Research on Optimal Replenishment and Pricing Decision of Vegetable Commodities

Fangyin Lu[1], Yurui Wu[2], Yining Tian[3], Yaoyao He[4], Wei Huo*[4]

* Corresponding author: 202113675@stu.neu.edu.cn

Fangyin Lu, 18341548901@163.com

Yurui Wu, wyr20041123@163.com

Yining Tian, 529672480@qq.com

Yaoyao He, 2862463871@qq.com

Wei Huo, 202113675@stu.neu.edu.cn

[1]School of Resources and Materials, Northeastern University at Qinhuangdao, 143 Taishan Road, Qinhuangdao, Hebei, China 066004

[2]School of Economics, Northeastern University at Qinhuangdao, 143 Taishan Road, Qinhuangdao, Hebei, China 066004;

[3]School of Foreign Language and Cultures, Northeastern University at Qinhuangdao, 143 Taishan Road, Qinhuangdao, Hebei, China 066004

[4]School of Control Engineering, Northeastern University at Qinhuangdao, 143 Taishan Road, Qinhuangdao, Hebei, China 066004

Abstract: In the fresh food superstore, the vegetable category has a shorter freshness period and needs to make the replenishment and pricing strategy of the day according to the historical sales and demand in order to get the maximum revenue. In this paper, for the decision-making problem of maximizing the revenue of superstores, we process the data, analyze The relationship between total sales volume of vegetable categories and cost-plus pricing, construct a superstore revenue maximization model based on the unit price and sales volume, set the constraints and parameters, and give the optimal replenishment and pricing decisions.

Keywords: Replenishment and Pricing Strategies , XGBoost, Timing Prediction, Genetic Algorithms

1 Introduction

In the context of today's society, fresh produce superstores face a number of challenges as a retail industry providing vegetable ingredients [1-3]. Vegetable items have a short shelf life and most cannot be re-sold the next day, so there is a need to balance replenishment and sales volumes as much as possible to minimize losses [4-8]. Supermarkets are often stocked between

ICBBEM 2024, March 29-31, Wuhan, People's Republic of China

DOI 10.4108/eai.29-3-2024.2347451

3:00 and 4:00 a.m., and accurate information about individual items and prices is not available, so merchants need to make same-day replenishment decisions without knowing the exact price of the item [9-11]. Vegetables are usually priced using the "cost-plus pricing" method, for the quality of vegetables that may deteriorate during transportation, supermarkets sell them at a discount. Therefore, demand analysis is very important in making replenishment and pricing decisions.

2 Model Building

2.1 Description of Notation

The meanings of the notations are shown in the Table 1.

Table 1. Description of Notation.

Notation	Description
Pi	Selling price of a category on day i
ci	Wholesale price of a category on day i
ri	Average profit per kilogram as a percentage of wholesale price for a category on day i
loss_i	Wastage rate of the ith individual item
sell	Sales volume
replenishment	Restocking volume
profit	Supermarket revenue

2.2 Data Collection and Pre-Processing

2.2.1 Data Collection

To facilitate the validation and evaluation of our model, we searched for information on a superstore available on the Internet, where the data was obtained from:https://www.kaggle.com/datasets/roopacalistus/superstore.

2.2.2 Data Pre-Processing

In order to understand and simplify the data more deeply and prepare well for modeling, we used a series of data processing methods while collecting relevant data: data cleaning and data aggregation processing.

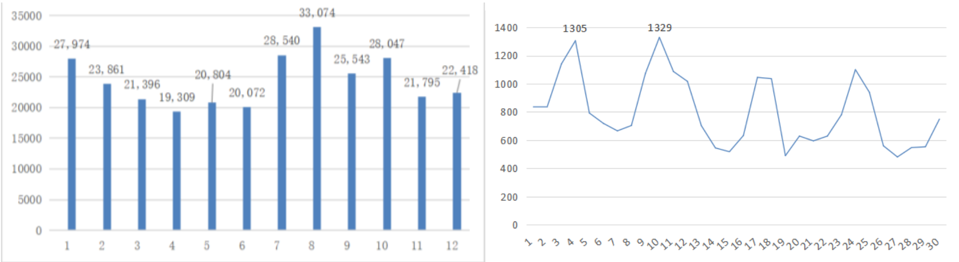

Figure 1. Monthly Statistics of Supermarket Sales Orders and Sales order statistics for September 2022.

3 Wholesale Price and Sales Volume Forecasts

3.1 Wholesale Price Forecasting Model Based on LSTM Time Series

LSTM (Long Short-Term Memory Network) is a method for predicting data with time-series characteristics, which can solve the shortcomings of limited memory capacity in RNN, which can't connect to the early input information, and is suitable for dealing with data with long-term dependence. In this paper, wholesale prices have seasonal changes over time, which can be predicted using LSTM.

The core structure of LSTM consists of three main gate structures: forgetting gates, input gates, output gates, and a cell state. Information can be selectively passed through the gate structure, which is used to remove or add information to the cell state, which contains a sigmoid neural network layer [and point-by-point multiplication. The sigmoid layer outputs a probability value between 0 and 1 that describes how much information each part can transmit. In this question, data features that need to be memorized in the wholesale price (e.g., seasonal changes) can be passed all the way through, while irrelevant information is truncated.

Based on the structure of the LSTM network, a prediction model is built, where the formula for each LSTM unit is as follows:

In the first step, the information to be discarded is decided.f_t is the forgetting gate function for selecting information in the cell state, C is the sigmoid activation function, and x_t is the input at the current moment.

$$f_t = C\left(W_f[h_{t-1}, x_t] + b_f\right) \tag{1}$$

In the second step, decide the values that need to be updated. sigmoid layer decides the values that need to be updated and those that do not need to be updated. tanh layer creates a vector of candidate values ta. which in turn yields:

$$i_t = C\left(W_i[h_{t-1}, x_t] + b_i\right) \tag{2}$$

$$C_t = \tanh\left(W_c[h_{t-1}, x_t] + b_c\right) \tag{3}$$

The two required elements for renewal are shown:

In the third step, the cell state is updated. Multiply the old state with f_t, discard the information identified as needing to be discarded, and add the new information added to get the new candidate value.

$$C_t = f_t C_{t-1} + i_t C_t \tag{4}$$

In the fourth step, the output is obtained based on the cell state. First the output part of the cell state is determined by the sigmoid layer, then the cell state is processed with tanh to get a value between -1 and 1, which is then multiplied with the output of the sigmoid gate to get the output:

$$O_t = C（W_i[h_{t-1}, x_t] + b_{oi}） \tag{5}$$

$$h_t = O_t \tanh(C_t) \tag{6}$$

For the prediction of wholesale prices, it uses the lagged features of the prices as inputs, and create the lagged data of the past day, week or month as new features. In this paper, the step order is set to 2, and the formula for the lagged feature is expressed as follows:

$$X_{t-n} = f(X_n X_{t-1}, \dots, X_{t-n=1}) \tag{7}$$

Where X_t denotes the eigenvalue at time t, X_{t-n} denotes the eigenvalue at time t-n, and f denotes the generating function of lagged features. Through the construction of lagged features, the dynamic change information of the time series can be incorporated into the features to improve the prediction accuracy of the model.Forecast wholesale prices for the next 7 days as shown in the table1.

Table 1. Prediction results.

date	wholesale price (yuan)
July 1, 2023	8.412
July 2, 2023	8.960
July 3, 2023	9.216
July 4, 2023	8.971
July 5, 2023	7.877
July 6, 2023	7.425
July 7, 2023	7.044

According to Figure 2, LSTM has a good fitting effect on the changes in wholesale prices, especially for recent wholesale price changes, which almost overlap with the real prices. It has effectively extracted the characteristics of wholesale price changes and can be used as a prediction model for the next 7 days.

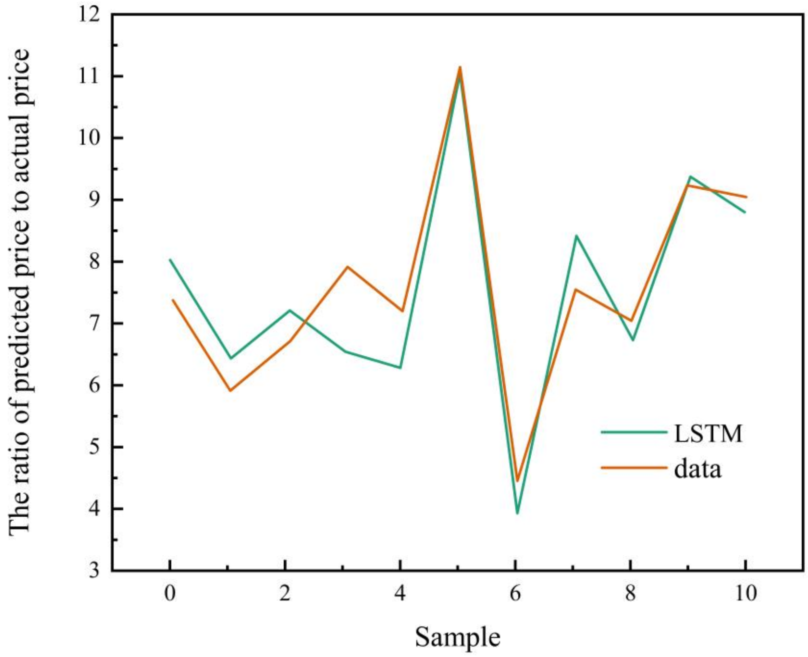

Figure 2. Comparison between Actual and Predicted Values.

3.2 XGBoost-based sales volume forecasting

XGBoost is based on decision trees and minimizes the loss function by iteratively adding new trees. The loss function consists of two parts, a prediction error and a regular term that controls the model complexity to prevent overfitting.

(1) Optimization objective

Assume that the model has k decision trees, i.e:

$$\hat{y}_i = \sum_{i=1}^{k} f_k（x_i）f_k \in F \tag{8}$$

In the above equation, is the feature vector of the ith input sample, is the predicted value of sales calculated after mapping relationships, and is the set of all mapping relationships.

XGBoost updates the model by minimizing the loss function, which is solved using the gradient boosting method. In each iteration, XGBoost computes the first and second order derivatives of the loss function with respect to the current model predictions, and then constructs a new decision tree based on these derivatives. The optimization objective as well as the loss function is:

$$L(t) = \sum_{i=1}^{n} l\,[y_i, \hat{y}(t-1)_i + f_t(x_i)] + \Omega(f_t) \tag{9}$$

In the above equation, L（t）is the objective function at the tth iteration,n is the number of samples, l[.] is the loss function, $\hat{y}(t-1)_i$ is the predicted value of the model at the t-1st iteration, $f_t(x_i)$ is the newly added function, and $\Omega(f_t)$ is the regular term.

(2) Obtained by performing a second-order Taylor expansion of L(t) and the operation of removing the constant term:

$$\tilde{L}(t) \cong \sum_{j=1}^{n}[\,(\sum_{id_j}^{n} d_i)\, w_j + \frac{1}{2}(\sum_{i\in I_j}^{n} g_i + \lambda) w_j^2] + \gamma N \tag{10}$$

Where d_i is the first-order derivative of $\hat{y}(t-1)_i$, g_i is the second-order derivative of $\hat{y}(t-1)_i$, N is the number of leaf nodes, I_j is the set of samples on each leaf node, w_j^2 is the square of the modulus of the fraction of each leaf node, and λ and γ are the weighting coefficients to prevent overfitting.

4 Develop Daily Replenishment Totals and Pricing Strategies for The Coming Week

4.1 Establishment of Objective Function

In a business environment, determining the appropriate selling price is a critical decision that not only affects the profitability level of a firm, but is also closely related to market competition, consumer demand and a variety of other factors. Among them, the average attrition rate.

$$\text{loss} = \frac{1}{N} = \sum_{i=1}^{n} \text{loss}_i \tag{11}$$

n denotes the number of individual items in the category class. The depletion rate for the edible mushrooms category was calculated to be 8.759%.

The replenishment volume, i.e., the amount of goods purchased by the superstore each day, including sales and wastage, is calculated by the following formula:

$$\text{repienishment} = \frac{\text{sell}}{1-\text{loss}} \tag{12}$$

Superstore revenue refers to the total income or profit earned by a commercial supermarket or retail store over a given period of time. It includes all revenues earned from the sale of goods, less all costs and expenses associated with the operation, and is calculated as follows:

$$\text{profit} = \text{sell} \times \text{P-repienishment} \times \text{c} \tag{13}$$

In this paper, a mathematical planning model of a superstore is formulated and solved using a genetic algorithm aimed at determining the optimal selling price of vegetables with a view to maximizing its total revenue.

The objective function of planning:

$$\max \text{profit} = \sum_{i=1}^{7} \text{sell}(P_I)/(1-\text{loss}) \times c_i \tag{14}$$

Decision variable: The main decision variable in the model is the daily sales price, denoted as P_i. The output of the model is the average profit per kilogram as a percentage of the wholesale price. Since the previous section has calculated the average profit per kilogram as a percentage of the wholesale price, this paper selects its maximum value.

Parameters and Data:The input parameters of the model include the daily wholesale price (denoted as aaaa) and the wastage rate (denoted as bbbb) The wholesale price is the price at which the superstore purchases the goods from its suppliers, while the wastage rate reflects the percentage of goods lost. There is also a key input data, which is the daily sales forecast based on the unit price of sales and wholesale price.

Objective Function: At the heart of the model is the objective function, which describes what the superstore wants to achieve - maximizing total revenue. The total revenue is determined by the number of units sold per day, the difference between the unit price and the wholesale price, and the wastage rate. The goal of the model is to find an optimal set of unit prices such that the predicted number of sales maximizes the total revenue for a given wholesale price and wastage rate.

Constraints: In order to make the selling prices are reasonable, the model also includes an important constraint: the selling unit price must be higher than the wholesale price. This ensures that the superstore does not sell items at a loss.

Our objective is to find a unit-of-sale strategy x_1, x_2, ... $x_7$to maximize the cumulative revenue over 7 days. The daily revenue is determined by a combination of the number of sales, the difference between the sales unit price and the wholesale price, and the attrition rate.

4.2 Genetic Algorithm Solving

To solve the above planning model, a genetic algorithm can be used. Genetic algorithm is a heuristic search method for optimization and search problems that mimics the process of natural selection. We can consider each possible price strategy x_1, x_2, ...x_7 as a chromosome, generate new chromosomes by selection, crossover and mutation operations, and evaluate their fitness according to the objective function. The specific steps are as follows:

Encoding and Chromosomes: In GA, solutions are encoded as "chromosomes". For our problem, each chromosome can be a seven-dimensional vector of seven-day sales prices. Thus, a chromosome might look like this: x_1, x_2, ...x_7, where each x represents the sales price on day i.

Generation of initial population:At the beginning of the algorithm ,an initial population is required. This population can be generated randomly where each chromosome is a possible solution.

Evaluation of fitness: The "fitness" of each chromosome is based on its expected objective function value. In this case, the fitness function is equal to our expected gain. The higher the fitness, the better the chromosome (sales strategy).

Selection operation:The selection operation simulates the principle of survivalist superiority. Chromosomes are selected as the basis for the next generation, and the chance of selection is proportional to its fitness.

Crossover (pairing) operation: This simulates the mating process in heredity. Two chromosomes are selected and combined to produce a new offspring. For example, a break point can be selected and parts of the two chromosomes exchanged, resulting in two new chromosomes.

Mutation operation:In order to increase the diversity of a population and avoid premature convergence, chromosomes may undergo mutation. As a result, a part of the chromosome will be randomly changed.

Termination criteria: The genetic algorithm continues to iterate until a specific termination criterion is met, such as reaching a predetermined maximum number of iterations, reaching a certain fitness threshold, or the change in fitness is less than a specific value.

Based on the above method, we calculate the optimal replenishment quantity and pricing strategy as well as superstore revenue for the coming week.

The convergence diagram of the model is shown in Figure 3, after the fifth iteration, the model has converged, indicating a fast convergence speed and good performance. The predicted results are shown in Figure 4.

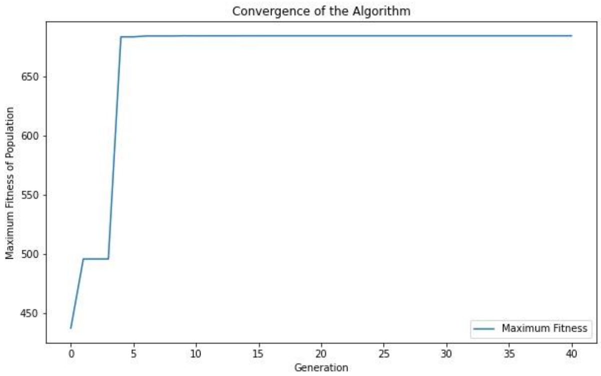

Figure3. Convergence diagram of genetic algorithm model.

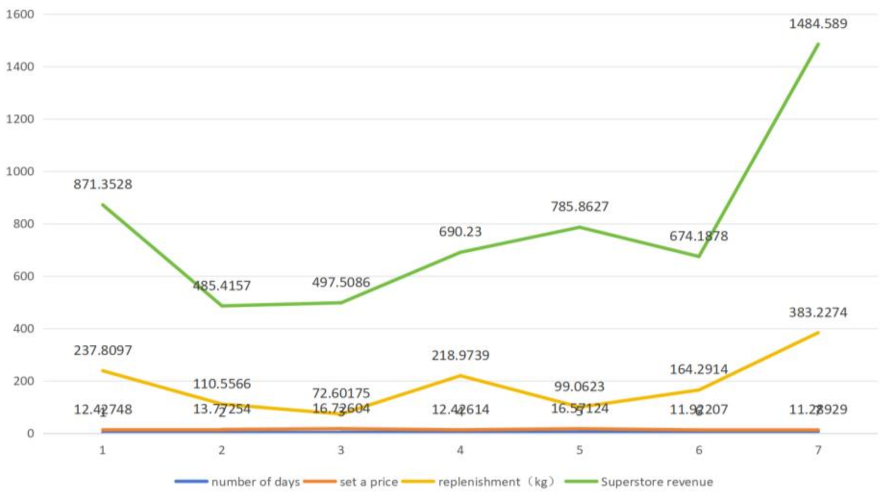

Figure4. Projected results.

5 Conclusion

In this paper, a model is constructed based on historical wholesale prices to predict the wholesale price of vegetable categories in the coming week. Then, the XGBoost sales volume prediction model is constructed based on sales volume, sales unit price, and wholesale price. Since the superstore revenue needs to optimize the sales unit price and sales volume at the same time in order to obtain greater profits, this paper constructs a superstore revenue maximization model based on the sales unit price and sales volume, and by setting constraints and parameters and solving with heuristic algorithms, it obtains the sales volume and sales unit price corresponding to the maximization of the revenue every day, which is very effective.

Acknowledgments. This work was supported by the Research Project on Graduate Education and Teaching Reform of Hebei Province of China (YJG2024133).

References

[1] Morales, Juan Benito. Calanche, et al. "An Approach to the Spanish Consumer's Perception of the Sensory Quality of Environmentally Friendly Seabass." Foods (Basel, Switzerland) 10.11(2021).

[2] Liu, Chang, et al. "An Edge Computing Visual System for Vegetable Categorization." 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA) IEEE, 2019.

[3] Hossain, Mohammad Alamgir, M. Quaddus, and N. Islam. "Developing and validating a model explaining the assimilation process of RFID: An empirical study." Information Systems Frontiers 18.4(2016):645-663.

[4] Sarkar, P.. "Organic coatings as a short term intervention for extending shelf life of fruits/vegetables." (2014).

[5] Daneshi, Mohammad, et al. "Effect of refrigerated storage on the probiotic survival and sensory properties of milk/carrot juice mix drink." Electronic Journal of Biotechnology 16.5(2013):5-5.

[6] Alabi, Da Wen. "Transport phenomena and their effect on microstructure of frozen fruits and vegetables." Trends in Food Science & Technology 101.1(2020).

[7] Mishra, Bhawana, et al. "Edible coatings for postharvest preservation of fresh fruits and vegetables. " Research (2006).

[8] Jafarzadeh, Shima., et al. "Application of bio-nanocomposite films and edible coatings for extending the shelf life of fresh fruits and vegetables." Advances in colloid and interface science 291(2021):102405.

[9] Pramatari, Katerina C.. "Efficient store replenishment through Internet-based information-sharing and collaborative supply-chain practices." (2006).

[10] Dharmawardane, Chethana, V. Sillanp, and J. Holmstrm. "High-frequency forecasting for grocery point-of-sales: intervention in practice and theoretical implications for operational design." Operations Management Research (2021):1-23.

[11] Priyamvada, and A. Kumar. "Modelling retail inventory pricing policies under service level and promotional efforts during COVID-19." Journal of cleaner production 381(2022):134784.