

Vegetable Automated Pricing Model Based on Particle Swarm Algorithm

Bosi Wang *

Faculty of Economics, Fudan University, Shanghai, China, 200433

* Corresponding Author Email: 22300680049@m.fudan.edu.cn

Abstract. Fresh grocery supermarkets face challenges such as the short shelf life of vegetables, quality degradation over time, and difficulty in selling products the following day. Considering these issues, they restock early in the morning without complete knowledge of the specific items or purchase prices, based on historical sales and demand data. To address this issue, this paper first visualizes the data to demonstrate a positive correlation between sales volume and time. Next, to address the issue of predicting the total daily restocking quantity and pricing strategy from July 1 to July 7, 2023, the document utilizes data from the previous three years to establish an ARIMA time series forecast model to predict the costs and sales volume of six vegetables for the first week of July 2023. Finally, to devise the replenishment quantity and pricing strategy for individual items on July 1, 2023, the paper filters the data for all vegetable items from June 24th to 30th, 2023, and combines it with the particle swarm optimization algorithm to obtain the results. This study is pragmatic, providing a feasible solution to the problem of automatic pricing for fresh vegetables.

Keywords: Prediction Mode, ARIMA Time Series Model, Particle Swarm Optimization.

1. Introduction

Fresh vegetable supermarkets face challenges such as short shelf lives of vegetables, quality degradation over time, and difficulty selling produce the following day [1]. Decisions are made based on historical sales and demand data, with typical strategies including "cost-plus pricing" for setting prices and offering discounts on goods that have suffered transportation damage or are of inferior quality [2].

For inventory control models in fresh food supermarkets, genetic algorithms that use binary encoding have been the traditional approach. However, other researchers have turned to the Apriori algorithm, a data mining technique, to identify correlations in datasets and optimize procurement decisions [3]. While genetic algorithms with binary encoding have their strengths, the Apriori algorithm can provide valuable insights into demand, purchasing behavior, and more. The use of these techniques can inform inventory management decisions, improve supply chain efficiency, and lead to better customer satisfaction and profitability in the fresh food market [4]. Researchers have also constructed an automatic pricing model for an e-commerce website based on deep learning and BP neural network technology. The model can generate recommended prices for products by analyzing both structured and unstructured data provided by sellers, taking into account the information about the products [5].

This article utilizes the ARIMA time series forecasting model to predict the vegetable procurement and pricing strategy for July 1st to 7th, 2023. Furthermore, an improved particle swarm algorithm is used to develop the restocking plan for vegetable products and pricing strategies for July 1st, 2023. Particle swarm optimization converges to the optimal solution more quickly than genetic algorithms for high-dimensional optimization problems. By utilizing these analytical models and algorithms, accurate predictions of demand and pricing can be made, enabling fresh food supermarkets to optimize their inventory and pricing strategies for maximal profit.

2. The basic functions of the model

2.1. The Structure of the ARIMA Time Series Model

ARIMA (Autoregressive Integrated Moving Average) is a widely used time series forecasting model that analyzes and predicts future trends based on past observations and patterns. It is a type of statistical model that uses a combination of autoregressive (AR), moving average (MA), and differencing (I) techniques to model the behavior of time series data and make forecasts [6].

ARIMA model primarily focuses on the analysis of the time series data and aims to identify the underlying patterns and relationships among the data points [7]. The model uses past observations to predict future values of the series by estimating the parameters of the model using optimization techniques like maximum likelihood estimation [8].

The ARIMA model consists of an AutoRegressive (AR) component, an Integrated (I) component, and a Moving Average (MA) component.

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} \quad (1)$$

where Y_t is the observed value at time point t . c is the constant term. $\phi_1, \phi_2, \dots, \phi_p$ are the AutoRegressive coefficients, representing the influence of past p observed values on the current value. p is the order of the AutoRegressive part. ϵ_t is the white noise error term. $\theta_1, \theta_2, \dots, \theta_q$ are the Moving Average coefficients, representing the influence of past q error terms on the current value. q is the order of the Moving Average part.

In summary, the ARIMA model's formula combines the AutoRegressive, Integrated, and Moving Average components to describe the trend and seasonality in a time series.

The fundamental idea of an ARIMA model is to determine the optimal parameters - differencing order (d), autoregressive order (p), and moving average order (q) - that need to be used on the given time series data. The differencing parameter plays a crucial role in removing the trend component and converting the non-stationary data into stationary data, which facilitates the process of making predictions for future values. By adjusting these parameters correctly, the ARIMA model can make accurate predictions and become a useful tool in time series analysis. The structure of the ARIMA model is shown in Figure 1.

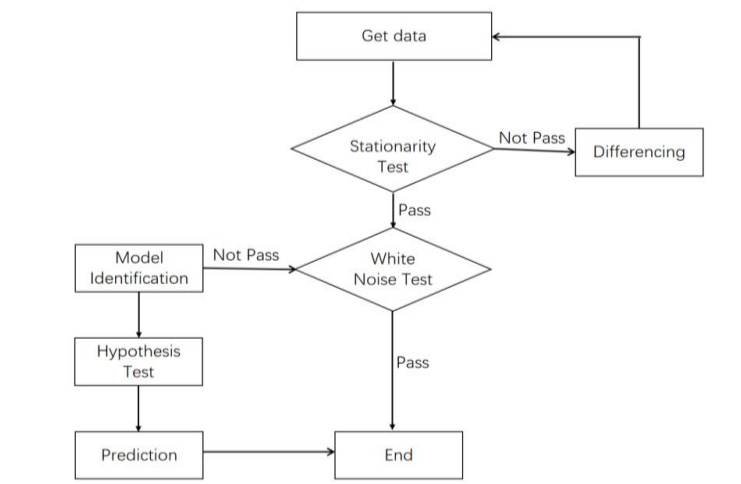


Figure 1 The workflow of the ARIMA Time Series Model

The workflow of the ARIMA time series forecasting model consists of three main steps:

(1) Data preparation: The time series data is first analyzed for the presence of trend, seasonality, and noise. The data is then transformed and pre-processed to make it stationary and ready for modeling.

(2) Model building: Based on the analysis of the data, a suitable ARIMA model is selected, and the parameters are estimated. The model is then fitted to the data to make predictions and check the accuracy of the model.

(3) Model evaluation: The performance of the model is evaluated based on the accuracy of the predictions and the goodness of fit of the model. If the model is found to be accurate and reliable, it can be used to make predictions for future values of the time series data.

2.2. The structure of Particle Swarm Optimization

Particle Swarm Optimization (PSO) is a stochastic optimization algorithm that is inspired by the social behavior of bird flocking or fish schooling. It is a population-based optimization technique that uses a group of particles moving in the search space to find the optimal solution [9].

The basic principle of PSO is to simulate the social behavior of particles in a group by updating their position and velocity [10]. Each particle represents a potential solution to the problem, and its position and velocity are updated based on its own best-known position and the best-known position of the group. By continuously adjusting the position and velocity of the particles, the PSO algorithm searches for the optimal solution [11].

The workflow of PSO consists of the following steps:

- (1) Initialization: Initialize the swarm of particles randomly in the search space.
- (2) Evaluation: Evaluate each particle's fitness by calculating the objective function value.
- (3) Update personal best position: Update each particle's personal best position based on its fitness value.
- (4) Update global best position: Update the best position of the swarm based on the personal best positions of each particle.
- (5) Update velocity and position: Update the velocity and position of each particle based on the personal best and global best positions.
- (6) Repeat steps 2 to 5 until the convergence criteria are met.

The workflow of the Particle Swarm Optimization is shown in Figure 2.

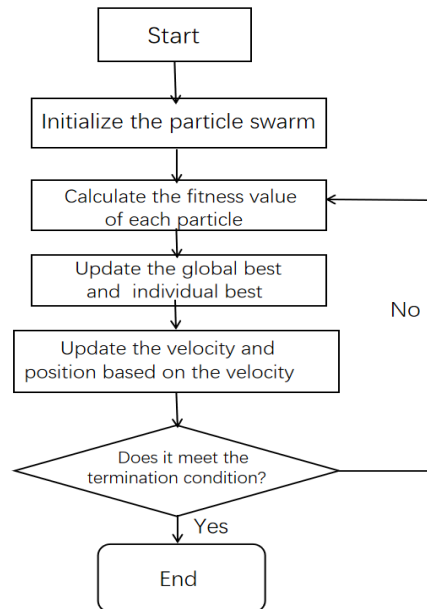


Figure 2 The workflow of the Particle Swarm Optimization

Assuming there are N particles in a D -dimensional search space, each representing a solution.

$$X_{id} = (x_{i1}, x_{i2}, \dots, x_{iD}) \quad (2)$$

where X_{id} represents the position of the i -th particle.

$$V_{id} = (v_{i1}, v_{i2}, \dots, v_{iD}) \quad (3)$$

where V_{id} represents the velocity and direction of movement of the i -th particle.

The two core elements of the algorithm are the particles' velocity and position. Velocity represents the direction and distance the particle will move in the next iteration, while position is a solution to

the problem being solved. Therefore, the first two steps of the solution involve determining the velocity and position of the i -th particle.

$$P_{id,pbest} = (p_{i1}, p_{i2}, \dots, p_{iD}) \quad (4)$$

where $P_{id,pbest}$ represents the best position found by the i -th particle during the search (individual best solution). The historical best fitness value of the individual is f_p .

$$P_{d,gbest} = (p_{1,gbest}, p_{2,gbest}, \dots, p_{i,gbest}) \quad (5)$$

where $P_{d,gbest}$ represents the optimal position found by the entire population during the search (global best solution). The historical best fitness value of the population is f_g .

In the next iteration, each particle continuously iterates its distance and direction of movement based on inertia direction, individual best direction, and swarm best direction, ultimately converging towards the swarm's optimal solution.

3. Results

3.1. The interrelationships among the sales volume of various vegetable categories.

This article explores the relationships among six different categories of vegetables. The Pearson correlation coefficient plays an important role in the study of linear correlation and is widely used in random processes and statistics. By using correlation analysis, the degree of linear closeness between two variables can be measured, and the closer the number is to 1, the higher the degree of linear correlation. In this article, the Pearson correlation coefficient among the six categories of vegetables was calculated using a Python program [12]. The numbers 1 to 6 represent aquatic root vegetables, edible fungi, eggplants, chili peppers, leafy vegetables, and cauliflower vegetables, respectively. The specific coefficient set is shown in Figure 3.

From the Pearson correlation coefficients, it can be seen that aquatic root vegetables are highly correlated with chili peppers, and leafy vegetables are highly correlated with chili peppers. It is speculated that this result may be closely related to the dietary habits of the location where the fresh supermarket is located.

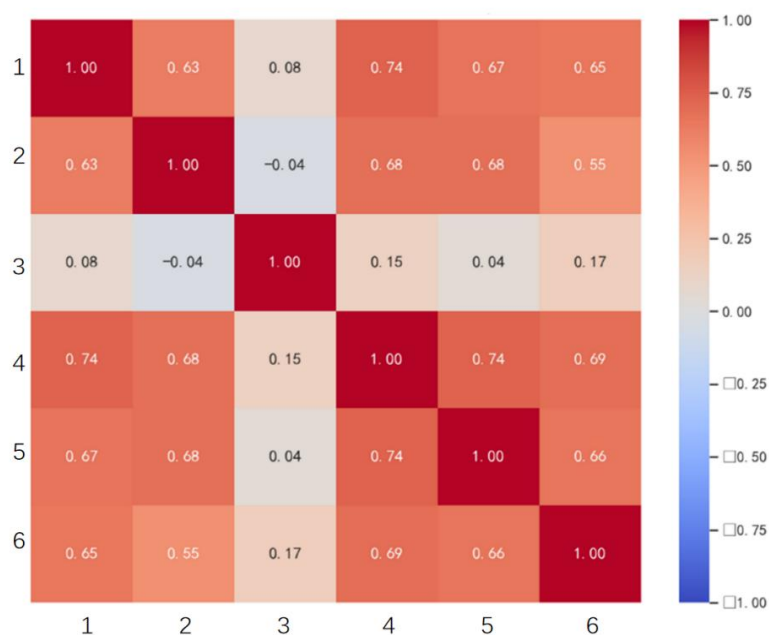


Figure 3 The set of Pearson correlation coefficients among the six categories of vegetables

3.2. Predict the daily replenishment total and pricing strategy of each vegetable category in the next week (July 1-7, 2023)

As the data for each year in the three years have similar changes, the July data from the previous three years are used to estimate the July data for 2023. To further predict the daily replenishment total and pricing strategy of each vegetable category in the next week (July 1-7, 2023), this article first conducts time series smoothness tests on the sales volume data of each vegetable category in July for the three years of 2020, 2021, and 2022, and the results are shown in Table 1.

Table 1 Results of time series smoothness test for sales volume data of vegetable categories

Date	ADF statistic	Critical value	P-value
2020.7 time series	-12.27	-1.9416	0.001
2021.7 time series	-19.3344	-1.9416	0.001
2022.7 time series	-18.2224	-1.9416	0.001

It was found that the ADF statistic of the July data for the three years was significantly lower than the critical value, and the corresponding P-value was much lower than 0.05. This indicates that the time series is smooth and suitable for observing the general pattern or trend of the data. Therefore, this article chooses to use the ARIMA time series forecasting model to predict the daily replenishment total and pricing strategy of each vegetable category in the next week (July 1-7, 2023).

After performing program calculations using Python on the above model and input data, the optimal strategy was obtained, resulting in the optimal daily replenishment total and pricing, as well as the daily maximum profit, shown in Tables 2, 3, and 4, respectively.

Table 2 Table of the optimal daily replenishment total for each vegetable category

(Unit: kilograms)	7.1	7.2	7.3	7.4	7.5	7.6	7.7
Cauliflower	19.593	18.135	16.500	15.965	15.795	16.004	16.130
Leafy vegetables	133.03	141.57	143.01	143.13	143.21	142.01	142.78
Chili peppers	81.663	80.970	79.998	78.005	78.996	79.004	79.226
Eggplants	21.184	20.195	21.699	20.759	19.999	20.696	19.894
Edible fungi	45.636	47.786	48.654	49.686	48.656	48.657	49.563
Aquatic root vegetables	16.157	16.008	15.982	16.566	17.003	16.736	17.213

Table 3 Table of optimal pricing for each vegetable category

(Unit: yuan)	7.1	7.2	7.3	7.4	7.5	7.6	7.7
Cauliflower	11.660	11.293	10.982	11.679	11.657	11.432	11.334
Leafy vegetables	9.501	9.463	9.125	9.478	9.465	9.392	9.270
Chili peppers	8.564	8.463	8.241	7.999	8.625	8.908	8.657
Eggplants	6.408	6.344	6.027	6.275	6.486	6.160	6.492
Edible fungi	10.675	10.996	10.473	10.984	10.038	10.200	10.837
Aquatic root vegetables	13.852	12.675	12.998	13.359	13.206	13.295	13.800

Table 4 Table of daily profits for each vegetable category

(Unit: yuan)	7.1	7.2	7.3	7.4	7.5	7.6	7.7
Cauliflower	66.429	55.282	45.471	54.493	53.574	50.877	49.780
Leafy vegetables	617.88	650.28	612.71	660.05	658.85	643.60	630.89
Chili peppers	164.92	155.65	136.45	114.67	165.40	184.57	165.74
Eggplants	40.424	37.314	33.532	37.139	39.943	34.883	39.998
Edible fungi	200.28	224.28	203.81	232.48	183.15	190.64	224.61
Aquatic root vegetables	85.324	66.610	71.596	80.089	79.829	80.135	90.903

The table results were obtained by establishing a nonlinear programming model with a 50% discount rate for vegetables, which was based on the average vegetable loss rate in previous years. However, in reality, the discount rate and loss rate of vegetables can vary greatly by month and season. For instance, July is a high-loss month for vegetables, and the actual discount rate should be higher than the set value. Additionally, from the table contents, it can be seen that leafy vegetables have high procurement quantities and profits, which may be related to local dietary habits.

3.3. Restocking Plan for Vegetable Products on July 1st

In this article, data preprocessing was carried out on all individual items from June 24-30 in the first week, resulting in 49 items whose inventory quantities changed during that week. As the task requires the selection of 27-33 vegetables and the development of a strategy, the Analytic Hierarchy Process (AHP) was first used to comprehensively evaluate two indicators: daily average sales volume and average profit margin for each item, by combining common knowledge and statistical rules. Additionally, based on practical experience, a matrix of relative importance between the two indicators was set manually:

$$criteriaMatrix = \begin{bmatrix} 3 & 1 \\ 3 & 1 \end{bmatrix} \quad (5)$$

The 49 individual items were numbered 1, 2, 3, and so on, and their two indicators were calculated respectively. The data was then normalized and used to calculate the total sales volume and daily average profit as indicators. After further processing the data for 33 vegetable items, the mean, standard deviation, minimum value, and maximum value were obtained for each item's total sales volume, selling price, average purchase price, and daily average sales volume, as shown in the following Table 5:

Table 5 Table of Mean, Standard Deviation, and Extreme Values

Variable	Obs	Mean	Std. dev.	Min	Max
Sales volume	33	60.5859	46.9309	11.683	226
Selling price	33	7.8530	5.3966	1.8796	24
Average purchase price	33	5.3153	4.1926	1.4529	18
Daily average sales volume	33	18.3445	15.3811	2.1837	77.8633

Through the above data, it can be observed that the daily profit range for the 33 individual items is relatively large, with a wide distribution range. However, the maximum purchase price for vegetable items is relatively high, while the rest of the values are relatively stable. This shows that there are significant differences in profit rates between different individual items.

Further research on the data revealed that there were five sets of data that had problems with one of the indicators being too small and difficult to process, which did not meet the expected requirements. Therefore, this article used the AHP indicators on the remaining 16 sets of data that were selected after screening the initial 49 sets of data and found the five sets of data with higher quality to replace the previous five sets of data that did not meet the conditions for the next step of Particle Swarm Optimization algorithm modeling.

To solve the problem of function optimization, the Particle Swarm Optimization (PSO) algorithm was introduced as a potential means of achieving the objective of finding the optimal solution. The PSO algorithm holds significant advantages over other methods of optimization, such as being relatively simple to implement. Furthermore, PSO can solve optimization problems with multiple constraints, thereby increasing the algorithm's utility as a search method [12]. Two constraints were introduced in this context to limit the upper and lower limits of price and sales volume, respectively, thus facilitating the process of refining the optimization search space. Through the utilization of these constraints, the optimal solution could be effectively and efficiently identified while satisfying the specified constraints.

Firstly, simple processing was carried out on the predicted data, and the grey prediction model was used. However, due to the relatively stable purchase price for each vegetable item, the model's performance was not very good. Furthermore, some individual item data had a large variance. Therefore, for the daily cost data, data with a variance smaller than 0.1 were replaced with the average value, and data with a variance greater than 0.1 were replaced with the median value. The results of the calculation are shown in the following Table 6:

Table 6 Replenishment quantity and pricing strategy for individual items on July 1st

Name	Pricing (yuan)	Replenishment quantity (kg)	Revenue (yuan)	Name	Pricing (yuan)	Replenishment quantity(kg)	Revenue (yuan)
Xixia Mushroom	26.40	4.62	43.309	Long Eggplant	13.20	4.21	24.256
Amaranth	4.19	9.82	14.511	Baby Choy Sum	5.72	4.90	12.713
Yunnan lettuce	10.12	1.84	6.643	Guizhou Lettuce	4.91	20.93	22.483
Choy sum	4.15	19.85	30.743	Yunnan Brassica	4.57	22.20	33.083
Shanghai green	8.80	3.86	15.608	Spinach	6.06	7.00	11.733
Wood ear lettuce	5.87	5.93	14.440	Loofah	14.51	3.00	2.543
Lotus Rhizome Strips	22.10	4.00	5.769	Bird's Eye Chili	6.35	23.84	93.330
Purple eggplant	6.60	16.35	43.369	Luffa	2.84	15.59	18.134
Green eggplant	6.60	2.68	6.421	Facing Heaven Chili	5.42	16.93	31.952
Broccoli	13.65	15.98	82.995	Ginger	5.19	7.00	17.557
Lotus root	15.84	6.03	28.072	Cordyceps Flower	3.97	2.29	2.577
Baby bok choy	7.20	15.64	37.212	King Oyster Mushroom	5.53	15.00	31.824
Sweet Potato	5.87	4.96	11.996	Hangzhou Chili	6.04	2.14	5.400
Serrano Peppe	12.42	7.53	32.168	Red Pepper	20.78	2.00	13.806
Broccoli	14.08	5.36	21.294	Enoki Mushroom	2.07	16.57	10.148
Napa Cabbage	5.27	9.63	22.423	Seafood Mushroom	3.02	13.29	14.163
Wuhu Green Pepper	5.72	17.71	38.478				

Upon observation of the table above, it is evident that the data prediction results are good. From the table, it can be observed that there are significant differences in vegetable pricing between different categories, with cauliflower and eggplant pricing being relatively high. Additionally, there are significant price differences between different SKUs within the same category. For instance, there is a price difference between Xixia flower mushrooms and seafood mushrooms. Fresh food supermarkets should use the model results to develop a reasonable daily procurement plan to minimize vegetable loss and waste and maximize profits.

4. Conclusions and outlooks

This article utilizes the ARIMA time series model and conducts nonlinear programming. Due to the strong stationarity property of the given time series, the accuracy of time series prediction is ensured, which leads to more reliable optimal decision-making. Meanwhile, the particle swarm algorithm is used, taking into account the discount rate and transportation loss rate, to balance the profit and improve the accuracy. However, the purchasing pricing strategy obtained from each run of the particle swarm algorithm may differ, and it may only converge to local optimal solutions. After multiple attempts, the best result was selected, but there may still exist even better purchasing pricing strategies. It can be considered to combine particle swarm optimization with genetic algorithm, leveraging the strengths of both algorithms to enhance global search capability and identify the global optimal solution.

This article starts with the research on automatic pricing and replenishment decision-making strategies for vegetable products. By utilizing ARIMA time series forecasting models, particle swarm

optimization algorithms, and other models, and combining them with known historical data from the past three years, it predicts and plans future procurement and pricing schemes for supermarkets. It has strong practical applications. With appropriate improvements and the inclusion of more quantified variable factors, more complex mathematical models can be obtained. This has broad application prospects, not only for predicting fresh vegetable procurement in supermarkets but also for extending to applications such as predicting flower sales in flower shops, fruit sales in fruit stores, breakfast sales in breakfast shops, etc. The use of mathematical models aims to maximize profits for decision-makers.

References

- [1] Chen Jie, Lou Qunfeng, et al. Analysis of vegetable production and marketing in and characteristics of the Shanghai urban agglomeration [J]. *Chinese Vegetables*, 2016, 8: 4-9.
- [2] Liu Huanhuan, Zhao Huifeng, et al. Selection of vegetable logistics operation mode based on AHP[J]. *Journal of Jiangxi Agricultural University: Social Sciences Edition*, 2012, 11(1): 40-45.
- [3] Zhang Ai, Zhou Jianqin. Research on fruit inventory control in fresh supermarkets based on genetic algorithm [J]. *Management Science and Engineering*, 2019, 8: 113.
- [4] Zhao Xiaoli, Ma Chenqi, et al. Research on automatic pricing and replenishment decision-making for vegetable products [J]. *Advances in Applied Mathematics*, 2023, 12.
- [5] Huang He. Research on an Automatic Pricing Model for Products Based on Deep Learning[J]. *Modern Commerce and Industry*, 2019, 40(09): 188-190.
- [6] Xu X, Du Y, Lie J. An improved ARIMA model based on regularized Gaussian basis function and its application to stock price forecasting [C]. Wuhan Zhicheng Times Cultural Development Co., Ltd.. *Proceedings of 2022 International Conference on Financial Market and Enterprises Management Engineering (FMEME 2022)*. School of Mathematics, Sun Yat-sen University; School of Management, Shanghai University of International Business and Economics; School of Mathematics, Beijing Normal University, 2022:10.
- [7] Gao L D, Liang Y J, Xu H B. Predict Gold Price Trend Based on ARIMA Model [C]. Eliwise Academy. *Proceedings of the 2022 International Conference on Computing Innovation and Applied Physics (CONF-CIAP 2022) (part 2)*, 2022:9.
- [8] Zhang Y. Analysis and Prediction of Influencing Factors of Total Retail Sales of Consumer Goods in China——Based on Multiple Linear Regression Models and ARIMA Models [C]. Wuhan Zhicheng Times Cultural Development Co., Ltd.. *Proceedings of 4th International Conference on Global Economy and Business Management (GEBM 2022)*. School of MinZu University of China, 2022:13.
- [9] Yin Xiaoye. Research on Function Optimization Solution Method Based on Particle Swarm Algorithm[J]. *Journal of Integrated Circuits*, 2022, 39(09): 210-211.
- [10] Li Nana. Research on Multi-objective Optimization Problem Based on Improved Particle Swarm Algorithm [D]. Guizhou Minzu University, 2023.
- [11] Miao Jianjie, Li Debo, Li Huijun, et al. Research on Solution Method for Nonlinear Equation System Based on Improved Particle Swarm Algorithm [J]. *Environmental Engineering*, 2023, 41(S2): 851-856.
- [12] Zeng Shouzhen, Luo Dandan. Probability Language TOPSIS Multi-attribute Decision-making Method Based on Pearson Comprehensive Correlation Coefficient [J]. *Journal of Systems Science and Mathematical Sciences*, 2021, 41(01): 126-143.
- [13] Research on function optimization using the Particle Swarm Optimization algorithm by Yin Xiaoye [J]. *Journal of Integrated Circuits*, 2022,39(09): 210-211.