Optimization Of Vegetable Sales Strategies and Daily Replenishment in Supermarkets

Yikai Wang *, Ziang Xu, Haoxiang Ye

School of Information Engineering, Guangdong University of Technology, GuangZhou, China, 510006

* Corresponding Author Email: hearthstoneggb@163.com

Abstract. Vegetable products in the superstore have a shorter freshness period, the quality will deteriorate with the increase of sales time, and most varieties can only be sold on the same day. In order to develop a reasonable pricing and replenishment program, this paper first establishes a polynomial regression model to determine the relationship between the total sales volume of each vegetable category and the cost-plus pricing, taking the daily pricing of each category as the independent variable, the daily sales volume of each category as the dependent variable, and taking the polynomial order as 3 to construct the regression function. Then the LSTM model is constructed to predict the confidence interval of the total sales volume of each category in the next seven days, and finally this paper constructs a nonlinear programming model to maximize the revenue of the day as the objective to construct the revenue function, with the constraint that the total daily replenishment volume of each category needs to satisfy the confidence interval of the total sales volume, and then the simulated annealing algorithm is used to find the optimal total daily replenishment volume and pricing strategy in the next seven days.

Keywords: Polynomial Regression, LSTM, Nonlinear Programming.

1. Introduction

Vegetable commodities in the supermarkets have a short freshness period, and the quality will deteriorate with the increase in sales time, and most varieties can only be sold on the same day. Merchants need to take into account the many varieties of vegetables, different places of origin, purchase transaction time and other factors, according to the historical sales and demand for each commodity, in the case of not knowing exactly the specific single product and the purchase price, to make the replenishment decision of the day of each vegetable category. In order to study how to help merchants achieve the purpose of maximizing revenue and reducing losses, and to develop a reasonable pricing and replenishment scheme, based on the existing research on automatic pricing and replenishment decision-making problems of vegetable categories by Li Zongyou and others [1-4], this paper, based on the data of sales flow and loss rate of six vegetable categories from July 1, 2020 to June 30, 2023 of a supermarket on the National Data Network. A polynomial regression model is constructed to determine the relationship between the total sales volume of each vegetable category and the cost-plus pricing, an LSTM model is constructed to predict the confidence intervals of the total sales volume of each category in the next seven days, a nonlinear planning model is established, and a simulated annealing algorithm is used to solve the model, so as to formulate the optimal daily replenishment volume and pricing strategy.

2. Polynomial regression modeling and solution

2.1. Principle and establishment of the model

Polynomial regression models for performing regression analysis of nonlinear relationships[5] , a polynomial function is introduced on the basis of linear regression model so that nonlinear data can be better fitted, therefore, in this paper, a polynomial regression model is used to analyze the relationship between total sales and cost-weighted pricing. The general form of the polynomial regression model is as follows:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + ... + \beta_n x^n + \varepsilon \quad (1)$$

where y is the predicted value, x is the input feature, β is the model parameter, n is the polynomial number of times, and ε is the error term.

The key to polynomial regression modeling is choosing the right number of polynomials n. If the number is chosen too low, the model may not capture the nonlinear relationships in the data; if the number is chosen too high, the model may overfit the data, resulting in poor performance on new data. In this paper, we use the mean square deviation and p values to assess the performance and fit of the model, the p values are solved using the SPSSPRO software, and the mean square deviation formula is as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - y_i)^2, i = 1, 2, ..., n \quad (2)$$

where MSE is the mean square error, n is the sample size, Y_i is the actual value of the ith observation, and y_i is the observed value of the ith observation.

2.2. Solving the model.

The mean squared deviation, p-value and regression equation obtained by solving using Python are shown in Table 1.

Table 1. Mean squares, p-values and regression equations by category

kind	mean square error	p-value	regression equation
philodendron	0.03	0.035	$f(x) = 782.81 - 256.58x + 36.05x^2 - 1.69x^3$
cauliflower	0.06	0.042	$f(x) = 13.65 - 2.09x - 0.25x^2 - 0.01x^3$
Aquatic rhizomes	0.06	0.043	$f(x) = 65.21 - 25.58x + 3.05x^2 - 1.9x^3$
eggplant	0.05	0.052	$f(x) = -4.21 + 14.18x - 1.97x^2 + 0.07x^3$
capsicum	0.04	0.035	$f(x) = 195.16 - 19.64x + 0.82x^2 - 0.01x^3$
edible fungi	0.02	0.035	$f(x) = 61.41 - 69.43x - 6.23x^2 - 1.64x^3$

The calculation results show that the mean squared error of the regression equations for each category is close to 0, and the p-value is close to the significance level of 0.05, which indicates that the polynomial regression model has a strong explanatory power and a good performance and fitting effect.

3. LSTM modeling and solving

3.1. Modeling

LSTM is a special variant of recurrent neural network RNN, compared with RNN, LSTM adds three gates of input gate, forgetting gate and output gate and an internal memory unit, and decides to update or discard the data through the logic control of the gate unit, which can overcome the shortcomings of RNN that the influence of the weight is too large, and it is prone to gradient disappearance and explosion, so that the network can be better and faster convergence, and it can improve the prediction accuracy effectively [6-8]. The structure of LSTM is shown in Figure 1, where X_i represents the input of the nerve cell at the moment of t , h_t represents the state value of the nerve cell at the moment of t. δ represents the neural network layer with the activation function of *sigmoid* , and *tanh* represents the feedforward network layer with the activation function of *tanh*.

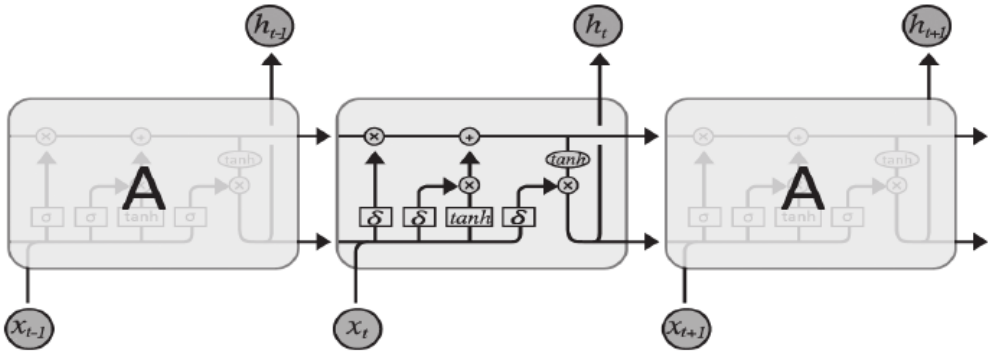


Figure 1. Schematic diagram of LSTM structure

In this paper, Matlab is used to build and solve the LSTM model, the steps to build and solve the LSTM model using Matlab are as follows.

*Step*1: Data preprocessing. Sieve out the outliers with large fluctuations in the amount of data to reduce the impact on the training data; and then normalize the data to limit the value to the [0,1] interval to reduce the impact of the magnitude on the calculation and accelerate the convergence speed of the algorithm. The normalization formula is as follows.

$$X = \frac{\left(X - X_{min}\right)}{\left(X_{max} - X_{min}\right)} \tag{3}$$

Where X is the original data, X_{min} is the minimum value of the data and X_{max} is the maximum value of the data.

*Step*2: Construct the LSTM network architecture and its training options. Determine the number of input sequence features, the number of output sequence responses and the number of implicit units in the LSTM layer, determine the parameters such as the maximum number of training rounds, the gradient queue, the initial learning rate, etc., and train using the Adam optimizer that comes with Matlab .

*Step*3: Predicting future time steps. The test set data is entered into the predictAndUpdateState function for prediction, and then the predicted values are compared with the test data, and a line graph is plotted to analyze the degree of fit.

*Step*4: Use observations to update the network state. First initialize the network state, then make predictions step by step using the training data, and finally compare the predicted data with the test data.

*Step*5: Prediction Data. The future data is predicted based on the trained model, and the confidence intervals for the predicted data are calculated by calculating the mean squared error.

3.2. Analysis of model fit.

Matlab was used to train the model, and the true value was compared with the predicted value, resulting in a line graph of the model fit as shown in Figure 2,Figure 3 and Figure 4, it can be observed that the model fit is good, reflecting the model's generalization ability and robustness.

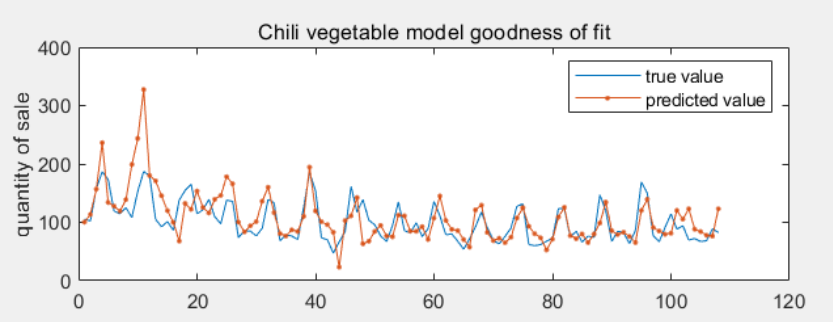


Figure 2. Plot of model fit superiority for chill vegetable categories

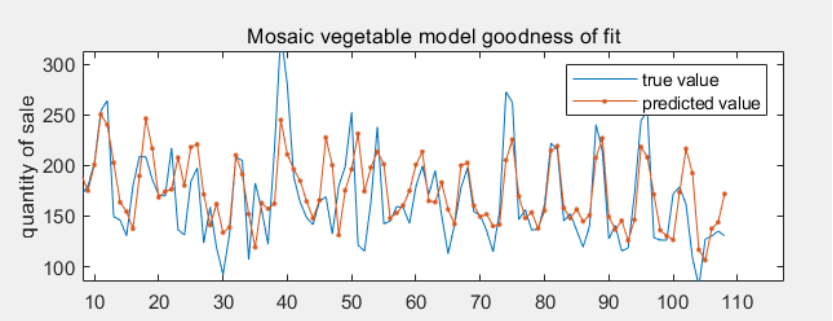


Figure 3. Plot of model fit superiority for mosaic vegetable categories

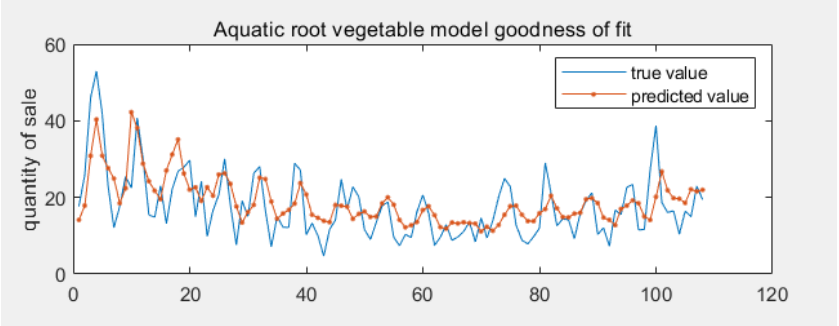


Figure 4. Plot of model fit superiority for aquatic vegetable categories

3.3. Solving the model

The results of the solution using Matlab are shown in Table 2 and Table 3.

Table 2. Confidence intervals for different categories for the next 7 days

	Predicted replenishment minimum, predicted replenishment maximum			
Kind	July 1	July 2	July 3	July 4
philodendron	163, 249	122,194	119,191	74,172
cauliflower	14,18	12,14	12,15	11,33
aquatic rhizomes	25,30	22,25	22,24	11,23
eggplant	13,17	17,21	15,16	14,27
capsicum	116,127	111,120	110,113	114,238
edible fungi	20,88	78,81	76,77	75,77

Table 3. Confidence intervals for different categories for the next 7 days

	Predicted replenishment minimum, predicted replenishment maximum		
Kind	July 5	July 6	July 7
philodendron	19,158	35,162	19,158
cauliflower	37,41	29,32	37,41
aquatic rhizomes	60,103	68,72	60,103
eggplant	31,37	30,31	31,37
capsicum	218,300	213,219	218,300
Edible fungi	159,189	144,145	143,143

4. Nonlinear planning modeling and solution

4.1. Revenue maximization function

Nonlinear programming is a mathematical optimization method for solving optimization problems containing nonlinear constraints and nonlinear objective functions[9] . In nonlinear programming, the objective is to find a set of values of decision variables such that the objective function reaches a minimum or maximum value while satisfying a set of constraints. Based on the purpose of revenue maximization, in this paper, revenue maximization is determined as the objective function, the daily pricing and daily replenishment quantity of each category as the decision variables, and the constraints are that the range of values of the replenishment quantity of each category should satisfy the confidence interval of each category, and the nonlinear programming model is constructed. The revenue maximization function is shown below.

$$\omega_{MAX}=\sum_{i=6}^{6}[\chi_i\rho(\chi_i)(1-\alpha)+\nu\chi_i\rho(\chi_i)\alpha-\gamma(1-\varepsilon)\zeta] \tag{4}$$

Where ω_{MAX} is the total profit, χ_i is the daily pricing of each category, $\rho(\chi_i)$ is the daily sales volume of each category, α is the ratio of the sales volume of the discounted category to the total sales volume of the category, i.e., the discount rate, ε is the loss rate, ν is the ratio of the daily discounted price of the category to the daily pricing of the category, i.e., the discounting strength, and ζ is the replenishment volume.

The results of the solution using Python are shown in Table 4.

Table 4. Correlation coefficients of the revenue maximization function

Item Name	discount rate	Discount	attrition rate	Category purchase price
philodendron	0.0513	0.417	0.11	3.86
cauliflower	0.0536	0.275	0.14	6.78
Aquatic rhizomes	0.0987	0.326	0.11	5.92
eggplant	0.0264	0.054	0.08	8.42
capsicum	0.0355	0.228	0.07	9.45
edible fungi	0.0264	0.397	0.04	7.21

4.2. Simulated annealing algorithm.

Simulated annealing algorithm is a global optimization algorithm for finding optimal solutions in a search space[10] . It is inspired by the process of solid annealing, in which a substance has high entropy at high temperatures and then gradually cools to reach a stable state. In optimization problems, the simulated annealing algorithm searches through stochasticity to avoid local minima and gradually reduces the stochasticity to converge to the global optimal solution. Under the premise of obtaining the confidence interval and the revenue maximization function, this paper uses the simulated

annealing algorithm to find the most appropriate total daily replenishment and pricing strategy for the next seven days. The simulated annealing algorithm modeling steps are as follows:

Step1: Define the objective function: the objective function of the problem needs to be explicitly defined, i.e., the revenue maximization function.

Step2: Select initial solution: randomly generate or select an initial solution.

Step3: Define the energy function: usually, the energy function represents the change in the value of the objective function, i.e. the regression equation.

Step4: Define the method for calculating the probability of accepting a differential solution: use the Metropolis criterion or the Boltzmann distribution to determine the method for calculating the probability of accepting a differential solution. The higher the temperature, the higher the probability that the method accepts the differential solution.

Step5: Setting the initial temperature and stop temperature: Initialize the initial temperature and stop temperature. The initial temperature is set to a higher value and the stop temperature is set to a lower value.

Step6: Define the stop condition:The conditions for determining the stopping of the algorithm are reaching the maximum number of iterations and decreasing the temperature below the stopping temperature.

5. Results

5.1. Total daily replenishment and pricing strategy

The results of solving the model using SPSSPRO are shown in Table 5.

Table 5 Pricing strategy and total daily replenishment for each category for the coming week

kind	Pricing strategy, total daily replenishment						
	July 1	July 2	July 3	July 4	July 5	July 6	July 7
philodendron	9.82, 192	9.98, 187	7.52,131	6.83,167	6.52,60	9.48,159	6.52,143
cauliflower	8.8,14	6.4,12	5.096,12	9.5,21	14,40	15.64,31	11.23,27
Aquatic rhizomes	7.29,29	5.29,23	5.53,22	4.32,29	4.98,80	10.22, 61	9.76,68
eggplant	8.9,12	10.33,17	5.33,16	8.86,14	4.61,30	10.11,29	5.39,28
capsicum	10,122	5.599,111	3.2,110	2.48,230	2.97,237	6.05, 217	6.7, 225
edible fungi	8.54,78	8.87,81	8.77,77	8,77	8.2,77	6.08,185	7.57,154

5.2. Analysis of results

Calculations show that in the coming week, vegetable commodities such as foliage, peppers and mushrooms are highly popular, with relatively high total daily replenishment and relatively low pricing, while cauliflower, aquatic roots and eggplant are less popular, with relatively low total daily replenishment and relatively high pricing.

6. Conclusion

Vegetable products in supermarkets have shorter freshness periods, and their quality will deteriorate with the increase of selling time, and most of them can only be sold on the same day. In order to help merchants achieve the purpose of maximizing the day's revenue and reducing losses, and develop a reasonable pricing and replenishment program, this paper first establishes a polynomial regression model to obtain the relationship between the total sales volume of each vegetable category and the cost plus pricing, and then constructs an LSTM model to predict the confidence interval of the total sales volume of each category for the next seven days, and finally constructs a nonlinear programming model to maximize the day's revenue as the goal of the construction of revenue function. Finally, a nonlinear programming model is constructed, and the revenue function is constructed with

the objective of maximizing the revenue of the day, and the constraint is that the total daily replenishment amount of each category should satisfy the confidence interval of the total sales amount.

References

[1] Li, Zongyou, Zhao Xiaoli, Ma Yinyi, Xu Haoran, Li Binggeng, & Yu Kunping. 2011 A new species of the genus Pterostilbene (Coleoptera, Staphylinidae, Staphylininae) from China. (2023). A study on automatic pricing and replenishment decision problem for vegetable products. Advances in Applied Mathematics, 12, 4539.

[2] Chen Jun, Kang Sha. Joint decision making for pricing and inventory replenishment of agricultural products for dual-channel sales[J]. Industrial Engineering, 2023, 26(3): 39-46.

[3] Jiang, Cherry-Mei, Mou, Jin-Jin. Joint decision making of inventory and pricing of fresh processed products based on perceived freshness. Highway Transportation Science and Technology, 2020, 37(3): 151-158.

[4] Jiang Dan. Research on Consumer Behavior Analysis and Market Dynamic Pricing Model of Agricultural Products [D]. Northeast Agricultural University, 2014.

[5] Niu, X. Y., & Zou, J. F. (2021). (2021). Optimal subsampling algorithm for nonparametric local polynomial regression estimation. Systems Science and Mathematics, 42(1), 72.

[6] GUAN Tao,CHEN Purui,YU Hao. ARIMA-LSTM time series probabilistic prediction method for high arch dam construction simulation parameters[J]. Journal of Hydropower Generation:1-12[2023-09-09]:1-12.

[7] Deng Lijun,Yuan Jinbo,Liu Jian,Shang Wentian. Detection of abnormal wind speed fluctuation based on SSA-LSTM[J]. Coal Science and Technology:1-9[2023-09-09]:1-9.

[8] MAO Baihai,QIN Wu,XIAO Xianbin,ZHENG Zongming. SOH estimation of lithium-ion battery based on LSTM&GRU-Attention multiple joint model[J]. Energy Storage Science and Technology:1-10[2023-09-09]:1-10.

[9] Guan, Chao, Zhang, Zeqiang, Zhu, Lixia, & Mao, Lili. (2019). A mixed-integer nonlinear programming model and a two-stage improved simulated annealing algorithm for a two-story aisle layout problem. China Mechanical Engineering, 30(08), 975.

[10] Fan, Bin, Li, Chidian, & Zhou, Cheng. (2022). Optimization method of nonlinear earthwork deployment scheme based on improved simulated annealing algorithm. Civil Engineering Information Technology, 14(5), 13-20.