Forecasting Analysis of Agricultural Product Logistics Demand Based on Informer Neural Network: A Case Study of Central China Region (Postprint)

Authors: ZUO Min, Hu Tianyu, Dong Wei, Zhang Kexin, Zhang Qingchuan

Date: 2023-05-15T00:00:00+00:00

Abstract

Ensuring the stability of agricultural product logistics is critical to safeguarding livelihood issues. Forecasting agricultural product logistics demand is an important guarantee for the rational planning of agricultural product logistics stability. However, forecasting agricultural product logistics demand is inherently complex, as the forecasting process is influenced by various factors. Therefore, to ensure the accuracy of agricultural product logistics demand forecasting, it is necessary to consider multiple influencing factors. This study takes agricultural product logistics demand as the research object, constructs a neural network model for forecasting agricultural product logistics demand using the Informer neural network, and forecasts the agricultural product logistics demand of Henan Province, Hubei Province, and Hunan Province in Central China. Simultaneously, Long Short-Term Memory (LSTM) networks and Transformer neural networks are used to forecast the agricultural product logistics demand of the three Central China provinces, and the prediction results of the three models are compared. The comparison results demonstrate that the average percentage prediction test error of the Informer neural network model constructed in this study is 3.39%, which is lower than the 4.43% and 4.35% of the LSTM and Transformer neural network models, respectively. Moreover, the predicted values for the three provinces using this Informer neural network model are relatively close to the actual values: the 2021 forecasted value for Henan Province is 4185.33, the actual value is 4048.1, with an error of 3.389%; the 2021 forecasted value for Hubei Province is 2503.64, the actual value is 2421.78, with an error of 3.380%; the 2021 forecasted value for Hunan Province is 2933.31, the actual value is 2836.86, with an error of 3.340%. This demonstrates that the model yields relatively accurate results for forecasting agricultural product logistics demand in the three Central China provinces. The forecasted values

for the three provinces in 2023 are higher than those for 2021. Therefore, based on the logistics transportation supporting facilities in 2021, it is imperative to ensure logistics transportation efficiency and strengthen logistics transportation capacity to meet the growing logistics demand in Central China.

Full Text

Forecast and Analysis of Agricultural Products Logistics Demand Based on Informer Neural Network: Take the Central China Area as an Example

ZUO Min, HU Tianyu, DONG Wei*, ZHANG Kexin, ZHANG Qingchuan
(College of E-commerce and Logistics, Beijing Technology and Business University, Beijing 100048, China)

Abstract: Ensuring the stability of agricultural products logistics is key to safeguarding people' s livelihood. Forecasting agricultural products logistics demand is crucial for rational planning of logistics stability. However, such forecasting is complex due to multiple influencing factors. This study employs an Informer neural network to construct a forecasting model for agricultural products logistics demand, using data from Henan, Hubei, and Hunan provinces in Central China as a case study. The model' s performance is compared against LSTM and Transformer networks. Results show the Informer model achieves an average prediction error of 3.39%, outperforming LSTM (4.43%) and Transformer (4.35%). Predicted values for 2021 closely match actual values: Henan (predicted: 4185.33, actual: 4048.10, error: 3.389%), Hubei (predicted: 2503.64, actual: 2421.78, error: 3.380%), and Hunan (predicted: 2933.31, actual: 2836.86, error: 3.340%). The model accurately forecasts logistics demand for the three provinces. Predictions for 2023 indicate continued growth, necessitating improved logistics efficiency and capacity to meet rising logistics demand in Central China.

Keywords: agricultural products logistics; demand forecasting; Informer neural network; long short-term memory (LSTM); Transformer; deep learning

---

1 Introduction

With China' s rapid economic development, people' s income and consumption capacity have continuously improved, leading to increasingly diversified demand for agricultural products. Consumer expectations have shifted from simply having enough to eat to eating well, with high-quality agricultural products becoming more popular. This raises the bar for logistics service quality. Currently, agricultural products logistics faces two major challenges. First, China has a large flow of agricultural products, and transporting perishable goods to consumers quickly is a significant logistics challenge, especially during urban emergencies where transport capacity by road, rail, water, or air is limited. Rational planning of agricultural products logistics is therefore crucial for safeguarding people' s livelihood. Second, agricultural products logistics is highly seasonal and cyclical, with strong timeliness requirements during harvest seasons. For example, fruits and vegetables have short shelf lives, and inadequate logistics capacity planning leads to product losses and material waste. Accurate prediction of agricultural products logistics demand facilitates rational logistics planning, ensures long-term sustainable economic growth, and improves living standards.

In recent years, scholars have proposed innovative and practical methods for logistics demand forecasting. Yao et al. [1] selected data on fresh agricultural products output from 2012–2019 and used the grey prediction model GM(1,1) to forecast Liaoning province' s fresh agricultural products demand, achieving high prediction accuracy that reflects demand trends for 2020–2024. Wang and Li [7] applied the GM(1,1) model to forecast China' s agricultural cold chain logistics demand using 2014–2019 data, predicting 2020–2025 demand volumes. Zhang [2] used grey correlation analysis to construct a logistics demand system and built multivariate grey prediction, optimized neural network, and principal component regression models combined with Shapley value method to forecast Shandong province' s fresh agricultural products logistics demand for 2020–2025. Wang and Yan [3] constructed a Genetic Algorithm-Back Propagation (GA-BP) neural network model, leveraging genetic algorithms' global search capabilities to forecast Beijing' s urban agricultural cold chain logistics demand. Huangfu [5] identified six agricultural products closely related to cold chain logistics and used grey prediction models to forecast Mianyang city' s agricultural output for 2020–2029. Liu [6] applied grey prediction models to Liaoning' s fresh agricultural products cold chain logistics demand forecasting. Huang et al. [8] used GM(1,1) and BP neural network models to simulate and predict Guangdong province' s logistics demand, finding the BP neural network more accurate and stable. Xu et al. [9] compared five methods for Shandong province logistics demand forecasting, finding LSTM achieved the highest prediction precision. Van der Laan et al. [10] analyzed order planning processes for humanitarian logistics. Lu and Park [11] established a logistics demand potential model from spatial economic principles. Baisariyev et al. [12] evaluated the Bootstrap method for aviation spare parts forecasting. Zeng et al. [13] used grey prediction models for rural logistics demand forecasting in Guangdong province.

Agricultural products logistics demand forecasting is influenced by multiple factors, increasing prediction difficulty. Existing research primarily uses regression analysis for multiple influencing factors, but large data volumes and extensive historical data often yield unsatisfactory results. Many studies employ long time series historical data as the forecasting foundation. Time series analysis effectively reveals development patterns and dynamic relationships between phenomena, widely applied in hydrological forecasting [14-16], environmental pollution control [17,18], astronomy [19,20], and oceanography [21,22], making it suitable for agricultural products logistics demand prediction.

2 Data and Methodology

2.1 Data Sources

Data were collected from the *Henan Statistical Yearbook*, *Hunan Statistical Yearbook*, *Hubei Statistical Yearbook* (2017-2021), and the 2021 national economic and social development statistical bulletins of the three provinces. These data sources provide statistical support for forecasting agricultural products logistics demand in Central China' s three provinces.

2.2 Data Extraction

Based on 2017-2021 data, this study identified primary influencing factors on agricultural products logistics demand in the three provinces, considering local economic development level, per capita consumption capacity, industrial structure, logistics development level, agricultural supply factors, and humanistic factors. Sixteen indicators were selected as independent variables: regional GDP G1 (100 million yuan), per capita disposable income G2 (yuan), per capita consumption expenditure G3 (yuan), value-added of primary industry G4 (100 million yuan), value-added of secondary industry G5 (100 million yuan), value-added of tertiary industry G6 (100 million yuan), total crop sowing area G7 (thousand hectares), grain output G8 (10,000 tons), per capita grain output G9 (kg), annual agricultural products output G10 (10,000 tons), planting and breeding area G11 (thousand hectares), freight volume G12 (10,000 tons), railway operating mileage G13 (10,000 km), and highway mileage G14 (10,000 km).

Nine categories of agricultural products were selected: grain, meat, eggs, milk, oil crops, tea, fruits, forest products, and aquatic products. Forest products consist of seedlings and logs. Annual agricultural products output is the sum of these nine categories' annual outputs. Wood output volume uses a conversion coefficient of 0.8 (10,000 cubic meters = 8,000 tons) [23]. Planting and breeding area comprises total crop sowing area, orchard area, tea plantation area, aquatic products breeding area, and oil crops sowing area.

The dependent variable S (agricultural products logistics demand) is calculated as the product of each province' s permanent population and per capita agricultural products consumption [24], as shown in formula (1). Per capita agricultural products consumption is the sum of per capita consumption of nine product categories: grain, edible oil, vegetables and edible fungi, meat, poultry, aquatic products, eggs, milk, dried and fresh fruits, and sugar.

2.3 Data Normalization

Since input data values differ significantly across magnitude and dimension, maximum-minimum normalization was applied to eliminate dimensional effects and ensure comparability, as shown in formula (2). This preprocessing removes adverse effects from outlier samples.

---

3 Prediction Model Construction

Deep learning algorithms have been widely applied in agricultural disease identification, medical diagnosis, geological exploration, and industrial equipment fault diagnosis. However, time series forecasting requires strong extrapolation capabilities as distributions may change over time. Informer, an improved Transformer neural network, employs multi-head probabilistic sparse self-attention mechanisms to allocate larger weights to important features, reducing time complexity while improving prediction speed through generative decoding. This addresses long-term dependency issues in time series data, providing better extrapolation capability. Therefore, this study selected the Informer neural network to forecast agricultural products logistics demand in Central China' s three provinces.

Figure 1 [Figure 1: see original paper] illustrates the Informer neural network model for forecasting agricultural products logistics demand in Central China' s three provinces. The model consists of an encoder and decoder. Informer constructs its architecture through multi-head probabilistic sparse self-attention layers, multi-head attention mechanisms, and masked multi-head probabilistic sparse self-attention layers. The attention mechanism resembles human selective visual attention, selecting critical information from numerous inputs to solve long-term time series forecasting problems and improve prediction accuracy.

The attention mechanism calculation is shown in formula (3), involving three vectors: Query vector (Q ^{L_Q×d}), Key vector (K ^{L_K×d}), and Value vector (V ^{L_V×d}), where L_Q, L_K, and L_V represent linear transformation layers for queries, keys, and values, respectively, and d is input dimension. Let q_i, k_i, v_i denote the i-th row of Q, K, V. The attention for the i-th Query is defined as a probabilistic kernel smoother:

$$A(q_i, K, V) = \sum_{j=1}^{L_k} \frac{k(q_i, k_j)}{\sum_{l=1}^{L_k} k(q_i, k_l)} v_j = \mathbb{E}_{p(k_j|q_i)}[v_j]$$

The sparsity metric for the i-th Query is given by formula (4):

$$M(q_i, K) = \ln \sum_{j=1}^{L_k} e^{\frac{q_i k_j^T}{\sqrt{d}}} - \frac{1}{L_k} \sum_{j=1}^{L_k} \frac{q_i k_j^T}{\sqrt{d}}$$

Based on this metric, ProbSparse self-attention is obtained by allowing each key to attend to only u dominant queries, as shown in formula (5), where $\bar{Q}$ represents the sparse matrix of top-u queries:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{\bar{Q}K^T}{\sqrt{d}}\right)V$$

The Informer encoder processes input time series of agricultural products logistics demand indicators for the three provinces, capturing dependencies between demand indicators and influencing factors. The encoder comprises two identical stacks (Figure 2 [Figure 2: see original paper]). Stack1 is the main stack receiving the entire input sequence, while Stack2 performs the same operations but receives only half the input slice. Each stack consists of encoding and distillation layers. The encoding layer includes multi-head probabilistic sparse self-attention, feed-forward neural networks, residual connections, and normalization operations, as shown in formula (6):

$$O = \text{LayerNorm}(x + \text{Sublayer}(x))$$

where O represents layer normalization for each sublayer, Sublayer denotes processing functions for multi-head sparse self-attention and feed-forward networks, and LayerNorm is the normalization function.

The distillation mechanism in Informer enhances network robustness and reduces memory usage. It applies 1D convolution along the time dimension, followed by ELU activation and pooling to halve input length. The distillation layer has one fewer layer than the encoding layer. The distillation process from layer j to j+1 is shown in formula (7):

$$x_{j+1} = \text{MaxPool}(\text{ELU}(\text{Conv1d}([x_j]_{AB})))$$

where $[\cdot]_{AB}$ includes multi-head probabilistic sparse attention operations and attention block key operations, Conv1d denotes 1D convolution, MaxPool is max pooling, and ELU is the activation function calculated as formula (8):

$$\text{ELU}(x) = \begin{cases} x & \text{if } x > 0 \\ e^x - 1 & \text{if } x \leq 0 \end{cases}$$

The decoder input comprises two parts: encoder output features regarding logistics demand risks, and masked target demand sequences with zero placeholders. These connect to multi-head attention mechanisms and a fully connected layer to output predicted demand X_{T+1}.

4 Experiments and Results Analysis

4.1 Experimental Method

PyTorch served as the experimental platform. Detailed environment parameters are listed in Table 2 . Informer was used to forecast agricultural products logistics demand in Central China' s three provinces with 5-fold cross-validation. The dataset was divided into five approximately equal subsets; four subsets served as training data and one as test data in each iteration, repeated five times with different training/test splits.

Historical demand data and 16 influencing factors (including province and year) were used as inputs, with corresponding provincial agricultural products demand (10,000 tons) as output. LSTM and Transformer models served as benchmarks. LSTM, an excellent RNN variant, solves vanishing gradient problems and predicts long time series effectively. Transformer computes correlations directly without hidden layer propagation, enabling parallel computation and efficient GPU utilization.

4.2 Results Analysis

Training results for Informer, LSTM, and Transformer are shown in Figures 3 [Figure 3: see original paper]–5 [Figure 5: see original paper], respectively. Informer demonstrates superior performance, with loss function values approaching zero after 180 iterations. LSTM performs poorly, with large initial loss values that only approach zero after 350 iterations. Transformer shows moderate performance, with loss approaching zero around 150 iterations, increasing slightly, then stabilizing near zero after 280 iterations.

Table 3 presents absolute and percentage errors for each training/test set and 5-fold cross-validation averages. Informer' s absolute error is 30.37 in training sets and 36.61 in test sets, indicating small differences between predicted and actual values. After 5-fold cross-validation, Informer achieves average percentage errors of 3.07% (training) and 3.39% (testing). LSTM yields 4.37% (training) and 4.43% (testing), while Transformer achieves 3.07% (training) and 4.35% (testing). Informer' s lower test error demonstrates higher prediction accuracy and practical value.

Informer predictions for 2021 closely match actual values: Henan (predicted: 4185.33, actual: 4048.10, error: 3.389%), Hubei (predicted: 2503.64, actual: 2421.78, error: 3.380%), and Hunan (predicted: 2933.31, actual: 2836.86, error: 3.340%). These results validate the model' s effectiveness.

Using this validated model, 2023 demand forecasts are: Henan province 4217.13, Hubei province 2521.47, and Hunan province 2974.65. All three provinces show higher predicted values for 2023 compared to 2021, indicating that infrastructure improvements and enhanced transportation capacity are needed to meet growing logistics demand.

4.3 Ablation Experiment

To validate the scientific selection of 16 influencing factors, an ablation experiment was conducted using 2021 Henan data. Predictions using 15 factors (removing one factor each time) were compared against those using all 16 factors. The absolute error using 16 factors is 3.43%, lower than the 3.51% error when using 15 factors (Table 4), confirming that the 16-indicator selection is scientifically sound.

---

5 Conclusion

This study constructed an Informer neural network model to forecast agricultural products logistics demand in Henan, Hubei, and Hunan provinces, with LSTM and Transformer as comparative models. The 2021 prediction results demonstrate that Informer outperforms both alternatives, achieving closer alignment with actual values and providing practical value for rational logistics planning in Central China' s three provinces.

The 2023 forecasts suggest continued logistics demand growth, requiring infrastructure development and improved emergency logistics coordination mechanisms. A limitation of this study is the high data indexing difficulty and substantial analytical workload due to numerous indicators. Future research should focus on data simplification while maintaining predictive accuracy.

Conflict of Interest Statement: The authors declare no conflicts of interest regarding this research.

References

[1] YAO G X, YANG Y, XU J, et al. Forecasting of logistics demand for agricultural products under the background of COVID-19—A case study of Xinghua city[J]. Logistics sci-tech, 2021, 44(9): 91-95.

[2] ZHANG J L. Logistics demand forecast of fresh agricultural products in Shandong province based on combination model[D]. Jinan: Shandong University, 2021.

[3] WANG X P, YAN F. Prediction of cold chain logistics demand for agricultural products in Beijing based on GA-BP model[J]. Mathematics in practice and theory, 2019, 49(21): 17-27.

[4] SHYLA M K, SHIVA K K, KUMAR D R. Image steganography using genetic algorithm for cover image selection and embedding[J]. Soft computing letters, 2021, 3: ID 100021.

[5] HUANGFU H J. Research on cold chain logistics demand forecast of fresh agricultural products in Mianyang city[D]. Mianyang: Southwest University of Science and Technology, 2021.

[6] LIU W B. Research on demand forecast of cold chain logistics of fresh agricultural products in Liaoning province[J]. China circulation economy, 2022(4): 11-13.

[7] WANG M, LI X. Demand forecasting of agricultural cold chain logistics based on metabolic GM (1, 1) model[J]. IOP conference series: Earth and environmental science, 2021, 831(1): ID 012018.

[8] HUANG L, XIE G, ZHAO W, et al. Regional logistics demand forecasting: A BP neural network approach[J]. Complex & Intelligent Systems, 2021: 1-16.

[9] XU X Y, YANG H M, LYU X K, et al. Comparative research on forecast of logistics demand in Shandong Province based on different models[J]. Packaging engineering, 2022, 43(23): 207-215.

[10] VAN DER LAAN E, VAN DALEN J, ROHRMOSER M, et al. Demand forecasting and order planning for humanitarian logistics: An empirical assessment[J]. Journal of operations management, 2016, 45: 114-122.

[11] LU B, PARK N. A study on model of regional logistics requirements prediction[J]. Journal of Korean navigation and port research, 2012, 36: 553-559.

[12] BAISARIYEV M, BAKYTZHANULY A, SERIK Y, et al. Demand forecasting methods for spare parts logistics for aviation: A real-world implementation of the Bootstrap method[J]. Procedia manufacturing, 2021, 55: 500-506.

[13] ZENG M L, LIU R M, GAO M, et al. Demand forecasting for rural E-commerce logistics: A gray prediction model based on weakening buffer operator[J]. 2022, 2022: ID 9846253.

[14] ANSHUKA A, CHANDRA R, BUZACOTT A J V, et al. Spatio temporal hydrological extreme forecasting framework using LSTM deep learning model[J]. Stochastic environmental research and risk assessment, 2022, 36(10): 2961-2986.

[15] LIU J J, YUAN X, ZENG J H, et al. Ensemble streamflow forecasting over a cascade reservoir catchment with integrated hydrometeorological modeling and machine learning[J]. Hydrology and earth system sciences, 2021, 26(2): 265-278.

[16] XU C W, WANG Y Z, FU H, et al. Comprehensive analysis for long-term hydrological simulation by deep learning techniques and remote sensing[J]. Frontiers in earth science, 2022, 10: ID 875145.

[17] DU S D, LI T R, YANG Y, et al. Deep air quality forecasting using hybrid deep learning framework[J/OL]. arXiv:1812.04783, 2018.

[18] JANARTHANAN R, PARTHEEBAN P, SOMASUNDARAM K, et al. A deep learning approach for prediction of air quality index in a metropolitan city[J]. Sustainable cities and society, 2021, 67: ID 102720.

[19] PENG J H, CAO H L, ALI Z, et al. Intelligent reflecting surface-assisted interference mitigation with deep reinforcement learning for radio astronomy[J]. IEEE antennas and wireless propagation letters, 2022, 21(9): 1757-1761.

[20] SEDAGHAT N, ROMANIELLO M, CARRICK J E, et al. Machines learn to infer stellar parameters just by looking at a large number of spectra[J]. Monthly notices of the Royal Astronomical Society, 2021, 501(4): 6026-6041.

[21] LIU H L, LIN P, ZHENG W P, et al. A global eddy-resolving ocean forecast system in China-LICOM Forecast System (LFS)[J]. Journal of operational oceanography, 2021, 16(1): 15-27.

[22] ZHU X M, ZU Z Q, REN S H, et al. Improvements in the regional South China Sea operational oceanography forecasting system (SCSOFSv2)[J]. Geoscientific model development, 2022, 15(3): 995-1015.

[23] Comprehensive Encyclopedia[EB/OL]. (2020-10-29)[2023-01-30]. https://www.yebaike.com/22/1574584.ht

[24] ZHOU H Y, ZHANG S H, PENG J Q, et al. Informer: Beyond efficient transformer for long sequence time-series forecasting[J/OL]. arXiv:2012.07436 [cs.LG], 2020.

[25] SONG Y, CAI Z P. Normalized method of intrusion detection data based on information theory in big data environment[J]. Journal of Wuhan university (natural science edition), 2018, 64(2): 121-126.

Note: Figure translations are in progress. See original paper for figures.