International Journal of New Developments in Engineering and Society
ISSN 2522-3488 Vol. 7, Issue 9: 31-36, DOI: 10.25236/IJNDES.2023.070906

Research on the optimization of vegetable sales and pricing strategy of single products

Jinlun Yu*, Jiawei Li, Siyuan Wang

School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an, 710049, China
**Corresponding author: 2211410868@stu.xjtu.edu.cn*

Abstract: *The comprehensive analysis of sales optimization and individual pricing strategies for vegetable commodities is crucial for the profitability of fresh food supermarkets. Drawing from transaction records of specific vegetable items in a supermarket, this study reveals that different vegetable categories exhibit cyclical patterns in sales over time and certain inter-category correlations exist. Through Pearson correlation analysis, we identified strong sales-time correlations between Shanghai green and Yunnan choy sum (correlation coefficient of 0.94), bamboo leaf vegetables and sweet potatoes (0.96), and Yunnan lettuce and Yunnan choy sum (0.96), suggesting promising sales combinations for supermarkets. An algorithmic approach was employed to derive a functional relationship between sales volume and pricing for different vegetable categories, which generally follows an exponential relationship, albeit with variations among categories. Leveraging transactional big data, we used LSTM time series forecasting to provide optimal restocking quantities and pricing decisions for various vegetable categories from July 1-7, 2023. Ultimately, by offering insights into sales combinations, optimal restocking volumes, and pricing decisions, this paper aids fresh food supermarkets in achieving higher profits and holds practical significance for enhancing the market circulation of fresh produce.*

Keywords: *Commodity sales and pricing strategy, Correlation Analysis, LSTM model*

1. Introduction

With the continuous improvement of social productivity and the increasing market competition, people have higher and higher requirements for the variety and quality of fresh vegetables. A single product on the market can no longer meet the current demand, replaced by a wide variety of different specifications for the sale of a combination of goods[1] ; At the same time, vegetable commodities are easy to deteriorate and difficult to store, which makes vegetable management face great challenges. Therefore, it is of great significance to reasonably analyze the market demand and formulate reasonable replenishment and pricing strategies in advance to improve the quality and quantity of vegetables sold.

Since Sepp Hochreiter and Jürgen Schmidhuber jointly proposed Long Short-term Memory Networks (LSTM) in 1997, LSTM network has been widely used in natural language processing, speech recognition, time series prediction and other fields. At the same time, based on the transaction data of vegetable commodities in a supermarket, this paper analyzes the distribution law and relationship between the sales volume of vegetable categories and individual products, and correlates the characteristics of vegetable purchase price and consumption rate, and uses LSTM time series prediction model to obtain the replenishment volume and sales price in the next few days, which provides a reference for the formulation of replenishment and pricing strategies of supermarkets, and greatly improves the fresh operating profit of supermarkets.

2. Mining the relationship between vegetable sales and time

2.1 Extraction and integration of data features

Due to the huge amount of detailed data of vegetable sales, we first carried out data feature extraction and integration. The time dimension is divided into three years: the first year is from July 1st, 2020 to June 30th, 2021, the second year is from July 1st, 2021 to June 30th, 2022, and the third year is from July 1st, 2022 to June 30th, 2023. Then we deleted the transaction record of "the sales volume of some single products was recorded as negative due to customer return", and obtained the sales volume

ISSN 2522-3488 Vol. 7, Issue 9: 31-36, DOI: 10.25236/IJNDES.2023.070906

of 251 kinds of single products in different months through statistical analysis. Observation of the data reveals that the annual sales volume of many single products was 0. We thought that this data was not conducive to the subsequent extraction of digital features, so we deleted it.

2.2 Distribution and interrelation of vegetable categories

It is considered to draw the curve of the sales volume of each vegetable category over time in three years and the column chart of the sales volume of each category in three years. Observe the change of the sales volume of each vegetable category with time, and analyze its seasonal and periodic change trend; At the same time, the sales volume differences of different categories of vegetables are compared, so as to draw a conclusion.

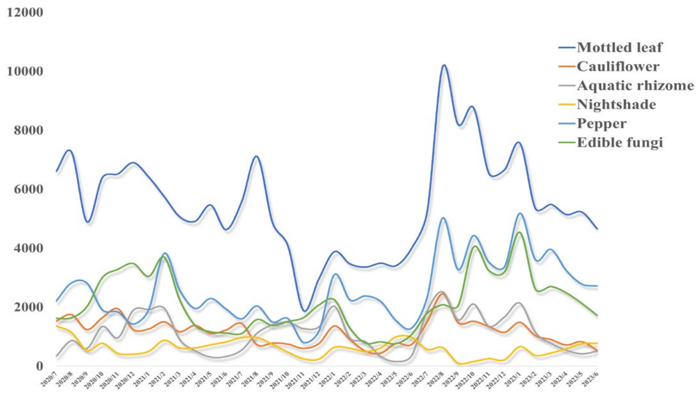
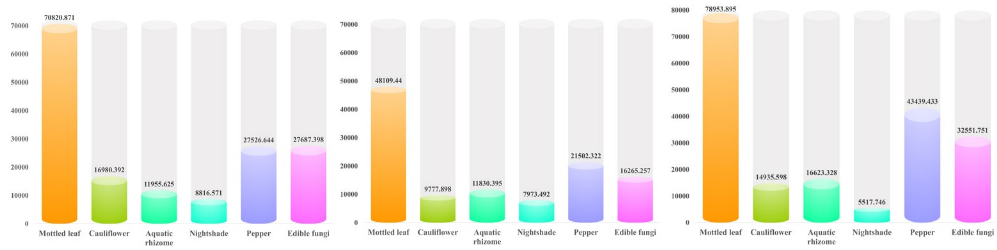


Figure 1: Curve of sales volume of various vegetable categories over time in three years

The following conclusions can be obtained by observing and analyzing Figure 1: Generally speaking, the sales volume of each vegetable category has a certain annual periodicity and seasonal periodicity within three years; In the second year, the sales of various vegetable categories were significantly lower than those in the other two years. Considering that the region adopted relatively strict epidemic control measures in the second year, resulting in a decline in sales.

By analyzing the changes of vegetable sales over time, it can be seen that the sales of most vegetable categories have declined to some extent in winter and spring, and have rebounded significantly in summer and Autumn: around August every year, the sales of various vegetable categories have increased by different extents, especially in the third year, which may be due to the bumper harvest of most vegetables and sufficient reserves at the supply end; In the following months, the sales volume of various vegetable categories fluctuated to some extent, especially in the second year, the sales volume of mottled leaf dropped sharply; Around February of the next year, the sales of various vegetable categories have rebounded to some extent. Considering that it is the Spring Festival, the demand of customers for various vegetables has driven the sales of vegetable products of supermarkets; From March to April, the sales of various vegetable categories fell to varying degrees.

Comparing the sales change curves of different vegetable categories, it can be seen that the change trends of "Edible fungi and Aquatic rhizomes" and "Cauliflower and Nightshade" are similar. It is speculated that the two groups of vegetable categories have certain similarities in consumer demand and the supply cycle of supermarkets. Considering that supermarkets may adopt the promotion scheme of bundle sale; In addition, in the winter of the third year, the sales of pepper and edible fungi were highly similar. Therefore, it is believed that market has adopted a new sales mix to improve profits.



(a) (b) (c)

Figure 2: Column chart of sales volume of each vegetable category in three years

ISSN 2522-3488 Vol. 7, Issue 9: 31-36, DOI: 10.25236/IJNDES.2023.070906

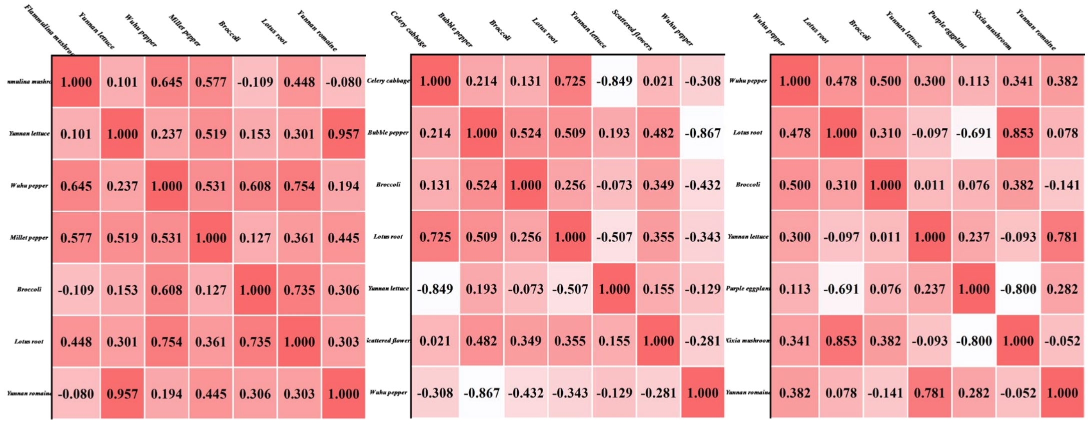
The following conclusions can be obtained by observing and analyzing Figure 2:

The sales of various vegetable categories in the three years showed some similarities: the mottled leaf accounted for the vast majority of the market share, the sales of pepper and edible fungi accounted for about half of the sales of the mottled leaf, the sales of cauliflower and aquatic rhizomes accounted for about one quarter of the sales of the mottled leaf, and the market share of nightshade was the least.

It is speculated that the local supply of mottled leaf, peppers and edible fungi is sufficient, and the residents' dietary structure has a certain preference, and there is a large demand for this kind of vegetables in daily life; However, the yield of cauliflower, aquatic rhizome and nightshade may be limited due to regional climate, and the residents' demand for such vegetables is small.

2.3 Distribution and relationship of vegetable single products

Considering that the sales volume of single vegetable products meets the continuity requirements, the Pearson correlation coefficient among vegetable menu products is calculated by calling the "Pearson function" in Excel, and the correlation heatmap is drawn. The correlation heatmap of some vegetable items is shown in Figure 3:



(a) *(b)* *(c)*

Figure 3: 3-year correlation heatmap of some categories

Analyzing Figure 3, we can see that the closer the Pearson correlation coefficient is to 1, indicating that the sales volume curve of the two is more consistent with time. Considering that they may have high similarities in growth cycle, producer supply and consumer demand. Based on this, supermarkets can make reasonable sales plans according to the combination of highly correlated items, so as to improve revenue.

3. Research on the relationship between vegetable types and pricing

3.1 Extraction and integration of data features

In order to study the relationship between the sales volume and pricing of vegetable categories, based on the sales flow details of each single item from July 1, 2020 to June 30, 2023, the daily sales volume and pricing statistics of each vegetable category were obtained. Among them, vegetable sales volume p is defined as:

$$p = \sum n \tag{1}$$

$$q = \frac{\sum n \times w}{\sum n} \tag{2}$$

Where n is the sales volume of individual vegetables in this category.

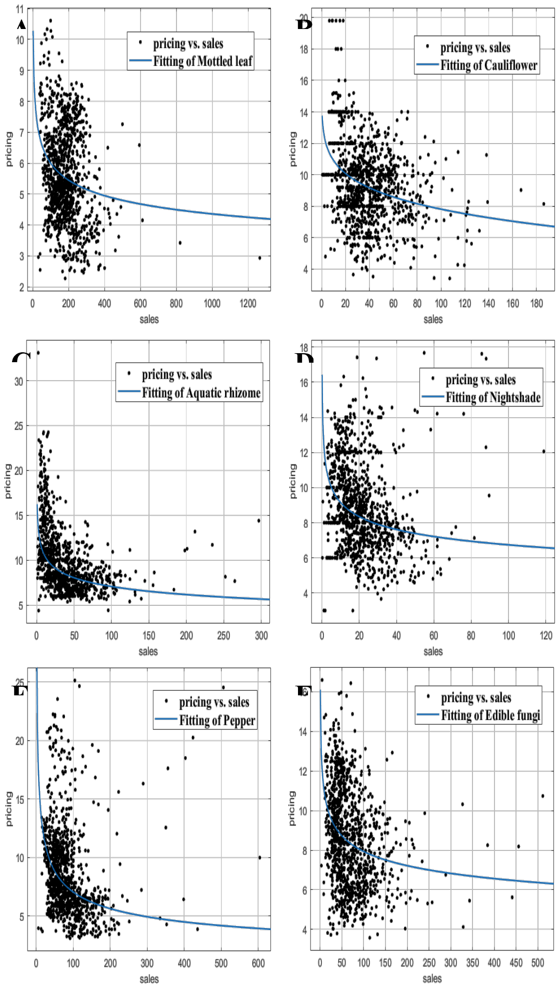
Vegetable category pricing q (weighted pricing) is defined as:

Where w is the pricing price of each vegetable in this category.

ISSN 2522-3488 Vol. 7, Issue 9: 31-36, DOI: 10.25236/IJNDES.2023.070906

3.2 Analyzing the relationship between vegetable sales and pricing

Based on the statistical data obtained in 3.1, the "daily sales pricing" scatter diagram of each vegetable category is drawn. We guess the possible functional relationship between sales volume and pricing, and use the fitting toolbox "curve fitting toolbox" in Matlab to fit solution. The graphs and relationship equations between total sales and pricing of the six vegetable categories were obtained as shown in Figure 4 and Table 1, where x-axis is sales and y-axis is pricing.



(A: Mottled leaf, B: Cauliflower, C: Aquatic rhizome, D: Nightshade, E: Pepper, F: Edible fungi)

Figure 4: Fitting curve of the relationship between sales and pricing of various vegetable categories

Table 1: Fitting function of the relationship between sales and pricing of various vegetable categories

ID	Function	R^2	RMSE
Mottled leaf	$f(x) = 11.35x^{-0.14}$	0.06	1.47
Cauliflower	$f(x) = -5.56x^{0.15} + 18.68$	0.11	2.34
Aquatic rhizome	$f(x) = 20.61x^{-0.11} - 5.47$	0.98	0.47
Nightshade	$f(x) = 12.42x^{-0.13}$	0.96	0.53
Pepper	$f(x) = 31.49x^{-0.32}$	0.98	0.47
Edible fungi	$f(x) = 14.9x^{-0.14}$	0.09	2.31

Comparing the fitting results of different vegetable categories, the fitting degrees of different vegetable categories are quite different: the fitting effects of aquatic rhizome, nightshade and pepper are significantly better than the other three categories, and the goodness of fit of pepper is R^2=0.98, RMSE=0.47, which better expresses the relationship between sales and pricing; However, the goodness of fit of cauliflower: R^2=0.11, RMSE=2.34, the fitting effect is poor. The correlation between the sales volume and pricing of aquatic rhizome, nightshade and pepper is strong, and it is less affected by market and human factors, showing a certain functional relationship.

ISSN 2522-3488 Vol. 7, Issue 9: 31-36, DOI: 10.25236/IJNDES.2023.070906

4. Prediction of total replenishment quantity and pricing strategy

The vanilla LSTM incorporates changes by Gers et al.[2] and Gers and Schmidhuber[3] into the original LSTM[4] and uses full gradient training. A schematic of the vanilla LSTM block can be seen in Figure 5. It features three gates (input, forget, output), block input, a single cell (the Constant Error Carousel), an output activation function, and peephole connections. The output of the block is recurrently connected back to the block input and all of the gates.

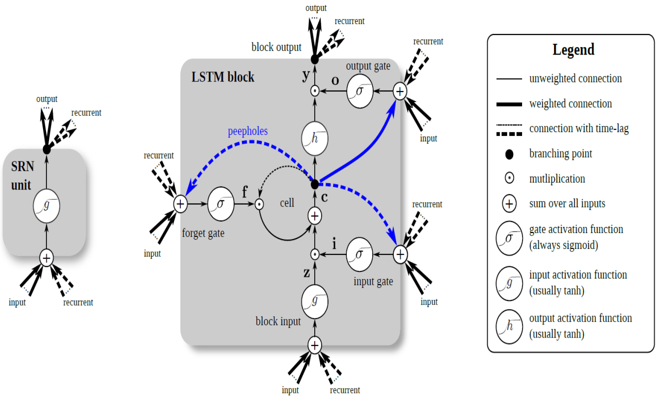


Figure 5: Detailed schematic of the Simple Recurrent Network (SRN) unit (left) and a Long Short-Term Memory block (right) as used in the hidden layers of a recurrent neural network.[5]

By associating the data given by the sales flow details and wholesale prices of each commodity of the supermarket from July 1, 2020 to June 30, 2023, it is found that the sales volume, sales unit price and wholesale price of different items of the same category are different in the same day. Therefore, this paper sums the sales unit price and wholesale price of different items of the same category according to the weight of their daily sales volume in the category. Further, the sales volume, sales unit price and wholesale price data of different vegetable categories from July 1, 2020 to June 30, 2023 are obtained.

Put the above data into the LSTM time series prediction model to obtain the sales volume and wholesale price of each vegetable category from July 1 to July 7, 2023.

Finally, the replenishment volume and pricing strategy of categories from July 1 to July 7 are predicted, as shown in Table 2 and Table 3:

Table 2: Replenishment quantity of vegetable categories from July 1st to 7th

Replenishment quantity (kg)	Pepper	Mottled leaf	Aquatic rhizome	Edible fungi	Cauliflower	Nightshade
July 1st	114.49	172.87	1.39	45.32	21.01	24.06
July 2nd	113.43	219.42	12.09	50.45	14.14	27.68
July 3rd	120.41	73.09	30.65	42.49	9.43	21.29
July 4th	70.53	35.01	17.40	46.63	12.75	17.54
July 5th	107.90	120.30	36.49	36.97	12.96	15.29
July 6th	110.34	198.37	42.89	48.45	31.09	17.09
July 7th	110.71	185.50	51.93	59.59	31.34	14.24

Table 3: Vegetable pricing strategy from July 1st to 7th

Pricing (Yuan/kg)	Pepper	Mottled leaf	Aquatic rhizome	Edible fungi	Cauliflower	Nightshade
July 1st	4.20	6.31	25.92	6.85	18.59	8.85
July 2nd	3.89	7.94	24.45	6.68	19.71	8.54
July 3rd	5.88	8.20	23.64	6.24	19.81	8.00
July 4th	6.24	8.47	25.10	7.81	11.58	7.14
July 5th	6.32	7.13	24.20	7.39	17.10	9.43
July 6th	3.67	6.28	22.05	8.05	11.80	9.52
July 7th	7.24	7.47	22.53	7.57	16.56	10.77

ISSN 2522-3488 Vol. 7, Issue 9: 31-36, DOI: 10.25236/IJNDES.2023.070906

5. Conclusions

This paper starts from the optimization of vegetable sales and item pricing strategy, and based on the transaction data of vegetable items of a supermarket, the following conclusions are drawn:

The time distribution of the sales volume of different vegetable categories has certain annual periodicity and seasonal periodicity: the sales volume of each vegetable category in the second year is significantly lower than that in the other two years; The sales volume of most vegetable categories declined in winter and spring, and rebounded significantly in summer and autumn. There is also a certain degree of correlation between different vegetable categories: similar sales trends for "Edible fungi and Aquatic rhizome" and "Cauliflower and Nightshade".

Based on the vegetable products with high Pearson correlation coefficient, six best-selling sales combinations were obtained: purple eggplant and Yunnan lettuce (correlation coefficient 0.82); Chinese cabbage and Chinese Cabbage (correlation coefficient 0.89); Shanghai green and Yunnan lettuce (correlation coefficient 0.94); The correlation coefficient between bamboo leaf vegetable and sweet potato tip was 0.96; Brassica chinensis and Chinese cabbage (correlation coefficient 0.93); Yunnan lettuce and Yunnan lettuce (correlation coefficient 0.96).

Using the fitting algorithm, the functional relationship between the sales volume and pricing of various vegetable categories is obtained, which generally presents a power function relationship. The fitting effects of different vegetable categories have certain differences: the fitting effects of aquatic rhizome, nightshade and pepper are significantly better than those of mottled leaf, cauliflower and edible fungi. The correlation between the sales volume and pricing of the former is strong, and it is less affected by market and human factors.

The transaction big data is integrated, and the LSTM time series prediction method is used to determine the optimal replenishment volume and pricing decision of each vegetable category from July 1 to July 7, 2023, so as to maximize profits.

Finally, this paper provides supermarkets with scientific replenishment and pricing strategies in the form of best-selling sales mix, optimal replenishment volume and pricing decision, which helps supermarkets in the fresh category achieve higher profits, and has a certain positive significance in promoting the market circulation of fresh commodities.

References

[1] Wang S , Li Y , Jin X .Based on the multi-level optimization model of remanufacturing and pricing balance strategy research under TIFR subsidy in China[J].EDP Sciences, 2021. DOI:10.1051/e3sconf/202125702069.

[2] Gers F A, Schmidhuber J U, Cummins F. Learning to Forget: Continual Prediction with LSTM[J]. Technical Report IDSIA-01-99. 1999.

[3] Gers F A, Schmidhuber J U. Recurrent nets that time and count[J]. Neural Networks, 2000. IJCNN. 2000.

[4] Hochreiter S, Schmidhuber J U. LONG SHORT-TERM MEMORY[J]. Neural Computation 9(8):1735-1780. 1997.

[5] Greff K, Srivastava R K, Koutnik J, et al. LSTM: A Search Space Odyssey[J]. IEEE Trans Neural Netw Learn Syst. 2017, 28(10): 2222-2232.