Machine Learning-Based Food Price Analysis

1st Xuhui Zheng, 2nd Xiao Han*
295851@whut.cn, 593739052@qq.com

School of Management Wuhan University of Technology, Wuhan, 430080, China,

Abstract— In this paper, for a variety of food unit prices over a period of more than a year in the same region, the same kind of food with the same trend is studied by SOM neural network clustering algorithm, and the price curve is calculated smoothly by three indices to forecast food prices and trends. Food data comes from national data websites. The results show that there are recognizable categories of food products because of the policy, season and environmental factors received, and the same price trend of food products in the same category is predicted on the basis of which the prices of the next stage are predicted.

Keywords- SOM; Exponential smoothing; Trend forecasting; Food prices

1 Introduction

As the saying goes, people take food as their heaven, and food is an indispensable part of life, so the price of food is more closely related to everyone's life. The monitoring of retail food prices for urban residents is, on the one hand, a reflection of changes in CPI as a component of the CPI, and at the same time it can better represent market information closely related to the lives of residents. Changes in retail food prices for urban residents are influenced by many factors. Environmental factors and financial factors play an significant role. The relationship between changes and other economic issues has great unknowns and complexity. CPI is lagging data, while the retail price of food is observed in real time, with better timeliness. It has good practical significance for the prediction of its changes, which is helpful for timely detection of signs of economic problems, and then timely adjustment of market economic policies.

2 A Review of The Literature

As an important index to reflect market conditions and find economic problems, the retail price of food for urban residents has been studied by scholars in relevant fields. Professor Wang Xiahua (2012) has used the real value forecast and monthly mean forecast model to forecast food prices [1]. Wei Runrun (2016) pointed out that CPI and other data are more significant seasonal impact, and CPI index is relative to PPI lag, only studying CPI can’t reach the actual forecast of timeliness [2]. This study differs from previous scholars in that, under the conditions of USSS Modeler, Python, MATLAB and other tools, SOM and neural network-based clustering analysis are used

ICEMME 2022, November 18-20, Nanjing, People's Republic of China

DOI 10.4108/eai.18-11-2022.2326738

to try to process commodity data with time series from the perspective of clustering analysis [3]. Reference author Zhu Gemiao used SOM clustering method in empirical analysis based on consumer price index to cluster the consumption index situation between different provinces, make corresponding promotion, cluster for different kinds of food price changes in a certain period of time, pay attention to the similarity and synchronization of data between classes [4]. Reference Miao Kaichao points out in the research on agricultural product price prediction based on index smoothing model that agricultural product prices still show some regularity over time, and exponential smoothing method is a mathematically mature random time series prediction method, which has been widely used in health, economy and other fields. And food shows seasonality over time, this paper tries to combine exponential smoothing method into food price forecast with time series, taking into account seasonal factors to make price curve smoothing and prediction for time series [5]. Draw different conclusions from other scholars.

The first part of this article discusses the project background, research significance, research status and innovations of this article. In the second part, the data preprocessing, SOM neural network clustering algorithm and cubic exponential smoothing curve research methods are described. The third part is empirical research and analysis, which analyzes the results of the model, summarizes the trend and characteristics of the data curve, and analyzes the influencing factors. The fourth part is the conclusions and recommendations of the research. The research in this article can provide references for involved government departments, consumers, and food sellers to upgrade the ability to control future food price trends, understand market conditions and economic phenomena, provide theoretical support for decision-making, and enrich related theoretical research.

3 Research Methods

3.1 Sample selection and variable definition

The sample data used in this article is the price data of 42 types of urban residents' food in a specified period of time provided by the government price department. The time period of the data is 1 year, the time interval is 10 days, there are 42 evaluation objects and 39 time points of evaluation indicators.

According to the type of food contained in the data center set the corresponding m evaluation indicators, and according to the food price time series set n evaluation objects, the original data matrix is:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \cdots & \cdots & \cdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \tag{1}$$

Table 1 The Variable Definition

The name of the variable	The variable description
X	Evaluation indicator matrix
x_{ij}	The ith indicator of the jth evaluation method
r_{ij}	Elements in the standardized matrix

The data is processed as follows:

$$r_{ij} = \left(x_{ij} - \min_j(x_{ij})\right) \Big/ \left(\max_j(x_{ij}) - \min_j(x_{ij})\right) \tag{2}$$

This treatment is intended to minimize the magnitude gap between different food prices and to standardize the rate of change between different food prices relative to their own prices.

3.2 SOM clustering algorithm based on neural network

Self-Organizing Maps (SOM), which enables unsupervised learning clustering of data. Its idea is a neural network with only the input layer -- the hidden layer. A node in the hidden layer represents a class that needs to be clustered. Training takes the form of "competitive learning", where each input sample finds a node in the hidden layer that best matches it, called its activation node, also known as "winning neuron". The parameters of the activation node are then updated with a random gradient drop method. At the same time, the points adjacent to the activation node update the parameters appropriately based on their proximity to the activation node. The primary purpose of an SOM network is to convert input from any dimension into a discrete mapping of one or two dimensions, and to adapt the process in a topologically ordered manner. The learning algorithm used by SOM network is unsupervised clustering, which maps the input of any pattern to a one-dimensional or two-dimensional discrete graph at the output layer and keeps its topology intact. The learning process is divided into three main processes:

- Competition: For each input mode, neurons in the network calculate the values of their respective discrimination functions. This differential function provides a basis for competition between neurons. A specific neuron with the maximum value of the discrimination function becomes the winner.

- Cooperation: The winning neuron determines the spatial location of the topological neighborhood of the excited neuron, thus providing the basis for the cooperation of adjacent neurons.

- Weight adjustment: Excited neurons increase the value of the judgment function about the input mode by adjusting their synaptic weights appropriately, thus enabling the neuron to have an enhanced response to similar inputs later.

3.3 Smooth prediction of the third order index

On the basis of clustering, calculate the average price in the cluster, form a 1 x n price standard value matrix, and then think of this matrix as the price standard value of this cluster over time,

the time interval is consistent to 10 days, then the value of the time argument corresponding to the nth data can be set to an equally self-added array of n, then set T_i (i =0, 1, ..., m).

The three-time exponential smoothing algorithm can predict time series that contain both trends and seasonality, and is based on the one-time exponential and secondary exponential smoothing algorithms.

$$S_i = \alpha x_i + (1-\alpha)(S_{i-1} + t_{i-1}) \tag{3}$$

$$t_i = \beta(S_i - S_{i-1}) + (1-\beta)t_{i-1} \tag{4}$$

$$p_i = \gamma(x_i - S_i) + (1-\gamma)p_{i-k} \tag{5}$$

The α is the smoothing parameter, Si is the smooth value of the previous i data, the value is [0, 1], the α is more connected 1, the smoother the value is closer to the data value of the current time, the less smooth the data, the closer the α to 0, the smoother the value is closer to the smooth value of the previous i data, the smoother the data. α values can usually be tried several times more to get the best results. Secondary exponential smoothing preserves trend information, so that the predicted time series can contain trends from previous data. Secondary smoothing represents the trend after smoothing by adding a new variable t. Three-time exponential smoothing retains seasonal information on the basis of secondary exponential smoothing, which is a predictable time series with seasonality. Three exponential smoothing adds a new parameter p to represent the trend after smoothing. The values of the γ, α, β are between [0, 1], and can be tested several times to achieve the best results.

The selection of the initial values of S, t, p is not particularly large for the algorithm as a whole, the usual value is $S_0 = x_0$, $t_0 = x_1 - x_0$, when added up p = 0.

4 Empirical Research and Analysis

4.1 Analysis of SOM neural network clustering results

Before SOM processing, consult the existing documents to learn that China's salt received strict control [6], and ordinary food sales channels and related regulations are not the same, can be considered salt is a special category of food, its price can be considered in the macro-policy situation, can remain stable.

Table 2 Clustering results

classification	Included food
1	Live chicken, eggs, tofu, sugar, brown sugar
2	white sugar
3	Rapeseed oil, soybean oil, peanut oil, soybean blend oil, two specifications of fresh lamb, reeds, fresh milk
4	Fresh pork, celery, cabbage, garlic moss

5	leeks, apples
6	carp
7	Fresh beef, grass fish
8	Cucumbers, tomatoes, potatoes, beans
9	Cabbage, rape, radish
10	Bring fish, carrots, green peppers, peppers, bananas
11	Chicken, eggplant, watermelon, soy sauce, vinegar
12	Serve salt

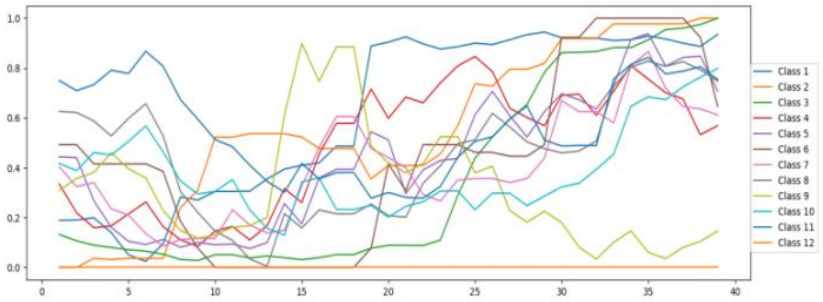

Figure 1 Average standardized curve stacking of 12 categories of food

The price level of 42 kinds of commodities was analyzed by SOM network clustering, when these fruits were divided into twelve categories, the commodities with similar price curves were divided into one category, and the performance of the clustering indexes was better.

- First, there was a slow downward trend, with prices rising steadily after the same turning point, only to fall slightly in the final stages, and overall price volatility was high.
- Similar to the first feature, the price fluctuates greatly, but is slower in the rising phase and does not end up in the descending phase
- Prices show a steady and slow upward trend, accelerating after the same critical time and finally remaining relatively stable
- Prices remain stable, remain constant for a long time, rise by a small margin, and remain stable after price increases.

Overall, from this curved trend, the vast majority of food as a whole shows an upward trend, a small portion of food may appear a slight decline in prices. The remaining small amount of food prices will remain unchanged. And most of the end of the curve will have a slight downward trend, but overall compared to 12 months ago, food prices are up.

4.2 Analysis of third-order exponential smoothing prediction

Since the price of salt for food category 12 has remained stable throughout the year, it can be presumed that their forecasts should remain as steady as the original prices. Three exponentially smoothed forecasts of the prices of the remaining 11 categories of food over time are given, with

price forecasts for the next three (10th, 20th, 30th days). When the α value is 0.3, the β value is 0.3, and the γ value is 0.5, the chart shows as follows:

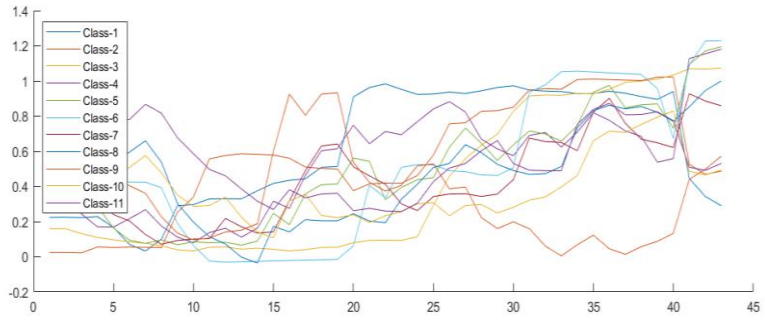

Figure 2 Smooth curves for the top 11 categories of food predict the accumulation chart

The forecast prices for the next three phases of the corresponding food are as follows:

Table 3 Clustering Results

Food name	**1st**	**2nd**	**3th**	**Food name**	**1st**	**2nd**	**3th**
Rapeseed oil	6.03	6.02	6.03	Tomato	3.43	3.60	3.70
Soybean oil	5.69	5.68	5.69	Potato	2.56	2.69	2.77
peanut oil	118.5	118.4	118.6	Carrot	1.95	1.93	1.95
Soybean blend oil	63.67	63.64	63.74	Green pepper	3.07	3.00	3.09
Fresh pork	14.58	14.68	14.77	Pepper	5.00	4.93	5.03
Fresh beef	18.48	18.36	18.28	Cabbage	2.01	2.04	2.07
Fresh lamb	21.94	21.93	21.97	Beans	6.38	6.77	6.99
Fresh lamb	19.55	19.54	19.57	Garlic moss	8.23	8.33	8.43
Live chicken	10.24	9.91	9.73	Leek	3.79	3.92	3.96
Chicken	8.32	8.28	8.38	Reeds	3.88	3.87	3.88
Egg	3.97	3.81	3.72	Apple	5.88	6.03	6.08
Hairtail	10.63	10.52	10.67	Banana	2.87	2.85	2.88
Grass carp	6.27	6.20	6.16	Watermelon	2.04	2.00	2.08
Bean curd	2.19	2.18	2.17	Serve salt	1.30	1.30	1.30
Chinese cabbage	1.27	1.35	1.45	Sugar	7.78	7.59	7.48
Rapeseed	2.60	2.77	2.99	White sugar	4.79	4.67	4.71
Cucumber	3.07	3.28	3.40	Brown sugar	5.69	5.52	5.43
Turnip	1.04	1.08	1.13	Soy sauce	5.21	5.20	5.21
Eggplant	3.20	3.15	3.26	Vinegar	3.50	3.48	3.51
Fresh milk	3.51	3.51	3.51	Carp	6.04	6.13	6.13
Sugar	7.78	7.59	7.48	Celery	4.03	4.08	4.13

In this exponential smoothing method prediction, the setting α is 0.3, β is 0.3, γ is 0.6, and the time period is half a year both 18 points in time. The price trend can be predicted by combining the recent price changes with the food situation. The observation curve can be easily found:

- The prices of the first and the second categories of food will be reduced at the next three points in time, and the reduction is larger, resulting in a larger reduction in the main factors for the seasonal cycle. The price trend of these two categories of food in each six months are a substantial increase and then a moderate decline, so the forecast results are more in line with the actual situation.

- The prices of the third, eighth and seventh categories of food will rise slightly in the future. Looking at its price trends, it can be seen that the prices of these three kinds food are a cycle of change every six or three months. The trend of change is now downwards, followed by an increase. In the end, the overall trend is upward. The predicted point in time is the second month of each forecast period (6 months in the cycle), and the forecast trend is in line with the change in the cycle.

- The prices of food products in categories 4, 5, 6 and 9 will rise substantially in the future.

- The prices of categories 10 and 11 will fall to varying degrees. Obvious periodic characteristics cannot be seen from these two types of price curves, but prices have recently reached their highest point and there is a clear descending trend.

- Salt is a special class and its price will remain constant.

4.3 Analysis of the effects of the seasons

By analyzing foods such as eggplants, tomatoes, potatoes, carrots, green peppers, peppers, beans and watermelons, we found a characteristic feature: prices fluctuate with the seasons. Eggplants, tomatoes, potatoes, green peppers, peppers, beans and watermelons are suitable for sale in summer, while carrots are suitable for autumn and winter sales. These fruits and vegetables are highly seasonal.

4.4 Analysis of the impact of national policies

The price of salt is strictly controlled by the government, so it can remain stable at all times, as long as government policy remains consistent.

4.5 Analysis of reality factors influence

There are only three types of food in the third category: cabbage, radish and rape. As can be seen from the figure, the characteristic feature of both food is that their prices are very volatile and unstable. References show that the main reasons for the price instability of these three kinds of food are as follows:

- Chinese cabbage and radish: 1. Frequent price changes, crop farmers at a loss;2. Frequent natural disasters, extensive management, decreased cabbage quality and yield;3. There are many sales links, and the yards are added steadily.

- Rape: 4. Affected by the decline in total rapeseed production and imports;5. Affected by the development of biodiesel production;6. Affected by panic buying by oil production enterprises.

The above practical factors are the causes of the price fluctuations of these two types of food, but also explain the commonality of these two types of food, proving that the classification has reference.

4.6 Analysis of impact of food supply chain and public demand

- Edible oils: including rapeseed oil, soybean oil, peanut oil, etc. These foods are mainly processed and sold by some large food companies on the domestic market. These companies have a certain ability to adjust, in the supply of raw materials and inventory without much change, prices will be relatively steady.

- Meat, poultry and eggs: With the current development of the breeding industry in the direction of large-scale, corporatization, and grouping, the supply of such foods will be continuously constant. Their fluctuations with time and seasons are small.

- Non-staple food: This type of food primarily includes some secondary processed foods such as sugar, soy sauce, and salt. Such foods are often necessities of life, so the demand is consistent and the price will not change much.

5 Conclusion and Suggestion

This paper analyzes the prices of 42 commodities in a single region over time in 13 months. SOM clustering analysis is used for standardized data, and the next price forecast is made for each category of analysis results. It is found that 42 kinds of food products show very evident differences and analogous phenomena in price curve, and make rational guess and analysis of this phenomenon combined with the reality. Food occupies the largest proportion of the CPI: 34.00% [7], which shows that the retail price of food for urban residents has an important impact on the results of the CPI. Being able to obtain market information reflected in food prices in a timely and correct manner is of great significance for keeping these food prices that have an important impact on the living standards of residents at a reasonable level of change.

Acknowledgment. It should be clarified that I would like to thank Mr. Han Xiao for his guidance on this paper. This paper is the research result of a national innovation and entrepreneurship training program for college students project under the guidance of teacher Han Xiao. The project number is S202110497051

References

[1] Wang Xiahua. Urban Residents' Food Retail Price Predictions and Recommendations. [J] Modern Trade Industry, 2012, 24 (17): 68-70

[2] Wei Runrun. Study on the Relationship between CPI and PPI Based on Seasonal Adjustment and Classification. [D] Jinan University, 2016.

[3] Yang Zhanhua, Yang Yan. Research and progress of SOM neural network algorithms. [J] Computer Engineering, 2006.
[4] Jug Miao. Empirical analysis based on the consumer price index. [D] Nanjing Normal University, 2018.
[5] Miao Kaichao. Research on agricultural price forecast based on exponential smoothing model. [D] Hefei University of Technology, 2009.
[6] Ni Yuzhu, Xu Yuyun. Research on the Construction of China's Salt Price Stability Mechanism. [J] Commercial Economic Research, 2017 (12): 20-23.
[7] Zhang Xuehui. Discussion on the composition of China's CPI. [J] Value Engineering, 2010, 29 (02): 130-131.