Journal of Computers Vol. 34 No. 3, June 2023, pp. 357-370
doi: 10.53106/199115992023063403027

Prediction and Early Warning Methods for Agricultural Commodity Price Based on SSA-LSTM

Dian Zhang, Yi-Qun Wang, Wen-Bai Chen*

School of Automation, Beijing Information Science and Technology University,
Beijing, China

iszhangdian@126.com,{wangyiqun, chenwb}@bistu.edu.cn

Received 1 March 2023; Revised 1 April 2023; Accepted 17 May 2023

Abstract. China is a large agricultural country. Fluctuations in the prices of agricultural products can have a significant impact on the income of farmers. It is also a barometer of the agricultural market. Accurate and effective price forecasting of agricultural products plays an important role in strengthening agricultural informatization. Therefore, it is important to explore the characteristics and laws of agricultural price fluctuations to stabilize agricultural market prices and protect farmers' incomes. This paper takes the price of pork among agricultural products as an example. This paper summarises several key factors that influence pork price fluctuations. Ultimately, this paper uses three pig prices, namely Outer Ternary, Inner Ter-nary and Black pig, and two feed ingredient prices, namely soybean meal, and maize, for a total of five indicators to forecast pork prices. This study uses the Sparrow Search Algorithm (SSA) to optimize the Long Short-Term Memory (LSTM) hyperparameters to enhance the forecasting capability of the LSTM. An early warning mechanism for pork prices was established to warn of pork price fluctuations. The experimental results verified the prediction accuracy of the proposed model and the effectiveness of the early warning mechanism.

Keywords: agricultural commodity prices, SSA-LSTM, early warning mechanisms

1 Introduction

China is a large agricultural country. Fluctuations in the prices of agricultural products can have a significant impact on farmers' incomes and are a barometer of the agricultural market. Accurate and effective price forecasting of agricultural products plays an important role in strengthening agricultural informatization. China is also a major producer and consumer of pigs, which account for around 60% of total meat consumption and stockpiles. The pork industry also plays a pivotal strategic role in China's national economic and social development and in the production and life of urban and rural residents. Based on this, this experiment will take pork prices as an example to study the prediction and early warning of agricultural prices.

In recent years, scholars have continued to study forecasting models related to agricultural prices. Kai Ye et al. used hog prices, corn prices and bean prices as data sets to predict pork prices, demonstrating that feed ingredient prices and hog prices can be used as features to predict pork prices [1]. However, this experimental data was weekly data with a large time span, which may lead to the problem of data discontinuity. Zhang selected 229 weekly average pork prices nationwide from 1 January 2008 to 1 July 2012 as samples for SVM training and testing [2], demonstrating that pork prices can also be predicted using machine learning methods. Wang et al. used Census X12- GM (1,1) combination model to forecast pork prices [3], but only for a single feature, without considering pork price influencing factors, which may have some bias in long-term forecasting. Yan Guo et al. used a mixed model of multiple methods to forecast corn prices [4], which proved that it is feasible to use a combination model to forecast prices. Tang et al. developed a multipole adaptive wavelet recurrent neural network model to predict prices by analyzing the variation pattern of prices [5]. This showed that the combined model had lower information loss compared to a single model. R. Murugesan used five variants of the LSTM to predict agricultural prices [6], demonstrating that it is feasible to predict agricultural prices using LSTM correlation models.

Agricultural early warning studies mainly include the impact of climatic conditions, environmental changes and other factors on agricultural production. Some scholars have used the method of constructing econometric models to analyze the mechanism of the impact of drought on wheat, rice and other food crops for early warning

* Corresponding Author

of food production [7-8]. Other scholars selected average temperature and other early warning indicators, established an early warning indicator system, and constructed an early warning model for agricultural production risk early warning [9-10].

The accuracy of LSTM prediction is closely related to the setting of its weights and thresholds. Bioheuristic algorithms are an effective method for LSTM parameter optimization. Biological heuristics include genetic algorithm (GA) [11], particle swarm optimization (PSO) [12], bat algorithm (BA) [13], grey wolf optimizer (GWO) [14] and sparrow search algorithm (SSA) [15]. The SSA is based on the foraging and anti-predation behaviour of sparrow populations and is used to solve the problem of random variation in model input weights and thresholds. It has high-performance global search capability, good stability and convergence accuracy [16]. SSA algorithm has fast convergence speed and powerful search capability, which overcomes the blindness and uncertainty of traditional neural network model parameter selection. Currently, SSA is widely used in non-linear time series data processing, such as load prediction [17], stock prediction [18], path planning [19], and optimization detection [20]. The sparrow search algorithm has significant advantages in terms of convergence speed and search accuracy, and its simple structure can accurately cope with complex problems.

Therefore, based on the previous research results, this paper combines the sparrow optimization algorithm and long and short-term memory networks. A pork price prediction model based on SSA-LSTM is proposed. The main purpose is to solve the problems of a single LSTM model with long training time, insufficient generalization ability and difficulty in handling large amounts of level data. This model uses the sparrow search algorithm to perform parameter search optimization on the learning rate and the number of neurons in the hidden layer of the long and short-term memory network to reduce the training cost of the LSTM [21]. The prediction results were compared with five other models to verify the validity of the model. Finally, a black warning model was built based on the prediction results to provide early warning on the prediction results and warning on the fluctuation of pork prices. And it provides reliable support for future pork price early warning research.

2 Data and Model Construction

2.1 Data Sources and Pre-processing

The acquisition of pork prices relies heavily on web crawlers, which can also be referred to as web robots. It has a wide area of application and is able to retrieve data efficiently. It crawls the data using the acquired link address, relies on the web link address to read it and then finds other web links. The whole process can be done independently. Crawler technology plays an important role in the current stage of network security and scientific research. [22]

The data used in this study comes from the China Pig website (https://www.zhuwang.cc/). This website has a huge amount of data on pork prices. The data covers the prices of different breeds of pigs and feed ingredients, and is relatively realistic and reliable. This study crawled three types of pig prices, namely, Outer Ternary, Inner Ternary and Black pig, and two types of feedstuff prices, namely, corn and soybean meal, to make reasonable pork price forecasts. The price data from 2 November 2021 to 3 November 2022 was obtained through data crawling. This includes six columns of 367 rows of data for hog prices (external three yuan), hog prices (internal three yuan), hog prices (ground hogs), corn prices, soybean meal prices and average wholesale pork prices. These data are recorded on a day-by-day basis and arranged in a time series. As shown in Table 1.

In the process of analyzing data, missing values in the data may reduce the statistical power of the study subject or even lead to erroneous results due to biased estimates. Therefore, filling of missing values is very important. The method used in this study is Multivariate feature imputation, which is mainly done in Python with the help of the sklearn library. For a missing value in a feature, the relationship between the feature with the missing value and other features is also considered. The other features are used as independent variables and the missing value feature is used as the dependent variable. The final modeling is done to predict the missing values in the missing value features. The predicted values from the last iteration are filled in the sequence by controlling the number of iterations.

Table 1. The data set used in this experiment

Serial number	Date	Outer Ternary price	Inner Ternary price	Black pig price	Corn price (yuan/0.01t)	Soybean meal price (yuan/0.01t)	Average wholesale price of pork
1	2021/11/2	16.43	16.26	15.83	27.39	38	23
2	2021/11/3	16.39	16.15	15.76	27.64	38.12	22.8
3	2021/11/4	16.26	16.08	15.61	27.39	38.15	23.02
4	2021/11/5	16.48	16.2	15.87	27.44	37.68	23.25
5	2021/11/6	17.08	16.71	16.32	27.38	37.59	23.49
…	…	…	…	…	…	…	…
363	2022/10/30	27.1	26.76	26.22	28.97	56.52	34.74
364	2022/10/31	26.59	26.21	25.62	28.93	56.12	35
365	2022/11/1	26.67	26.45	25.76	29.01	55.98	35.43
366	2022/11/2	26.83	26.45	25.82	29.18	55.97	35.37
367	2022/11/3	26.44	26.22	25.7	29.18	56.36	33.67468047

2.2 Long and Short-term Memory Network Model

Recurrent Neural Network (RNN) is a chained network structure consisting of connected recurrent units, which was proposed by Elman in 1990 [23]. RNN not only considers the input at the current moment in the learning process, but also relies on the hidden state of the previous moment. The RNN is therefore a recurrent neural network with a memory function. It is capable of handling problems with time-series information and has been widely used in fields such as machine translation [24], speech recognition [25] and speech synthesis [25].

The long short-term memory (LSTM) neural network model was first proposed by Hochreiter et al [26] in 1997. This model was later improved by Graves et al [27] in 2012 and is a widely known special RNN model.

The LSTM introduces a mechanism for the validity of information over long periods of time. The gradient disappearance and gradient explosion problems of RNN models are overcome by adding cells (cells) that store long-term valid data and introducing controllable self-looping gates (gates). The mathematical model of the LSTM principle is,

(1) Calculate the information to be saved by the input gate

$$C_t = f_t C_{t-1} + \sigma\left(W_i\left[h_{t-1}, x_t\right] + b_i\right)\tanh\left(W_c\left[h_{t-1}, x_t\right] + b_c\right). \quad \textbf{(1)}$$

where C_t is the cell state information in the LSTM input gate; C_{t-1} is the previous cell state information; W_C is the weight matrix; b_c is the bias parameter; and tanh is the activation function.

(2) Calculate the information to be removed from the forgetting gate

$$f_t = \sigma\left(W_f\left[h_{t-1}, x_t\right] + b_f\right). \quad \textbf{(2)}$$

where x_t is the input feature vector; σ is the sigmoid activation function; W_f is the weight matrix for training; h_{t-1} is the hidden layer information at the previous moment; b_f is the bias parameter; f_t is the weight of the retained information.

(3) Calculate the information to be output from the output gate

$$h_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right)\tanh\left(c_t\right). \quad \textbf{(3)}$$

where h_t is the output information of the output gate; W_o is the weight matrix for training; and b_o is the bias parameter.

The structure of the LSTM neural network is shown in Fig. 1.

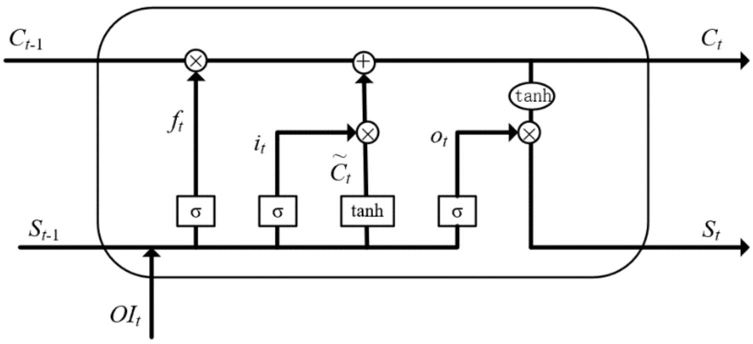

Fig. 1. LSTM neural network structure

LSTM has a strong memory capability when dealing with time series prediction. It is widely used in prediction scenarios with long time intervals and time lags in time series. LSTM adds more neural network layers to the neural network. It adds 3 modules with memory functions: memory units, input gates, output gates and forgetting gates. Each gate has a different function and selectively allows information to pass through.

2.3 Sparrow Search Algorithm

Sparrow Search Algorithm (SSA) solves the problem of random variation in model input weights and thresholds based on the foraging and anti-predation behavior of sparrow populations. It has high performance global search capability, stability, and good convergence accuracy.

To accomplish foraging, individual sparrows are usually divided into explorers and followers. In their natural state, individuals watch each other. Followers in a flock usually compete for the food resources of their high-feeding companions in order to increase their own predation rate. While foraging, all individuals are alert to their surroundings to prevent the arrival of natural predators.

In SSA, finders with good fitness values are given priority in the search for food. In addition, because the discoverer is responsible for finding food for the entire sparrow population and providing foraging directions for all joiners. As a result, discoverers are given a larger foraging search range than joiners.

During the course of each iteration, the discoverer's position is updated as described in Equation (6).

$$X_{i,j}^{t+1} = \begin{cases} x^t_{i,j} \cdot e^{\frac{-i}{\alpha \cdot iter_{\max}}} & \text{if } R_2 < ST \\ x^t_{i,j} + Q \cdot L & \text{if } R_2 \geq ST \end{cases} . \quad \textbf{(4)}$$

where t represents the number of current iterations. $X_{i,j}$ is a constant indicating the maximum number of iterations. $X_{i,j}$ denotes the location information of the ith sparrow in the jth dimension. $\alpha \in (0,1]$ is a random number. R_2 ($R_2 \in [0,1]$) and ST ($ST \in [0.5,1]$) denote the warning and safety values, respectively. Q is a random number obeying a normal distribution. L denotes a matrix of 1 × d, where each element in this matrix is all 1s. When $R_2 < ST$, this means that there are no predators around the foraging environment at this time and the finder can perform an extensive search operation. When $R_2 \geq ST$, this means that some sparrows in the population have spotted a predator and alerted the rest of the population and that all sparrows need to fly quickly to another safe place to forage.

The location update of joiners (followers) is described in equation (7).

$$X_{i,j}^{t+1} = \begin{cases} Q \cdot e^{\frac{X^t{}_{worst} - X^t{}_{i,j}}{i^2}} & \text{if } i > \dfrac{n}{2} \\ x_P{}^{t+1} + \left| X_{i,j}{}^t - x_P{}^{t+1} \right| \cdot A^+ \cdot L & otherwise \end{cases} . \tag{5}$$

where X_p is the optimal position currently occupied by the discoverer. X_{worst}, on the other hand, denotes the current worst position. A denotes a 1 × d matrix. Each element of the matrix is randomly assigned a value of 1 or -1 and $A^+ = A^T (AA^T)^{-1}$. When $i > \frac{n}{2}$, this indicates that the ith joiner with a lower fitness value is not getting food and is in a very hungry state. At this point, it needs to fly to other places to forage for more energy [28].

When aware of the danger, sparrow populations engage in anti-predatory behavior, the mathematical expression of which is shown in equation (8).

$$X_{i,j}^{t+1} = \begin{cases} X_{best}^t + \beta \cdot \left| X_{i,j}^t - X_{best}^t \right| & \text{if } f_i > f_g \\ X_{i,j}^t + K \cdot \left(\dfrac{\left| X_{i,j}^t - X_{worst}^t \right|}{(f_i - f_\omega) + \varepsilon} \right) & \text{if } f_i = f \end{cases} . \tag{6}$$

X_{best} is the current global optimum. β, the step control parameter, is a random number with a normal distribution with mean 0 and variance 1. $K \in [-1,1]$ is a random number. f_i is the current fitness value of the individual sparrow. f_g and f_w are the current best and worst fitness values, respectively. ε is the smallest constant, which is used to avoid zero in the denominator.

For simplicity, when $f_i > f_g$ means that the sparrow is at the edge of the population and is extremely vulnerable to predators. When $f_i = f_g$, sparrows in the middle of the population are aware of the danger and need to move closer to other sparrows to minimize their risk of predation. K indicates the direction of movement of the sparrows and is also a step control parameter.

2.4 SSA-LSTM Combined Model Building

Pork prices, as a time series, are influenced by many factors. It is strongly stochastic and volatile. In order to accurately predict pork prices, this paper constructs an SSA-LSTM model based on the LSTM model that performs well in time series processing. And it combines LSTM with the sparrow search algorithm with fast fusion, robustness and stability. This model retains the advantages of LSTM networks which are good at handling time-series data. SSA is also used to optimize the hyperparameters to ensure that the model can converge to the global optimum quickly and stably.

To ensure the stability of the training and the applicability of the model, the data in Table 1 was divided. The data was divided into a training set and a test set. The training set contains the first 220 rows of data and the test set contains the last 147 rows of data.

In LSTM, hyperparameters such as general learning rate, number of iterations and number of neurons have a large impact on the accuracy of prediction results. Therefore, the three parameters in LSTM are used as the optimization variables for SSA.

The pseudo-code for the SSA-LSTM optimization parameters is shown in Table 2. The steps of the optimization process are as follows.

Irandomly initialise the sparrow positions according to the range of values taken for the hyperparameters such as learning rate, number of training sessions and number of neurons of the LSTM network.

Update the discoverer, joiner and vigilante positions and calculate the fitness value in the iteration.

Output the global optimal sparrow position at the end of the iteration to obtain the optimal parameters of the LSTM network.

Table 2. Pseudo-code for SSA-LSTM optimization parameters

```
Algorithm 1. The core algorithm of SSA-LSTM
Input: train_data, train_label, test_data, test_label
Output: bestX, Convergence_curve, result
1: Sparrow population initialisation
2: for t ←M # This section is a change of position for the discoverer (explorer).
3:     do discoverer (explorer) location update.
4:     if r2 < 0.8 # Early warning values are low and no predators are present.
5:         do The discoverer changes position and updates the new fitness value.
6:     else # The early warning value is high. This indicates that the presence of a predator is threatening the safety of the population and that it needs to forage elsewhere.
7:         do The discoverer changes position and updates the new fitness value.
8:     end if
9:     for i ←pNum + 1to pop # This section shows the change of position of the joiners (followers).
10:        if  i > pop / 2 # This part of the sparrow is in a very hungry state (because their energy is very low, representing a very low adaptation value).
11:            do #Sparrows need to forage elsewhere.
12:        else #This section of followers forages around the best finders. During this time they are likely to have food fights, turning themselves into producers.
13:            do Determine if the boundary is exceeded and calculate the fitness value.
14:        end if
15:    Aware of the danger a part of the sparrow updates its position.
16:    Optimal solution update.
17: return bestX, Convergence_curve, result
```

3 Model Evaluation Indicators

In predictive regression tasks with continuous values, evaluation metrics are often used for validation in order to analyze the accuracy and certainty of the model. Common evaluation metrics are Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). When the magnitudes are different, RMSE, MAE and MSE are difficult to measure the effectiveness of the model. Therefore, R2_score is introduced in this paper.

3.1 R2_score

R2_ score, is coefficient of determination. It is a measure of the proportion of the total variation in the dependent variable that can be explained by the independent variable through the regression relationship.

$$R^2 = 1 - \frac{SSE}{SST} . \tag{7}$$

$$SST = SSR + SSE = \sum_{i=1}^{n} \left(y_i - \overline{y} \right)^2 . \tag{8}$$

$$SSR = \sum_{i=1}^{n} \left(\hat{y}_i - \overline{y} \right)^2 . \tag{9}$$

$$SSE = \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2 . \tag{10}$$

Of which, *SSE* stands for the sum of squared residuals. *SST* stands for the sum of squared total deviations. *SSR* refers to the regression sum of squares. y denotes true observations. $\overline{y}$ denotes the mean of the true observations. $\hat{y}$ denotes the fitted value.

SSR is the sum of squared regressions, which refers to the error of the estimate from the mean. It can reflect the sum of squared deviations from the degree of correlation between the independent and dependent variables.

SSE is the sum of squared residuals, which refers to the error of the estimated value from the true value. It can reflect the degree of model fit.

SST is the sum of squared total deviations and refers to the error between the mean and the true value. It can reflect the degree of deviation from mathematical expectations.

When R2_score = 1, the predicted and true values in the sample are exactly equal and there is no error. This also indicates that the better the independent variable explains the dependent variable in the regression analysis, the better the model fits the data [29]. When R2_score = 0, the numerator equals the denominator and each predicted value in the sample is equal to the mean value.

3.2 MAE

Mean Absolute Error (MAE), is the average of the absolute errors. It can better reflect the actual situation of the error of the predicted value.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right| . \tag{11}$$

Of which, y_i denotes true observations, $\hat{y}$ denotes the fitted value.

3.3 RMSE

Mean Square Error (MSE) is the square of the difference between the true value and the predicted value, then averaged over the sum. It is generally used to detect deviations between the predicted and true values of a model, and the use of RMSE is very common in regression. It is considered to be an excellent general error measure for numerical prediction, both in statistics and machine learning. A high RMSE indicates a large deviation between the predicted and actual values [30]. another important property of the RMSE is the fact that the error squared implies that greater weight is assigned to larger errors. In contrast, Root Mean Square Error (RMSE) is the root mean square error open sign. The Root Mean Square Error represents the sample standard deviation of the difference between the predicted value and the observed value.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2 . \tag{12}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} . \tag{13}$$

Of which, y_i denotes true observations, $\hat{y}$ denotes the fitted value.

3.4 MAPE

Mean Absolute Percentage Error (MAPE). Theoretically, the smaller the MAPE value, the better the prediction model fits and the better the accuracy.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| . \tag{14}$$

Of which, y_i denotes true observations, $\hat{y}$ denotes the fitted value.

4 Empirical Studies

4.1 Comparative Analysis of Model Predictions

Fig. 2 shows the pork price prediction results of the SSA-LSTM model. The prediction evaluation metrics for each prediction model are shown in Table 3. MLP is an earlier model applied to prediction, with values of 0.155565321, 5.565072, 4.893069 and -1.09843 for its four evaluation metrics. relatively high values of MAPE, RMSE and MAE indicate that MLP prediction errors are more pronounced. the LSTM model is the most classical methods, but in this experiment, the predictive ability was not significantly enhanced. The predictive power of the Bi-LSTM was significantly improved compared to the LSTM, with values of 0.074215161, 2.829802, 2.355783 and 0.440825 for the four evaluation metrics, respectively. GRU is an optimized model based on the LSTM. The values of MAPE, RMSE, MAE and R2 are 0.070951906, 2.611833, 2.23938 and 0.52365 respectively. The prediction ability of the model is significantly improved after applying the population intelligence algorithm to optimize the LSTM. the values of the prediction evaluation indicators of the IPSO-LSTM model are 0.05319370432748787, 2.1462281624115804, 1.7040217828007949 and 0.6878919777597307. This significantly outperformed the classical prediction method and illustrated the effectiveness of the population intelligence algorithm in optimizing the parameters of the LSTM model. The SSA-LSTM model proposed in this paper uses a new type of SSA to optimize the LSTM, and its prediction results are further improved compared with IPSO-LSTM. The values of MAPE, RMSE, MAE and R2 corresponding to the SSA-LSTM model are 0.001948241, 0.087956, 0.061099 and 0.999476 respectively. The examples of the four metrics show that the errors between the predicted and actual values of the SSA-LSTM are small. And the prediction accuracy of SSA-LSTM is higher than other models, and the model has a strong fitting ability.

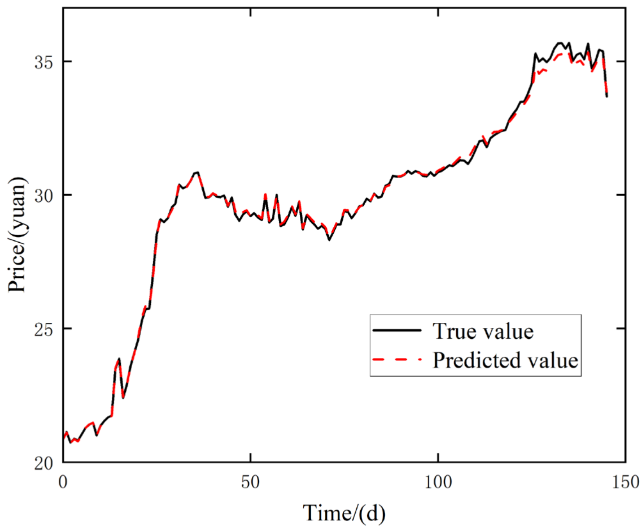

Fig. 2. Pork price forecasting results from SSA-LSTM model

Table 3. Predictive evaluation indicators for each model

	MAPE	RMSE	MAE	R2
SSA-LSTM	0.001948241	0.087956	0.061099	0.999476
IPSO-LSTM	0.05319370432748787	2.1462281624115804	1.7040217828007949	0.6878919777597307
GRU	0.070951906	2.611833	2.23938	0.52365
Bi-LSTM	0.074215161	2.829802	2.355783	0.440825
LSTM	0.13260507	4.77547	4.175694	-0.5452
MLP	0.155565321	5.565072	4.893069	-1.09843

According to the above analysis, the SSA-LSTM model is more effective and has higher prediction accuracy than the traditional network model when dealing with time series data.

4.2 Predicted Results

In order to verify the short-term and long-term forecasting effects of the SSA-LSTM model. Based on the model built in the previous section, this paper predicted the pork prices in the coming week as well as the next 30 days respectively. The actual prices were also checked through the internet. The comparison graphs between the real price and the predicted price are shown in Fig. 3 and Fig. 4.

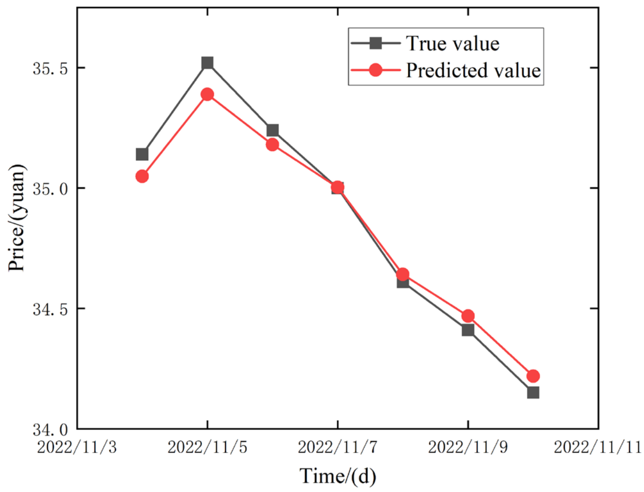

Fig. 3. Short-term forecast results for pork prices in a week's time

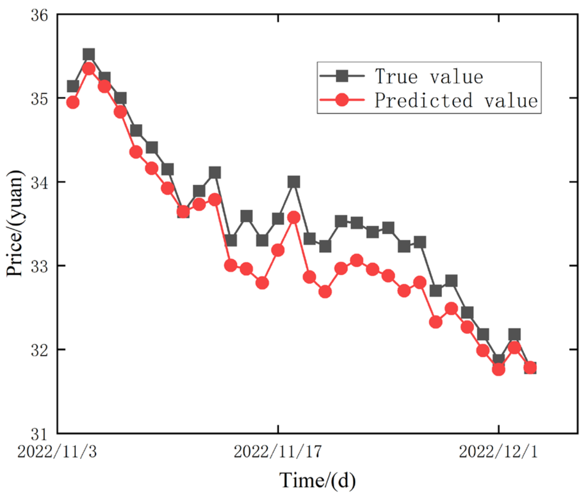

Fig. 4. Long-term forecast results for pork prices in one month's time

In Fig. 3 and Fig. 4, the square points represent the real pork price values and the circular points represent the predicted pork price values. As can be seen from the graphs, in the short-term forecast (predicting the price after one week), the pork price curve predicted using the SSA-LSTM model can basically fit the curve of the real price. In the long-term forecast (after 30 days), the SSA-LSTM model can be used to predict a pork price curve that largely mimics the real price trend. However, it is still clear from the long-term forecast change graph that the forecast curve does not fit the true price well.

5 Design of an Early Warning Mechanism for Price Fluctuations

5.1 Early Warning Mechanism Design

As shown in Fig. 5, the methods of economic early warning can be divided into five categories based on their mechanisms. Of these, the yellow warning method, where the warning is based primarily on warning signs. The black warning method, where the warning is based directly on the time series fluctuation pattern of the warning indicators. The red warning method, which estimates the warning situation after identifying the warning signs and various environmental and social factors. The green warning methods: mainly used for green crop warnings. The white warning methods, which rely on measurement techniques to warn of fixed influences.

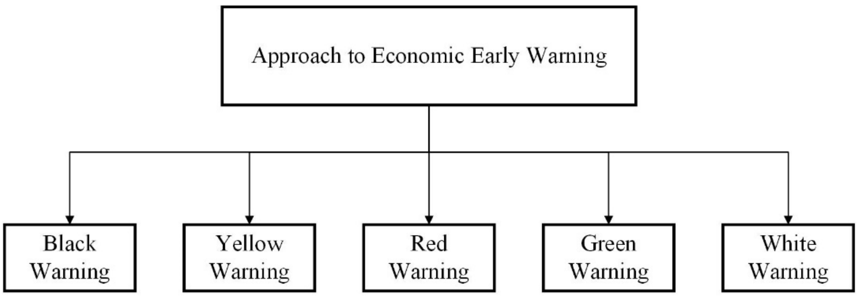

Fig. 5. Five economic warning methods

The price data in this study is mainly sourced from web crawlers and has a more complete time series. It is also able to analyse the volatility characteristics of the data better. Therefore, this study adopts an early warning method that fits better with the data, namely the black warning method. This method does not focus on what causes changes in the warning indicators. It focuses only on the changes in the warning Instance over their time series and from this, it looks for their intrinsic patterns, i.e., cyclical volatility characteristics.

An warning Instance is an indicator that responds to the presence or absence of an alarm in pork price fluctuations. It is mainly based on growth rate indicators. This is because the use of growth rate indicators not only addresses the impact of the unit, but also makes it easy to determine the warning limit and the warning degree. In economic indicators, growth indicators show growth in absolute terms and growth rate indicators tend to show volatility. Pork prices are growth indicators in terms of their overall trend. Therefore, in this study, volatility is used as an indicator of alertness, and the conversion is shown in equation (15).

$$R_t = \left(\ln P_t - \ln P_{t-1}\right) \times 100\% . \tag{15}$$

In equation (14), R_t represents volatility. P_t and P_{t-1} represent the prices at the previous and subsequent moments respectively.

A warning instance is a situation in which the economy is operating abnormally. Generally, warning instances are measured by growth rates or ratios. This study uses price volatility as a warning instance for pork price warning. This indicator reflects changes in the supply and demand of agricultural products. Under normal circumstances, prices of agricultural products fluctuate within a certain range, and market price risk occurs when prices fluctuate outside this range. Depending on the alarm situation, the corresponding light and the current status are determined. A negative heavy alarm corresponds to a white light, which means that excessive price in-creases. A negative light alarm corresponds to a blue light, which means that faster price increases. No alarm corresponds to a green light, which means that prices are stable. A positive light warning corresponds to a yellow light, which means that faster price drops. A positive heavy warning corresponds to a red light, which means that excessive price de-creases.

The warning limit serves as the boundary that classifies the presence or absence of an alert and is the basis for determining the level of alert. In a mature market economy, the price of a commodity accurately reflects market supply and demand. An equilibrium market does not strictly require prices to be fixed, but allows for slight upward and downward price fluctuations around the equilibrium point. The slight fluctuation of prices around the equilibrium point indicates that price signals are able to reflect the contrasting forces of supply and demand in the market in a timely and sensitive manner. Market participants are also able to adjust the level of supply and demand in time to bring the market back to equilibrium. As can be seen, the warning indicators of price volatility are bi-lateral with alarms, not no volatility is reasonable, therefore, it is necessary to specify the interval of normal price volatility. This study will use the plurality principle, the half principle and the minority principle to classify the alert limits. The warning degree refers to the level of severity of the harmful consequences caused by the occurrence of a risk (or hazard) and is determined after taking into account the nature, severity, controllability and scope of the consequences.

Under the influence of market economy conditions, the fluctuation of pork prices in China mainly fluctuates with the change of supply and demand in the market. When the price of pork is too high or too low it means that the pork market is in a state of being too cold or too hot. Therefore, when building a price warning model, it is necessary to find the equilibrium point of the market, and then use the equilibrium point as the centre to build a two-sided warning model with alarms. The equilibrium point is used to establish the warning limit range, with the negative warning limit range representing low pork prices, an overly cold market and a decrease in producer motivation. The positive warning limit range represents a high pork price, an overheated market and a reduced desire to buy. The warning Instance, warning limit, warning degree and warning light information in this experiment is shown in Table 4.

Table 4. Comparison table of warning situation and warning limit range, warning level

Warning instance	Warning limit range (%)	Warning degree	Warning limit division principle	State	Warning light
Negative heavy warning	(-∞, -1.01)	-2	Minority principle	Excessive price increases	White light
Negative light warning	[-1.01, -0.26)	-1	1/5 principle	Faster price increases	Blue light
non-warning	[-0.26, 0.4]	0	Moiety Rules	Price stability	Green light
Positive light warning	(0.4, 0.93]	1	1/5 principle	Faster price drops	Yellow light
Positive heavy warning	(0.93, +∞)	2	Minority principle	Excessive price decreases	Red light

The evaluation of the warning accuracy of the black warning model requires the use of the constructed black warning model to make predictions on the training set data. Find the warning limit range corresponding to the predicted value and get the warning degree corresponding to the risk level. At the same time, the corresponding real data values and risk levels for the predicted time period are obtained and compared with the predicted warning levels. This is used to determine whether the warning system can reliably issue warning messages.

5.2 Early Warning Experiments

The SSA-LSTM pork price forecasting method obtained by comparison in the previous paper is used to forecast pork prices. The predicted price data from the previous paper's short-term and long-term forecasts were selected to obtain the daily price volatility. And the predicted and actual warnings were judged with the help of the set warning limit intervals. Table 5 and Table 6 show the forecast price, actual price, volatility and warning information for the short-term and long-term forecasts respectively.

Table 5. Pork price volatility warning in short-term forecasts

True value	Volatility	Warning rank	Predicted value	Volatility	Warning rank	True value
2022/11/4	35.14			35.048325		
2022/11/5	35.52	0.46712	Yellow light	35.38955	0.420777299	Yellow light
2022/11/6	35.24	-0.34371	Blue light	35.18017	-0.25771066	Green light
2022/11/7	35	-0.29679	Blue light	35.002823	-0.21948619	Green light
2022/11/8	34.61	-0.48664	Blue light	34.64201	-0.44999892	Blue light
2022/11/9	34.41	-0.25169	Green light	34.46821	-0.21843525	Green light
2022/11/10	34.15	-0.3294	Blue light	34.21835	-0.31596661	Blue light

Table 6. Pork price volatility warning in long-term forecasts

True value	Volatility	Warning rank	Predicted value	Volatility	Warning rank	True value
2022/11/4	35.14			34.94727		
2022/11/5	35.52	0.46712	Yellow light	35.34841	0.495662833	Yellow light
2022/11/6	35.24	-0.34371	Blue light	35.13676	-0.26081714	Blue light
2022/11/7	35	-0.29679	Blue light	34.8348	-0.37483908	Blue light
2022/11/8	34.61	-0.48664	Blue light	34.35546	-0.60175536	Blue light
2022/11/9	34.41	-0.25169	Green light	34.15991	-0.24790501	Green light
2022/11/10	34.15	-0.3294	Blue light	33.922	-0.3035268	Blue light
2022/11/11	33.64	-0.65347	Blue light	33.6409	-0.36138438	Blue light
2022/11/12	33.89	0.321558	Green light	33.73054	0.115568794	Green light
2022/11/13	34.11	0.281015	Green light	33.78675	0.072312436	Green light
2022/11/14	33.3	-1.04375	White light	33.00199	-1.020629	White light
2022/11/15	33.59	0.376577	Green light	32.95948	-0.05597771	Green light
2022/11/16	33.3	-0.37658	Blue light	32.79248	-0.22060889	Green light
2022/11/17	33.56	0.337772	Green light	33.18396	0.515394887	Yellow light
2022/11/18	34	0.565696	Yellow light	33.57287	0.506025759	Yellow light
2022/11/19	33.32	-0.87739	Blue light	32.86285	-0.92832443	Blue light
2022/11/20	33.23	-0.11747	Green light	32.68848	-0.2310498	Green light
2022/11/21	33.53	0.390321	Green light	32.96492	0.365729958	Green light
2022/11/22	33.51	-0.02591	Green light	33.06324	0.129338346	Green light
2022/11/23	33.4	-0.1428	Green light	32.9537	-0.14412259	Green light
2022/11/24	33.45	0.064966	Green light	32.87701	-0.101187	Green light
2022/11/25	33.23	-0.28658	Blue light	32.69987	-0.23462876	Green light
2022/11/26	33.28	0.065298	Green light	32.79796	0.130080577	Green light
2022/11/27	32.7	-0.76356	Blue light	32.32637	-0.62898924	Blue light
2022/11/28	32.82	0.159082	Green light	32.48794	0.216523495	Green light
2022/11/29	32.44	-0.50577	Blue light	32.26611	-0.29755644	Blue light
2022/11/30	32.18	-0.34948	Blue light	31.98779	-0.37623738	Blue light
2022/12/1	31.87	-0.4204	Blue light	31.76092	-0.30911623	Blue light
2022/12/2	32.18	0.420398	Yellow light	32.01978	0.352526977	Green light
2022/12/3	31.78	-0.54321	Blue light	31.7825	-0.3230288	Blue light

As can be seen from Table 5 and Table 6, the short-term predicted values using SSA-LSTM are within 0.07yuan of the true values. The difference between the long-term predicted value and the true value using SSA-LSTM forecasting was within 0.3yuan. The difference between the predicted, actual price of pork and its corresponding volatility is not significant. There is some referenceability in classifying the volatility through the Warning limit range.

6 Conclusion

This study uses pork prices as an example to study the prediction and early warning of agricultural commodity prices. Relevant factors affecting pork prices were selected to produce a dataset for model training and testing. To address the problem that the selection of hyperparameters in LSTM networks is often based on subjective experience and existing research, this study proposes an SSA-LSTM model for pork price prediction. This model uses the sparrow search algorithm to optimize the parameters of the LSTM model, objectively explaining the origin of the model network structure and parameter settings. The process minimises the influence of human factors on the pork price forecasting process and improves the generalisation ability and forecasting effectiveness of the model. This study also developed a black warning model for pork price volatility. Based on the prediction results of the combined time series forecasting model, a black warning model was constructed to complete the study of the black warning mechanism. The price volatility is used to delineate the warning limit range, determine the degree of warning, and forecast the warning results with the signal light color. From the results, it can be concluded that the greatest thing about the black warning mechanism for pork prices is that the warning information is relatively simple, and only uses the data historical information change pattern to do related calculations. This does not need to take into account the influence of external factors and does not take into account information from other relevant indicators. This has some applicability and reference in the future for pork traders to develop price strategies and avoid price risks. Given the powerful self-learning capability, good generalization ability and high adjustability of the neural network, this study can be further improved in the future.

7 Acknowledgement

This research was funded by The Major Project of Scientific and Technological Innovation 2030 (2021ZD0113603).

References

[1] K. Ye, Y. Piao, K. Zhao, X. Cui, A heterogeneous graph enhanced LSTM network for hog price prediction using online discussion, Agriculture 11(4)(2021) 359.
[2] S. Xu, Proceedings of 2013 World Agricultural Outlook Conference, Springer, Berlin, Heidelberg, 2014.
[3] C. Wang, Z. Sun, Monthly pork price forecasting method based on Census X12-GM(1,1) combination model, PLOS ONE 16(5)(2021) e0251436.
[4] Y. Guo, D. Tang, W. Tang, S. Yang, Q. Tang, Y. Feng, F. Zhang, Agricultural Price Prediction Based on Combined Forecasting Model under Spatial-Temporal Influencing Factors, Sustainability 14(17)(2022) 10483.
[5] M. Tang, J. Zhang, A multiple adaptive wavelet recurrent neural network model to analyze crude oil prices, Journal of Economics and Business 64(4)(2012) 275-286.
[6] R. Murugesan, E. Mishra, A.H. Krishnan, Forecasting agricultural commodities prices using deep learning-based models: basic LSTM, bi-LSTM, stacked LSTM, CNN LSTM, and convolutional LSTM, International Journal of Sustainable Agricultural Management and Informatics 8(3)(2022) 242-277.
[7] C.S. Mesike, T.U. Esekhade, Rainfall variability and rubber production in Nigeria, African Journal of Environmental Science and Technology 8(1)(2014) 54-57.
[8] Q. Zhang, J. Zhang, C. Wang, L. Cui, D. Yan, Risk early warning of maize drought disaster in Northwestern Liaoning Province, Natural hazards 72(2014) 701-710.
[9] G.J. Husak, C.C. Funk, J. Michaelsen, T. Magadzire, K.P. Goldsberry, Developing seasonal rainfall scenarios for food security early warning, Theoretical and Applied Climatology 114(2013) 291-302.
[10] P. Ceccato, K. Fernandes, D. Ruiz, E. Allis, Climate and environmental monitoring for decision making, Earth Perspectives 1(1)(2014) 1-22.
[11] S. Mirjalili, Genetic algorithm, Evolutionary Algorithms and Neural Networks: Theory and Applications (2019) 43-55.
[12] R. Poli, J. Kennedy, T. Blackwell, Particle swarm optimization, Swarm intelligence 1(2007) 33-57.
[13] X.-S. Yang, A.H. Gandomi, Bat algorithm: a novel approach for global engineering optimization, Engineering computations 29(5)(2012) 464-483.
[14] S. Mirjalili, S.M. Mirjalili, A. Lewis, Grey wolf optimizer, Advances in engineering software 69(2014) 46-61.
[15] J. Xue, B. Shen, A novel swarm intelligence optimization approach: sparrow search algorithm, Systems science & control engineering 8(1)(2020) 22-34.
[16] Y. Ma, Y. Tang, B. Li, B. Qi, Residential high-power load prediction based on optimized LSTM network, in: Proc. 2020

International Conference on Artificial Intelligence and Computer Engineering (ICAICE), 2020.
[17] G.-C. Liao, Fusion of Improved Sparrow Search Algorithm and Long Short-Term Memory Neural Network Application in Load Forecasting, Energies 15(1)(2021) 130.
[18] F. Liu, P. Qin, J. You, Y. Fu, Sparrow Search Algorithm-Optimized Long Short-Term Memory Model for Stock Trend Prediction, Computational Intelligence and Neuroscience 2022(2022).
[19] G. Liu, C. Shu, Z. Liang, B. Peng, L. Cheng, A modified sparrow search algorithm with application in 3d route planning for UAV, Sensors 21(4)(2021) 1224.
[20] T. Liu, Z. Yuan, L. Wu, B. Badami, An optimal brain tumor detection by convolutional neural network and enhanced sparrow search algorithm, Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine 235(4)(2021) 459-469.
[21] X. Liu, H. Guo, Air quality indicators and AQI prediction coupling long-short term memory (LSTM) and sparrow search algorithm (SSA): A case study of Shanghai, Atmospheric Pollution Research 13(10)(2022) 101551.
[22] S.M. Mirtaheri, M.E. Dinçktürk, S. Hooshmand, G.V. Bochmann, G.-V. Jourdan, I.V. Onut, A brief history of web crawlers, <https://arxiv.org/abs/1405.0749>, 2014 (accessed 06.12.2022).
[23] J.L. Elman, Finding structure in time, Cognitive science 14(2)(1990) 179-211.
[24] S.K. Mahata, D. Das, S. Bandyopadhyay, MTIL2017: Machine Translation Using Recurrent Neural Network on Statistical Machine Translation, Journal of Intelligent Systems 28(3)(2019) 447-453.
[25] G.K. Anumanchipalli, J. Chartier, E.F. Chang, Intelligible speech synthesis from neural decoding of spoken sentences, BioRxiv (2018) 481267.
[26] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9(8)(1997) 1735-1780.
[27] A. Graves, Supervised sequence labelling, Springer Berlin Heidelberg 2012.
[28] T. Wumaier, C. Xu, H. Guo, Z. Jin, H. Zhou, Fault diagnosis of wind turbines based on a support vector machine optimized by the sparrow search algorithm, IEEE Access 9(2021) 69307-69315.
[29] Z. Jin, Y. Yang, Y. Liu, Stock closing price prediction based on sentiment analysis and LSTM, Neural Computing and Applications 32(2020) 9713-9729.
[30] M. Elsaraiti, A. Merabet, A Comparative Analysis of the ARIMA and LSTM Predictive Models and Their Effectiveness for Predicting Wind Speed, Energies 14(20)(2021) 6782.