# Automatic Pricing and Replenishment Strategy Model of Vegetable Commodities

Yang Wu * and Tianci Jia

School of Energy and Power Engineering, Jiangsu University, Zhenjiang, China

* Corresponding Author Email: 1454270841@qq.com

**Abstract.** In the fresh supermarket, the general vegetable goods as a short shelf life and rapid deterioration of products, its inventory loss is very serious. In the process of sales, its quality tends to deteriorate with the increase of sales time. In view of the short shelf life of fresh products, this paper establishes an automatic pricing and replenishment decision model of vegetable commodities based on prediction and optimization algorithm. Through the historical sales and demand replenishment of each commodity, the distribution law and mutual relationship of vegetable categories and individual products are analyzed, and the future pricing and replenishment plan of supermarket is formulated. At the same time, it looks forward to the future, and puts forward some opinions on the decision-making process of profit maximization.

**Keywords:** Correlation Model, LSTM Prediction, Bayesian Optimization, 0-1 Integer Nonlinear Programming.

## 1. Introduction

The supermarket conducts vegetable purchase transactions every day from 3:00 to 4:00 a.m., but it is not clear about the specific items and purchase prices. In order to make the same day vegetable replenishment decision, merchants need to consider the relationship between vegetable sales and time and the abundance of varieties available. In terms of pricing, the supermarket adopts the cost-based pricing method, and discounts the goods with transport damage and product phase change difference [1]. At the same time, due to the limited sales space, a reasonable sales mix is also very important. Based on the data of vegetable wholesale sales from July 1, 2020, to June 30, 2023, and reliable market analysis, this paper establishes the corresponding mathematical model to analyze the distribution law and mutual relationship of the sales volume of each category and single item of vegetables. The daily replenishment amount and pricing strategy of each vegetable category in the coming week are given by category. According to the single item replenishment plan, under the condition that the total number of available items is controlled, the single item replenishment quantity and pricing strategy with maximum profit are given in this work [2].

Firstly, the data is preprocessed to obtain the value of the conditional interval. Through Origin visual qualitative analysis, the sales distribution law of each category is obtained, and the relationship between the sales volume of each category is further determined. By establishing Spearman correlation model, the relationship between strength and weakness is obtained. SPSS was used for hierarchical clustering algorithm to classify the characteristics of each single product. 20 samples from vegetable items with annual sales volume of more than 1500 kg were selected as the research objects for correlation analysis, and the correlation between different vegetable items was analyzed.

After that, Matlab was used to clean the data, obtain the relationship between the total sales volume of each vegetable category and the cost-based pricing, make the price-sales scatter plot of each category, and fit the commodity demand curve of each category with curvefitting, that is, the functional relationship between pricing and sales volume. By using the LSTM single time series model of Bayesian optimization, Matlab software is used to solve the forecast of the total daily replenishment, and hyperparameter optimization is carried out to improve the performance of the model. Then, using the method of "cost-based pricing" and the function relationship of pricing and sales volume, the equivalent relationship between markup rate and wholesale price is obtained, and the optimization model is established.

Furthermore, the wholesale price of a single product is predicted, the total sales volume is analyzed and processed, and then the total revenue of the commercial surplus is taken as the objective function, the daily replenishment volume and pricing of a single product are taken as the decision variables, based on the restriction of the number of sold items, the requirements of the minimum display volume setting, the reasonable pricing range of a single product, and the maximum replenishment volume as the constraint conditions. On the premise of satisfying the market demand for various types of vegetable commodities as much as possible, a 0-1 integer nonlinear programming model was constructed to obtain the pricing strategy corresponding to the total amount of daily replenishment [3].

## 2. Model building and analysis

### 2.1. Analysis of distribution and relationship of vegetable categories

It is required to analyze the distribution rules and mutual relations of various vegetable categories. For the convenience of observation, first of all, the data are collected to obtain the total sales volume of different categories of vegetables. Considering the discount sales and return of goods, all the data in the attachment are retained here, and the daily sales volume of each category is visualized to obtain the distribution law of sales volume with the date, SPSS tested the normal distribution of the sales volume of each category, and the result rejected the null hypothesis. Spearman correlation analysis was used to solve the correlation coefficient, and the heat map of the mutual coefficient between vegetable categories was given.

#### 2.1.1 Descriptive statistics of category classification

Summarize the daily sales volume of different categories of vegetables to obtain the relationship between the total sales volume of each part, as shown in Figure 1. As can be seen from Figure 1, Mosaic vegetables account for the largest proportion of supermarket sales, 42.2%, and nightshade vegetables account for the lowest proportion of supermarket sales. Mosaic vegetables are the most popular among consumers. Other categories that accounted for a relatively high proportion were capsicum and edible fungi, accounting for 19.4% and 16.2%, respectively. It is secondary popularity of the flowering vegetables.

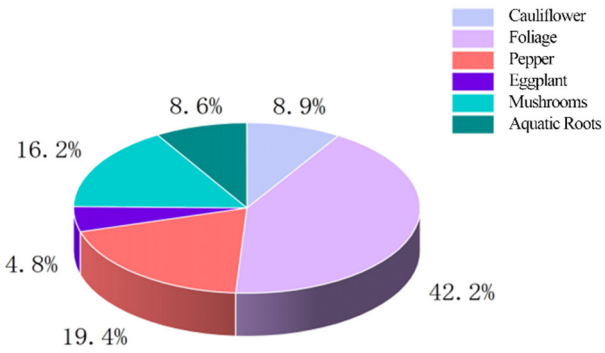



**Fig. 1** The proportion of total sales of different categories

The proportion of total sales of different categories is used to analyze the changes of sales of various categories over time, as shown in Figure 2.

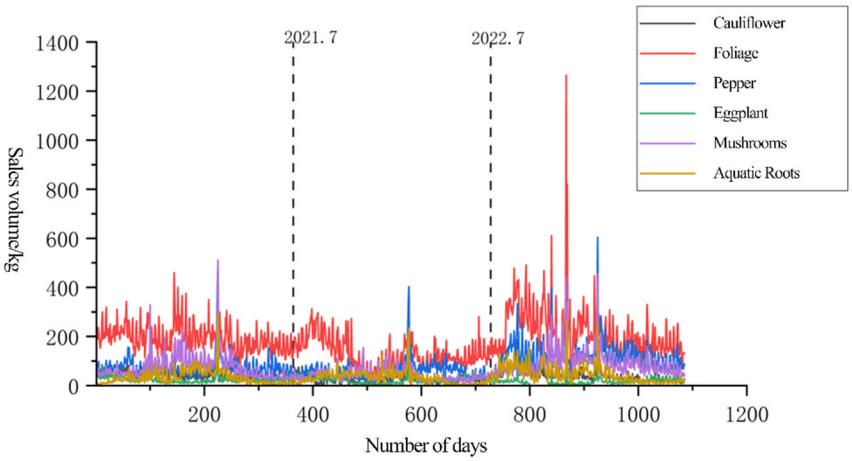



**Fig. 2** Changes of daily sales volume of different categories with the number of days

Note: In order to explore the cyclical change rule of various categories of sales, combined with the actual situation of the given data, the data is taken July every year as the time node, and the chart is divided into three time periods.

Due to the excessive daily sales data of various categories, the relationship between the sales volume of different categories of vegetables over time is preliminarily judged, so as to facilitate the accurate calculation of the correlation degree of different categories. Visualize the data, as shown in Figure 3.

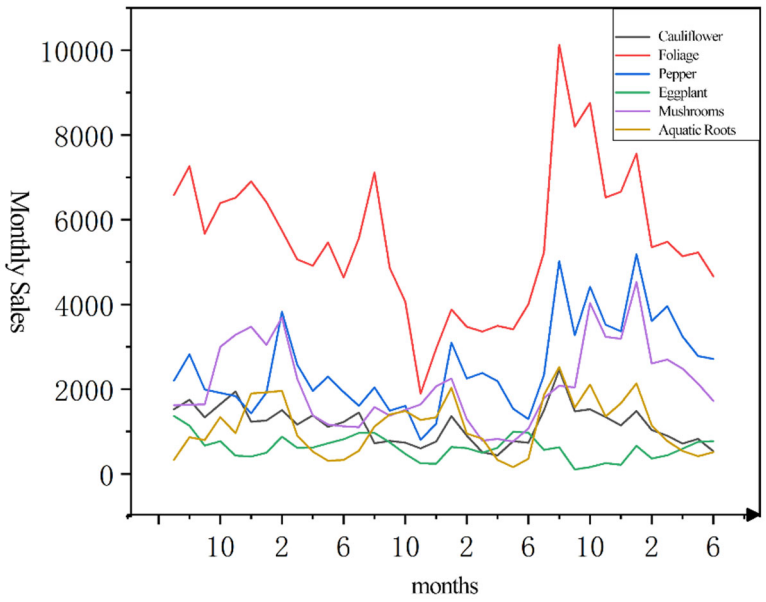



**Fig. 3** Changes of daily sales volume of different categories over months

### 2.1.2 The correlation of sales volume of different categories

It can be found in Figure 3 that the monthly sales volume of all vegetable categories fluctuates with different intensities over time, and the changes are similar. Flowering and leaf vegetables changed the most, and the sales data of chili, edible fungi, cauliflower and aquatic root vegetables were similar, and the changing trend was roughly the same. There is no obvious relationship between the sales volume of nightshade vegetables and other vegetables with the change trend of month, and the correlation may be poor. In order to get closer to the relationship between the sales volume of each category of vegetables, it is necessary to use the correlation coefficient method to judge. Firstly, the daily sales volume of six vegetable categories was calculated, and the Shapiro-Wilk test was used to test the normal distribution of the data. The data results are summarized as shown in the following table.

**Table 1.** Normality test of daily sales volume of different categories of vegetables

| Category | DF | Statistics | Conclusion at (5%) level |
|---|---|---|---|
| Cauliflower | 1084 | 0.90285 | Normality exclusion |
| Flowers and leaves | 1085 | 0.85254 | Normality exclusion |
| Chili | 1085 | 0.76975 | Normality exclusion |
| Capsicum | 1050 | 0.88696 | Normality exclusion |
| Edible fungi | 1085 | 0.78026 | Normality exclusion |
| Aquatic roots | 1085 | 0.8124 | Normality exclusion |

It can be seen that each category of vegetables does not conform to the normal distribution, so we use the Spearman correlation coefficient to give the relationship between the sales volume of each category of vegetables.

Spearman's correlation coefficient is:

$$\rho = 1 - \frac{\left(6 * \sum D^2\right)}{\left(n * (n^2 - 1)\right)} \tag{1}$$

Spearman's correlation coefficient ranges from -1 to +1. When the value is -1, it means that the two categories are completely negatively correlated. And +1 means a perfectly positive correlation. A value of 0 indicates no correlation between the two. Where, n represents the number of observed values, and $D$ is the difference between the ranking values of data of various categories. Bring cauliflower and Mosaic vegetables into the above step to obtain $\rho$=0.695. The results showed that the positive correlation between cauliflower and Mosaic was strong, which was consistent with the conjecture in the previous paper. Similarly, combine other categories in pairs and repeat the above steps to figure out the Spearman correlation coefficient between them, as shown in Figure 4.。

| Spearman coefficient | Cauliflower | Foliage | Pepper | Eggplant | Mushrooms | Aquatic Roots |
|---|---|---|---|---|---|---|
| Cauliflower | 1 | 0.695 | 0.31 | 0.076 | 0.462 | 0.427 |
| Foliage | 0.695 | 1 | 0.488 | -0.043 | 0.578 | 0.448 |
| Pepper | 0.31 | 0.488 | 1 | -0.17 | 0.49 | 0.316 |
| Eggplant | 0.076 | -0.043 | -0.17 | 1 | -0.447 | -0.467 |
| Mushrooms | 0.462 | 0.578 | 0.49 | -0.447 | 1 | 0.669 |
| Aquatic Roots | 0.427 | 0.448 | 0.316 | -0.467 | 0.669 | 1 |

**Fig. 4** Spearman phase relationship among different categories

According to the data in Figure 4, the positive correlation between flowering and cauliflower vegetables was the largest, followed by the positive correlation between aquatic roots and edible fungi. The negative correlation between all categories was not obvious, and the biggest negative correlation was between edible fungi and nightshade vegetables, and nightshade and Mosaic were basically unrelated.

### 2.2. Relationship between Sales volume and Pricing of vegetable categories

The relationship between total sales volume and cost-plus pricing of various vegetable categories is established above. After various regression fitting attempts on the data in the early stage, it is found that the fitting effect is not ideal. In order to give an ideal relationship between total sales and cost-plus pricing, it is necessary to further refine the data, taking the total sales and pricing of each vegetable category in each hour of each day as the smallest unit of the data set, and establish a linear regression model. The detailed steps are as follows.

Step1: Build a linear regression model. According to the requirements of the topic, "total sales" is taken as the independent variable of the model, and "pricing" is taken as the dependent variable of the model.

Step2: The least square estimation, using the least square method to estimate the model, to get the estimated value of the undetermined coefficient.

Step3: Interpret the results and analyze the overall trend of cost-plus pricing changes with the total sales volume. The scatter plot of total sales volume and cost-plus pricing of each vegetable category was drawn with Matlab software, and the unknown parameters of the linear regression model were determined with least square estimation, and the functional relationship between total sales volume and cost-plus pricing was given, as shown in the figure below.

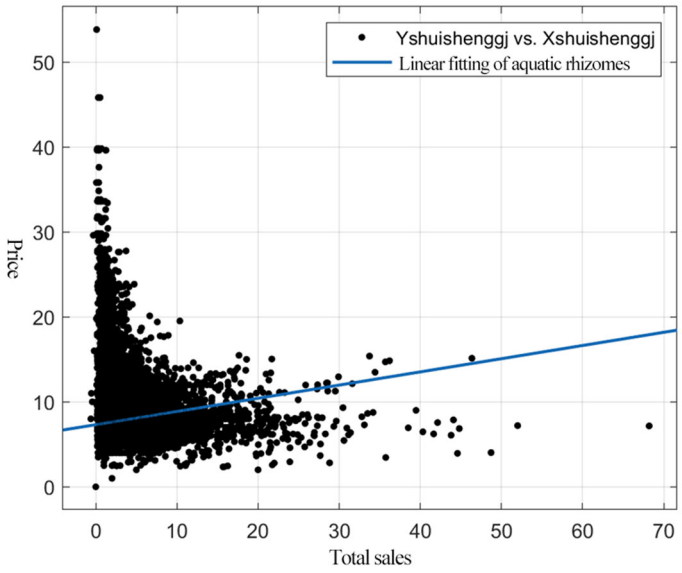



**Fig. 5** The linear fitting diagram of aquatic roots

After analyzing the total sales volume and cost-based pricing of six categories and is summarized in the following table.

**Table 2**. Fitting relationship between total category sales and cost-plus pricing

| Category | Fitting equation |
|---|---|
| Cauliflower | $f(x) = -0.01034x + 5.682$ |
| Flowers and leaves | $f(x) = 1.534x + 8.143$ |
| Chili | $f(x) = 0.1552x + 7.314$ |
| Capsicum | $f(x) = 0.04035x^3 - 0.8215x^2 + 4.008x + 4.176$ |
| Edible fungi | $f(x) = 0.002771x^2 - 0.1809x + 9.659$ |
| Aquatic roots | $f(x) = -0.0001126x^3 + 0.015x^2 - 0.4953x + 11.08$ |

Considering that supermarkets make replenishment plans by category, in order to give the total daily replenishment in the coming week, it is necessary to conduct statistics on the previous replenishment volume. On this basis, a single time series model of LSTM based on Bayesian optimization is constructed to determine the final daily replenishment volume [4]. In this paper, RMSE (root mean square error), MSE (mean square error), determination coefficient R2, MAE (mean absolute error) and MAPE (mean relative error) were selected as objective functions to evaluate parameter performance. MSE (mean square error), which measures the average square value of the difference between the predicted and true value of vegetable replenishment. For a dataset with n samples, the calculation method is:

$$\mathrm{MSE} = \frac{\sum (yi - \hat{y}i)^2}{\mathrm{n}} \tag{2}$$

Where *yi* is the actual value and *ýi* is the predicted RMSE (root mean square error), representing the mean difference between the predicted and the true value:

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} \tag{3}$$

R2(coefficient of certainty), which evaluates how well the model fits the observed data on a scale of 0 to 1. The closer the value is to 1, the better the model fits the data. The closer it is to 0, the worse the fitting effect. Its expression is:

$$\mathrm{R2} = 1 - \frac{\mathrm{SSR}}{\mathrm{SST}} \tag{4}$$

SSR is the sum of squares of residuals, which represents the difference between the predicted value and the real value of replenishment. SST represents the total sum of squares and represents the difference between the true value and its mean. MAE (Mean Absolute Error), which measures the mean absolute difference between the predicted value and the true value, is calculated for a dataset with n samples:

$$\mathrm{MAE} = (1/\mathrm{n}) * \Sigma \,|\, \mathrm{yi} - \hat{\mathrm{y}}\mathrm{i} \,|\, \tag{5}$$

MAPE (Mean Relative Error), which measures the mean relative difference between the predicted and true values, is calculated for a dataset with n samples:

$$\mathrm{MAPE} = (1/\mathrm{n}) * \ \Sigma(|\,(\mathrm{yi} - \hat{\mathrm{y}}\mathrm{i})/\mathrm{yi}\,|) * 100 \tag{6}$$

Using Matlab software to write a program, using the trained model to predict the training set, verification set and test set, and get the daily replenishment results of each category in the next 7 days. Take Mosaic vegetables as an example and the results are shown in the following figure.

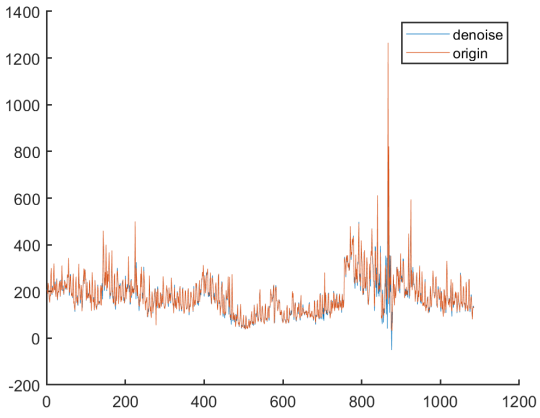



**Fig. 6** Comparison of noise reduction of flower and leaf category data

Aiming at the determination of the total replenishment strategy of each vegetable category in the next week, this paper adopted the LSTM single time series model based on Bayesian optimization to solve the problem. In the training process, the model showed strong fitting ability, and the calculation ability of the model was verified and analyzed, as shown in the following figure.

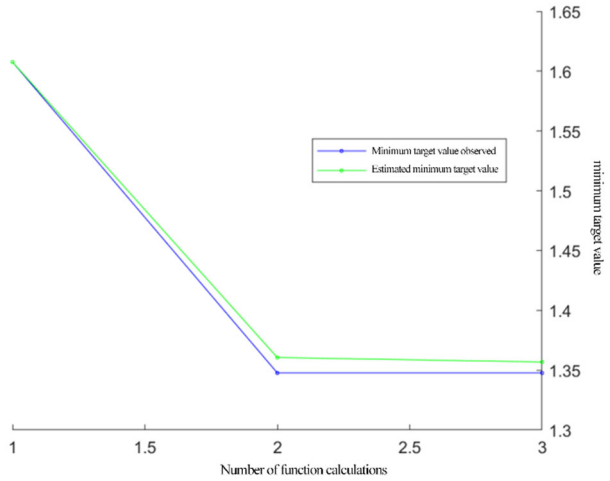



**Fig. 7** Minimum target value and calculation times of function

As can be seen from the figure above, the function used in this model can obtain the final estimated minimum target value in the case of two calculations, and the running time of the model is short, the efficiency is high, and the calculation accuracy is high. For Mosaic vegetables, the relative error between the estimated value and the observed value is 0.676%, indicating a high precision of the model.

### 2.3. Establishment of daily replenishment and pricing strategy

Under the premise of meeting the market demand for various types of vegetable commodities as far as possible, the daily replenishable quantity of vegetables and the pricing strategy are constructed. In order to maximize the profit of supermarket, 49 vegetable items are obtained after extracting simplified data, and they are taken as the main research object [5]. The objective function constructed is as follows:

$$\text{Max} = \sum_j x_j \left[ \left( p_j - k_j \right) n_j - s_j n_j k_j \right] \tag{7}$$

Where $x_j$ indicates whether the JTH item is a saleable item among 49 items. If it is a saleable item, $x_j$ is 1; otherwise, it is 0. $k_j$ represents the wholesale price of item j, $p_j$ represents the selling price of item j, $s_j$ represents the depletion rate of item j, and $n_j$ represents the sales volume of item j. Due to the lack of sales space of vegetable commodities, if the supermarket wants to further develop the replenishment plan of a single product, the total number of available items needs to be controlled within a certain range, such as 27-33. Then there are the following inequality constraints:

$$27 \le x_j \le 33 \tag{8}$$

Due to the limited sales space of vegetable products, it is necessary to meet the constraint condition that the minimum display quantity of each product can meet the minimum display quantity of 2.5kg.

$$n_j \ge 2.5 \tag{9}$$

The total supply range of all types of market demand is set to $X$, and the maximum supply $n_{max}$ must meet:

$$X \le \sum n_j \le n_{max} \tag{10}$$

The constraint of pricing is:

$$p_{\min} \le p_j \le p_{\max} \tag{11}$$

Where $p_{max}$ and pmin are respectively the maximum and minimum price of each item during the sales period. The maximum excess return model based on 0-1 integer nonlinear programming is constructed as follows:

$$\begin{aligned} &\text{Max} = \sum_j x_j \left[ \left( p_j - k_j \right) n_j - s_j n_j k_j \right] \\ &\text{s.}t. \begin{cases} 27 \le \text{x}_\text{j} \le 33 \\ \text{n}_j \ge 2.5 \\ X \le \sum n_j \le n_{\max} \\ p_{\min} \le p_j \le p_{\max} \end{cases} \end{aligned} \tag{12}$$

The results of the maximum quotient excess return model based on 0-1 integer nonlinear programming are as follows:

**Table 3**. Replenishment and pricing based on the model.

| Category | Total replenishment | Price |
|---|---|---|
| Eggplant | 2.50 | 25.76 |
| Houttuydata | 2.50 | 28.26 |
| Amaranth | 2.50 | 27.58 |
| Suizhou Bubble green | 2.50 | 16.74 |
| Dragon Eggplant | 2.50 | 10.93 |
| Sweet cabbage | 2.50 | 6.43 |
| Wuhu Green pepper | 2.50 | 22.84 |
| Steak mushroom | 2.50 | 23.09 |
| Xixia Mushroom | 2.50 | 17.50 |
| Pleurotus eryngii | 2.50 | 12.98 |
| Fungus | 2.50 | 4.45 |
| Chinese mustard | 2.50 | 38.93 |
| 7 color pepper | 2.50 | 13.01 |
| Lotus seed | 2.50 | 16.32 |
| Yunnan lettuce | 2.50 | 16.63 |

## 3. Summary

A correlation model was used in this paper, which can find out the correlation between variables by analyzing historical data, so as to predict the changes of factors such as demand and sales volume. Using the single time series model to solve the problem can predict the future sales volume according to the historical sales data, and it has a certain accuracy for the prediction of vegetable demand. Such models can take into account factors such as seasonality and trend. Demand-oriented vegetable pricing replenishment strategy is a dynamic process. It can quickly and accurately identify market changes, so as to give the best current pricing replenishment strategy. In the actual sales process, merchants can determine and adjust the parameter values of the model according to the changes of vegetable market conditions and consumer demand, and constantly obtain the latest pricing and replenishment plans.

# References

[1] Qiao Xue. Joint replenishment pricing strategy of fresh products considering loss of sales [D]. The Southeast University, 2021. DOI: 10.27014/d.cnki.gdnau.2021.003764.

[2] Mao L. Research on pricing strategy and production and marketing model of vegetable wholesale market from the perspective of supply chain [D]. Zhongnan Forestry University of Science and Technology, 2022. DOI: 10.27662/d.cnki.gznlc.2022.000680.

[3] Yang Shuai, Huang Xiangmeng, Wang Junbin. Research on joint optimization strategy of fresh food shelf distribution and pricing [J]. Supply Chain Management, 2022, 3 (08): 49-59. DOI: 10.19868/j.cnki.gylgl.2022.08.005.

[4] Liang Yuanshun. Research on pork price prediction based on feature selection and LSTM hybrid model[D]. Fuyang Normal University, 2023.DOI: 10.27846/d.cnki.gfysf.2023.000238.

[5] Pan Xiaofei, Xie Zhiheng, Wang Shuyun. Optimal decision of fresh-keeping effort and pricing of fresh products considering loss avoidance [J]. Journal of Highway Traffic Science and Technology, 2022, 39(06): 177-185+190.