Trends in Social Sciences and Humanities Research
ISSN: 2959-9938
DOI: 10.61784/tsshr3015

LSTM MODEL ENHANCED BY KOLMOGOROV-ARNOLD NETWORK: IMPROVING STOCK PRICE PREDICTION ACCURACY

XiaoXuan Yao
School of Mathematics and Statistics,Guangxi Normal University, Guilin 541006, Guangxi, China.
Corresponding Email: xiaoxuan.yao@qq.com

Abstract: This study addresses the accuracy limitations of traditional LSTM models in stock price prediction by proposing an innovative hybrid model, the LSTM-KAN model. Combining the classical Long Short-Term Memory (LSTM) network with the Kolmogorov-Arnold Network (KAN), this model aims to enhance the performance of the LSTM model in predicting complex financial time series by leveraging the highly nonlinear expressive power of KAN. Through empirical analysis of historical stock data, a comparative study is conducted to examine the differences between the LSTM-KAN model and the basic LSTM model in terms of prediction error, stability, and generalization capability. The results demonstrate that the LSTM-KAN model significantly reduces prediction errors in most cases, improving prediction accuracy and providing new perspectives and tools for stock market analysis.
Keywords: LSTM; Kolmogorov-Arnold Network; Stock price prediction; Time series analysis; Nonlinear models

1 INTRODUCTION

As global financial markets become increasingly complex and volatile, accurate stock price prediction has become a critical factor for investor decision-making. Traditional prediction methods, such as technical analysis and fundamental analysis, often fall short when faced with high-dimensional, nonlinear financial time series data that contain significant noise. In recent years, the rapid development of machine learning and deep learning technologies has brought new breakthroughs to this field. Among these, Long Short-Term Memory (LSTM) networks, a special form of recurrent neural networks, have been successfully applied to various prediction problems, including stock price prediction, due to their excellent performance in handling sequential data.
However, while LSTM excels at capturing long-term dependencies, its prediction accuracy and generalization ability still have room for improvement in the highly nonlinear financial market environment. The Kolmogorov-Arnold Network (KAN), a novel neural network based on the Kolmogorov-Arnold representation theorem, offers strong nonlinear expressive power with relatively low model complexity, providing a new approach to solving such complex problems.
Currently, scholars at home and abroad have extensively explored the application of LSTM in stock price prediction and have achieved certain results. Some studies have attempted to combine LSTM with other machine learning algorithms or deep learning models to improve prediction accuracy, such as ensemble learning and convolutional neural networks (CNN)[1]. Despite this, research on combining KAN with LSTM for stock price prediction remains relatively sparse, especially in terms of model construction, parameter optimization, and performance evaluation, leaving considerable room for further exploration.
Therefore, this study aims to design and implement an innovative hybrid model, the LSTM-KAN model, to integrate the powerful memory capability of LSTM with the nonlinear expressive advantages of KAN, overcoming the limitations of a single model in handling complex financial time series data. Specific research objectives include: constructing the LSTM-KAN hybrid model framework, exploring the effective integration mechanism of the two models; validating the improvement in prediction accuracy, stability, and generalization capability of the LSTM-KAN model using historical stock data; analyzing the specific impact of KAN's introduction on the performance of the LSTM model, and exploring the underlying theoretical basis.

2 THEORETICAL FOUNDATIONS

2.1 Long Short-Term Memory Networks (LSTM)

LSTM is a special type of recurrent neural network (RNN) proposed by Hochreiter and Schmidhuber in 1997. It aims to address the vanishing and exploding gradient problems commonly encountered by traditional RNNs when dealing with long-term dependencies. LSTM introduces memory cells (cell states), input gates, forget gates, and output gates to control the flow of information, allowing the model to effectively learn patterns in long sequence data. Memory cells can accumulate and retain important information over long periods, while the gate mechanisms control the reading, writing, and forgetting of information, thus enhancing the model's ability to handle sequential data[2].

2.2 Kolmogorov-Arnold Representation Theorem

The Kolmogorov-Arnold representation theorem is a significant result in mathematics, proposed by Andrey Kolmogorov and Vladimir Arnold. It asserts that any continuous multivariable function can be composed of a series of simple functions, formally represented as nested single-variable functions. Equation (1) represents the mathematical formulation of the Kolmogorov-Arnold representation theorem. This theory inspired the design of the Kolmogorov-Arnold Network (KAN), where each node can be regarded as a highly nonlinear mapping. The network approximates complex function relationships by combining these mappings. The advantage of KAN lies in its potential for efficient expression and the compactness of its model structure. KAN is based on a supervised learning task aimed at approximating a function f, which maps the inputs x of all data points to their outputs y. This method uses the Kolmogorov-Arnold theorem to decompose any multivariate function into a series of single-variable functions and summations. The equation indicates that for each input dimension x_p, there is a univariate function h_p that aggregates the outputs of these univariate functions, as expanded in Equation (2).

$$f(x_1, \ldots, x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p} \left(x_p \right) \right) \tag{1}$$

$$f(x) = \sum_{i_{L-1}=1}^{n_{L-1}} \phi_{L-1,i_L,i_{L-1}} \left(\sum_{i_{L-2}=1}^{n_{L-2}} \cdots \left(\sum_{i_2=1}^{n_2} \phi_{2,i_3,i_2} \left(\sum_{i_1=1}^{n_1} \phi_{1,i_2,i_1} \left(\sum_{i_0=1}^{n_0} \phi_{0,i_1,i_0} \left(x_{i_0} \right) \right) \right) \right) \cdots \right) \tag{2}$$

2.3 Structure and Characteristics of KAN

KAN's core lies in its structural design, which differs from traditional neural networks by treating the activation function as a part of the model for learning. This means that each connection in the network not only has weight parameters but also has an activiation function that is trainable, allowing the network to automatically discover the most suitable non-linear transformation. This provides greater flexibility in handling non-linear problems, especially in scenarios with complex data distributions or highly non-linear relationships. The left side of Figure 1 shows the activation symbols flowing through the network, while the right side illustrates the activation function parameterized as B-splines, enabling switching between coarse and fine-grained grids[3].

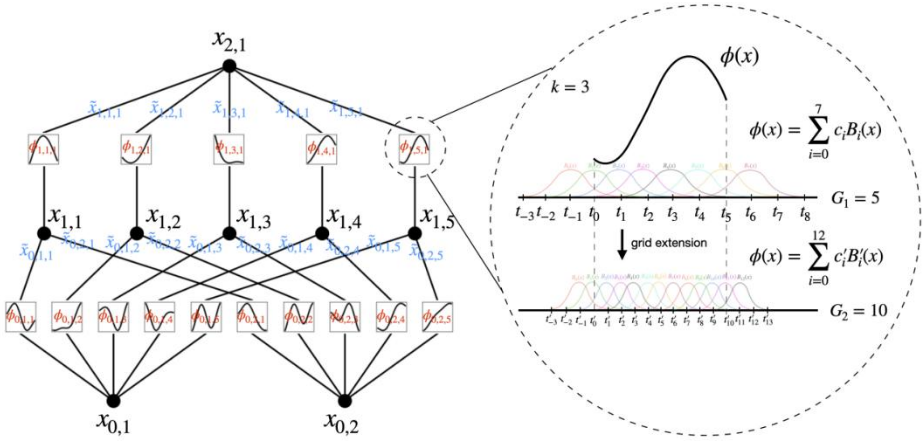

Figure 1 KAN Structure Diagram

2.4 Overview of Stock Price Prediction Methods

Stock price prediction is a significant topic in financial engineering and econometrics, involving various statistical and machine learning methods. Early methods primarily included time series analysis (such as ARIMA models), moving averages, and exponential smoothing. With the increase in computational power and data volume, machine learning methods have become increasingly popular, including support vector machines (SVM), random forests, and neural networks. In recent years, deep learning techniques, particularly RNNs and their variants (such as LSTM and GRU), have gained attention for their ability to better capture sequential data patterns[4].

3 CONSTRUCTION OF THE LSTM-KAN MODEL

3.1 Model Design and Parameter Configuration

The core design of the LSTM-KAN model is to combine the memory capabilities of LSTM in handling sequential data with the advantages of KAN in expressing complex nonlinear relationships. The LSTM layers are responsible for capturing long-term dependencies in time series, while the KAN layers further refine this information by utilizing flexible basis function activation and piecewise polynomial weights to adapt to the highly nonlinear patterns in stock price prediction. The model's parameter configuration includes input size, hidden size, number of layers, output size, dropout ratio, and whether to use batch-first mode for the LSTM. Additionally, the KAN layer's configuration involves grid size, the order of piecewise polynomials, scaling noise, and scaling of activation functions. These parameters collectively determine the model's complexity and adaptability.

3.2 Code Implementation

In the code implementation, the LSTM-KAN model class is first defined, inheriting from nn.Module. During model initialization, the LSTM layer is set up, and the KAN model (via the KAN class) is innovatively included as a subsequent processing layer for the LSTM output, replacing the conventional fully connected layer. In the forward propagation function, the final hidden state of the LSTM is used as input to the KAN layer, designed to leverage KAN's characteristics to further enhance nonlinear expression in predictions[5].

3.3 KAN Layer Implementation

The KAN layer's implementation is based on piecewise polynomial weights and optional independent scaling factors. It internally manages grid generation, weight initialization, and regularization loss calculation. The computation of piecewise polynomial weights considers grid steps, scaling noise, and the influence of the base activation function, achieved through methods like curve2coeff. Additionally, the KAN layer supports the computation of regularization losses, including L1 regularization and entropy regularization terms, to prevent overfitting and promote the model's generalization capabilities.

3.4 Model Flexibility and Future Research

By combining LSTM with KAN, the LSTM-KAN model can handle both the long-term dependencies of time series data and the nonlinear complexities inherent in financial data, potentially achieving lower prediction errors in stock price forecasting tasks. The model's flexibility and customizability, such as adjusting grid size and selecting activation functions, provide extensive room for adaptation to different market characteristics and datasets. Future research can further explore model performance under different parameter settings and how the depth and complexity of the KAN layer affect prediction accuracy and computational efficiency.

4 EXPERIMENT DESIGN AND DATA PROCESSING

4.1 Experiment Objectives

The aim of this chapter is to evaluate the performance of the LSTM-KAN model compared to the pure LSTM model in stock price prediction tasks. The effectiveness of the KAN layer in enhancing nonlinear expression capabilities will be assessed. The study uses stock data from the Shanghai Stock Exchange Composite Index (SSE) spanning from January 5, 1998, to June 2, 2020, to explore the models' potential in predicting actual financial data.

4.2 Data Preprocessing

First, daily trading data of the SSE within the specified date range were collected, including eight variables: opening price, closing price, highest price, lowest price, price change, price change percentage, trading volume, and trading amount. Outliers and missing data were removed to ensure data integrity. All numerical features were normalized to the same scale to facilitate model training. A fixed random seed was used to ensure reproducibility of the experiments. The dataset was split into a training set (80%) and a test set (20%). The input for the models consists of data from the previous five days (sequence_length=5), predicting the closing price on the sixth day.

4.3 Model Construction and Training

A basic LSTM model was constructed with an input dimension (input_size) of 8 and an output dimension (output_size) of 1 (i.e., the predicted closing price), using a five-day data sequence as input. Based on this, an LSTM model with an integrated KAN layer was constructed. The parameter configuration for the KAN layer, as described in previous chapters, optimizes its flexibility and efficiency in nonlinear expression. Both models were trained for 100 epochs, recording the loss for each epoch, and optimized using the Adam optimizer.

4.4 Model Evaluation

The performance of the models was evaluated primarily through Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). These metrics directly reflect the deviation between the predicted and actual values. The RMSE and

MAE of the LSTM and LSTM-KAN models on the test set were compared to analyze the performance improvement brought by the KAN layer. Additionally, the trend of loss during the training process was observed to assess the models' convergence and the risk of overfitting.

5 RESULTS ANALYSIS AND DISCUSSION

5.1 Summary of Experiment Results

In this study, we trained both a basic LSTM model and an LSTM-KAN model, using trading data from the Shanghai Stock Exchange Composite Index (SSE) from January 1998 to June 2020 to evaluate their predictive performance. Data preprocessing included normalization and splitting the dataset into 80% training and 20% testing sets. The models used the previous five days of trading data to predict the closing price on the sixth day. During training, we recorded the loss for each epoch. After 100 epochs, we assessed the models' prediction accuracy by calculating the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) on the test set.

5.2 Comparison of Results

The loss variations over epochs for both LSTM and LSTM-KAN models are plotted, with blue representing the LSTM model and yellow representing the LSTM-KAN model, as shown in Figure 2. Initially, the LSTM model's loss decreases rapidly, but from the fourth epoch onwards, the LSTM-KAN model's loss decreases faster, showing better performance. A comparison of the LSTM and LSTM-KAN predictions against actual values on the test set is illustrated in Figure 3. The blue line represents the LSTM model, and the yellow line represents the LSTM-KAN model. When zoomed in, it is evident that the LSTM-KAN model predictions are closer to the actual values. As shown in Table 1, the basic LSTM model stabilizes during training, with a final RMSE of 0.065467 and an MAE of 0.052623 on the test set. This indicates that the LSTM model can learn the patterns in historical data relatively well but may have limitations when dealing with complex nonlinear relationships. The LSTM model combined with the KAN layer also shows good convergence characteristics during training. However, compared to the basic LSTM model, the LSTM-KAN model reduces the test set RMSE to 0.008226 and the MAE to 0.005742, demonstrating a significant performance improvement. This confirms the KAN layer's ability to handle nonlinear financial data, particularly in capturing complex price fluctuation patterns.

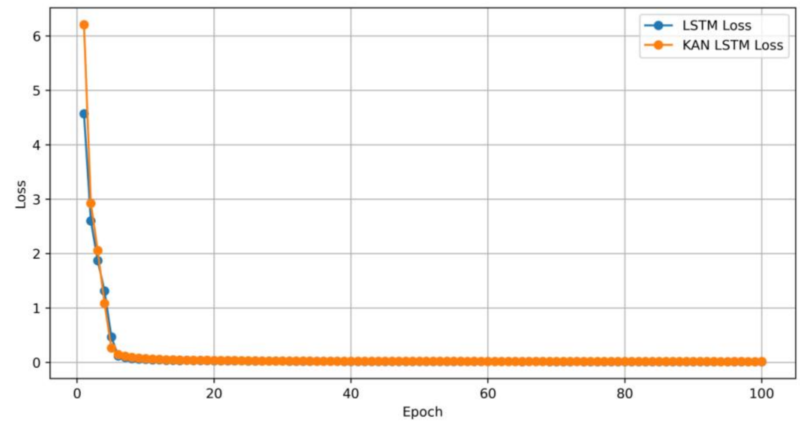

Figure 2 Loss Comparison

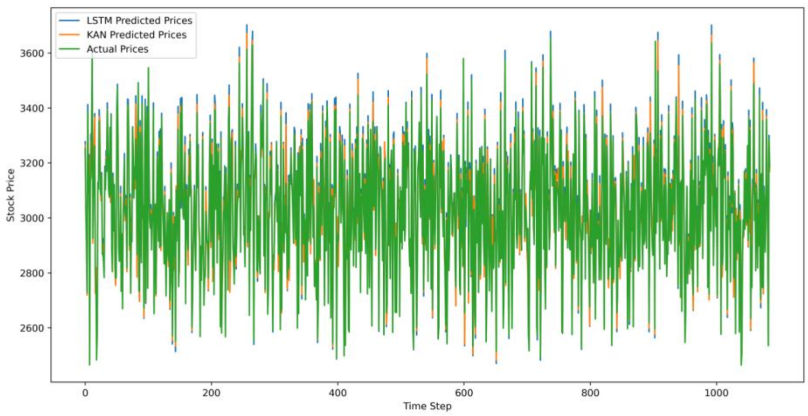

Figure 3 Predicted vs Actual Stock Prices

Table 1 Metrics Comparison

Metrics	LSTM	LSTM-KAN
RMSE	0.065467	0.008226
MAE	0.052623	0.005742

5.3 Discussion

In analyzing the results, we found that the LSTM-KAN model significantly reduced prediction errors on the test set, indicating that the introduction of the KAN layer indeed enhanced the model's nonlinear expression capability. However, it is important to note that the increased training complexity of the model might lead to a risk of overfitting. By observing the loss curves during training, we found that the LSTM-KAN model exhibited strong generalization ability, maintaining good performance on the test set despite the increased model complexity. Additionally, the parameter configurations of the KAN layer, such as grid size and the order of piecewise polynomials, significantly impact the model's performance, requiring further research and optimization. Finally, while the experimental results performed well on SSE data, the effectiveness might be influenced by the dataset, necessitating validation on other markets or datasets.

6 CONCLUSION AND FUTURE DIRECTIONS

6.1 Conclusion

This study explored the potential of the Kolmogorov-Arnold Network (KAN) to enhance prediction accuracy by comparing the performance of a basic LSTM model and an LSTM-KAN model in the task of stock price prediction. Through empirical analysis of historical data from the Shanghai Stock Exchange Composite Index (SSE) from January 1998 to June 2020, we concluded the following:

1) The LSTM-KAN model outperformed the traditional LSTM model in stock price prediction, significantly reducing prediction errors and improving prediction accuracy.
2) The introduction of the KAN layer enhanced the model's nonlinear expression capability, allowing it to better capture complex stock price fluctuation patterns, especially in handling nonlinear and complex financial time series data.
3) Although the training complexity of the LSTM-KAN model increased, reasonable parameter configurations and training strategies enabled the model to maintain good generalization ability, avoiding the risk of overfitting.

6.2 Future Prospects

Despite the positive results of this study, several future research directions deserve further exploration:

1) Application to Different Markets and Financial Products: Apply the LSTM-KAN model to different markets or various types of financial products to test its universality and robustness across different datasets.
2) Optimization of KAN Layer Parameters: Further study the optimization strategies for KAN layer parameters to find the best parameter combinations, enhancing the model's efficiency and prediction accuracy.
3) Incorporation of External Information: Explore the inclusion of additional macroeconomic indicators, news sentiment analysis, and other external information to improve the model's prediction accuracy and practical applicability.
4) Development of Real-time Prediction Systems: Develop real-time prediction systems to enable the model to adapt promptly to market dynamics, which is crucial for high-frequency trading and real-time risk management.
5) Improving Model Interpretability: Work on enhancing the interpretability of the model to make it more understandable and actionable for end-users.

COMPETING INTERESTS

The author have no relevant financial or non-financial interests to disclose.

FUNDING

This article was supported by the project titled "Identification, Measurement, and Governance of Relative Poverty in Rural Guangxi from the Perspective of Rural Revitalization" (Project No. 23BTJ001) funded by the Guangxi Philosophy and Social Science Office. The project titled "Reform of the Evaluation Model of Physical Education Entrance Examination in Guangxi Based on Data Mining" (Project No. 2022ZJY2311) funded by the Guangxi Zhuang Autonomous Region Admissions Examination Institute.

REFERENCES

[1] Chen S, Ge L. Exploring the attention mechanism in LSTM-based Hong Kong stock price movement prediction. Quantitative Finance, 2019, 19(9): 1507-1515.
[2] Kim T, Cho S. Predicting residential energy consumption using CNN-LSTM neural networks. Energy, 2019, 18272-81.
[3] Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljačić, Thomas Y. Hou, Max Tegmark. KAN: Kolmogorov-Arnold Networks. arXiv preprint arXiv: 2024. 19756.
[4] Wenchao G, Zhigang L, Chuang G, et al. Stock price forecasting based on improved time convolution network. Computational Intelligence, 2022, 38(4): 1474-1491.
[5] Zhang, Zhiping, Wang, et al. Design of financial big data audit model based on artificial neural network. International Journal of System Assurance Engineering and Management, 2021, (prepublish): 1-10.