ISSN 2616-5775 Vol. 4, Issue 6: 86-96, DOI: 10.25236/AJCIS.2021.040615

Garlic price forecast based on the combined model of time-frequency decomposition and neural network

Yuliang Feng[1, *]

[1]Department of Computer Science, Shandong Agricultural University, Tai'an, China
**Corresponding author*

Abstract: *China's garlic production accounts for more than 70% of the world's total output, and it is one of the most profitable agricultural products exported by China. Since 2003, garlic prices have fluctuated sharply and frequently, which has brought severe challenges to the healthy and sustainable development of China's garlic industry. Based on the analysis and comparison of various price prediction models, this paper constructs an EEMD-GRU combined prediction model. The model first uses the EEMD decomposition method to decompose the garlic price series on multiple time scales, and obtains 12 IMF eigenmode functions, which are classified and integrated to obtain high-frequency, intermediate and low-frequency terms. The analysis found that the low-frequency and intermediate-frequency items have a high contribution rate to the garlic price, and their contribution rates add up to 94%. Import the reconstructed garlic price sequence into the GRU model, select 50 training times and 30 hidden neurons to predict the garlic price, and obtain the prediction results. The prediction results obtained by the EEMD-GRU combined model are combined with ARIMA and ARIMA through evaluation indicators such as MSE. Comparing the prediction results obtained by ARIMA-SVR and LSTM models, it is found that the EEMD-GRU combined model has the highest trend prediction accuracy, the smallest prediction error, and the prediction effect is significantly better than the ARIMA, ARIMA-SVR and LSTM models. This research can help garlic-related industries to make a reasonable market allocation.*

Keywords: *garlic industry big data, price fluctuation characteristics, multiple time scales, price forecasting model, price forecasting system*

1. Introduction

China is the world's largest producer and consumer of garlic. Garlic has become a pillar industry for farmers in the main garlic producing areas to increase their incomes and become rich. However, after 2000, due to the increase in garlic planting area and the influence of natural disasters and public opinion speculation, garlic prices fluctuated more and more significantly, and large fluctuations such as "garlic you are cruel" and "garlic you are cheap" appeared, which directly affected With the benefits of garlic industry practitioners, garlic farmers and garlic practitioners "talk about the price". Predicting garlic prices and helping industry stakeholders make scientific decisions are very important for stabilizing the garlic industry.

Early agricultural product price forecasts were mostly time series models, such as ARMA and the ARIMA model derived later. Liu Feng et al. used it to predict the monthly price data of cabbage, and the results showed that the ARIMA (0, 1, 1) model had a better prediction effect [1]; Hu Yang et al. also used the ARIMA (0, 1, 1) model to predict Nanjing City The price of green peppers has achieved better results [2]. With the development of technology, machine learning has shown greater potential in price prediction. Ji Dawei took China's commodity prices as the research target and used the improved BP network to predict prices. The results showed that the improved BP neural network had a small gap between the predicted value and the actual value, and it had better results [3].

Due to the limitations of the model itself, a single prediction model often fails to obtain the desired effect of price prediction. The combined model can meet the requirements for further improvement of prediction accuracy by integrating the advantages of multiple models. Zheng Wei combined the seasonal index adjustment method with the HGWO-SVR algorithm to predict the monthly price index of China's agricultural product wholesale. The results show that the combined model has a good effect [4]; Cao Shuang and others use the overall price of agricultural products as a research Object, constructed a SVM-ARI-MA agricultural product price forecasting model based on wavelet decomposition, extracted four changes in trends, and found that the combined forecasting model has a higher forecasting effect than the traditional forecasting model [5]; Yu Weijia used cucumber prices as research Objects, the LASSO regression model is introduced to screen the affected prices, and a combined model (L-BPNN) combining the Lasso regression method and the BP neural network is constructed. The results show that the effect of this model is better than the Lasso regression model, the BP neural network model and the RBF neural network. Network model [6]. Based on the research results of the combination model, it is found that the combination model has a significant advantage in the price prediction of agricultural products.

According to domestic and foreign literature, the forecasting effect of machine learning is often stronger than that of traditional time series models, and the effect of combined models is better than that of single models. The forecasting method of the model is mostly direct forecasting of agricultural product prices, with less processing of redundant items in agricultural product prices, and fewer forecasting models for daily prices.

2. Method

2.1. Ensemble empirical mode decomposition method

Empirical Mode Decomposition (EMD) is a time-frequency decomposition method that can be used to process non-stationary time series, while Ensemble Empirical Mode Decomposition (EEMD) is based on empirical mode decomposition by inserting Gaussian white noise sequences, which effectively suppresses the edge effect and scale mixing phenomenon in the EMD method. The EEMD method keeps the components of the final decomposed IMF (Intrinsic Mode Function) physically unique. Since the EEMD method does not need to set any basis functions in advance, and the added white noise amplitude has little effect on the final result, this method has better adaptability.

The basic steps of EEMD are as follows:

First, add a new set of white noise sequence to the original signal, and get the new signal sequence as $X(t)$, as in formula (1):

$$X(t) = x(t) + \omega(t) X(t) = x(t) + \omega(t) \tag{1}$$

Secondly, perform EMD decomposition to obtain a set of IMF components, as shown in formula (2):

$$X(t) = \sum n{-}1_{j=1} IMF_{j(t)} + r(t) X(t) = \sum j = 1n - 1 IMF_{j(t)} + r(t) \tag{2}$$

Where: $IMF_{j(t)}$ represents the j-th IMF component; $r(t)$ is the trend item.

Then repeat the first two steps m times for the original signal, adding different white noise each time to obtain a new set of IMF components, as shown in formula (3):

$$X_i(t) = \sum n{-}1_{j=1} IMF_{j(\ )} + r(t) X(t) = \sum j = 1n\ IMF_{j(i)t} +_i \tag{3}$$

The IMF components obtained each time are integrated and averaged so that the added white noise cancels each other out as the final result. After the above 4 steps, the eigenmode function on each time scale can be obtained.

2.2. Recurrent Neural Network

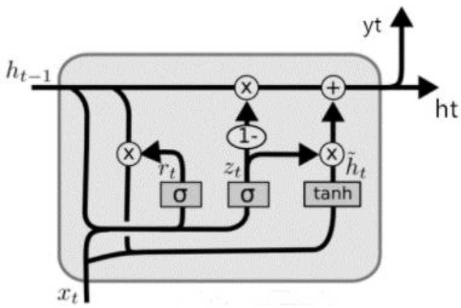

Figure 1: GRU structure diagram

Recurrent Neural Network (RNN) is a type of neural network characterized by various cycles in the network structure. Through connection, RNN can store the relationship between the input at the current

ISSN 2616-5775 Vol. 4, Issue 6: 86-96, DOI: 10.25236/AJCIS.2021.040615

moment and the output at the previous moment. GRU is a variant of RNN network. GRU can solve the long dependency problem in the RNN network. There are two doors in the GRU model: the update door and the reset door. The specific structure is shown in Figure 1.

z_t and r_t respectively indicate the update gate and reset gate. The update gate is used to control the extent to which the previous state information is brought into the current state. The larger the value of the update gate, the more state information from the previous moment is brought in. The reset gate controls how much information from the previous state is written to the current candidate set. The smaller the reset gate, the less information is written in the previous state.

GRU training process:

The parameters to be learned are W_r, W_z, W_h, W_o. Among them, the first three parameters are all spliced (the vectors in the subsequent calculation process are also spliced), so they need to be separated separately during the training process:

$$W_r = W_{rx} + W_{rh} \tag{4}$$

$$W_z = W_{zx} + W_{zh} \tag{5}$$

$$W_{\tilde{h}} = W_{\tilde{h}x} + W_{\tilde{h}h} \tag{6}$$

Input to the output layer:

$$y_t^i = W_o h \tag{7}$$

The output of the output layer:

$$y_t^o = \sigma(y_t^i) \tag{8}$$

After getting the final output, you can write the loss of network transmission. The loss of a single sample at a certain moment is:

$$E_t = \frac{1}{2}(y_d - y_t^o)^2 \tag{9}$$

Then the loss of a single sample at all moments is:

$$E = \sum_{t=1}^{T} E_t \tag{10}$$

After calculating the partial derivatives for each parameter, the parameters can be updated and iterated in sequence until the loss converges. As a recurrent neural network model, GRU uses various gate functions to retain important features, so the features will not be forgotten during long-term propagation.

2.3. Experimental data

The experimental data for price prediction is the price of garlic in China from May 22, 2003 to June 14, 2021. The data comes from the garlic industry chain big data platform, Garlic World and the International Garlic Trade Network.

2.4. Evaluation Index

The generalization performance of the model is the core key index for evaluating the prediction model, and the prediction accuracy is the numerical index for the generalization performance of the model. For this reason, the prediction error ratio (ErrorRate) and the prediction accuracy rate (AccuracyRate) are selected as evaluation indicators.

The prediction error ratio is a commonly used judgment index, which is defined as the difference between the actual output of the neural network and the expected output divided by the value of the expected output. The smaller the value, the smaller the prediction error and the generalization performance of the model. The stronger. The formula is as follows:

ISSN 2616-5775 Vol. 4, Issue 6: 86-96, DOI: 10.25236/AJCIS.2021.040615

$$ErrorRate = \frac{|\, price_{prediction} - price_{real} \,|}{|\, price_{real} \,|} \tag{11}$$

$price_{prediction}$ is the actual output value of the neural network and $price_{real}$ is the expected output value.

The accuracy rate of forecast rise and fall is to compare the rise and fall predicted by the neural network with the expected rise and fall. If the prediction and the expected mark are the same, the prediction is correct. The formula is as follows:

$$AccuracyRate = \frac{n_{correct}}{n} \tag{12}$$

n is the total number of prediction samples, and $n_{correct}$ is the number of samples that are predicted correctly.

The calculation formulas of MSE, MAE and RMSE are as follows:

$$MAE = \frac{1}{N}\sum_{t=1}^{N} |\, y_t - \hat{y}_t \,| \tag{13}$$

$$MAPE = \frac{1}{N}\sum_{t=1}^{N} \frac{|\, y_t - \hat{y}_t \,|}{y_t} \tag{14}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N} (y_t - \hat{y}_t)^2} \tag{15}$$

Among them, y_t and $\hat{y}_t$ (t=1, 2, 3...N) respectively represent the actual value and the predicted value at time t, and N is the total number of training samples.

3. Garlic price prediction based on time-frequency decomposition and neural network

3.1. Overview of garlic price fluctuations

First, the garlic price process from 2003 to 2021 can be divided into 5 cycles through the peak and trough method. The five cycles are May 2003-May 2009, June 2009-June 2011, and July 2011. Month-July 2013, August 2013-May 2018, June 2018 to May 2021 (see Figure 2).

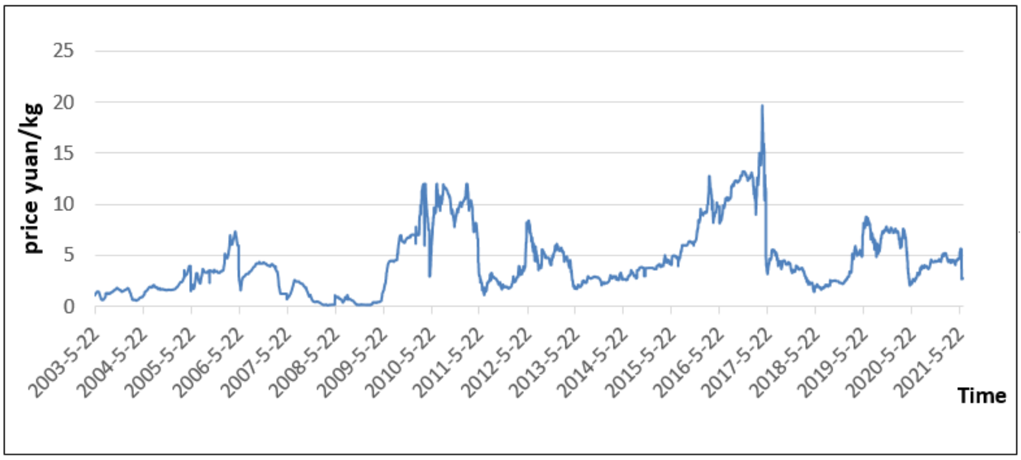

Figure 2: Jinxiang garlic price from 2003 to 2021

The first cycle: From May 2003 to May 2009, garlic prices rose first and then fell. Since May 2003, the price of garlic slowly increased to the peak price (7 yuan/kg) in May 2006, and then the price slowly dropped to the freezing point of May 2008.

ISSN 2616-5775 Vol. 4, Issue 6: 86-96, DOI: 10.25236/AJCIS.2021.040615

The second cycle: June 2009-June 2011, during which garlic prices fluctuated greatly. The price of garlic rose rapidly from less than 2 yuan/kg in June 2009 to 5 yuan/kg in May 2010, and then fell to 2 yuan/kg in June 2011 after a high level of shock.

The third cycle: From July 2011 to July 2013, the price of garlic first rose and then fell, and there was a small price peak. It started to rise from July 2011 to 4 yuan/kg in May 2012, and then the price continued to fall to 1.5 yuan/kg in July 2013.

The fourth cycle: from August 2013 to May 2018, the price of garlic fluctuated to its highest point, and then fell off a cliff. The price of garlic has risen all the way from August 2013 to the highest price of 18.6 yuan/kg in May 2017. After the "garlic you are cruel" phenomenon, the price of garlic plummeted and continued to drop to 2.1 yuan/kg.

The fifth cycle: From June 2018 to May 2021, the price slowly rises to the highest point in July 2019 at 8 yuan/kg, after which the price has been stable at around 6-8 yuan/kg.

3.2. The basic idea of constructing EEMD-GRU combined model

Methods such as ARIMA and SVR are difficult to deal with redundant items in garlic price. Based on the advantages of EEMD algorithm and GRU model, a combination model of EEMD-GRU price prediction is proposed. The model prediction step diagram is shown in Figure 3.

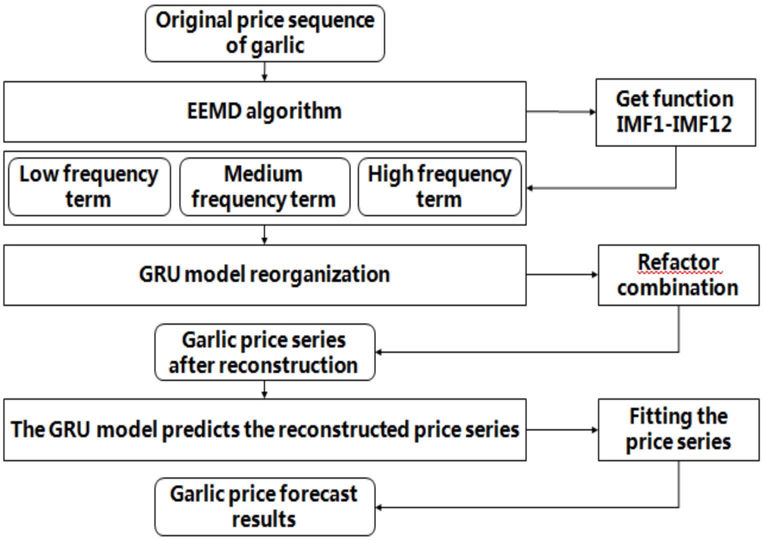

Figure 3: Garlic price forecasting steps

The specific implementation process is:

(1) Input the original price sequence of garlic, and use the EEMD algorithm in python to get the modulus functions IMF1 to IMF12.

(2) Reconstruct and integrate IMF1 to IMF12 to obtain low-frequency, intermediate, and high-frequency items, and calculate the correlation coefficients of low-frequency, intermediate, and high-frequency items with the original price series of garlic and the contribution rate of this item.

(3) Reconstruct the low-frequency, intermediate, and high-frequency items through the correlation coefficient and contribution rate to obtain a new garlic price sequence.

(4) Choose the most suitable training times and the number of neurons, establish a GRU model, and get the price prediction result.

(5) Compare the EEMD-GRU prediction results with the results obtained from the ARIMA, ARIMA-SVR, and LSTM models to verify the performance of the EEMD-GRU model.

3.3. Multi-time scale decomposition of garlic price

Using python software programming to realize the EEMD decomposition of Jinxiang garlic prices from May 22, 2003 to June 14, 2021, where IMF1 to IMF12 are 12 eigenmode functions with frequency

ISSN 2616-5775 Vol. 4, Issue 6: 86-96, DOI: 10.25236/AJCIS.2021.040615

from high to low, and the results are shown in Figure 4. Show.

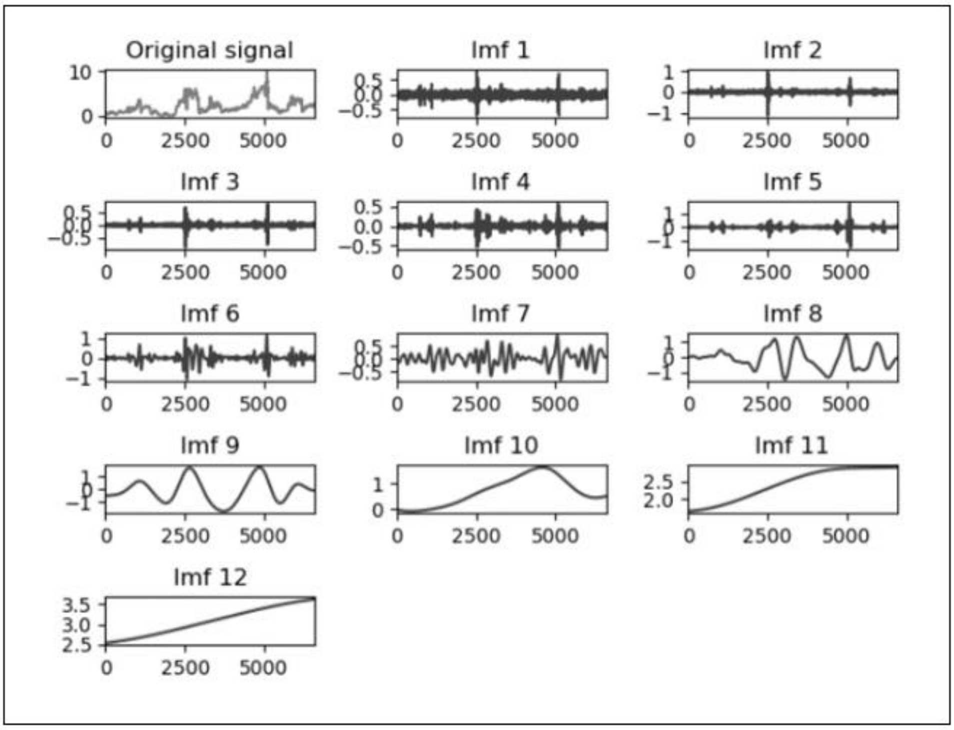

Figure 4: EEMD decomposition result diagram of garlic price

Use SPSS2.0 software to obtain three categories of items from IMF1 to IMF2 according to the Pearson correlation coefficient. IMF1, IMF2, IMF3, and IMF4 are classified into one category, namely high-frequency items, and IMF5, IMF6, IMF7, and IMF8 are one category , Is an intermediate frequency item, IMF9, IMF10, IMF11, IMF12 are a category, which is a low frequency item.

3.4. Garlic price reconstruction

In order to further analyze the characteristics of each sub-item data and analyze its relationship with the original sequence, according to the trend of high-frequency items, medium-frequency items, and low-frequency items, the correlation degree, period, and price of each pair of garlic between each sub-item and the original sequence are calculated. The actual contribution rate of the sequence is shown in Figure 5 to Figure 7.

High-frequency items (IMF1-IMF4) are short-term fluctuations, which refer to short-term series fluctuations caused by uncertain factors. According to Table 1 and Figure 5, the period of the high-frequency term is very short, about 5 days. The correlation coefficient between the high-frequency term and the initial sequence is 0.0895, which is not statistically significant; the analysis of the original sequence is 5.93%, and the analysis of garlic prices is relatively small. High-frequency items include the uncertainty of natural environment (such as weather changes in a short period of time), market environment and consumer psychological factors, and reflect the impact of short-term noise on garlic prices.

Table 1: Period and contribution rate table of each sub-item

	Correlation coefficient with original sequence	**cycle**	**Contribution rate of this item**
high frequency	0.0895	5 day	5.93%
Medium frequency	0.6412	219 day	42.47%
Low frequency	0.7792	1915 day	51.61%

ISSN 2616-5775 Vol. 4, Issue 6: 86-96, DOI: 10.25236/AJCIS.2021.040615

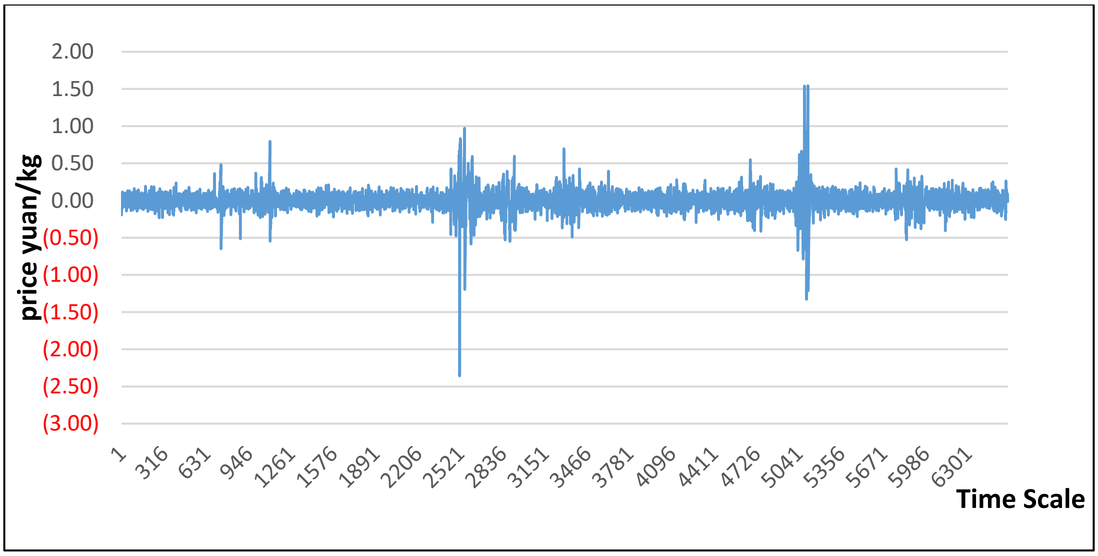

Figure 5: High-frequency component diagram after garlic price reconstruction

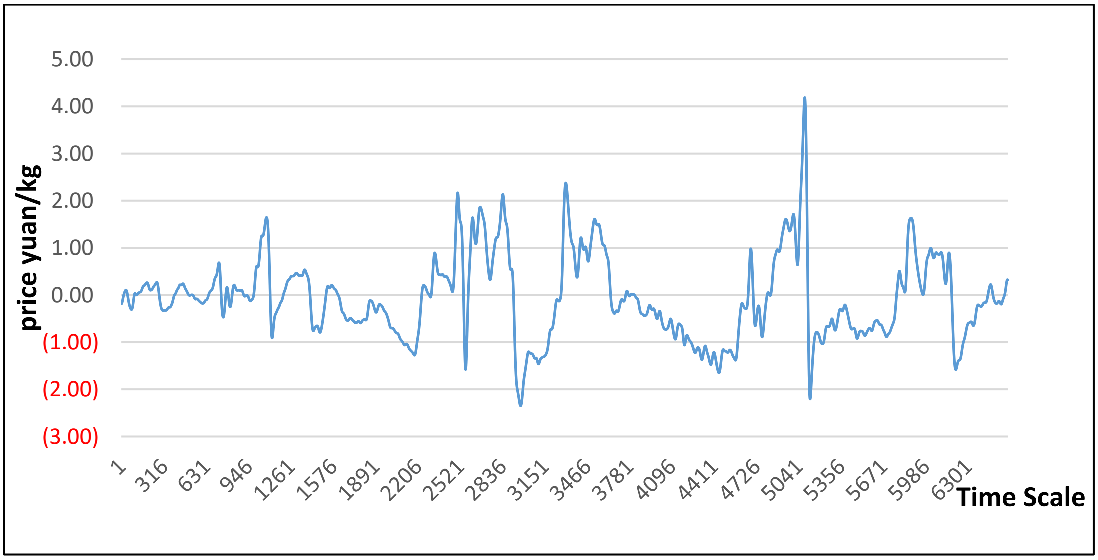

Figure 6: Medium frequency component diagram after garlic price reconstruction

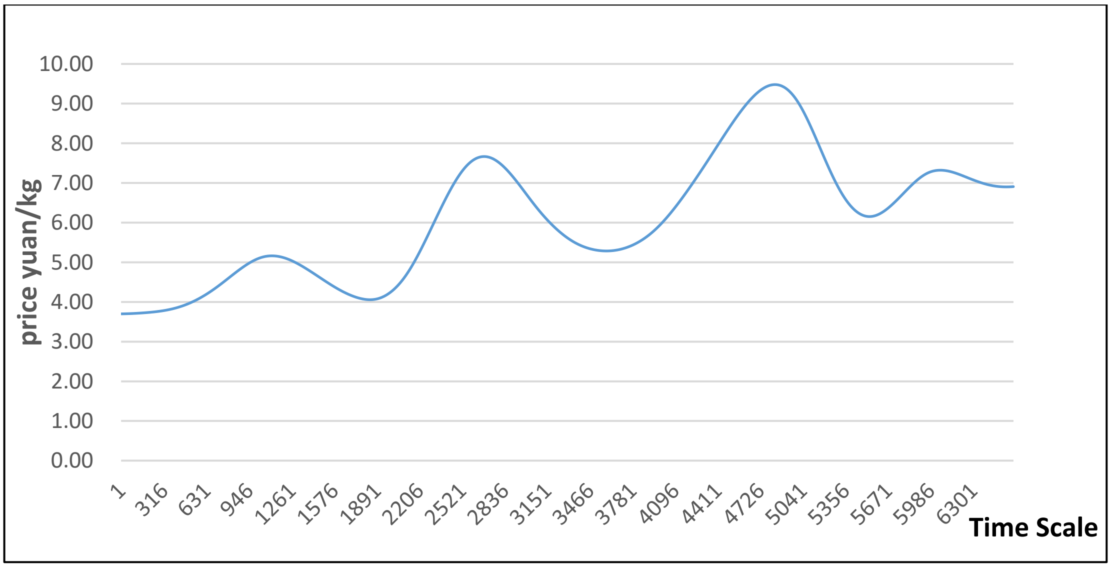

Figure 7: Low-frequency component diagram after garlic price reconstruction

ISSN 2616-5775 Vol. 4, Issue 6: 86-96, DOI: 10.25236/AJCIS.2021.040615

Intermediate frequency items (IMF5-IMF8), that is, major event items, reflect certain major events or garlic price fluctuations under related policies. From Table 1 and Figure 6, the average period is 219 days, and the correlation coefficient between the intermediate frequency term and the initial sequence is 0.6412, which is statistically significant; the analysis power of the initial sequence is 42.47%, which has a higher analysis of garlic prices force. The intermediate frequency item is basically consistent with the overall fluctuation of garlic price, with large and frequent fluctuations, which also reflects that major events will have a greater impact on the fluctuation of garlic prices. For example, in 2016, due to the increase in garlic planting area, the price of garlic also began to rise sharply. It was not until the large-scale launch of new garlic in May 2017 that the price of garlic began to decline.

As shown in Table 1 and Figure 7, the average period of the low-frequency item (IMF9-IMF12) is 1915 days, and the correlation coefficient between the low-frequency item and the original sequence is 0.7792, which is statistically significant; the analysis of the original sequence is 51.61%, and the analysis of the price of garlic Greater intensity. The low-frequency item is the cyclic period of garlic price fluctuation, that is, the long-term trend item, and it is also the most important component of garlic price fluctuation. Over time, garlic prices will gradually return to the volatility trend of the trend item.

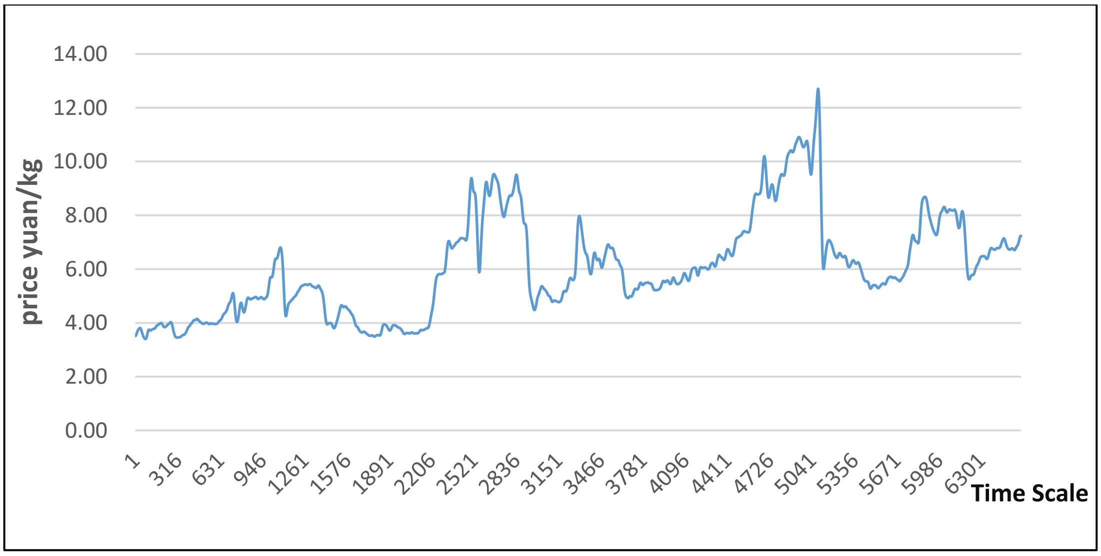

Figure 8: Garlic price series after reconstruction

It can be seen from Table 1 that the contribution rate of the high-frequency item to the garlic price series is 5.93%, and the high-frequency item is the impact of short-term noise on the garlic price, which is a redundant item. Remove the high-frequency items, combine the low-frequency items with the intermediate-frequency items to obtain a new garlic price sequence. The reconstructed garlic price sequence is shown in Figure 8. The reconstructed garlic price series will be imported into the GRU model.

3.5. EEMD-GRU model predicts garlic prices

The price of garlic is reconstructed from the intermediate frequency items and low frequency items obtained by EEMD decomposition to obtain a new price sequence, and the price sequence is imported into the GRU model for training. In order to accurately predict the real model, the number of training times with the smallest training error is selected using the criterion of empirical risk minimization. After comparison, it is found that the effect of model training 50 times is the best. If the number of training times is too high, over-fitting will occur, which will affect the generalization ability of the model. Figure 9 is an effect diagram of 50 times training.

The black line in Figure 9 represents the reconstructed price sequence, and the gray line represents the fitted price sequence of the GRU. The GRU model correspondingly weakens the volatility of garlic prices. As the number of training increases, the fit of the model is correspondingly improved. In order to ensure the generalization ability of the model and avoid over-fitting, this study selects price data with 50 training times for data prediction.

ISSN 2616-5775 Vol. 4, Issue 6: 86-96, DOI: 10.25236/AJCIS.2021.040615

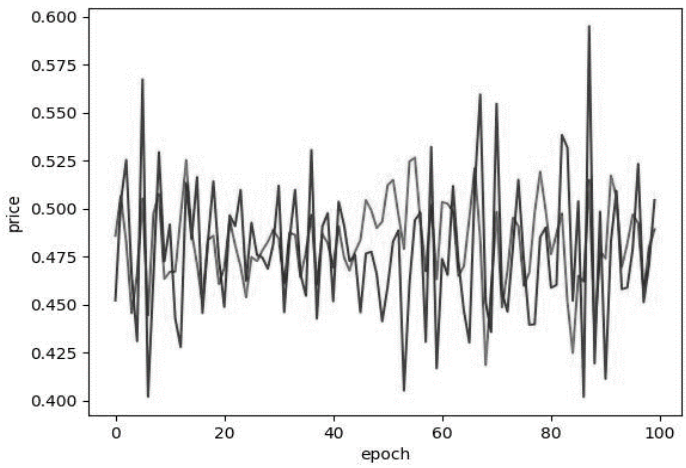

Figure 9: Effect picture after GRU model training 50 times

The prediction performance of the model is also closely related to the number of hidden neurons. For this reason, the same strategy as the number of iterations is used to determine the optimal parameter value. To this end, the prediction effects of 10, 20, 30, and 50 hidden neurons are selected respectively for comparison, and the prediction is judged by the ratio of prediction errors and the accuracy of predictions.

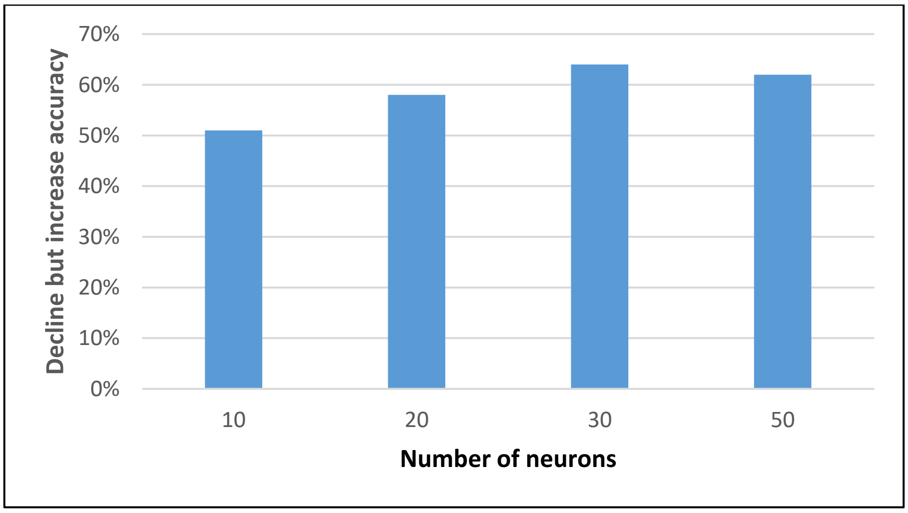

Figure 10: The percentage error of prediction effect of different neurons

It can be seen from Figure 10 that under the condition of 50 training times, the error percentage is the highest when the number of neurons is 10, which is 35%; the error percentage when the number of neurons is 20 is 22%, and the number of neurons is 30 The error percentage of is 11%, the prediction error is the smallest; the error percentage of 50 neurons is 14%, which is higher than the prediction effect of 30 neurons, so too many neurons will not get better Effect. Therefore, 30 is the most appropriate number of neurons.

It can be seen from Figure 11 that under the condition of 50 training times, the accuracy rate of the increase of 10 neurons is 51%; the accuracy of decrease of the number of neurons 20 is 58%; the number of neurons is The accuracy rate of the decline of 30 is 64%; the accuracy of decline of the number of neurons is 62%, and the accuracy of decline of the condition of the number of neurons of 30 has the best prediction effect, because of the time of price prediction The scale is "day", so a prediction effect with an accuracy rate of 60%-70% is a better prediction effect.Based on the judgment of the error percentage and the falling accuracy rate index, this experiment believes that the EEMD-GRU combined prediction model is trained for 50 times and the number of neurons is 30. The prediction effect is the best.14 forecast data are obtained from the forecast, and then the forecast data is restored. The comparison chart of the predicted price and the actual price after the restoration is shown in Figure 12.

ISSN 2616-5775 Vol. 4, Issue 6: 86-96, DOI: 10.25236/AJCIS.2021.040615

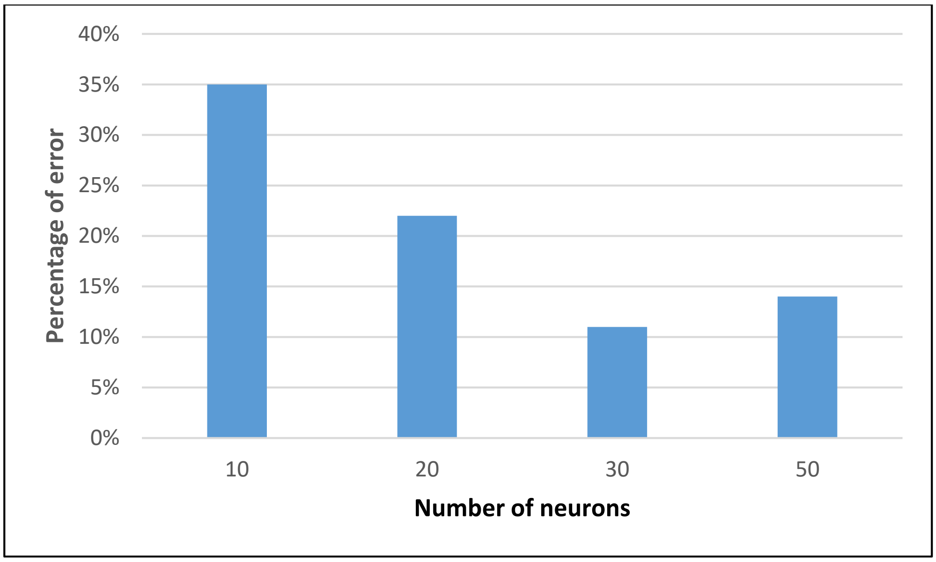

Figure 11: Accuracy rate graph of the fall and rise of different neurons

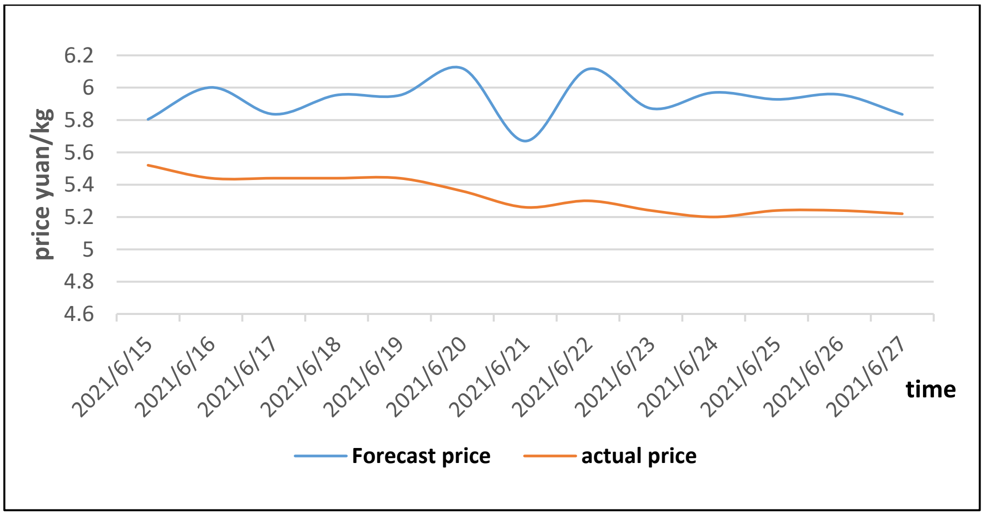

Figure 12: Comparison of predicted price and actual price of EEMD-GRU combined model

It can be seen from Figure 12 that the predicted result of the EEMD-GRU combined model is roughly the same as the actual price. The difference between the predicted price and the actual price is between 0.4-0.7 yuan and the predicted price is generally high. The actual price trend is a slight decline. The price predicted by the combined model is generally around 5.9 yuan/kg, and the actual price has a relatively obvious trough on June 21. The forecast result also has a relatively obvious trough. Experiments show that the combined forecast the prediction effect obtained by the model is close to the actual price fluctuation.

The predicted price fluctuations of the EEMD-GRU combined model are close to the actual garlic price fluctuations. In order to further verify the advantages of the EEMD-GRU combined model, the ARIMA model, the ARIMA-SVR combined model and the LSTM model, which is also a cyclic neural network, are selected as the models to be compared. The comparison table of various indicators is shown in Figure 13.

As shown in Figure 13, the MSE indicators of EEMD-GRU, LSTM, and ARIMA-SVR are basically the same. The lowest MSE indicator is the EEMD-GRU model, which is 0.0010. The MSE indicator of ARIMA is quite different from the first three, which is 0.0026. ; The MAE indicators of EEMD-GRU, LSTM, and ARIMA-SVR are 0.0251, 0.0282, and 0.0285, respectively. The MAE of EEMD-GRU is the lowest; the MAE of ARIMA is the highest at 0.0577, while the RMSE indicators of ARIMA-SVR and EEMD-GRU the values are close, and the indicators of LSTM and ARIMA are far from the former two. Through the results of these three indicators, it can be concluded that the EEMD-GRU prediction model has a more stable prediction effect, and the predicted price fluctuations are smaller than that of the LSTM and ARIMA-SVR algorithms. Therefore, EEMD-GRU is more conducive to the trend forecast of garlic prices, and there will be no excessive fluctuations in the forecast price. In general, the prediction effect of the EEMD-GRU combined model is better than the ARIMA, ARIMA-SVR, and LSTM models.

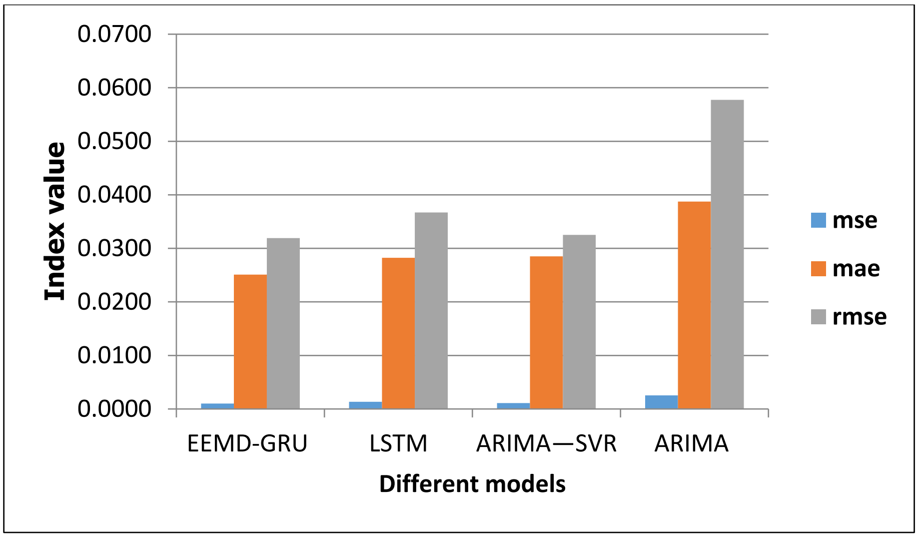

Figure 13: Comparison of MSE, MAE, and RMSE of different combination models

4. Conclusion

Construct an EEMD-GRU combined model to forecast garlic prices. First, use the EEMD method to decompose the garlic price to obtain the high-frequency, intermediate and low-frequency items, among which the high-frequency items have the lowest contribution rate; then reconstruct the garlic price sequence and import it into the GRU model to obtain the prediction results. The results show that the EEMD-GRU combination the accuracy of the model's prediction of garlic price rise and fall is 64%, and the MSE, MAE, and RMSE indicators of the EEMD-GRU model are the best, which means that the prediction error of the model is lower. Compared with the ARIMA, ARIMA-SVR, and LSTM models, the EEMD-GRU model has better prediction effects and is more suitable for Chinese garlic price prediction.

References

[1] Liu Feng, Wang Rujing, Li Chuanxi. The application of ARIMA model in agricultural product price forecasting [J]. journal6, 2009, 45(025):238-239.
[2] Hu Yang, Zhang Chaoyang. Corn price forecast in Hebei Province based on ARIMA model [J]. Agriculture and Technology, 2020, 40(23):4.
[3] Ji Dawei. Construction of market price prediction model based on improved BP network [J]. Automation Technology and Application, 2020, v.39; No.300(06): 62-65.
[4] Zheng Wei, Wang Canqiang, Li Weide. Agricultural product price prediction model based on seasonal index adjustment and HGWO-SVR algorithm [J]. Statistics and Decision, 2018(19): 5.
[5] Cao Shuang, He Yucheng. SVM-ARIMA agricultural product price prediction model based on wavelet decomposition [J]. Statistics and Decision, 2015(13): 92-95.
[6] Yu Weige, Wu Huarui, Peng Cheng. Research on the combined model of vegetable short-term price prediction based on Lasso regression and BP neural network [J]. Smart Agriculture (Chinese and English), 2020, 2(03): 108-117.