Research on automatic vegetable pricing and replenishment decision-making based on improved ant colony algorithm

Zihao Wang[1,*], Kehang Mu[2], Yuhan Ma[2]

[1]School of Electrical Engineering and Automation, Anhui University, Hefei, Anhui, China, 230601

[2]School of Computer Science and Technology, Anhui University, Hefei, Anhui, China, 230601

*Corresponding author: w1927427834@outlook.com

Abstract. In daily life, vegetable products have a short shelf life and a high perishability rate. As time goes by, their quality will deteriorate and cause waste. Therefore, in order to maximize benefits and reduce waste, the ant colony algorithm and ARIMA are combined The prediction model was combined and optimized, and the Ant Colony-ARIMA prediction model was established, and the relevant data of the sales flow details and wholesale prices of each commodity in a supermarket from July 1, 2020 to June 30, 2023 were used for training, optimization and prediction. Analyze and compare with various data from June 24, 2023 to June 30, 2023. The results show that the average relative errors of sales and pricing of the Ant Colony-ARIMA model are lower than the actual ones, indicating that the established Ant Colony-ARIMA model can accurately predict sales and pricing, and provide a high-quality solution for supermarkets and merchants. Accurate prediction method.

Keywords: Ant colony algorithm, ARIMA prediction model, Automatic replenishment and pricing.

1. Introduction

With the development of the global economy, people's shopping needs are also increasing, in order to meet the growing shopping demand, the variety and quantity of supermarket products are becoming more and more complete. However, the problem that arises is that the replenishment and pricing of supermarket goods becomes difficult due to the gradual increase in the number of goods. The traditional manual management method of supermarket replenishment can no longer meet the needs of the times [1-3]. Traditional pricing method is to analyze the relationship between agricultural product supply chain, agricultural product price prediction and price fluctuation factors and pricing based on ARIMA model, or establish a pricing mechanism, formulate acceptance criteria, and establish a supervision mechanism to formulate a reasonable pricing scheme [4-5]. However, there are still differences between the prediction results of the above methods and the actual situation, and this paper proposes the wholesale price prediction results using the AIRMA model on the basis of the predecessors, and inputs the prediction results into the ant Colony Algorithm model, and uses the cost-plus pricing method [6] to perform secondary calculations to obtain the prediction results, and verifies them on the 3-year sales data of a supermarket, and the results far exceed other methods in accuracy.

2. Automatic vegetable pricing and replenishment decision model

Due to the large number of vegetable products, in order to facilitate analysis and discussion, this article divides vegetables into six categories according to their characteristics, namely: Anthophyllum, Cauliflowers, Aquatic roots, Nightshades, Capsicum and Edible fungi. Discuss at the category level.

2.1. Cost-plus pricing model

In supermarkets, cost-plus pricing is usually used to determine pricing. The meaning of cost-plus pricing is that the price at which the product is sold is sufficient to cover the cost of production and sales and get a certain return. The commonly used pricing model for cost-plus pricing is:

$$Y = C(1+w) \tag{1}$$

Y represents pricing, C represents total cost, and w represents cost markup rate.

Since vegetables are subject to loss during transportation, the vegetable loss rate also needs to be considered, so the total cost C should include wholesale cost C_b and loss cost C_l [7]. Then the final cost pricing model is:

$$Y = (C_l + C_b)(1+w) \tag{2}$$

2.2. Autoregressive Integrated Moving Average model

Autoregressive Integrated Moving Average (ARIMA) models are a broad class of stochastic process models, including autoregressive terms (AR), integral terms (I), and moving average terms (MA). where AR involves a regression model of lagging time series values, I is used to eliminate non-stationary, and MR involves an error term as a linear combination of previous error terms[8].

The ARIMA model can be expressed as: $ARIMA(p,d,q)$ where: p is the order of the autoregressive part, d is the order of the difference, q is the order of the moving average part, the general form of the ARIMA model is:

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \tag{3}$$

Among them, $\varphi_1, \varphi_2, \cdots, \varphi_p, \theta_1, \theta_2, \cdots, \theta_q$ is a parameter and ε_t is white noise.

Through Formula (4), relying on existing time series data and adjusting parameters p, d, and q can accurately predict relevant data within a certain period of time.

3. Results

3.1. Results of the relationship between vegetable sales and pricing

Vegetable market prices have a medium-term impact on seed demand forecasts. In turn, it affects the sales of vegetables[9].

To accurately formulate replenishment totals and pricing strategies, the relationship between total sales volume and cost-plus pricing should be determined. This article uses the data related to the sales details and wholesale prices of each commodity in a supermarket from July 1, 2020 to June 30, 2023 as a sample for analysis.

According to relevant data, it is found that there is a certain regression relationship between total sales volume and cost-plus pricing. Therefore, in the relational expression, daily cost-plus pricing y is selected as the dependent variable, and total sales volume x is selected as the dependent variable for analysis. The related results of mosaic and leaf categories are shown in Table 1 and Table 2.

Table 1 Significance result table of mosaic and leaf linear equations of one variable

Model		Sum of square	df	Mean Square	F	Sig.
Anthophyllum	regression	135.570	1	135.570	62.379	0.000
	residual	2353.703	1083	2.173	.	.
	total	2489.273	1084	.	.	.

Table 2 Significance results table of coefficients of mosaic and leaf linear equations

Model		Unstandardized coefficient		t	Sig.
		B	Standard Error		
Anthophyllum	(Constants)	6.353	0.105	60.471	0.000
	Total sales	-0.004	0.001	-7.898	0.000

According to Table 1, the value of the F statistic is 62.379, and the corresponding p value is less than 0.05. Therefore, the null hypothesis is rejected and the equation is considered to be significant.

The linear relationship between the cost-plus pricing of mosaic and leafy vegetables and the total sales volume is significant. According to Table 2, the value of the statistic is 60.471, and the corresponding value is less than 0.05. Therefore, the null hypothesis is rejected and it is believed that the cost-plus pricing of cauliflower vegetables has a significant impact on the total sales volume.

Therefore, regression analysis can be performed on the total sales volume and cost-plus pricing data of each category to obtain the relevant regression equation results. The relevant results are shown in Table 3.

Table 3 Various regression equation tables

Category	Regression equation
Anthophyllum	$y_1 = 6.353 - 0.004x_1$
Cauliflowers	$y_2 = 10.661 - 0.032x_2$
Aquatic roots	$y_3 = 11.294 - 0.042x_3$
Nightshades	$y_4 = 9.595 - 0.036x_4$
Capsicum	$y_5 = 9.674 - 0.012x_5$
Edible fungi	$y_6 = 9.482 - 0.013x_6$

3.2. Time series forecasting of wholesale prices

In order to find the cost-plus pricing strategy, it is necessary to know what the total cost c is. It is known that the total cost C is composed of the wholesale cost C_b and the loss cost C_l, and the loss cost C_l can be determined by the loss rate l and the wholesale cost C_b. Therefore, the final total cost c can C obtained by determining the wholesale cost C_b.

There are already wholesale cost data of relevant categories in the past three years and recent loss rate data. Therefore, wholesale cost data for the future period can be used for time series prediction using the wholesale cost data of the past three years and using the ARIMA model to conduct time series forecasts, and perform ADF, ACF and PACF test, get $p=5,\ d=1,\ q=0$, The forecasted wholesale cost data for the next seven days, that is, July 1 to July 7, is shown in Table 4.

Table 4 Wholesale cost table of various categories in the next 7 days

Category	July 1st	July 2nd	July 3rd
Anthophyllum	3.6962	3.6534	3.6253
Cauliflowers	7.8079	7.8871	7.8144
Aquatic roots	11.9878	11.9383	11.8872
Nightshades	4.4443	4.4408	4.4350
Capsicum	5.9397	5.9684	5.9234
Edible fungi	4.6938	4.6761	4.7006

（continued table）

Category	July 4th	July 5th	July 6th	July 7th
Anthophyllum	3.622	3.6138	3.6212	3.6264
Cauliflowers	7.8213	7.7883	7.7855	7.7957
Aquatic roots	11.9793	11.9479	11.9517	11.9445
Nightshades	4.4356	4.4338	4.4353	4.4351
Capsicum	5.9511	5.9225	5.9383	5.9371
Edible fungi	4.6968	4.7020	4.6899	4.6942

3.3. Maximum benefit solution using ant colony algorithm

Using the regression equation of each category as a constraint, the final profit P can be solved as the objective function:

$$\max P = \sum_{i=1}^{6}\sum_{j=1}^{7} P_{it} = \sum_{i=1}^{6}\sum_{j=1}^{7} (y_{it} - c_{it})x_{it}$$

$$s.t.\begin{cases} y_{it} = \alpha_i + \beta_i x_{it} \\ x_{it}, y_{it}, P_{it} \geq 0 \\ \omega_{ij} \in R \\ i = 1, 2, \cdots, 6 \\ t = 1, 2, \cdots, 7 \end{cases} \tag{4}$$

Among them, i is the category and t is the time.

Obviously this is a single-objective optimization problem, and the global optimal solution obtained during these seven days is not the optimal solution on a certain day. Therefore, in order to obtain this result, this article decided to use the ant colony algorithm to solve it.

Ant colony algorithms are abstracted from ant foraging. To make the ant colony algorithm more efficient at dealing with real-world optimization problems, the researchers programmed artificial ant pheromones by mimicking the tracking pheromones of natural ants. Pheromones guide the ants, and they leave pheromones behind as they walk. In the ant colony algorithm, by continuously traversing the path, the pheromone volatilization characteristics were artificially set, and the pheromone concentration was higher in the path with shorter distance or faster convergence speed. In addition, because the pheromone concentration of the better path is higher than that of the worse path, and the selection bias is more toward the optimal path, the pheromone concentration of the short-distance, fast-shrinking path continues to increase, and the pheromone concentration of the long-distance, slow convergence path continues to increase. Pheromone concentrations dissipate, and eventually there is always a path to rapid convergence[10]. The probability of ant k moving from i to j is as follows:

$$P_{ij}^{k} = \begin{cases} \dfrac{[\tau_{ij}(t)]^{\alpha}[\eta_{ij}(t)]^{\beta}}{\sum_{s \in allow_k} [\tau_{ij}(t)]^{\alpha}[\eta_{ij}(t)]^{\beta}} & s \in allow_k \\ 0 & s \notin allow_k \end{cases} \tag{5}$$

The update rules are as follows:

$$\tau_{ij} \leftarrow (1-\rho)\tau_{ij} \forall (i, j) \in E \tag{6}$$

$$\tau_{ij} \leftarrow \tau_{ij} + \Delta\tau_{ij} \forall (i, j) \in Tours^{(k)} \tag{7}$$

The specific solution process is shown in Figure 1:

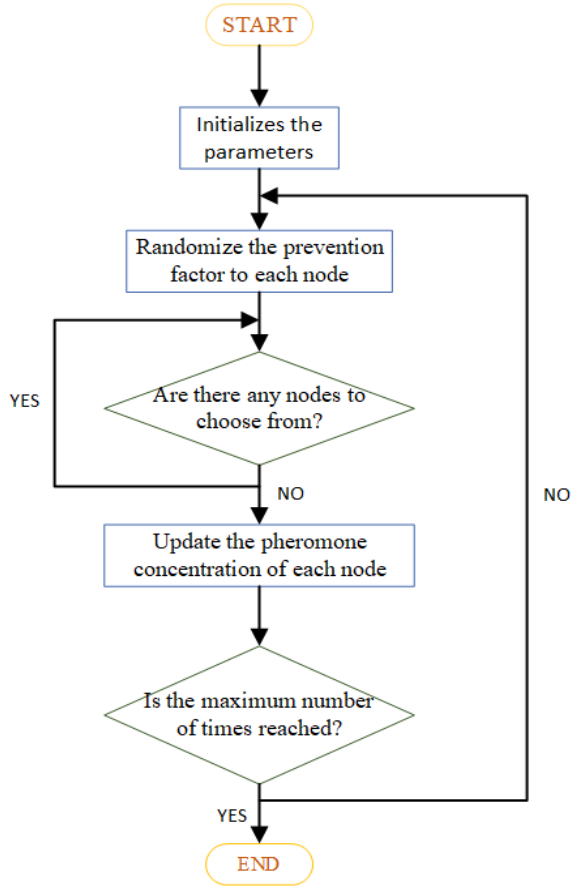

Figure 1 Ant colony algorithm flow chart

By using the ARIMA model to solve the predicted wholesale prices from July 1st to July 7th, and bringing them into the model, there is recent certain data on the loss rate, so the loss rate is considered to be a fixed value here.

The calculated maximum value of supermarket revenue, as well as the total daily replenishment volume and pricing strategy of each vegetable category this week are shown in Table 5.

Table 5 Daily Income Statement

Date	Profit
July 1st	732.49
July 2nd	1046.76
July 3rd	706.06
July 4th	611.66
July 5th	320.23
July 6th	806.54
July 7th	892.12

Among them, the total daily replenishment volume and pricing strategy of nightshade vegetables from July 1 to July 7 are shown in Table 6.

Table 6 Daily replenishment and pricing strategy table of nightshade

Date	Pricing strategy	Total replenishment(kg)
July 1st	0.4619	43.0959
July 2nd	0.9274	46.6327
July 3rd	0.5609	81.2315
July 4th	0.8684	68.0027
July 5th	0.0011	73.4227
July 6th	0.3977	76.2643
July 7th	0.9319	69.7305

For comparison, the total revenue of each category of vegetables in the past week (June 24 to June 30) was calculated. The calculation results are shown in Table 7.

Table 7 Total revenue of each category in the past week

Category	Profit
Anthophyllum	1120.43
Cauliflowers	394.51
Aquatic roots	469.07
Nightshades	321.38
Capsicum	1175.69
Edible fungi	507.33

Comparing the income before and after optimization, it was found that the results after optimization by this algorithm increased the income by 28.27% compared with before optimization.

4. Conclusion

The relationship between cost-plus pricing and total sales volume was analyzed, and it was found that there was a strong relationship between the data, and relevant regression tests were conducted. The results were found to be significant, so it was determined that there was a certain regression relationship. When the past wholesale cost data is known, the ARIMA model is used to predict the wholesale cost data for the next week, and when the loss rate is known, the corresponding cost-plus pricing model is established, and the corresponding relationship formula is obtained. Then the total profit in the next week is used as the objective function, and the relationship between each regression equation and the cost-plus pricing model is used as a constraint. The ant colony algorithm is used to adjust the corresponding parameters, and finally the best profit in the next seven days and the corresponding daily price of each category are obtained. Comparing the replenishment and pricing strategies with those before optimization, it was found that the revenue increased by 28.27%. This model effectively improves the final revenue and has a certain degree of universality. It can be applied to many fields only by knowing the relevant data, such as enterprise product sales, collection of related goods, etc.

References

[1] Cheng Zhongxin. Design and implementation of supermarket automatic replenishment system [D]. Shandong University, 2016.

[2] Zheng Couple. Research on dynamic bulk inventory replenishment of community group buying retailers [D]. Central South University of Economics and Law, 2022.

[3] LI Chunxiao. Research on the design of automatic replenishment system of unmanned supermarket [D]. Hebei University of Science and Technology, 2019.

[4] Maurisha. Research on pricing strategy and production and marketing mode of vegetable wholesale market from the perspective of supply chain [D]. Central South University of Forestry and Technology, 2023.

[5] Fang Li. Research on the price setting method of vegetable raw materials in college logistics [J]. University Logistics Research,2021(10): 22-24.

[6] Angel. Research on the pricing of scientific and technological search services based on cost-plus pricing method [J]. Library Research and Work, 2021(10): 25-31+24.

[7] Arivazhagan, R., Dinesh, et al. VEGETABLES WASTAGE ANALYSIS IN WHOLESALE POINTS OF TAMILNADU [J]. Journal of Critical Reviews, 2020, 7(17).

[8] Chodakowska E, Nazarko J, Nazarko Ł, et al. ARIMA Models in Solar Radiation Forecasting in Different Geographic Locations [J]. Energies, 2023, 16(13)

[9] Yihang Z, Yinglei Z, Jingjin Z, et al. Spring onion seed demand forecasting using a hybrid Holt-Winters and support vector machine model [J]. PloS one, 2019, 14(7).

[10] Dan L, Xiulian H, Qi J. Design and optimization of logistics distribution route based on improved ant colony algorithm [J]. Optik, 2023, 273