The research model of forecasting agricultural carbon emissions based on ARIMA-LSTM

Wenqiang Qin[1,3], Keqiang Wang[1,4], Qingsong Yan[2,5]

[1]Qinghai University School of Finance and Economics
[3]Department of Computer Technology and Application, Qinghai University

[3]1602186500@qq.com
[4]3472613677@qq.com
[5]1599537583@qq.com

Abstract. To accurately forecast the volume of agricultural carbon emissions in China, this paper proposes a forecasting model that combines the Autoregressive Integrated Moving Average Model (ARIMA) with Long Short-Term Memory networks (LSTM) to better handle both linear and non-linear parts in time series. Using the data on China's agricultural carbon emissions collected from 2000 to 2023 as the research subject, the data from 2000 to 2016 were first used as the training set, and the data from 2017 to 2023 were used as the test set. The Spearman's rank correlation coefficient method was utilized to study the correlation between various types of carbon emissions. Then, the established ARIMA-LSTM forecasting model was applied to predict the emissions in the test samples and make comparisons. The Absolute Relative Error (ARE) was adopted as the evaluation metric for different models, comparing the prediction errors of the ARIMA-LSTM model with those of the ARIMA and LSTM single models through analysis. The results show that the positive correlation between diesel carbon emissions and film carbon emissions was the highest; the absolute relative error of the ARIMA-LSTM model did not exceed 7%, which is significantly lower than the other two comparison models. This research model achieves high-accuracy predictions of carbon emissions, which is beneficial for the control and management of the greenhouse effect.

Keywords: Agricultural Carbon Emissions, Autoregressive Integrated Moving Average Model, Long Short-Term Memory Network, Forecasting, Error.

1. Introduction

Agricultural carbon emissions are one of the main causes of global climate change. Accurately predicting agricultural carbon emissions can assess their impact on climate change and facilitate the adoption of environmental protection and climate change response measures [1]. Governments and policymakers can formulate relevant policies based on the predicted volume of agricultural carbon emissions, promoting environmentally friendly and sustainable agricultural production and encouraging low-carbon agriculture. The forecasting of agricultural carbon emissions is of significant practical importance. Mathematical algorithms play a crucial role in this area, mainly used in developing models to predict emission trends Statistical models (such as ARIMA, VAR) or machine learning algorithms (such as neural networks, support vector machines) can utilize historical data to predict future emission

DOI: 10.54254/2755-2721/75/20240521

trends [2-3]. Mathematical algorithms are also applied in large-scale data mining and feature extraction, extracting relevant features from various aspects, establishing prediction models, and analyzing influencing factors. Mathematical optimization algorithms can optimize carbon emissions in agricultural production processes to achieve emission reduction and carbon neutrality goals. Time series analysis methods can handle historical data to establish dynamic models for predicting and regulating future emissions. In summary, through forecasting models, spatial analysis, data mining, optimization control, and other means, mathematical algorithms play a vital role in understanding and managing agricultural carbon emissions, supporting sustainable agricultural development and mitigating climate change. This study aims to establish an ARIMA-LSTM model for predicting agricultural carbon emissions, using historical data on agricultural carbon emissions as the research subject to improve the accuracy of the forecasting model.

2. Model Introduction

2.1. ARIMA Model

The ARIMA(p,d,q) model is a method for time series analysis and forecasting that combines three components: Autoregressive (AR), Integrated (I), and Moving Average (MA). The ARIMA(p,d,q) model can predict future values based on historical data on agricultural carbon emissions, where 'p' represents the order of the autoregressive terms, 'd' represents the degree of differencing, and 'q' represents the order of the moving average terms [4-5]. The basic modeling process is as follows: Firstly, test the time series for stationarity and use differencing to transform a non-stationary series into a stationary series. Then, determine the differencing parameter (d), and identify the autoregressive parameters (p) and moving average parameters (q) by examining the Partial Autocorrelation (PACF) and Autocorrelation Function (ACF) graphs. The model is specified as follows:

$$\begin{gathered} \phi(B)\nabla^d x_t = \varphi(B)\varepsilon_t \\ E(\varepsilon_t) = 0 \\ V_{ar}(\varepsilon_t) = \sigma_\varepsilon^2 \\ E(\varepsilon_t \cdot \varepsilon_s) = 0, s \neq t \\ E(x_t \cdot \varepsilon_t) = 0, \forall s < t \end{gathered} \tag{1}$$

Where $\phi()$、$\varphi()$ respectively represent the polynomials of the p-th order autoregressive coefficients and q-th order moving average coefficients; $\nabla^d = (1-B)^d$ denotes the differencing operation; B is the lag operator, and $B^n x_t = x_{t-n}$; ε_t is the random error series, and $E()$、$V_{ar}()$ distribution represents the expected value and variance of .

2.2. LSTM Model

Long Short-Term Memory (LSTM) networks, a variant of Recurrent Neural Networks (RNNs) within deep learning, are specifically designed to address long sequence dependency issues [6-7]. Given the persistent nature of agricultural carbon emissions over time, choosing an LSTM model is appropriate due to its ability to capture and remember long-term dependencies, which is particularly suitable for forecasting agricultural carbon emissions. Its principle structure is as follows in Figure 1:

DOI: 10.54254/2755-2721/75/20240521

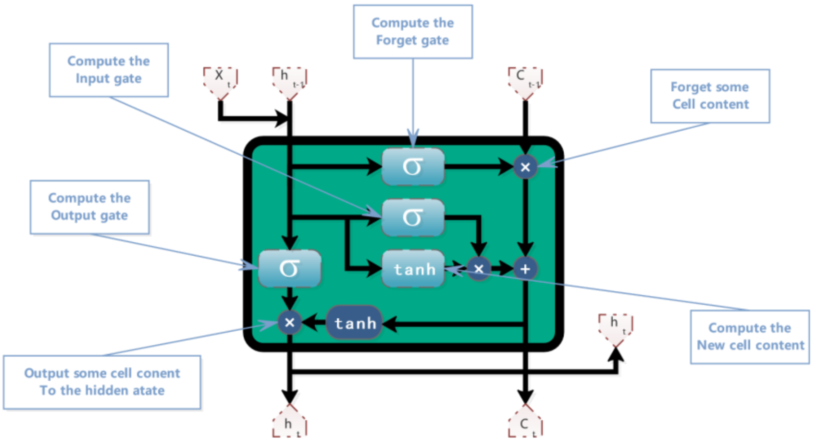

Figure 1. The Principle Structure of the LSTM Model

2.3. *ARIMA-LSTM Combined Model*

Initially, collect and organize time series data, ensuring there is sufficient historical information. Preprocess the data, including handling missing and outlier values and conducting stationarity tests. Use tools like ACF and PACF to select the order of the ARIMA model, fit the ARIMA model, and determine the order of differencing, autoregression, and moving average. Normalize the time series data appropriately to ensure model stability. Construct the LSTM model, deciding on the number of layers and the number of hidden units. Split the data into training and test sets for model training. Use the ARIMA model to predict future trends to obtain the ARIMA trend component and the LSTM model to predict non-linear patterns to obtain the LSTM prediction part. Combine the ARIMA and LSTM prediction results to obtain the final comprehensive forecast [8]. Compare the combined model predictions with the actual values in the test set, and evaluate performance using metrics such as RMSE and MAE. Adjust the ARIMA and LSTM parameters based on performance evaluation results to optimize model accuracy. Use the optimized ARIMA-LSTM combined model to predict future time series. The process is illustrated in the following Figure 2:

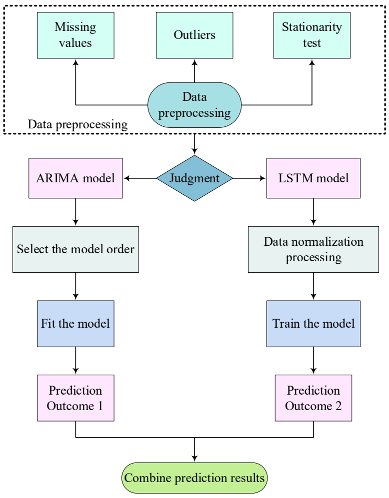

Figure 2. ARIMA-LSTM Combined Process Flowchart

3. Data Collection and Analysis

This study focuses on the agricultural carbon emissions in China. Data on various types of agricultural carbon emissions and the total agricultural carbon emissions from 2000 to 2023 were collected from the "China Statistical Yearbook" and Google. Partial data is shown in Table 1 below:

DOI: 10.54254/2755-2721/75/20240521

Table 1. Partial Data on Agricultural Carbon Emissions

Year	Fertilizer Carbon Emissions (10,000 tons)	Film Carbon Emissions (10,000 tons)	……	Irrigation Carbon Emissions (10,000 tons)	Total Agricultural Carbon Emissions (10,000 tons)
2000	3690.652	691.761028	……	1434.204154	7325.295408
2001	3786.149	750.730148	……	1445.637773	7536.004502
2002	3861.977	792.931608	……	1448.448043	7688.136276
……	……	……	……	……	……
2021	4483.913	1182.84739	……	1856.649201	9322.857666
2022	4283.038	1173.8965	……	1874.989105	9108.90884
2023	4098.689	1127.83918	……	1883.482966	9000.453789

The Spearman correlation coefficient is a non-parametric method used to measure the correlation between two variables. It calculates correlations based on rankings rather than actual data values. The Spearman correlation coefficient ranges from -1 to 1, where -1 indicates a perfect negative correlation, 1 indicates a perfect positive correlation, and 0 indicates no linear correlation. It is also useful for discovering non-linear relationships because it is based on rankings rather than actual numerical values. By using the Spearman correlation coefficient method to analyze the correlation between different types of carbon emissions, the thermodynamic correlation results are shown in Figure 3 below:

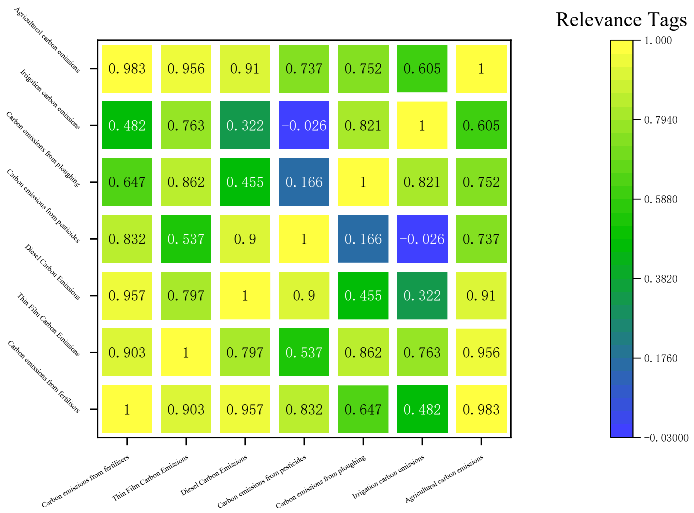

Figure 3. Correlation Results

Analysis from Figure 3 indicates that the positive correlation between diesel carbon emissions and film carbon emissions is the highest, at 0.957. The correlations between film carbon emissions and fertilizer carbon emissions, as well as between pesticide carbon emissions and fertilizer carbon emissions, are also significant, at 0.903 and 0.832, respectively. This is almost consistent with the development of the agricultural industry in China, leading to a significant increase in carbon dioxide emissions.

4. Analysis of Forecasting Results

Based on the above analysis, the collected sample data were divided into training and test sets, with agricultural carbon emission data from 2000 to 2016 serving as the training set and data from 2017 to

DOI: 10.54254/2755-2721/75/20240521

2023 serving as the test set. An ARIMA-LSTM model for predicting agricultural carbon emissions was established, and the Absolute Relative Error (ARE) was introduced as the model's average metric, as shown in the following formula:

$$ARE = \left|\frac{\hat{y}_i - y_i}{y_i}\right| \times 100\%$$

Where $\hat{y}_i$ represents the forecasted output value for the i -th forecast sample, and y_i represents the actual value for the i -th sample.

Using MATLAB 2023b, the prediction results for the test samples were calculated and are shown in Figure 4 below:

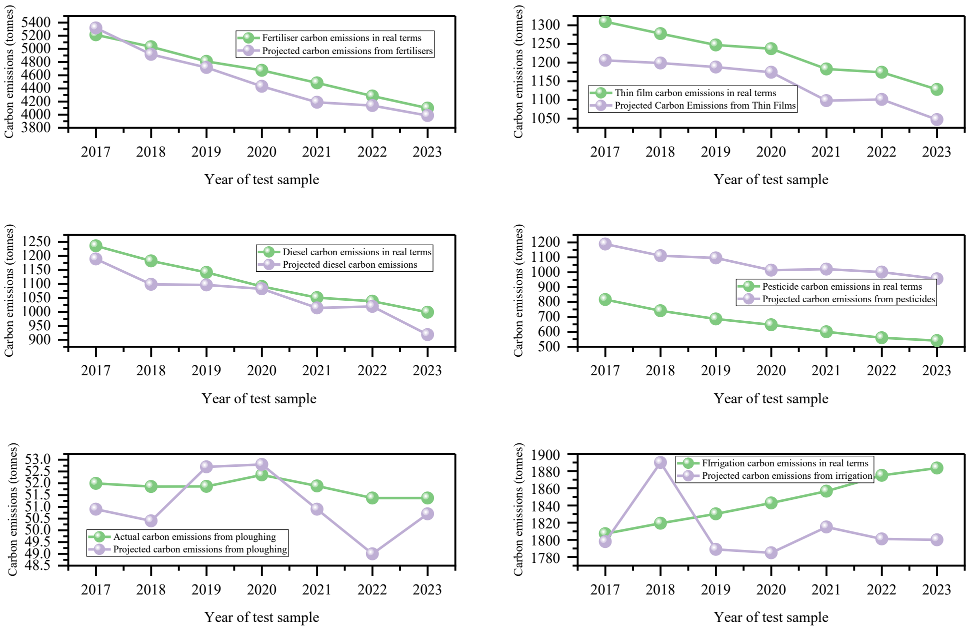

Figure 4. Comparison of Forecast Results for Test Samples

Analysis from Figure 4 indicates that the ARIMA-LSTM model's predictions for the carbon emissions of various agricultural carbon sources from 2017 to 2023 show a nearly consistent trend in forecast fluctuations and exhibit good predictive performance, making it suitable for application in agricultural carbon emissions forecasting.

Additionally, to evaluate the forecasting accuracy of the established combined model from multiple dimensions, the forecast results of the ARIMA-LSTM model for agricultural carbon emissions from 2017 to 2023 were compared with those of the ARIMA and LSTM single models. The comparison of forecasting errors between different models is shown in Figure 5 below:

DOI: 10.54254/2755-2721/75/20240521

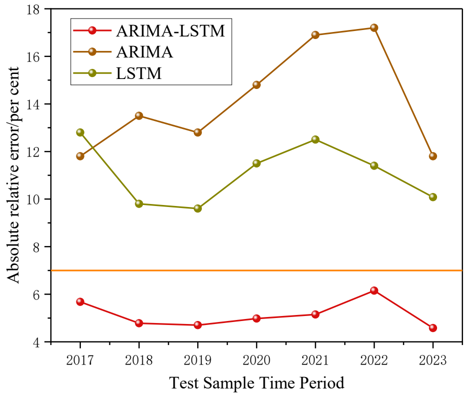

Figure 5. Comparison of Forecast Errors Between Different Models

Analysis from Figure 5 reveals that the forecasting error of the established ARIMA-LSTM carbon emission prediction model is less than 7%, and the error fluctuation trend is relatively stable, with no significant changes in forecasting error. In comparison to the ARIMA and LSTM models, the forecasting errors of these single models are significantly higher than those of the ARIMA-LSTM model, and the magnitude of fluctuation in forecasting error is greater, making them unsuitable for predicting China's future carbon emissions.

5. Conclusion

(1) By using the Spearman correlation coefficient method to analyze the correlation between carbon emissions from six agricultural carbon sources in China, it was found that the positive correlation between diesel carbon emissions and film carbon emissions was the highest, at 0.957.

(2) The ARIMA-LSTM model for forecasting agricultural carbon emissions was established and used to predict the test samples. The findings indicate that the predictions of this model closely match the actual values. Moreover, when comparing the errors of the ARIMA-LSTM model with those of the ARIMA and LSTM models individually, the forecast error of the ARIMA-LSTM model is significantly lower than that of the other comparative models, substantially enhancing the precision of the predictions.

References

[1] Huang, H. P., Li, Z. X., Huang, D., et al. (2024). Estimation, analysis, and prediction of agricultural carbon emissions in Jiangxi Province under the "dual carbon" targets. Journal of Ecology and Rural Environment, 40(02), 179-190.

[2] Chen, H. F., Wang, H., Li, Y., et al. (2024). Short-term wind speed forecasting research combining two-step decomposition and ARIMA-LSTM. Acta Energiae Solaris Sinica, 45(02), 164-171.

[3] Guo, L., & Guo, W. W. (2024). Prediction research of high calorific value of power coal based on SVR and random forest models. Energy Engineering, 44(01), 35-42.

[4] Siami-Namini S, Tavakoli N, Namin A S. A comparison of ARIMA and LSTM in forecasting time series[C]//2018 17th IEEE international conference on machine learning and applications (ICMLA). IEEE, 2018: 1394-1401.

[5] Dey B, Roy B, Datta S, et al. Forecasting ethanol demand in India to meet future blending targets: A comparison of ARIMA and various regression models[J]. Energy Reports, 2023, 9: 411-418.

[6] Cahuantzi R, Chen X, Güttel S. A comparison of LSTM and GRU networks for learning symbolic sequences[C]//Science and Information Conference. Cham: Springer Nature Switzerland, 2023: 771-785.

DOI: 10.54254/2755-2721/75/20240521

[7] Sherif K, Azmy M, Alakkari K, et al. Deep Learning in IoT: An LSTM Approach for NDVI Forecasting[C]//2023 3rd International Conference on Electronic Engineering (ICEEM). IEEE, 2023: 1-7.
[8] Yu, H., Chen, Q. J., Yue, K. D., et al. (2023). Prediction of mine water inrush volume based on the ARIMA-LSTM model. Inner Mongolia Coal Economy, (24), 1-3. DOI:10.13487/j.cnki.imce.024619.