Price And Replenishment of Vegetables Based on NARX Neural Network

Zuhong Tang *

Hunan University of Science and Engineering, Yongzhou, China, 425100

* Corresponding Author Email: 2669411780@qq.com

Abstract. The research on the replenishment and pricing of vegetable commodities is helpful for supermarket to make more scientific purchase and sale decisions. Based on this, this paper mainly uses NARX neural network to study the replenishment and pricing of vegetable commodities. Firstly, it preprocesses the missing and abnormal values of the data, constructs a correlation analysis model to study the relationship between the sales of various categories of vegetables, then establishes a NARX neural network model to solve the daily replenishment total, and finally constructs a single objective programming model to get the pricing strategy, which provides a theoretical basis for supermarket to make scientific pricing and replenishment decisions.

Keywords: Correlation Analysis, NARX Neural Network, Single Objective Programming Model, Pricing Strategy.

1. Introduction

In recent years, with the improvement of people's material living standards, more and more people began to pay attention to balanced diet nutrition, which promoted the development and growth of vegetable market. For the choice of vegetable varieties, people tend to be fresh, nutritious and tasty, but the shelf life of vegetables is generally short, and the appearance will deteriorate with the increase of storage time, so how to price and replenish vegetables has become a big problem that puzzles merchants. Merchants should make replenishment decisions for each vegetable category on the same day without knowing the specific items and purchase prices, and generally adopt the "cost plus pricing" method (unit cost+unit cost×Cost-profit ratio) to price vegetables, and discount the goods that are damaged by transportation and poor in appearance. There is often a certain correlation between the sales volume of vegetable commodities and time; Reasonable sales mix can make full use of the sales space of supermarket.

In this paper, the correlation analysis model [1] is used to analyze the distribution law and relationship of the sales volume of various vegetable categories, then the neural network model [2] is used to analyze the replenishment strategy of various vegetable categories, and then the single objective model programming model [3] and genetic algorithm [4] are used to solve the pricing strategy of commodities,so that supermarket can make better decisions.

The advantages of these models are as follows: the NARX neural network used adds delay and feedback mechanism, which is suitable for time series prediction and can be used to solve nonlinear series prediction problems in many fields. Genetic algorithm is easy to use, has strong algorithm compatibility, and can be easily combined with other heuristic algorithms to improve the performance of the algorithm.

2. Research design

2.1. Model assumptions

(1) The sales volume of each category has a certain distribution law, which can be analyzed.

(2) The historical average can be used to represent the indicators of known accessories.

(3) There is a certain relationship between sales volume and cost-plus pricing, which can be fitted and analyzed.

2.2. Symbol description

The specific definition of variable symbols in this paper is shown in Table 1.

Table.1. Variable symbol definition

Symbol	Meaning
U_j	The average unit price of normal sales of vegetables
V_j	The average unit price of discount sales of vegetables
$A_{t,j}$	Forecast total sales volume
$A^{'}_{t,j}$	Sales volume without discount
$A^{''}_{t,j}$	Sales volume at a discount
$\bar{M}_j$	Average discount rate of vegetable categories
$\bar{W}_j$	Average wholesale price of vegetable categories
K_j	Loss rate of vegetable categories
P_j	Pricing of vegetable categories
N_j	Historical selling price of vegetables categories

Note: t: Day; j: Vegetable categories

2.3. Data introduction

The data source for this article is www.mcm.edu.cn. They are: commodity information of six vegetable categories distributed by supermarket; related data of sales flow details and wholesale prices of commodities of supermarket from July 1, 2020 to June 30, 2023; recent loss rate data of each commodity.

Through the analysis of the above data, in order to prevent the subsequent model establishment from being adversely affected by the missing or abnormal data, this paper firstly uses normal distribution 3σ principle to filter the abnormal values in the attachment, and calculate the average value μ and standard deviation σ, to judge whether each data value is in $(\mu-3\sigma,\mu+3\sigma)$ Within, not as an abnormal value, it will be eliminated. Secondly, the cubic spline difference method is used to fill in the missing values in the attachment, and the specific formula is:

$$S_i(x)=a_i+b_i(x-x_i)+c_i(x-x_i)^2+d_i(x-x_i)^3, i=1,2,...,n-1 \tag{1}$$

and use Matlab software for interpolation calculation, get the completed data.

3. Distribution law and relationship of sales volume

3.1. Statistical analysis of data

From Table 2 and Table 3, it can be seen that although the daily and monthly sales volume of various categories of vegetables are not equal, the average value of each category of vegetables is in the same order, and the order of mosaic > capsicum > edible mushrooms > cauliflower > aquatic rhizomes > eggplant is also relatively consistent. Among them, the coefficient of variation of the monthly sales volume of each vegetable category is relatively small, which shows that the dispersion degree of the monthly sales volume of each vegetable category is small, that is, it is better to analyze the classification law by the monthly sales volume of each vegetable category.

Table.2. Statistics of daily sales of various vegetable categories

category	average value	extreme difference	variance	variable coefficient	Maximum sales range
Cauliflower	38.530	185.523	514.164	58.9%	[177, 187]
Mosaic	182.969	1234.175	7430.307	47.1%	[1250, 1270]
Capsicum	84.413	598.165	2855.409	63.3%	[590, 605]
Eggplant	21.364	118.679	173.155	61.6%	[110, 120]
edible mushrooms	70.126	508.124	2351.269	69.1%	[500, 512]
Aquatic rhizomes	37.402	295.866	983.273	83.8%	[290, 300]

Table.3. Statistics of monthly sales of various vegetable categories

category	average value	extreme difference	variance	variable coefficient	Maximum sales range
Cauliflower	1160.179	2018.145	198339.456	38.4%	[2440,2460]
Mosaic	5514.472	8240.309	2963371.461	31.2%	[10120,10140]
Capsicum	2544.129	4379.664	1117095.036	41.5%	[5170,5190]
Eggplant	623.105	1260.233	80789.332	45.6%	[1360,1380]
edible mushrooms	2113.520	3770.104	940215.4013	45.9%	[4520,4540]
Aquatic rhizomes	1127.259	2366.182	416839.294	57.2%	[2510,2530]

3.2. Distribution law of sales volume

Sum the monthly sales volume of each vegetable category from July 2020 to June 2023, and label the dates in order to get the total monthly sales volume of each vegetable category. The results are shown in Table 4:

Table.4. Monthly sales volume of various vegetable categories

serial number	date	Cauliflower	Mosaic	Capsicum	Eggplant	edible mushrooms	Aquatic rhizomes
1	2020.07	1522.555	6577.078	2198.049	1365.551	1617.836	332.056
2	2020.08	1748.658	7261.517	2822.344	1139.393	1636.524	867.18
3	2020.09	1336.066	5668.902	1992.413	671.692	1642.881	798.931
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
34	2023.04	715.518	5140.509	3240.117	591.339	2481.835	539.653
35	2023.05	825.848	5225.556	2783.96	760.167	2128.108	420.267
36	2023.06	533.189	4660.197	2714.933	769.315	1718.985	510.787

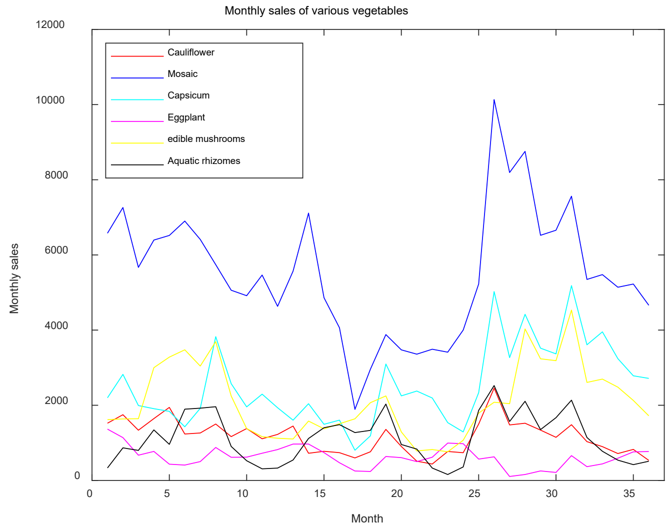

Figure 1. monthly sales distribution of various vegetable categories

As can be seen from Figure 1, from the overall trend, the monthly sales volume of most vegetable categories changes with the change of months. Among them, the monthly sales volume of mosaic is significantly higher than that of other categories, and its sales volume is greatly influenced by the month, with the largest sales volume in July every year and relatively small sales volume in October-November every year. The sales of cauliflower, capsicum, edible mushrooms and aquatic rhizomes are high from November to February of the following year, and the sales in other months fluctuate within a certain range. The monthly sales volume of eggplant is lower than that of other categories, which is less affected by the month and more stable.

3.3. Relationship between sales volume

In this paper, the distribution law of monthly sales volume of various vegetable categories is considered, and the normality tests are carried out on cauliflower, mosaic, capsicum, eggplant, edible mushrooms and aquatic rhizomes respectively. It is found that the P-P diagram is approximately a diagonal straight line, which obeys the normal distribution. Therefore, Pearson correlation coefficient [5] is used to measure the relationship between various vegetable categories.

From Table 5, it can be seen that the absolute value of Pearson correlation coefficient of cauliflower, mosaic, capsicum, edible mushrooms and aquatic rhizomes is between 0.4 and 0.8, and there is a strong correlation; The absolute value of Pearson correlation coefficient between cauliflower and eggplant, mosaic and eggplant, capsicum and eggplant is between 0 and 0.2, and the correlation is very weak.

Table.5. Correlation of various vegetable categories(Pearson coefficient)

category	Cauliflower	Mosaic	Capsicum	Eggplant	edible mushrooms	Aquatic rhizomes
Cauliflower	1	0.747	0.425	0.058	0.429	0.472
Mosaic	0.747	1	0.624	-0.035	0.546	0.484
Capsicum	0.425	0.624	1	-0.191	0.576	0.459
Eggplant	0.058	-0.035	-0.191	1	-0.409	-0.466
edible mushrooms	0.429	0.546	0.576	-0.409	1	0.653
Aquatic rhizomes	0.472	0.484	0.459	-0.466	0.653	1

4. The relationship between the sales volume and the average unit price

The average unit price of normal sales of vegetables is U_j ,and the average unit price of discount sales of vegetables is V_j .Fitting the relationship between the total sales volume and the average unit price of each vegetable category [6] , and the results are shown in the following table:

From Table 6, it can be seen that the decisive coefficient between the total sales volume and the average unit price of mosaic, capsicum, edible mushrooms and aquatic rhizomes is not discounted. R^2 Both are greater than 0.8, that is, regression can reduce the variation of dependent variables by more than 80%, and the regression effect is good. Determinant coefficient of total sales volume and average unit price of eggplant and mosaic R^2 Between 0.75 and 0.80, that is, regression can reduce the variation of dependent variables by more than 75%, and the regression effect is good.

From Table 7, we can see the decisive coefficient of the relationship between the total sales volume of cauliflower and capsicum and the average unit price when selling at a discount. R^2 If it is greater than 0.5, the relationship between the fitted total sales volume and the average unit price can be explained to some extent, which can measure the relationship between sales and pricing.

Table.6. the relationship between the total sales volume and the average unit price of normal sales of vegetables.

category	Regression equation of total sales volume and average unit price	R^2
Cauliflower	$y_1 = 15.193x_1^3 - 1.543x_1^2 0.041x_1$	0.762
Mosaic	$y_2 = 126x_2^3 - 23.448x_2^2 + 1.272x_2$	0.823
Capsicum	$y_3 = 19.597x_3^3 - 1.444x_3^2 + 0.031x_3$	0.841
Eggplant	$y_4 = 12.617x_4^3 - 1.816x_4^2 + 0.272x_4$	0.756
edible mushrooms	$y_5 = 24.557x_5^3 - 2.163x_5^2 + 0.049x_5$	0.812
Aquatic rhizomes	$y_6 = 11.596x_6^3 - 1.002x_6^2 + 0.022x_6$	0.856

Table.7. The relationship between the total sales volume and the average unit price of discount sales of vegetables.

category	Regression equation of total sales volume and average unit price	R^2
Cauliflower	$y_1 = -2.151x_1^3 + 0.857x_1^2 - 0.053x_1$	0.529
Mosaic	$y_2 = 10.747x_2^3 - 1.428x_2^2 + 0.042x_2$	0.486
Capsicum	$y_3 = 5.090x_3^3 - 0.548x_3^2 + 0.014x_3$	0.592
Eggplant	$y_4 = -0.078x_4^3 + 0.687x_4^2 - 0.072x_4$	0.467
edible mushrooms	$y_5 = 2.951x_5^3 - 0.234x_5^2 + 0.005x_5$	0.459
Aquatic rhizomes	$y_6 = 3.258x_6^3 - 0.277x_6^2 + 0.005x_6$	0.366

5. Forecast the total replenishment amount

In order to analyze the total daily replenishment from July 1, 2023 to July 7, 2023, the known data are regarded as a series of numbers arranged in chronological order,and uses the time series to predict the total daily replenishment.

5.1. Stationarity test

In this paper, the data are tested for stationarity, and the output result out1=1 is obtained by using ADF unit root stationarity test. The original hypothesis is not rejected, and the data is stable, so it can be predicted by time series.

5.2. Overview of NARX neural network

NARX neural network [7] is mainly composed of input layer, hidden layer, output layer and input-output delay, in which the number of nodes in input layer is set by the number of input values, and the number of nodes in output layer is set by the number of predicted values. In this paper, the parameters of NARX neural network are set as default values, that is, the input delay is set to 1:2, the output delay is set to 1:2, and the number of hidden neurons is set to 10.

5.3. Accuracy analysis of NARX neural network

Because neural networks are prone to over-fitting, the predicted values of training sets, verification sets and test sets are linearly fitted [8] , and the fitting results are shown in Figure 2:

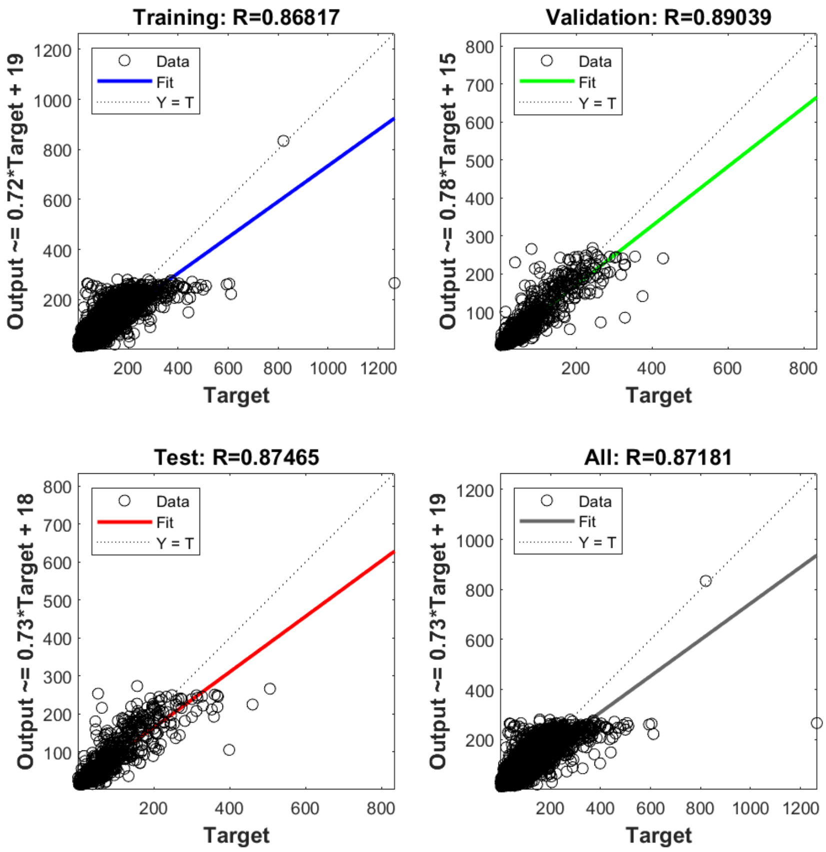

Figure 2. Regression curves of training set, verification set, test set and comprehensive verification

As can be seen from Figure 2, the fitting coefficients of the training set, the test set and the verification set are R=0.86817, 0.87465 and R=0.89039, respectively, and their values are close to 1, which shows that the training effect and test effect of this model are good, and the prediction accuracy is high. The whole model comprehensively verifies that the fitting coefficient of regression parameters is R=0.87181, and the fitting effect is good. To sum up, the time series neural network model has high accuracy and good feasibility.

5.4. NARX neural network solution

In this paper, the daily sales volume of each vegetable category is calculated by using the data, and forecast the expected sales volume of each vegetable category from July 1, 2023 to July 7, 2023 is shown in Table 8. The daily replenishment quantity is calculated by the expected sales volume /(1-loss rate), and the loss rate is the average loss rate of each vegetable category every day. The results are shown in Table 9.

Table.8. Expected daily sales of various vegetable categories.

date	Cauliflower	Mosaic	Capsicum	Eggplant	edible mushrooms	Aquatic rhizomes
2023.07.01	22.422	139.258	69.343	21.813	47.697	13.423
2023.07.02	25.278	141.394	71.422	24.656	49.811	16.633
2023.07.03	28.043	143.518	73.418	27.448	51.988	19.742
2023.07.04	30.712	145.617	75.423	30.136	54.121	22.714
2023.07.05	33.298	147.687	77.426	32.739	56.231	25.570
2023.07.06	35.808	149.725	79.431	35.266	58.318	28.323
2023.07.07	38.252	151.727	81.440	37.723	60.386	30.983

Table.9. Daily replenishment quantity of each vegetable category

date	Cauliflower	Mosaic	Capsicum	Eggplant	edible mushrooms	Aquatic rhizomes
2023.07.01	24.685	27.830	30.873	33.812	36.659	39.423
2023.07.02	156.747	159.151	161.543	163.905	166.235	168.529
2023.07.03	75.376	77.637	79.806	81.985	84.163	86.342
2023.07.04	23.392	26.441	29.436	32.318	35.120	37.819
2023.07.05	52.013	54.318	56.692	59.019	61.319	63.595
2023.07.06	14.821	18.366	21.799	25.081	28.236	31.275
2023.07.07	24.685	27.829	30.873	33.811	36.659	39.423

6. Pricing strategy of each vegetable category

In order to study the pricing strategy of each vegetable category when the profit of the supermarket is the maximum, this paper establishes a single-objective programming model [9] based on the above two results and the known data.

6.1. Single-objective programming model construction

By calculating various data of vegetable categories: $\bar{M}_j, \bar{W}_j, A_{t,j}, A'_{t,j}, A''_{t,j}, N_j$,and using these data to construct the profit relationship from July 1, 2023 to July 7, 2023.The total profit is P, the pricing of each category is P_j, there are:

$$P = \sum_{j=1}^{6} (A'_{t,j} P_j + A''_{t,j} \bar{M}_j P_j - A_{t,j} \bar{W}_j), t = 1, 2, ..., j \tag{2}$$

Let the relationship between the total sales volume and the average unit price of normal sales of vegetables is as follows $f_j (j = 1, 2, ..., 6)$,and the relationship between the total sales volume and the average unit price of discount sales of vegetables is as follows $f'_j (j = 1, 2, ..., 6)$.

Decision variables: pricing of each category $P_j (j = 1, 2, \ldots, 6)$

Objective function: The objective is to maximize the total income, so the objective function is:

$$\max P = \sum_{j=1}^{6} (A'_{t,j} P_j + A''_{t,j} \bar{M}_j P_j - A_{t,j} \bar{W}_j), t = 1, 2, ..., 7 \tag{3}$$

Constraints:

Constraints on the relationship between the total sales volume and the average unit price of normal sales :

$$f_j(P_j) = A'_{t,j}, j = 1,2,...,6 \tag{4}$$

Constraints on the relationship between the total sales volume and the average unit price of discount sales of vegetables:

$$f'_j(P_j\bar{M}_j) = A''_{t,j}, j = 1,2,...,6 \tag{5}$$

Predicted total sales volume constraint:

$$A'_{t,j} + A''_{t,j} = A_{t,j} \tag{6}$$

Average wholesale price constraint:

$$P_j \geq \bar{W}_j \tag{7}$$

Maximum historical selling price constraint:

$$P_j \leq \max N_j \tag{8}$$

According to the above conditions, a single objective programming model is established as follows:

$$\max P = \sum_{i=1}^{6}(A'_{t,j}P_j + A''_{t,j}\bar{M}_{t,j}P_j - A_{t,j}\bar{W}_{t,j})$$
$$s.t.\begin{cases} f_j(A'_{t,j}) = b_jP_j, j = 1,2,...,6 \\ f'_j(A''_{t,j}) = b'_jP_j\bar{M}_{t,j}, j = 1,2,...,6 \\ A'_{t,j} + A''_{t,j} = A_{t,j} \\ P_j \geq \bar{W}_{t,j} \\ P_j \leq \max N_j \end{cases} \tag{9}$$

6.2. Single-objective programming model solution

Genetic algorithm [10] is a heuristic algorithm based on natural principle. Its essence is to evolve the final optimal solution or quasi-optimal solution from generation to generation according to the principle of survival of the fittest through population search technology.According to the principle of survival of the fittest, the optimal solution or the most quasi-optimal solution is finally obtained.

In this paper, genetic algorithm is used to solve the pricing results of each vegetable category when the profit of quotient exceeds the maximum, as shown in tables 10 and 11:

Table.10. Pricing of vegetable categories when the profit reaches the maximum during normal sales.

date	Cauliflower	Mosaic	Capsicum	Eggplant	edible mushrooms	Aquatic rhizomes
2023.07.01	10.35	5.41	6.74	6.08	9.14	9.21
2023.07.02	11.66	5.49	6.95	6.88	9.55	11.41
2023.07.03	12.94	5.57	7.14	7.66	9.97	13.54
2023.07.04	14.17	5.65	7.34	8.40	10.38	15.58
2023.07.05	13.07	5.57	7.15	7.74	9.99	13.84
2023.07.06	12.16	5.50	6.97	7.18	9.63	12.50
2023.07.07	11.38	5.42	6.79	6.71	9.30	11.42

Table.11. Pricing of vegetable categories when the profit reaches the maximum during discount sales.

date	Cauliflower	Mosaic	Capsicum	Eggplant	edible mushrooms	Aquatic rhizomes
2023.07.01	8.65	3.41	3.70	4.59	7.01	8.65
2023.07.02	9.75	3.46	3.81	5.19	7.32	9.75
2023.07.03	10.81	3.52	3.92	5.78	7.64	10.81
2023.07.04	11.84	3.57	4.03	6.34	7.96	11.84
2023.07.05	10.92	3.52	3.92	5.84	7.66	10.92
2023.07.06	10.16	3.47	3.82	5.42	7.38	10.16
2023.07.07	9.51	3.42	3.73	5.07	7.13	9.51

7. Conclusions

In this paper, the correlation model is firstly used to describe the relationship between vegetable categories and the relationship between sales volume and average unit price is fitted by curve. Then the replenishment decision of supermarket in the future is given by using NARX neural network. Finally, the single objective programming model is constructed to describe the pricing strategy of vegetable categories in supermarket, which will enable supermarket to make more scientific replenishment and pricing decisions.

On the basis of the above, in order to make better replenishment and pricing decisions of vegetable commodities, the following three kinds of data are collected to make better commodity pricing.

(1) Climatic condition

The weather conditions will directly affect the price of vegetables Under the climatic conditions of long sunshine time and high temperature and humidity, the growth cycle of vegetables is short and the yield of vegetables is high; When encountering heavy rain and high temperature, vegetables are easy to rot and difficult to transport, and their prices will rise Therefore, the trend of vegetable replenishment and pricing can be judged by weather conditions.

(2) Market demand

Market demand is related to the income level of consumers, the price of goods and other factors. In general, the higher the price of goods, the less the amount of goods purchased will be, and the smaller the market demand will be. Therefore, after knowing the market demand, we can better understand the changes of consumers' demand for vegetables.

(3) Competitor pricing

The sales volume of vegetables has a certain relationship with the pricing of competitors. For the same type of vegetables with the same quality,Consumers tend to choose the ones with lower prices first, so the pricing of vegetables should be adjusted in time. Therefore, after knowing the pricing of competitors, we can better price all kinds of vegetables.

References

[1] Zhang Lukang, Wang Hairong, Gu Keyan, etc. Research on passenger volume forecasting model of transportation hub station based on Pearson correlation [J]. Heilongjiang Transportation Science and Technology, 2023, 46 (03): 137-139.

[2] Liu Da, Luo Weiping. River water level prediction based on time series neural network [J]. Journal of Wuhan Textile University, 2022, 35 (04): 43-47.

[3] Jiao Huirong, Zhu Jinghua and Xi Impressively. Material allocation model based on entropy method and single objective programming [J]. Science and Technology and Innovation, 2021 (06): 41-42.

[4] Chen Xi, Liu Jiao. Research on Immune Genetic Algorithm in Image Segmentation Design [J]. Electronic Fabrication, 2023, 31 (17): 103-105.

[5] Zhao Yuanshang and Lin Weifang. Research on Typical New Energy Output Scenarios Based on Pearson Correlation Coefficient Fusion Density Peak and Entropy Weight Method [J/OL]. China Electric Power: 1-10 [2023-10-14].

[6] Zhu Lijia. Calculation of Shield Center Plane Coordinate and Shield Center Curve Fitting for Transition Curve of Ratio Change of Ultra-high in Rail Transit [J]. Urban Rail Transit Research, 2023, 26 (08): 142-146.

[7] He Chaojie. Research on clustering data fusion based on NARX neural network time series prediction model [D]. Chongqing University of Posts and Telecommunications, 2017.

[8] Yao Lan, Rong Heting, Chu Chao, etc. Sales forecast of industrial supply chain based on artificial neural network [J]. Computer and Digital Engineering, 2021, 49 (10): 2057-2061 +2144.

[9] Hu Ruoyan and Liu Meihan. Research on crowdsourcing task pricing based on single-objective programming [J]. Computer Knowledge and Technology, 2018, 14 (14): 238-239.

[10] Zhu Yu. Research on low-carbon logistics distribution path based on improved genetic algorithm [D]. Shanxi University of Finance and Economics, 2023.