

Automatic Pricing and Replenishment Decision-Making for Vegetable Commodities Based on Bi-directional Long Short-Term Memory Recurrent Neural Networks and Markov Prediction Models

Yirou Jiang^{1, *}, Xi Li¹

¹College of Physics and Electronic Information, Yunnan Normal University, Kunming, China

*Corresponding author: 2054087713@qq.com

Abstract: Reasonable pricing and replenishment strategies are crucial for vegetable superstores to maximise profitability. In this paper, we firstly analyse the distribution pattern and interrelationship of each category of vegetables through Pearson correlation coefficient, and then use MATLAB tools to analyse and predict the data through the time series analysis of Long Short-Term Memory Recurrent Neural Networks (LSTM), and solve the optimal daily replenishment and pricing of each category in the coming week; lastly, we firstly use the small-period sample prediction method, i.e., the Markov prediction model, to analyse the data, and then construct a multi-objective planning model to further determine the optimal pricing strategy. Finally, we use the small-period sample prediction method, i.e., Markov prediction model, to analyse the data and then construct a multi-objective planning model to further determine the optimal pricing strategy.

Keywords: Person Correlation Coefficient, BiLSTM Model, Markov Prediction Model, Particle Swarm Optimisation Algorithm.

1. Introduction

In the supply-side superstores of the vegetable industry, there are many factors that affect the ultimate profitability of the superstores: on the one hand, there is a certain relationship between the sales volume of vegetable goods and the time of day; on the other hand, the final sales pricing of vegetables will be different from month to month due to the influence of the quality of the product and so on. Therefore, the development of a relatively reasonable pricing and replenishment strategy is conducive to maximise the profitability of supermarkets [1]. In this paper, we study the sales data of a superstore over the past three years, establish a mathematical model, and solve the following problems: firstly, we analyse the distribution pattern of each category of vegetables and its relationship with each other through the Pearson correlation coefficient method; then, we use MATLAB tools to analyse and predict the data through the time series analysis of the long and short-term memory recurrent neural network (LSTM), so as to solve the optimal daily replenishment total and pricing of each category in the coming week. We use the MATLAB tool to predict the total daily replenishment and pricing of each category in the coming week by time series analysis. At the same time, using the daily replenishment total as the decision variable and the supermarket revenue as the objective function, we constructed an optimisation model through the method of "cost-plus pricing" to solve the daily replenishment total and the corresponding pricing strategy; finally, we firstly used the small-period sample prediction method, i.e., the Markov prediction model, to analyse the data, and then constructed a multi-objective planning model to further formulate the replenishment total of individual products. Finally, we use the small-period sample prediction method, i.e. Markov prediction model, to analyse the data, and then construct the

multi-objective planning model to further formulate the replenishment plan of individual products, so as to solve the total daily replenishment and the corresponding pricing strategy.

2. Modelling and Solving

2.1. Pearson correlation coefficient

In statistics[2], Pearson correlation coefficient, also known as PPMCC or PCCs, commonly expressed as r or Pearson's r in articles) is used to measure the correlation (linear correlation) between two variables X and Y , with a value between -1 and 1 . The coefficient is widely used in the natural sciences to measure the degree of correlation between two variables, mainly to measure the linear correlation between variables. The Pearson's correlation coefficient is calculated as follows.

$$r = \frac{Cov(X,Y)}{S_X S_Y} \quad (1)$$

$$S_X = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} \quad (2)$$

$$S_Y = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}} \quad (3)$$

The closer the correlation coefficient is to 1 or -1 , the larger the absolute value of the correlation coefficient is, the stronger the correlation is; the closer the correlation coefficient is to 0 , the weaker the correlation is. The Pearson correlation coefficient between the data of each category is then calculated, and in order to visualise the structure of , the

matrix heat map is plotted as shown in Figure 1.

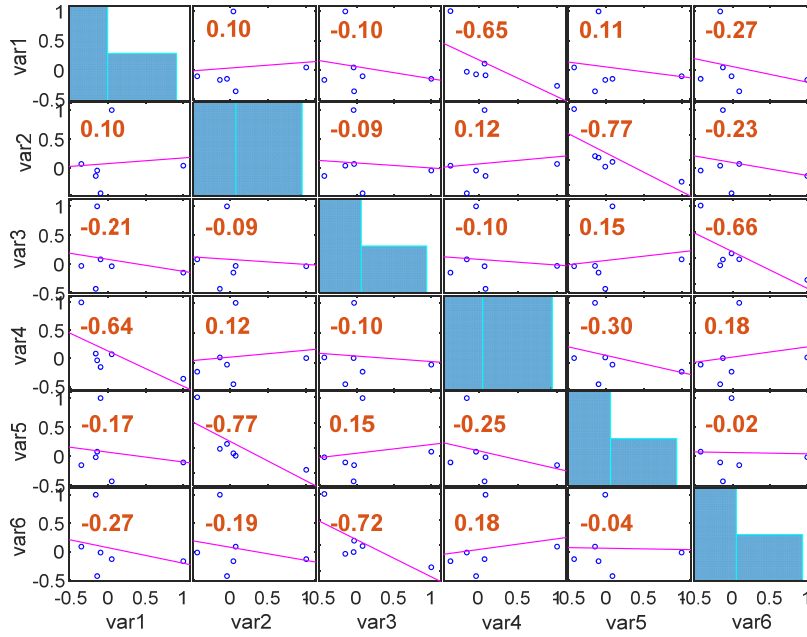


Figure 1. Matrix heat map

The floral and foliage category showed a strong correlation with chilli, aquatic roots and edible mushrooms, while the deep water roots and eggplant had a weak correlation, with an overall negative linear correlation now seen in the amount of individual items within the same category. The six categories can be initially divided into two parts, with the five categories of foliar, floriferous, chilli, edible mushrooms, and aquatic plants as one category, possessing more than moderate correlation, while eggplant has a weak correlation with the former.

2.2. Optimised and improved LSTM model based on WSO white shark algorithm

The formula for calculating total sales and profit is as

follows:

$$S_{total\ sales} = \sum n \cdot p_{sale} \tag{4}$$

$$S_{profit} = \sum n \cdot (p_{sale} - p_{stock}) \tag{5}$$

Cost plus percentage $\alpha = \frac{S_{total\ sale}}{S_{total\ sale} - S_{profit}} - 1$. Here, P_{stock} is the incoming stock price per unit weight n for the day.

Upon investigation, the daily sales volume of the six major vegetable categories in the superstore for the last 30 days is shown in Figure 2:

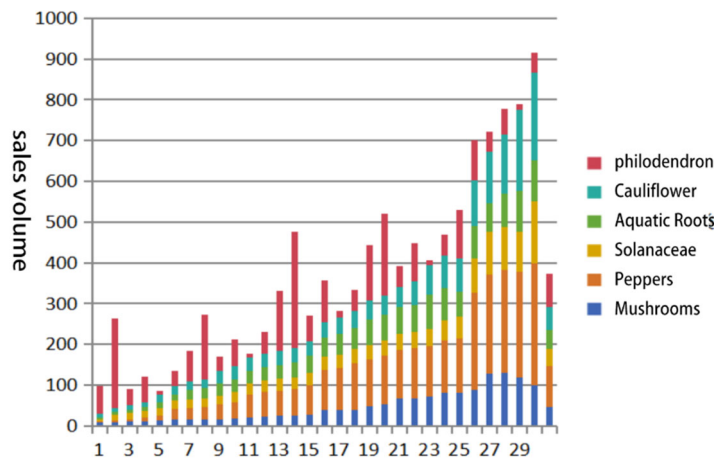


Figure 2. Daily sales volume of the six major vegetable categories for the last 30 days

Considering that 1-7 July 2023 is a complete week and there are no legal holidays, then weekdays and rest days also significantly affect sales volume. In order to ensure the

maximum revenue of the superstore, we optimise the improved BiLSTM model based on the WSO White Shark algorithm, and use the MATLAB tool to predict the sales data

within the last 30 days through the long Short-term Memory Recurrent Neural Networks (LSTM) for time-series analysis[3], and use the average (LMSE) as the loss function of the model to predict the data of the last 7 days and get the sales volume M_1, M_2, \dots, M_7 , And in order to mine the cycle

information, the average sales volume over 30 days is directly used to carry out by calculating the cycle weights $\beta_1, \beta_2, \dots, \beta_7$, the weights are obtained by using the same period data of previous years is as shown in Table 1.

Table 1. Weight during a week

1 July	2 July	3 July	4 July	5 July	6 July	7 July
0.175	0.127	0.110	0.130	0.143	0.159	0.155

The weighting of the number of days of the week gives A_1, A_2, \dots, A_7 , which is calculated as

$$A_n = \text{Mean}(\text{Sales volume in thirty days}) \cdot \beta_n \quad (6)$$

where $n=1,2,\dots,7$.

The final predictions are linearly weighted

$$D_n = 0.5M_1 + 0.5A_n, n = 1,2, \dots,7 \quad (7)$$

Finally, the predicted sales volume for one week is obtained; finally, a regression model is fitted to obtain the pricing strategy.

2.3. Markov prediction model and particle swarm optimisation algorithm

Firstly, we analyse the data of single product, for a total of 251 types of single product, and calculate its average revenue, as shown in Figure 3.

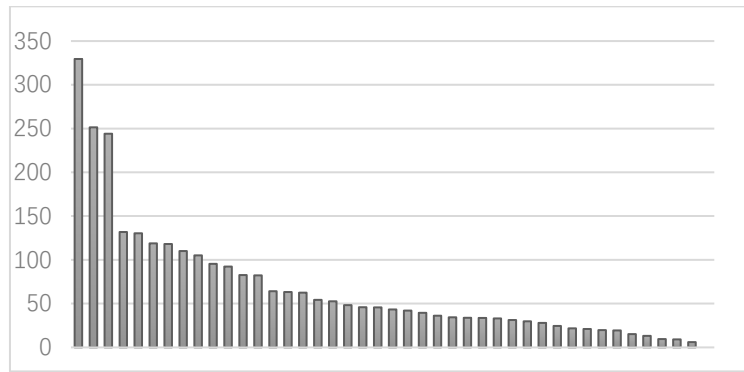


Figure 3. Histogram of average daily returns by category

The vast majority of individual products have an average daily return of ¥10 or less, but there are also higher-returning

products, which are further analysed in terms of average unit price and number of daily returns, as shown in Figure 4.

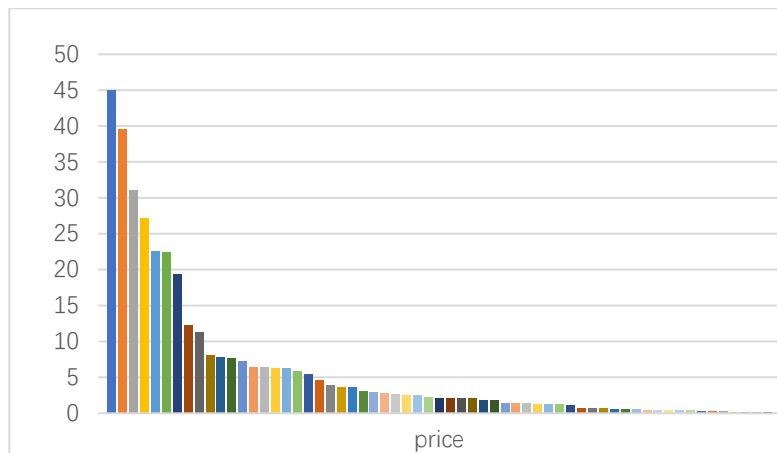


Figure 4. Price histogram

By filtering the data, we can filter out the 4 individual items represented by local spinach that have no transaction information. Then we use profit to select the category with profit contribution rate of 27-33, and give the replenishment

quantity and pricing strategy according to the single product and its corresponding data, and the results are shown in Table 2.

Table 2. Weight during a week

philodendron	cauliflower	Aquatic rhizomes	eggplant	chilli	edible mushroom
83.89956	48.55992	38.77241	37.9893	108.2921	49.71332

The problem can be solved by small sample cycle prediction method with dynamic programming model[4], here we use Markov prediction model[1] and particle swarm optimisation algorithm[5].

Before solving the problem, further cleaning of the problem data is required due to the constraints that exist in this real-world problem: (1) the order quantity of each item is required to meet the requirement that the minimum display weight is greater than 2.5kg. (2) a small sample of forecasts, with the vegetable items controlled to be 27-33. (3) constraints on the varieties that can be sold from 24-30 June 2023 Due to the short shelf life of vegetables, most of the individual items cannot be sold on the next day if they are not sold on the same day, so we introduce the wastage rate data to correct the shelf life, and establish the correlation between the wastage rate and the shelf life, which can be used to further filter out the non-compliant individual items. Intuitively, the lower the wastage rate is, the longer the shelf life is, and the smaller the average daily sales are in the case of queue value $T=2.5$.

The attrition rate was analysed first: the mean attrition rate was 9.43 percent with a standard deviation of 0.052.

Modelling accordingly:

$$B = \begin{cases} 1, & S > 9.43\% \\ -\frac{S}{S_m} + 2, & S \leq 9.43\% \end{cases} \quad (8)$$

where the wastage rate is S , the sample average wastage rate is S_m , and B is the corrected shelf-life, whose wastage rate $S \geq 0$.

Using the processed data, the relationship between replenishment and pricing for each individual vegetable item was calculated, resulting in 251 linear regression models that served as constraints on the equation. Using the solution of the assignment problem, for 251 individual vegetable items, different combinations of vegetables need to be selected each day. Therefore, it is necessary to introduce the variable $X_i = 0$ or 1 where $i=1, 2, 3, \dots, 251$. X_i is a 0-1 variable, where a value of 1 means that the vegetable item is sold and a value of 0 means that it is sold on the same day. Therefore, the switching variable X_i summation to represent the total number of individual items control constraint. That is,

$$27 \leq \sum_{i=1}^{251} X_i \leq 33 \quad (9)$$

For each individual item ordered to satisfy the minimum display quantity constraint, which can be expressed as the total amount of each individual item $x_i, i=1,2,3,\dots,251$. When a need is greater than 2.5 kg, then it can be expressed as $x_i > 2.5, i=1,2,3,\dots,251$.

For the calculation of profit, we follow our previous understanding and set y_i to be the profit margin. Thus, the pricing of each individual product can be expressed as

$$z_i = (1 + y_i)c_i \quad (10)$$

where z_i denotes the pricing of item i , and c_i denotes the stock price of item i . Therefore, The final total profit is expressed as $y_{total} = \sum_{i=1}^{251} y_i v_i$.

In summary, the optimisation model constructed is shown below.

$$\max Y_{total} = \sum_{i=1}^{251} x_i w_i v_i \quad (11)$$

$$\begin{cases} z_i = (1 + y_i)c_i \\ y = kx + b \end{cases} \quad (12)$$

$$\begin{cases} 27 \leq \sum_{i=1}^{251} x_i \leq 33 \\ x_i \geq 2.5, i = 1, 2, 3 \dots 251 \end{cases} \quad (13)$$

Since the planning model we have built is solved directly, whether using lingo or matlab, we will face a huge amount of computation and consume a lot of time. Therefore, for the multivariate planning model we built, we solved it using multi-objective particle swarm optimisation algorithm. The total profit of ¥722.335 is finally obtained for a total of 33 categories of individual products.

3. Model Evaluation

The advantages of the model proposed in this paper are: BiLSTM is suitable for dealing with long-term dependencies and nonlinear relationships; it can better handle large-scale data and high-dimensional data; it can capture long-term patterns and short-term fluctuations in serial data; Markov predicts that when the cumulative sales at the present moment are known, it can be used to study the sales of a shop at a certain point in the future, which is very practical for solving optimisation models. However, a small sample of data may be overfitted; the selection and adjustment of model parameters are complicated; In addition, in the analysis of outliers, the processing will be highly subjective, if there is a clear method of dealing with outliers, the impact of this factor can be reduced. In addition, factors such as predicted sales volume may be non-linear, while in this paper they are analysed as linear factors, which will lead to certain errors. The optimisation class model is a comprehensive tool for selecting the most reasonable solution from all possible solutions to reach the optimal goal. The model includes several steps. such as data preprocessing, statistical analysis, and correlation analysis. This can be extended to Dynamic programming (English: DP), a method used in mathematics, management science, computer science, economics, and bioinformatics to solve complex problems by decomposing the original problem into relatively simple subproblems. It is used for finding the maximum value, resource allocation problems such as the Fibonacci series problem, the step-up problem, the maximum continuous substring problem, the knapsack problem, the make change problem, etc.; it can also be related to graph theoretic problems, and then used to solve the Shortest Path Problem (SPP) as well as the (Multi-)traveller

problem by using Dijkstra's algorithm, the Bellman-Ford algorithm, etc., so as to It can also be applied to a greater extent to the classification and regression of data. The aim of this paper is to help supermarkets to maximise their profits, satisfy market demands and improve market competitiveness.

References

- [1] Wang Yan. An analysis of the influencing factors of pricing and its pricing strategy in small and medium-sized supermarket chains[J]. Journal of Lanzhou Institute of Technology, 2014, 21 (03): 99-102.
- [2] Cheng Juanjuan. An empirical study on the relationship between research and teaching in colleges and universities--an analysis based on Pearson correlation coefficient[J]. Science and Technology in Chinese Colleges and Universities, 2022, No.410(10): 46-52.
- [3] J. H. Li, Q. Y. Wang, Y. L. Fan, et al. Whale optimization algorithm-bidirectional long and short-term memory neural network for prediction of mechanical remaining life of circuit breakers[J/OL]. High Voltage Technology:1-13[2023-09-15].DOI:10.13336/j.1003-6520.hve.20230013.
- [4] CHENG Yuandong; YU Kexin; LI Xianyang. Regional logistics demand forecasting based on weighted Markov-ARIMA modified model[J]. Journal of Shandong Jiaotong College, 2023,31 (03):22-28.
- [5] Li Weiwei. Unbiased grey Markov optimization model and its application in merchandising forecasting[C]. Harbin Institute of Technology, 2015.