Research on supermarket replenishment volume prediction based on LSTM and PSO algorithm

Ke Yan[1, *], Daxuanrui Deng[2]

[1]School of Economics and Management, Xinjiang University, Urumqi, China, 830046

[2]School of Mathematics and Systems Sciences, Xinjiang University, Urumqi, China, 830046

*Corresponding author: 13363043178@163.com

Abstract. In order to meet people's health needs and reduce economic losses, the new fresh suppliers carried out real, reliable and targeted market demand analysis so that they formulate automatic pricing and replenishment decision of vegetable commodities with the aim of providing fresh and nutritious vegetable commodities. This paper adopts the apriori model to explore the correlation between vegetables and reflects the relationship between total sales and cost plus pricing by the linear fitting method.This paper further uses the LSTM time series prediction model, time sliding sequence and PSO particle swarm algorithm to predict the various vegetable categories of daily supplement and pricing in the next week, so as to maximize the business super profit. The results showed that the price of each vegetable category had a significant negative correlation with the sales volume and the dynamic prediction of replenishment quantity and pricing data had obvious application effect, which effectively helped the supermarket to make pricing and replenishment decisions.

Keywords: LSTM algorithm apriori model PSO particle swarm optimization algorithm prediction model.

1. Introduction

The development of the times has spawned the fresh business super community, retail and electronic form while fresh commercial supermarkets also actively adapt to the needs of health, refinement and diversification. However, the fruits and vegetables rot every year,which results in a huge economic loss. The perishable characteristics of vegetable commodities seriously restrict the profit increase of supermarkets and the health demand of consumers [1].

With the object of better preserving vegetable commodities, the relevant research is mainly divided into technical and metrological types.The first is technical research, Cao Tingting et al. adopted light technology to maintain the quality of fruits and vegetables [2].The second type is a metrological study,Cao Yu et al. studied the problem of multi-product joint replenishment by establishing the average replenishment cost model[3].Kiil K et al. positively investigated the impact of automatic replenishment and expanded the current understanding of automatic replenishment programs[4].Wang and Lin proposed a model based on a new scale-free network for the purpose of optimizing the supply chain network replenishment path and analyzing the network static and dynamic performance of [5]. Sharma S et al. fully characterized the optimal replenishment strategy of the dual-product system and provided a method to calculate the optimal replenishment volume so that they further developed a heuristic multi-product system[6]. Particle swarm optimization (PSO) mentioned by Chen and Tan is one of the most popular group-based random algorithms used to solve complex optimization problems[7]. Feng and Li used fuzzy planning method to handle the joint replenishment problems with fuzzy requirements and transform them into PSO problem model, providing important ideas for optimizing replenishment decision [8].

In this paper, the linear quasi-legal to explore the relationship between total sales and cost plus pricing. Based on the method of time series model and PSO particle swarm optimization, this paper globally search and optimizes iteration by iteration, and then predict the daily replenishment and pricing of each vegetable category in the next week, making the supermarket the maximum profit.

2. Model building

2.1. Correlation study

The Apriori algorithm is a classical algorithm for data mining and association rules to mine to find frequent association relations between items in the dataset.The Apriori algorithm has the advantages of finding potential correlation and flexibility as well as unsupervised learning, which can be used to mine potential correlation in datasets and discover association rules.[9]

This paper first scans the dataset to determine the frequency of sales volume and finds the set of frequent items based on the minimum support threshold finding that the items exceed the threshold in the dataset.Candidates sets were generated using the frequent set and the dataset was scanned again to determine the frequency of each candidate set.The above procedure is finally repeated until a new set of frequent items cannot be generated.It is helpful for us to use frequent item sets to generate association rules,which can help understand the correlation between items and can be used for prediction.[10]

2.2. Linear fitting of cost-plus pricing, discount rate and total sales

"Cost plus pricing" method is the general pricing method of vegetables, transportation loss and poor quality goods are usually sold at a discount. The cost plus formula is not only the pricing method, but also the essence of pricing[11].Therefore, real, reliable and targeted market demand analysis has become an important condition for replenishment decision and pricing decision.

2.2.1 The relationship between cost-plus pricing and total vegetable category sales

In order to study the relationship between total sales of each category and cost plus pricing, the following models are constructed for linear fitting:

$$f(x) = ax_i + b \tag{1}$$

x_i:Cost-plus pricing for item i
$f(x)$: Total sales of vegetable i

2.2.2 The relationship between the total sales volume and the discount rate of each vegetable category

$$f(x) = pk_i + q \tag{2}$$

k_i:Cost-plus pricing for item i
$f(x)$: Total sales of vegetable i

2.2.3 Daily replenishment volume and pricing strategy of each vegetable category in the next week

Particle swarm optimization (PSO) is an optimization algorithm based on swarm intelligence. By simulating the movement and information exchange of particles in the search space, it can find the optimal solution. It has the advantages of strong global search ability, fast convergence speed, strong adaptability and parallel processing. In addition, this paper uses PSO algorithm to optimize the key parameters of LSTM, and adds random noise to the input layer of PSO-LSTM neural network, so that the data characteristics can match the network topology and strengthen the anti-interference of the network, and then gradually optimize the replenishment volume and pricing strategy, achieve iterative optimization, and approximate the optimal solution[12].

In this paper, in the particle swarm algorithm, the replenishment quantity, discount rate and original selling price of commodity i on day t are set to the dimensions of each particle respectively.

To this end, this paper constructs the following model under the goal pursuit and constraint of profit maximization:

(1) objective function:

$$\text{Maximize } Z = \sum_i \sum_{t=1}^{7} (P_{i,t} \times (\beta_0 + \beta_1 \times P_{i,t} - L_{i,t}) + P_{i,t} \times (1 - D_{i,t}) \times L_{i,t}) - C_i \times R_{i,t} \tag{3}$$

among:
$P_{i,t}$: Original sale price of item i on t day.
$D_{i,t}$:Item i at t.
$L_{i,t}$:Loss and sales of goods in t days.
C_i:Cost of goods..
$R_{i,t}$: Reamount of item i in t.
(2)constraint condition:
①Inventory update:

$$I_{i,t} = I_{i,t-1} - (\beta_0 + \beta_1 \times P_{i,t}) \quad \forall i,t \tag{4}$$

Among:
$I_{i,t}$: Goods i in t.
②Sales volume does not exceed the inventory:

$$\beta_0 + \beta_1 \times P_{i,t} \leq I_{i,t-1} \tag{5}$$

③Sales volume fluctuation range constraint:
To ensure that the actual sales volume (normal sales + discount sales) is within a fluctuating range of the expected sales volume, this paper assumes that using the time series to predict the expected sales volume of commodity i at day t is E_i,and the fluctuation range is $\propto$, the model is as follows:

$$E_{i,t} - \propto \times E_{i,t} \ll \sum_{t=1}^{7} (\beta_0 + \beta_1 \times P_{i,t}) \ll E_{i,t} + \propto \times E_{i,t} \quad \forall i,t \tag{6}$$

In this model, this paper uses the linear relationship between the price and sales volume of goods to further adjust the objective function and constraints. The supermarket can not only flexibly adjust the selling price, discount rate and replenishment quantity of goods under the premise of meeting the constraints,but use the particle swarm optimization algorithm to solve this model on aacount of maximizing the target function Z, the maximum profit of the quotient supermarket.

3. Results

3.1. Correlation study

To explore the correlation between each vegetable category, python is used to build the apriori model and draw the correlation map. Each circle represents an item set. One item set is composed of multiple vegetable combinations and double arrows indicate the correlation between the two. As can be seen from Figure 1,anthophyllom, cauliflower, aquatic rhizomes, edible mushrooms, chilli and eggplant vegetables are closely related. In the item set, these vegetables often exist in the form of combination, while when an item set appears, the second item set often appears, and then the third or even the fourth item set will appear. When edible anthophyllom and edible mushrooms exist in item 1, aquatic rhizomes, chilli and eggplant will appear in the form of item 2, while aquatic rhizomes, chilli and cauliflower will also appear in the form of item 3, and aquatic rhizomes, chilli, eggplant and cauliflower will also appear in the form of item 4. Among them, anthophyllom, chilli and aquatic rhizomes are the most popular among consumers.Consumers often consume vegetables in the form of combination and easy to consume other vegetable categories.



Figure 1. Correlation diagram of each category

According to Figure 1, the vegetables of each category are closely related and have strong research ability.

In order to further explore the correlation between vegetable categories, the correlation matrix uses the heat map to show the correlation strength.The darker the color, the stronger the correlation.The lighter the color, the weaker the correlation. According to Figure 2, eggplant vegetables have the weakest correlation with other vegetables, aquatic rhizomes and edible mushrooms, anthophyllom and chilli. Chilli and edible mushrooms , other vegetables are moderate.the correlation difference between other vegetables is small.

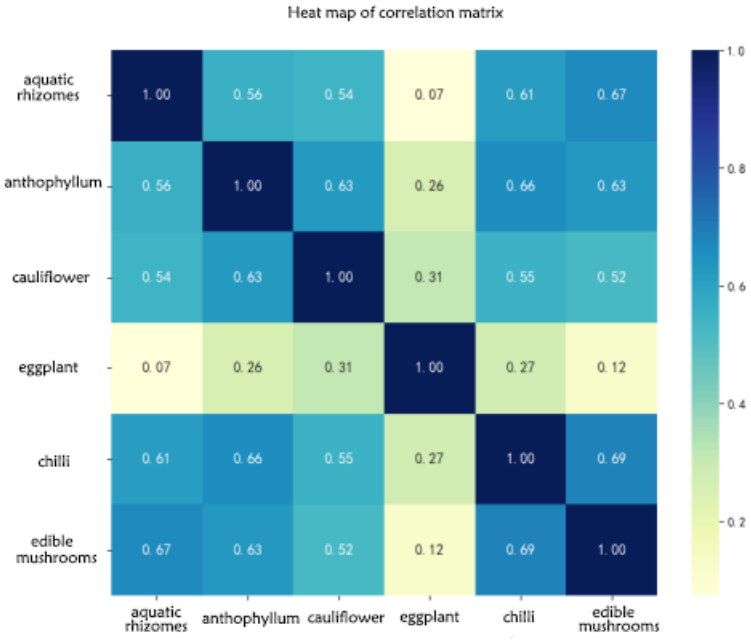

Figure 2. Thermal map of the category correlation matrix

Considering that there are many kinds of single varieties and the correlation relationship of many single products is weak, this paper deals with the relevant data according to the category to cut the set with weak correlation, select the set with strong correlation and studies the correlation relationship of different single products.

By figure 3, oyster mushroom and cordyceps flower, screw pepper and oyster mushroom, cordyceps flower and cayenne pepper, rugosa and oyster mushroom are the most correlated, there is

no correlation between cayyene pepper and green and red Hangzhou pepper.White beech mushroom and cayenne pepper, screw pepper, cordyceps flower, red pepper, white beech mushroom, small chilli pepper, rugosa and agaricus bisporus. Other correlations were moderate and showed small differences.

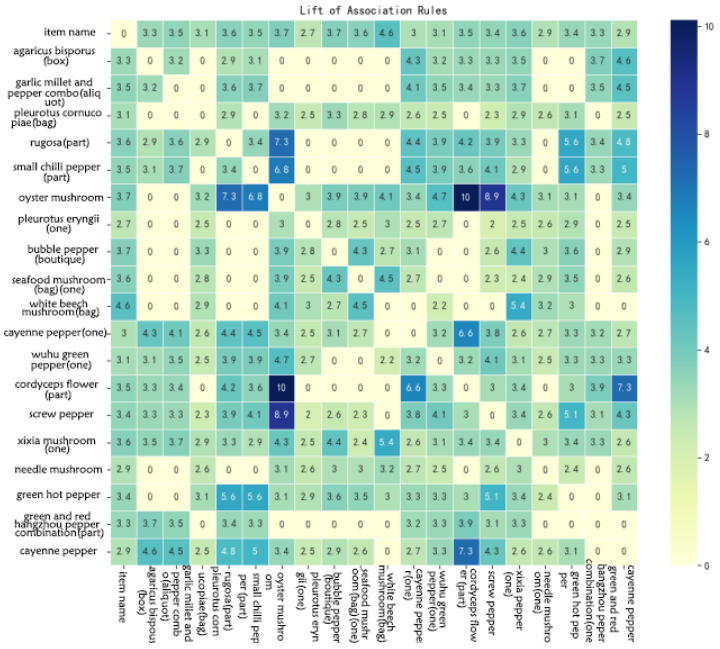

Figure 3. Heat map of the individual product correlation matrix

This shows that consumers are complementary between vegetable category consumption habits and need to buy vegetables often need to match different kinds of vegetables, which means that other related vegetables category will benefit when a vegetable category demand increase in that consumers need to buy a variety of vegetables to meet their needs. The seasonal yield fluctuations of different vegetable categories are usually related, resulting in similar seasonal characteristics among vegetable categories. Consumers usually choose different types of vegetables to obtain diversified nutrition, which affects the correlation between vegetable categories.

3.2. Supplemental volume and pricing

3.2.1 The relationship between the total sales volume of each vegetable category and the cost-plus pricing

Table 1 shows the correlation coefficient between the price and sales volume of each vegetable category. As can be seen from Table 1, it has a strong fit, the correlation coefficient of each vegetable category is less than 0,which revealed the price of each vegetable category is negatively correlated. Edible mushrooms had the strongest negative correlation and the weakest correlation for cauliflower.The negative correlation between selling price and sales volume has great reference value for pricing.Increase the price if it is low while lower the price if it is high.

Table 1. Fitting of the selling price and sales volume of each vegetable category

Category	a	b	R^2	RMSE
Aquatic rhizomes	-0.01462	0.7708	0.6379	0.3412
Anthophyllom	-0.03113	0.7586	0.6072	0.2926
Cauliflower	-0.009359	0.5381	0.7999	0.1533
Eggplant	-0.007657	0.5323	0.8143	0.1803
Chilli	-0.01226	0.5311	0.6255	0.2816
Edible mushrooms	-0.04205	1.024	0.6319	0.2238

3.2.2 The relationship between the total sales volume and the discount rate of each vegetable category

Table 2 shows the correlation coefficient between the price and discount rate of each vegetable category. As it can be seen from Table 2, it has a strong fit.The correlation coefficient of each vegetable category is greater than 0, which illustrates the discount rate of each vegetable category is positively correlated with the sales volume.The negative correlation of edible fungi is the strongest and the negative correlation of cauliflower is the weakest.The negative correlation between discount rate and sales volume can be further refined pricing. If the sales volume is low, the discount rate can be increased.If the sales volume is high, the discount rate can be reduced.

Table 2. Fitting of discount rate and sales volume of each vegetable category

Category	p	q	R^2	RMSE
Aquatic rhizomes	0.2472	0.7575	0.6931	0.3369
Anthophyllom	1.382	-0.2244	0.6722	0.2031
Cauliflower	0.002575	0.4545	0.6103	0.1536
Eggplant	0.2373	0.6367	0.7944	0.1846
Chilli	1.586	-0.4286	0.6751	0.2421
Edible mushrooms	1.861	-0.8318	0.7029	0.3077

Vegetable categories with high discount rate are often considered by consumers to have better cost performance.The high discount rate brings benefits to consumers, thus it stimulates consumers' interest in purchase and prompting them to buy the vegetable category. High discount rate can stimulate consumer consumption and strengthen the willingness to buy, causing consumers to buy the required vegetables in advance or to buy more quantity to obtain more discounts and thus increase the total sales volume.

3.2.3 Daily replenishment volume and pricing strategy of each vegetable category in the next week

In this model, this paper uses the linear relationship between the price and sales volume of goods to further adjust the objective function and constraints. The quotient supermarket can flexibly adjust the selling price, discount rate and replenishment quantity of goods under the premise of meeting the constraints.Additionally,they can use the particle swarm optimization algorithm to solve this model so as to maximize the target function Z.

In order to further improve the accuracy of the prediction model and eliminate the noise in the sales data, the VMD denoising method is adopted in the data processing, shown in Figure 4 is the picture after VMD noise reduction.

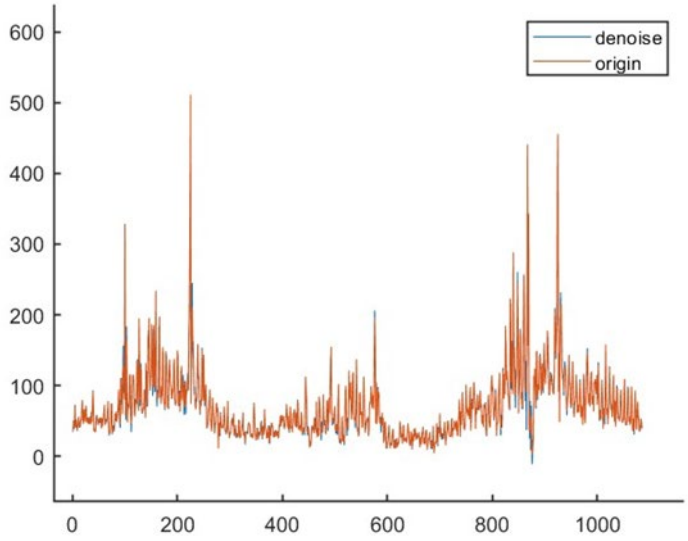

Figure 4. Picture after noise reduction in VMD

Figure 5 is a time series plot of the detrended data. Firstly,this paper uses ADF test to analyze the results.T value are significant,which reject the hypothesis that the sequence is unstable.By viewing the data comparison map before and after the difference, the data fluctuation range is not large, indicating that it is relatively stable.Then, the autocorrelation analysis was performed and the p and q values were calculated. Finally, the model white noise was tested according to the p value of q-statistic. Finally, the AIC and BIC were analyzed for comprehensive analysis to obtain the subsequent prediction results.

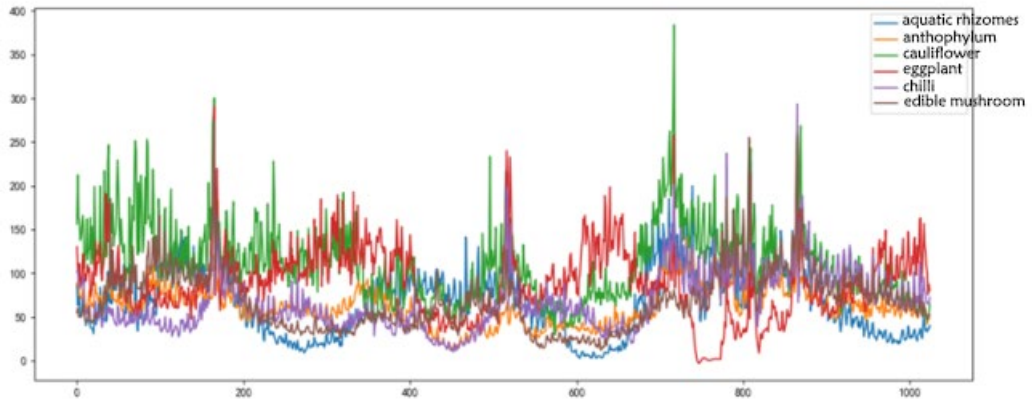

Figure 5. Time-series plot of Fig

Table 3 is the selling price based on time series prediction, data denoising, time sliding sequence and the PSO particle swarm algorithm. The basic price of different vegetable species varies greatly,thus the horizontal comparison is meaningless. Therefore, through the longitudinal comparison, the overall price of aquatic rhizomes vegetables increased.The anthophyllom are relatively stable and the price fluctuates slightly.Edible mushrooms first decrease and then increase, and then decrease again.The price of cauliflower fluctuates greatly, after rising first and then decreasing.The price of chilli vegetables is stable and unchanged.

Table 3. Pricing of each vegetable category in the next week Unit: yuan

Date	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Aquatic rhizomes	2.00	2.00	2.00	8.06	6.56	10.71	9.54
Cauliflower	20.21	20.04	20.48	19.56	19.63	19.62	21.47
Edible mushrooms	29.35	8.77	8.64	28.99	30.93	18.99	22.90
Eggplant	15.16	15.90	8.99	5.18	10.56	12.39	2.00
Anthophyllom	4.63	14.20	18.42	2.40	2.00	3.68	4.71
Chilli	2.00	2.00	2.00	2.00	2.00	2.00	2.00

Table 4 is a comprehensive replenishment volume based on the correlation of pricing decisions and vegetables. The replenishment quantity of aquatic rhizomes is the smallest, cauliflower, eedible mushrooms, eggplant, anthophyllom and chilli are large.The amount of cauliflower replenishment is the most stable.The replenishment quantity of edible mushrooms and chilli vegetables fluctuates greatly.

Table 4. Supplement amount of each vegetable category in the next week Unit: yuan

Date	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Aquatic rhizomes	2.40	5.18	9.44	5.42	1.00	1.00	5.53
Cauliflower	20.02	19.85	21.69	20.28	20.05	26.16	20.67
Edible mushrooms	26.35	11.54	1.00	7.71	12.92	1.00	17.13
Eggplant	19.28	20.95	4.48	10.04	3.89	8.21	14.77
Anthophyllom	4.98	1.00	1.00	8.83	25.55	1.00	23.18
Chilli	8.58	9.63	17.25	11.48	6.16	13.05	27.63

Specific data as above, this paper on the price and sales of more accurate forecast.These methods have certain feasibility in practical application and can provide important reference for super decision.However, it is important to note that in practical application there is the influence of other factors needed to consider,such as market competition, economic environment and so on, to further optimize the prediction results and develop more targeted strategy.

4. Conclusions

In view of the limitation of vegetable replenishment demand and price and other factors, this paper adopts the linear fitting method to reflect the relationship between total sales volume and cost plus pricing. The model application is good and the price of each vegetable category is negatively correlated with the sales volume.This paper further reduces the noise of the data, then it uses LSTM time series model and PSO algorithm to optimize the parameters and analyze the historical sales data.Based on the sales model and cyclical changes, this paper makes a continuous forecast and forecasts the replenishment volume and pricing data while predicting the dynamic changes of the replenishment sales volume.The experimental results show that the LSTM time series model and PSO algorithm well predict the replenishment volume and pricing data, which is of guiding significance for the optimization of business strategy, improving profitability and long-term sustainable development.

References

[1] Zhao Y, Zheng R. Combination of Fruit and Vegetable Storage and Fresh-Keeping with Postharvest Heat Treatment[J]. Journal of Chemistry, 2022, 2022.

[2] Cao Tingting, Zeng Kaifang, Deng Lili. Progress of application of light technology in storage and preservation of fruits and vegetables [J]. Food and fermentation industry, 48(10): 314-320.

[3] Cao Yu, Xu Rongliang, Hu Hanli, et al. Research on the joint replenishment strategy of vending machines in unmanned retail industry [J]. Chinese management science, 2021, 29(11): 179-190.

[4] Kiil K, Dreyer H C, Hvolby H H, et al. Sustainable food supply chains: the impact of automatic replenishment in grocery stores[J]. Production Planning & Control, 2018, 29(2): 106-116.

[5] Wang F, Lin L. Spare parts supply chain network modeling based on a novel scale-free network and replenishment path optimization with Q learning[J]. Computers & Industrial Engineering, 2021, 157: 107312.

[6] Sharma S, Abouee-Mehrizi H, Sartor G. Inventory management under storage and order restrictions[J]. Production and Operations Management, 2020, 29(1): 101-117.

[7] Chen W N, Tan D Z. Set-based discrete particle swarm optimization and its applications: a survey[J]. Frontiers of Computer Science, 2018, 12: 203-216.

[8] Feng J, Li C. PSO-optimization method of joint replenishment problem with fuzzy requirement[J]. Jisuanji Gongcheng yu Yingyong(Computer Engineering and Applications), 2012, 48(35): 238-242.

[9] Bai X , Jia J , Wei Q ,et al.Association rule mining algorithm based on Spark for pesticide transaction data analyses[J].International Journal of Agricultural and Biological Engineering, 2019(5).

[10] Zhang R, Chen W, Hsu T C, et al. ANG: a combination of Apriori and graph computing techniques for frequent itemsets mining[J]. The Journal of Supercomputing, 2019, 75: 646-661.

[11] Amaral J V, Guerreiro R. Factors explaining a cost-based pricing essence[J]. Journal of Business & Industrial Marketing, 2019, 34(8): 1850-1865.

[12] Ren X, Liu S, Yu X, et al. A method for state-of-charge estimation of lithium-ion batteries based on PSO-LSTM[J]. Energy, 2021, 234: 121236.