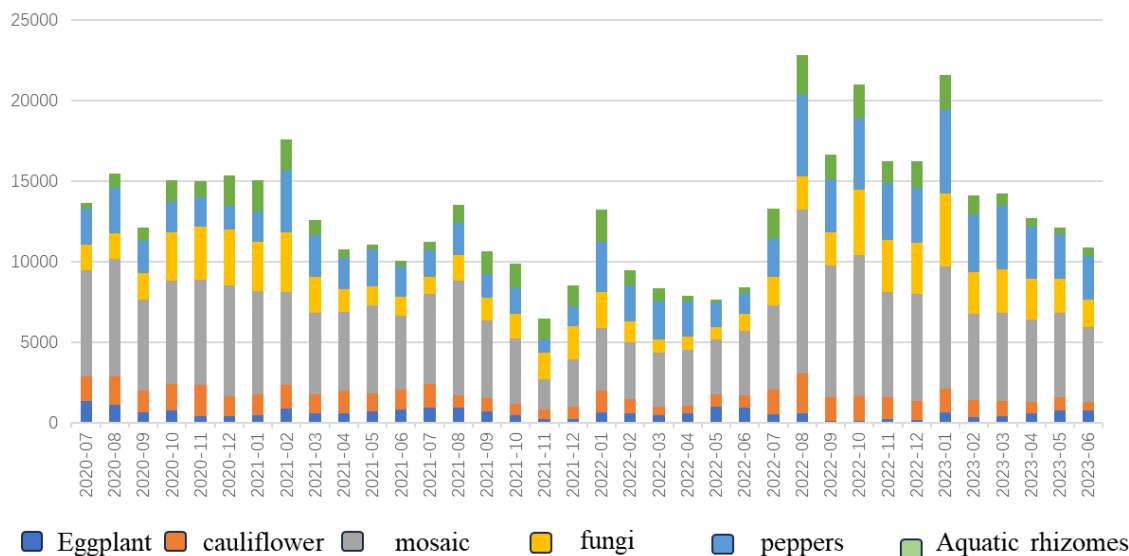


Figure 1: Summary analysis of various categories of vegetables

From the above graph we can see that for a single vegetable category, the flower and leaf data has the highest sales volume among the six vegetable varieties, while the eggplant data has the lowest sales volume. As a whole, the data fluctuations are more variable, especially during the period March 2021- June 2022, the data was found to have lower overall sales. We speculate here that it may be affected by factors such as the epidemic; after June 2022, the sales volume shows a rebound trend and even has a larger increase than before. Additionally, our findings revealed that the peak season is from September to January, and sales are generally low from March to July.

2.1.1 Research on the distribution pattern of single product sales volume based on Spearman correlation analysis

First, we analyzed the data over time with descriptive statistics. Including statistics such as sample size, maximum value, and average value were used to study the quantitative data as a whole.

The average sales of vegetables in a single category have increased over time, and the median sales have had a large increase over the previous period.

Spearman correlation analysis^[2] is a nonparametric statistical method used to assess the strength and direction of the relationship between two variables. It is based on the concept of ranking and solves the problem by ranking the values of each variable and then calculating the correlation between these rankings. From the above table we understand that the number of sales in a vegetable category is skewed with time, the distribution trend in the data set is to the left and it does not pass the test of normal distribution, also, the sales volume and time is a relationship between two continuous variables, so we decided to use Spearman correlation analysis.

The Figure 1 reveals a skewed sales trend in vegetables over time, violating normal distribution. Given the continuous relationship between sales volume and time, we applied Spearman correlation analysis. We segmented the data into three annual periods (2020-2021, 2021-2022, 2022-2023) using six vegetable categories. Analyzing Spearman coefficients within each category revealed:

Strongest positive correlation ($r=1$) in flowering & leafy veggies: red coral, red oak leaf, green butter. Strongest negative correlation ($r=-0.409$) in cauliflower: green peduncle loose flower, Zhejiang variety.

Highest positive correlation ($r=0.5712$) in aquatic roots: rhododendron, Honghu root veggies. The strongest positive correlation ($r=0.8786$) in eggplants: small peppers, and small wrinkled peels.

These results indicate a yearly increase in sales correlation within the six veggie categories, reflecting a growing sales abundance over time.

2.1.2 Test for seasonal effects of vegetable categories and individual categories based on ARIMA seasonal decomposition

ARIMA model is a common linear model in time series analysis, in which the seasonal ARIMA model is an ensemble product model for cyclically varying data. The model is expressed quantitatively by integrating the trend, periodicity, and random disturbances of the series with the help of statistical modeling. Considering the strong seasonality of the vegetable category, we guessed that its data might be cyclical, so we decided to use the ARIMA model for a closer analysis.

Based on the above analysis, we finally decided to choose a window with a period length of 7. Details about this is shown in Figure 2.

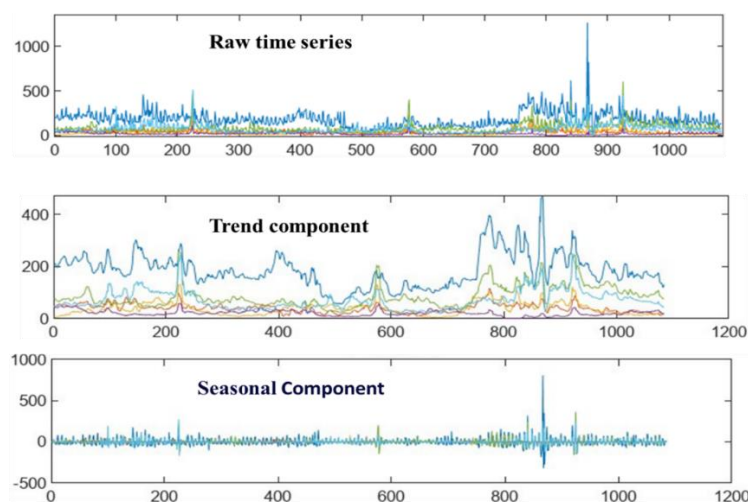


Figure 2: Diagram of time series data

The seasonality factor was tested as an evaluation index, and the final ranking relationship between the strength of the cycle in the six categories of vegetables is shown in the Table 3:

Table 3: Ranking of the Strength of the Vegetable Category Cycle

Periodic strength rankings	Vegetable categories
1	Aquatic rhizomes
2	Mosaic and leafy
3	edible mushrooms
4	Nightshades
5	Chili peppers
6	Cauliflower

Similarly, we derived that the top six relationships for the strength of the ranking cycle in a single category of vegetables at Table 4:

Table 4: Ranking of the cycle of vegetable single products

Periodic strength rankings	Vegetable items
1	Green butter
2	Cabbage leaves
3	Hongshan vegetable crush treasure tote bag
4	Red Coral (Coarse-leaved)
5	Fruit pepper (orange)
6	Lion's mane mushroom

2.2. Forecast of trends in vegetable sales

2.2.1 Prediction of cost data based on LSTM models

The relationship between sales volume and cost-plus pricing needs to be taken into account to analyze the combined pricing strategy for the next seven days, and given that the sample data is

partially missing, we decided to forecast the cost data based on the LSTM time series prediction model^[3].LSTM can effectively capture semantic associations between long sequences, selectively adding new information and forgetting previously accumulated information, mitigating the gradient loss or the explosion phenomenon, which can effectively help us solve the problem of missing data. The results obtained in Table 5 occupy important weights in the subsequent prediction of the sales volume of single vegetable products.

The core structure of the LSTM neural network consists of forgetting gates, input gates, cell states, and output gates, the following is a list of neural network parameters as well as an introduction:

Table 5: LSTM neural network parameters

parameter	meaning	parameter	meaning	parameter	meaning
x_t	Input at the current moment	f	Forgotten door control signal	C_{t-1}	The memory unit of the previous moment
$\tilde{C}_t$	Information candidate status	i_t	Enter the gate control signal	C_t	Memory unit at the current moment
h_{t-1}	The external state of the previous moment	o_t	Output gate control signal	h_t	The external state at the current moment

We take the data of the flower and leaf category as an example for Univariate Prediction. Considering the need to study the total daily replenishment and pricing strategy of each vegetable category in the coming week, we decided to set the moving step to 7. A sliding window of one week for rolling. RMSE and MAE are also used as evaluation metrics to observe the changes in their iterations. The results are shown in Figure 3 respectively:

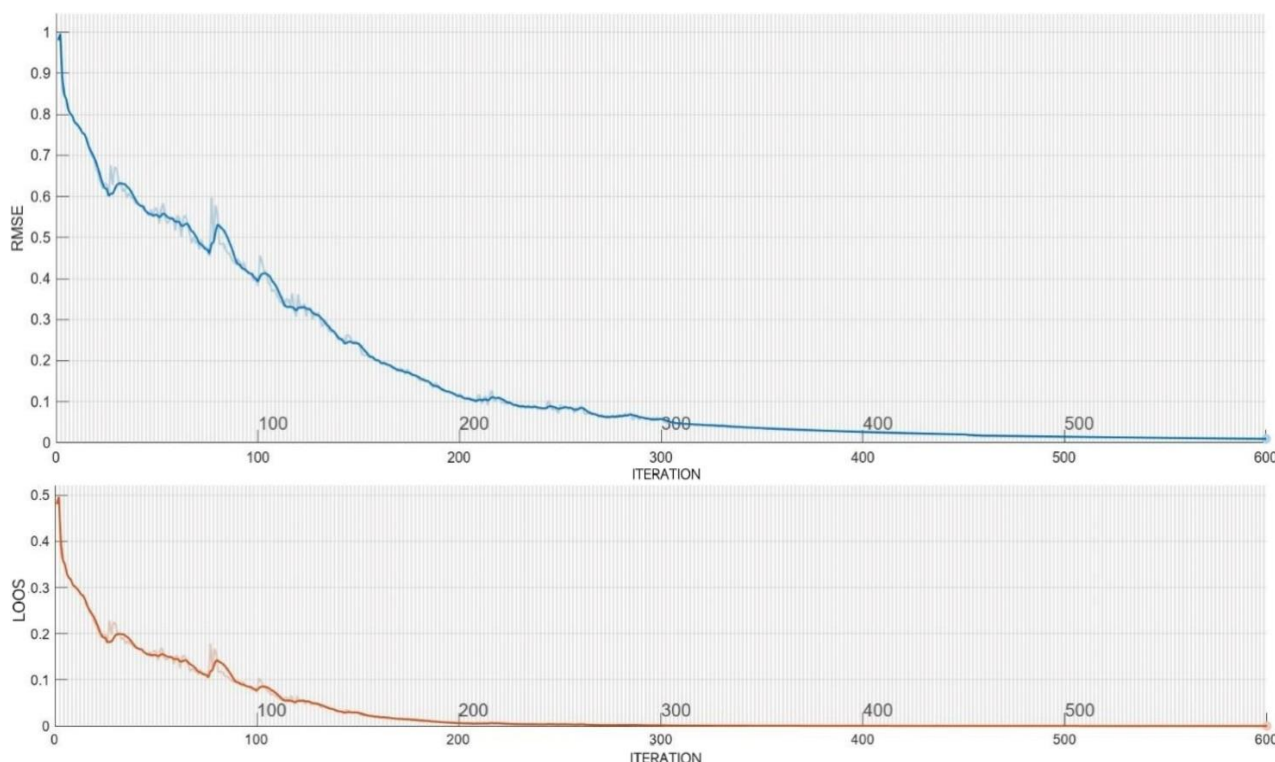


Figure 3: Visualization of forecasts for the next seven days

The results of using the fitted LSTM model to go for the prediction of future values respectively are shown in Table 6:

Table 6: LSTM neural network forecasts cost data for the next 7 days

time	Mosaic leave	cauliflowers	aquatic rhizomes	nightshades	peppers	Edible fungi
1	6.15	13.99	10.61	7.90	8.51	6.60
2	4.86	15.26	11.21	8.99	8.17	6.32
3	3.68	14.61	10.73	12.86	7.42	6.53
4	6.32	17.18	10.58	17.31	9.35	6.47
5	5.87	26.11	10.39	14.90	7.21	6.37
6	2.38	28.03	10.18	15.19	7.17	6.05
7	5.41	22.30	9.90	13.70	7.19	5.80

2.2.2 Relationship between total volume and cost-plus pricing based on least squares analysis

Supermarkets will replenish their stock of vegetables based on daily sales volume and customer demand, while superstores utilize a cost-plus pricing method for automatic pricing and replenishment decisions. Discounted sales are implemented for certain dishes, with high-sales-volume items typically not included in the discount strategy. The relationship between total sales volume and cost-plus pricing is therefore considered ^[4].

$$y = \alpha_0 + \alpha_1 x + \varepsilon \quad (1)$$

Where y is the dependent variable, x is the independent variable, and ε is the error term as a perturbation. Based on the above introduction, we thus obtain the fitting function represented by $L(a, b)$.

The value of the parameter K_n is finally obtained, through which the fitted curves of the six vegetable categories in the question are obtained, and the motion trend of the fitted curves is observed to obtain the fitted equations. We fit the primary function, quadratic function, and cubic function to the six vegetables, categories, and observe the fitting trend and the corresponding fitting equations, as shown in Figure 4:

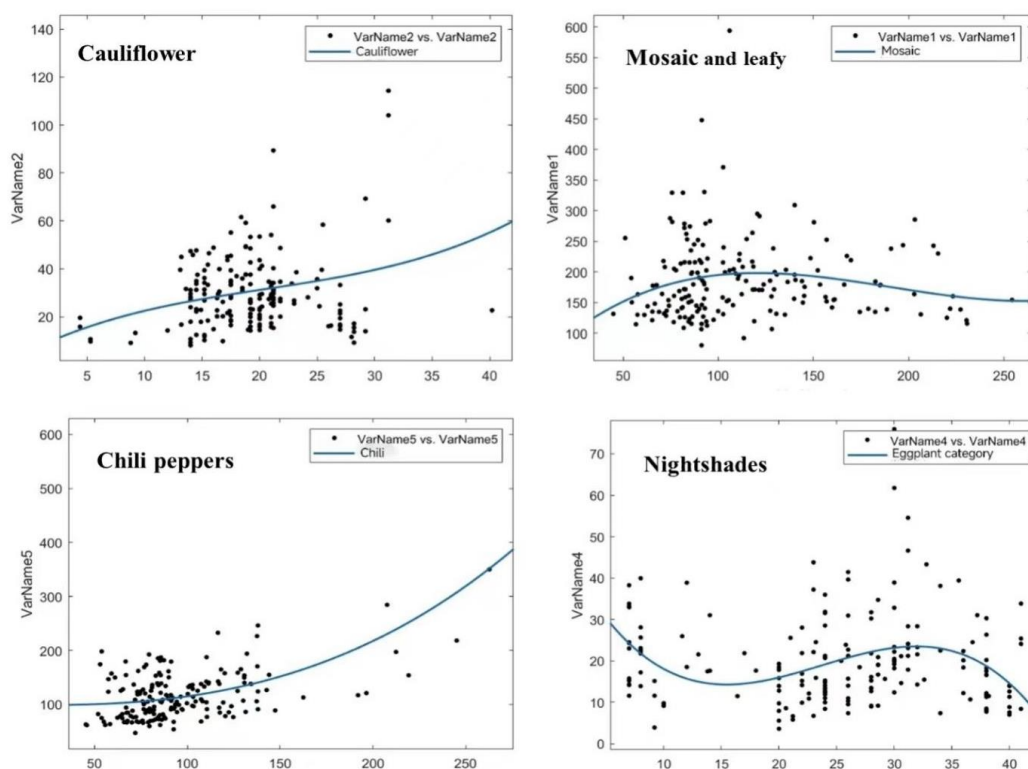


Figure 4: Scatter plot and fitting curve between total sales volume and pricing of vegetable categories

Equation fitting for vegetable categories can help merchants better understand market dynamics and make more accurate decisions , The fitting equations corresponding to kinds of vegetables are shown in Table 8:

Table 8: Fitting equations for six major vegetable categories

Vegetable categories	Fit the equation
Mosaic and leafy	$f(x) = -0.0192x^2 + 3.1591x + 37.5156$
Cauliflower	$f(x) = 0.0013x^3 - 0.0780x^2 + 2.326x^3 + 5.7276$
Nightshades	$f(x) = -0.0042x^3 + 0.296x^2 - 6.17x + 54.4$
Aquatic rhizomes	$f(x) = 0.004x^3 - 0.0573x^2 + 2.1116x + 22.6042$
Chili peppers	$f(x) = 0.0016x^2 - 0.1134x + 100.8717$
Edible mushrooms	$f(x) = -0.0009x^3 + 0.1991x^2 - 12.3786x + 294.3847$

2.2.3 Particle Swarm Optimization Algorithm-Based Replenishment Volume and Pricing Strategy for Individual Products

After some research, we found that the basic idea of the cost-plus pricing method is that the target price of the product = cost base + cost of ownership.

we can model the problem as an optimization problem to find the maximum profit gain. The daily replenishment quantity and pricing strategy are given by adding a single vegetable category selection as well as a minimum display quantity. Here we have decided to carry out the construction of a planning model based on a particle swarm optimization algorithm for vegetable item replenishment quantity and pricing strategy^[5].

For all particles, update the population best position: find the global best position, i.e. the optimal solution in the population, among the historical best positions among all particles.

$$V = c_1 \cdot randf \cdot (x_i pbest - x_i) + c_2 \cdot randf \cdot (x_i gbest - x_i) \quad (2)$$

$$x_i' = x_i + V \quad (3)$$

To speed up the convergence, according to the principle that velocity has inertia, the inertia is added to the velocity calculation, i.e., the velocity of the position parameter in the last iteration is added. The following formula, where w in a parameter of 0~1, usually with the increase of the number of iterations, gradually w reduces, because the closer to the back of the possible solution will be closer to the real solution, the slower the convergence of the iteration, so it is necessary to reduce w slow down the speed, or else it is easy to miss the optimal solution^[6].

$$V = V_{pre} \cdot w + c_1 \cdot randf \cdot (x_i pbest - x_i) + c_2 \cdot randf \cdot (x_i gbest - x_i) \quad (4)$$

3. Results and Discussion

Repeatedly updating the individual best and the population best and stopping the iteration when the maximum number of iterations is reached or a satisfactory solution is found. Programming calculations are performed using Matlab and the final output of the converged optimal solution of the algorithm is shown in Table 9 and Table 10, the prices of Cauliflower, Aquatic rhizomes, Cauliflower, Chili peppers, Edible mushrooms, Nightshades, Aquatic rhizomes, Nightshades , Edible mushrooms did not fluctuate, indicating the stability of the market, However, Mosaic and leafy prices showed a continuous rising trend during the observation period, showing strong market demand^[7].

Table 9: Pricing strategy data for the next 7 days

Vegetable categories	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Mosaic and leafy	122.26	122.12	122.30	122.36	122.38	122.42	122.44
Cauliflower	11.49	11.49	11.49	11.49	11.49	11.49	11.49
Aquatic rhizomes	82.27	82.27	82.27	82.27	82.27	82.27	82.27
Nightshades	27.83	27.84	27.84	27.84	27.84	27.85	27.85
Chili peppers	43.87	43.87	43.87	43.87	43.87	43.87	43.87
Edible mushrooms	39.02	39.02	39.02	39.02	39.02	39.02	39.02

Table 10: Purchase data for the next 7 days

Vegetable categories	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Mosaic and leafy	186.25	186.26	186.25	186.25	186.25	186.25	186.25
Cauliflower	23.74	23.74	23.74	23.74	23.74	23.74	23.74
Aquatic rhizomes	13.63	13.63	13.63	13.63	13.63	13.63	13.63
Nightshades	19.80	19.80	19.80	19.80	19.80	19.80	19.80
Chili peppers	74.02	74.02	74.02	74.02	74.02	74.02	74.02
Edible mushrooms	50.87	50.87	50.87	50.87	50.87	50.87	50.87

4. Conclusions

Through detailed preprocessing of vegetable commodity classification data and using Spearman correlation analysis, this study reveals a significant correlation between vegetable sales volume and time, and the sales volume shows a clear trend of increasing over time. Furthermore, this study used the ARIMA seasonal decomposition model to further explore the seasonal characteristics of vegetables and their classification, specifically pointing out that certain vegetable categories (such as aquatic roots and tubers) exhibit significant seasonal fluctuations.

In order to predict the future price trend of six major vegetables, this study applied the LSTM model, which provides a scientific basis for businesses' replenishment and pricing strategies by analyzing the patterns and trends of historical price data. In addition, through the analysis of the least squares method, this study compared the relationship between total sales volume and cost-plus pricing, and constructed a fitting curve based on the scatter plot of sales volume and vegetable prices, providing precise support for the decision-making of merchants.

This study also used the particle swarm optimization algorithm to construct a planning model for the replenishment quantity and pricing strategy of individual vegetable products. Through the iterative process of continuously updating the individual optimal solution and the group optimal solution, this study successfully predicted the pricing trend of each vegetable for the upcoming week. The research results show that the application of prediction models and optimization methods significantly improves the efficiency of supply chain management and helps to maximize profits.

The fitting equation in this study provides a tool for businesses to gain a deeper understanding of market dynamics, while the application of particle swarm optimization provides practical guidance for formulating replenishment and pricing strategies. Ultimately, the application of these methods is expected to significantly increase supermarket profits and reduce costs. This study not only provides scientific methodologies and tools for fresh food supply chain management, but also provides practical references for retailers in replenishment and pricing decisions.

References

- [1] Liu Tianqi and Lin Yonghong Optimization of Cold Chain Logistics Distribution Routes Based on Improved Genetic Algorithm [J] National Circulation Economy, 2024, (07): 52-55.
- [2] He Shuang. Analysis of factors affecting institutional evaluation results based on Spearman rank correlation and Topsis-fuzzy comprehensive judgment algorithm model [J]. Journal of Changchun Normal University, 2022, 41(10): 9-16.

- [3] Hu Yuyang, Zhang Cheng, Zheng Ming et al. Design and implementation of a sales forecasting system based on LSTM[J]. Computer and Network,2020,46(23):65-67.
- [4] HUANG Lixue, YAO Guanxin. Research on horizontal replenishment of fresh agricultural Produce in supermarket chains[J]. Logistics Science and Technology,2024,47(01):100-104.
- [5] Cai Xiao. Research on Hybrid Quantum Particle Swarm Optimization for Interference Management in Agricultural Product Logistics and Distribution [D]. Dongbei University of Finance and Economics, 2018
- [6] ZHANG Qin Yu, HU Zhigang, Xi Jin, et al. Optimization of back-propagation neural network based on particle swarm optimization algorithm to construct a near-infrared spectral prediction model for the freshness of refrigerated grass carp[J]. Journal of Food Safety and Quality Testing,2023,14(22):200-209.
- [7] Chen Changying. Research on logistics demand forecasting in Guangxi based on improved gray-Markov chain model[J]. Western Transportation Science and Technology, 2019(10):152-155.
- [8] C.Y. Han. Inventory commodity demand forecasting by LSTM model based on data differencing[J]. China Storage and Transportation,2023(09):152-153.
- [9] MING Xiaoju, ZHU Lan. Research on optimization technology of cold chain logistics distribution path for urban fresh food[J]. Packaging and Food Machinery,2022,40(02):76-81
- [10] Fan Bin, Yu Bin. Optimization of supplier replenishment strategy based on particle swarm optimization algorithm [J]. Statistics and Decision,2014,(12):44-46.