Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 16, No. 1, 2025

Time series anomaly detection hybrid model based on SARIMA and LSTM

Wenjuan Wu[a], Mingyuan Guo[b], Shangfei Wang[c], Jinying Han[d]

Department of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454003, China
[a]212209010029@home.hpu.edu.cn, [b]212209010004@home.hpu.edu.cn, [c]212209020022@home.hpu.edu.cn, [d]212209020042@home.hpu.edu.cn

Abstract: For complex time series data, it is difficult for short-time memory neural networks to capture multiple factors comprehensively, and seasonal differential autoregressive moving average model has limitations in dealing with nonlinear relationships and outliers. To solve these problems, a hybrid time-series data anomaly detection model combining seasonal differential autoregressive moving average model (SARIMA) and short and long-time memory neural network (LSTM) is proposed. First, SARIMA model was used to initially fit key performance index data to capture the linear trend and seasonal pattern in the series. Then the sliding window technique is used to convert the fitted residual data into supervised learning data format, and the input dimension of the LSTM model is determined accordingly. Finally, an improved sparse regularization multi-head attention mechanism is proposed to add into the LSTM model. This mechanism realizes the sparsity of attention weights by introducing L1 regularization into the standard multi-head attention mechanism, and takes the output residual of the SARIMA model as the input of the improved LSTM model for secondary prediction and anomaly detection. The proposed hybrid model is compared with the public data set, and the experimental results show that the SARMI-LSTM hybrid model has a good performance in the accuracy of anomaly detection.

Keywords: Time series anomaly detection hybrid model based; SARIMA; LSTM.

1. Introduction

Time series refers to a series of index values arranged in time sequence, widely existing in finance, power, control engineering and other fields. Each observation in the time series is accompanied by a corresponding time stamp, ensuring that the data is arranged in the order in which it was collected. In order to find outliers or outliers in data in a timely manner, anomaly detection [1-4] method in data mining has attracted much attention in recent years. The use of this method is helpful to find and eliminate faults in a timely manner, and is of great significance for ensuring the stability of the system.

At present, the time series modeling method based on deep learning [5-6] has shown good results in real-time state detection and fault detection of data, which not only reduces labor costs, but also improves work efficiency. In this context, a large number of deep learning research results have appeared in the field of time series analysis[7-8]; For example, Bai et al.[9] used a recurrent neural network based on attention mechanism to predict landslide displacement; Ren and Li[10] predicted the daily sea ice concentration on a subseasonal scale in the Arctic region by using deep learning models; Sheerza A[11] uses deep learning algorithms and multi-layer perceptrons to detect DDoS attacks effectively and efficiently; Nahid F A et al.[12] used convolutional Long Short-Term Memory (LSTM) to predict wind speed; Novel[13] proposes a framework based on long short-term memory network (LSTM) and deep generation model for predicting collective context anomalies; Xu et al. proposed an algorithm based on long short-term memory network model to predict the typhoon path; Zheng and Bai propose an algorithm based on Keras long short-term memory network model to predict spatial quality index; Based on the LSTM model, Karevan et al.[14] established the transformed Long and short memory network model (T-LSTM) and applied it to weather prediction in practice; Zhang et al.[15] constructed a Bi-LSTM model to predict PM2.5 concentration in Beijing. In addition, GRU (Gate Recurrent Unit) model, as a variant of LSTM model [16-18], has been applied to the prediction of groundwater level by Gharehbaghi Amin [19] et al., and has achieved good results.

Different model variants have different prediction effects, so it is very important to select the model reasonably according to the characteristics of the study sequence. Liu Ying and Chen Xudong constructed a prediction model of Chongqing AQI based on multidimensional multi-step LSTM, which is superior to BP neural network model, RF-BP neural network model and RF model in both prediction accuracy and fitting effect. Jasmir J[20] et al. used recurrent neural networks to classify public clinical trial protocols and achieved good results.

It can be seen that deep learning algorithm has certain advantages in sequence prediction, especially in the processing of time series data containing nonlinear factors. However, in the face of more complex time series analysis involving multiple influencing factors, especially when the data contains both linear and nonlinear components, these methods still have shortcomings, and a single prediction model often fails to fully capture all the characteristic information in the data.

In contrast, SARIMA (Seasonal Autoregressive Integrated Moving Average) model can effectively deal with seasonal and non-seasonal patterns in time series data. And it can adapt to the autocorrelation and non-stationarity of data. However, SARIMA model has limitations in dealing with nonlinear relationships and outliers, and is complicated in model selection and parameter estimation. In view of this, a time series anomaly detection model combining SARIMA and LSTM is proposed in order to extract more comprehensive information of complex time series data and reduce information loss. In order to deal with the problems of overfitting risk and seasonal feature capture failure, the sparse regularization multi-head attention mechanism is introduced. By applying sparse regularization to the multi-head attention mechanism, the model is able to focus on key features, thereby reducing overfitting and enhancing generalization.

The structure of this paper is as follows: The second section describes the basic principles of SARIMA and LSTM models, and introduces the attention mechanism; In the third section, the structural design of SARMI-LSTM combined model, the improvement of LSTM model and the introduction of sparse regularization multi-head attention mechanism are discussed in detail. The fourth section shows the experimental results and analysis. Finally, the fifth section summarizes the full text.

2. Related Work

2.1. SARIMA Model

The SARIMA model is a time series prediction model that is an extension of the ARIMA (Autoregressive Integrated Moving Average) model specifically designed to process time series data with seasonal patterns. The SARIMA model consists of a combination of the following three parts:

(1) The part of Seasonal: Similar to the AR and MA parts, is used to capture seasonal patterns in time series data.

(2) The part of autoregressive: Used to capture autocorrelations in time series data.

(3) The part of integrated: Used to differentiate time series data to make it smooth.

The ARIMA model is usually represented as ARIMA (*p*, *d*, *q*) and is one of the most common models used for time series prediction. Compared with the ARIMA model, the SARIMA model extends the period in which *s* is seasonal and is related to the seasonal period. SARIMA model is generally expressed as SARIMA (*p*, *d*, *q*) × (*P*, *D*, *Q*) [*s*], where *p* is the autoregressive order, *d* is the difference order, *q* is the moving order, *P* is the seasonal autoregressive order, *D* is the seasonal difference order, *Q* is the seasonal moving average order, and *s* is the seasonal period.

$$Y_t = \theta + \sum_{i=1}^{p} \alpha_i * Y_{t-1} + \sum_{j=1}^{q} \beta_j * u_{t-j} + \gamma S * Y_{t-s} + u_t$$

Where θ is a constant term; Y_t is a time series; α_i represents the ith parameter of the lag i; β_j is the jth parameter associated with the lag j, and u_t is the residual of time t. γS is the seasonal coefficient s.

By using the SARIMA model to fit the time series data, find the SARIMA parameters that best fit the data. After finding the optimal parameters, the model is fitted to the training data for subsequent analysis or prediction.

2.2. LSTM Model

With the rapid development of deep learning algorithms, LSTM has been widely used in various fields of society. As a variant of Recurrent Neural networks (RNN), LSTM inherits the ability of RNN to process time series data, and can use historical data to override the short-term memory and gradient disappearance/explosion problems of RNN. LSTM can learn and retain long-term dependent information through its unique memory module, so that the model has a better performance when processing time series data.

In recent years, models based on LSTM and its variants have been widely used in various research fields, solving many problems that traditional artificial intelligence algorithms are difficult to overcome. For example, Bi-LSTM, a bi-directional long short-term memory network, can not only leverage information from the past, but also capture information from the future to provide a fuller context. In addition, stacking LSTMS increases network depth by building multi-layer LSTM networks, helping models learn more complex features and patterns. When working with serial data, this deeper network structure can better capture long-distance dependencies in time series.

In conclusion, LSTM network model is widely used in various time series prediction research because of its good performance and ability to deal with the long-term dependence of time series

2.3. Attention Mechanism

Attention mechanism is an important part of deep learning, which is widely used in many fields such as natural language processing and computer vision. The concept of attention mechanisms originally originated in the field of computer vision[21], which mimics the workings of the human visual system by enabling neural networks to focus attention on important areas or features when processing input data. By autonomously and selectively focusing information, attention mechanisms can improve processing efficiency and performance. Therefore, combining attention mechanisms with other network models can help solve some problems that are difficult to deal with in traditional ways.

The study of multi-head attention mechanism is derived from self-attention mechanism and aims to further enhance the expressiveness and generalization ability of the model. This mechanism uses multiple independent attention heads, calculates the attention weights separately, and concatenates or weights the results to improve the model's attention to different information. The attention mechanism in deep learning has the advantages of autonomy, enhancing focus information, reducing other noise and redundant information, and autonomously allocating more attention to matching targets.

3. SARIMA-LTSM Hybrid Model

The performance of the LSTM model is closely related to the length of the input sequence. When the sequence is short, the ability of the model to capture complex dynamic features in the long sequence is easily limited. To overcome this limitation, a hybrid model combining SARIMA and LSTM is proposed. The design aims to combine the advantages of the two models to achieve comprehensiveness and accuracy of feature extraction.

In order to enhance the prediction ability of the model, the stack LSTM structure is introduced in the LSTM part. This approach on the one hand gives the model a deeper hierarchy and better learning ability, and on the other hand allows it to flexibly integrate with other types of neural networks or feature extraction methods to build more complex and powerful models. In the stacking process, each layer of neural network can capture the timing information and pass it to the next layer, so that the model can gradually integrate different dimensional features through layer by layer.

In order to accurately capture the timing features that are crucial to the prediction results, a sparse regularized multi-head attention mechanism is introduced in the stacked LSTM model. The mechanism assigns different weights to each temporal feature to enhance the sensitivity and capture ability of the model for the global key temporal features. This method can not only reduce the risk of overfitting, but also provide a more linear and regular prediction capability, so that the model has a stable prediction performance.

The overall frame diagram of the SARMIA-LSTM hybrid model is shown as follows:

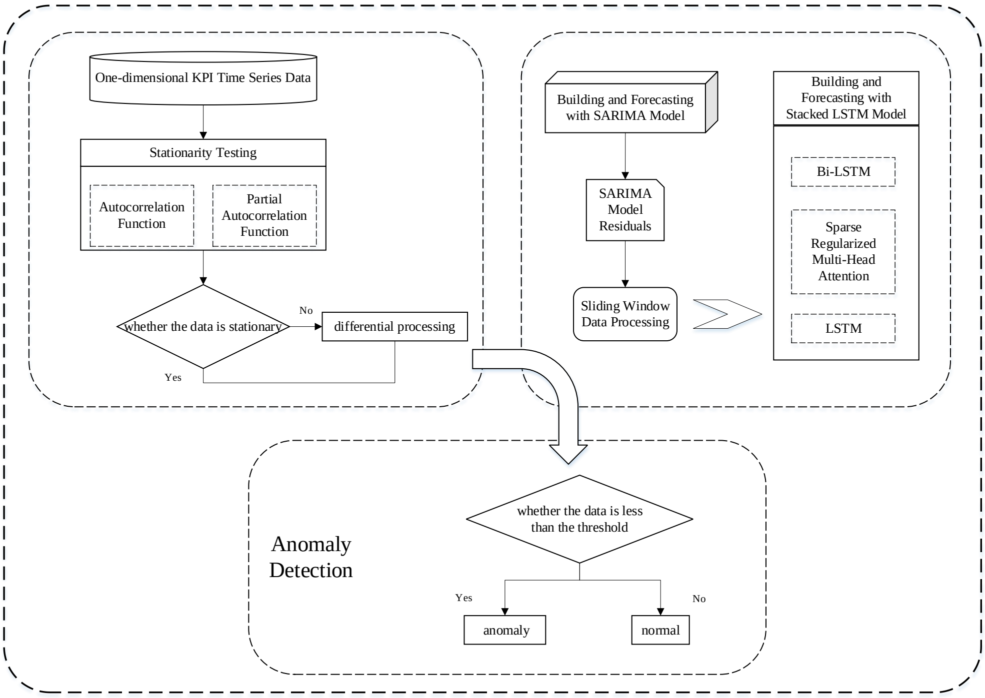

Figure 1. Model overall frame diagram

The SARMIA-LSTM hybrid model mainly includes the following components:

1)Input and output: The model takes a continuous time series as input, and transforms the original data into a fixed-length sub-sequence X by sliding window method, whose window size and step size determine the granularity of data transformation. For example, when the window size is set to 3, the sliding window can get a corresponding set of output y each time it moves, which changes the input data from unsupervised data to supervised data. In addition, after the processing process of the SARMI-LSTM model, the abnormal detection mechanism is integrated in this paper, and the normal or abnormal detection results of the time series data can be directly output by setting thresholds, thus realizing the dual functions of prediction and anomaly detection.

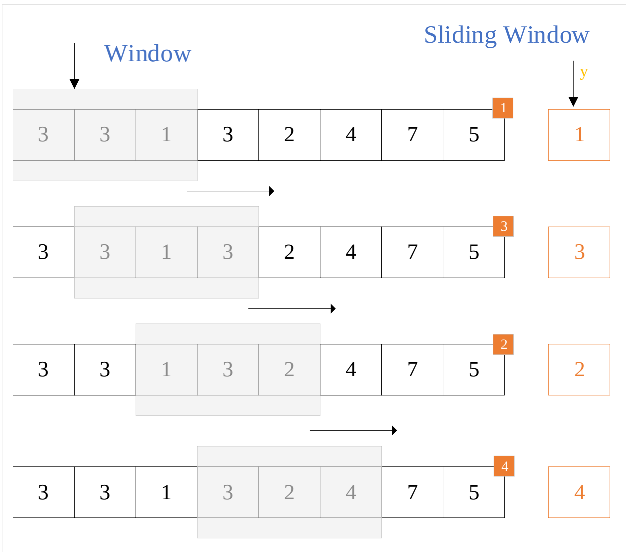

Figure 2. Sliding window

2) Stacked LSTM model structure: Each layer of LSTM model has the same number of hidden nodes, which ensures the consistency of the model structure and the efficiency of parameter optimization. By taking the output feature vector sequence of each layer as the input of the subsequent layer, the higher-level time series features are extracted layer by layer, so as to better capture the nonlinear characteristics of the data stream. At the same time, the attention mechanism is introduced to give different weights to different temporal features, which can enhance the learning ability of the model for globally important temporal features. The core components of the stacked LSTM structure include bidirectional LSTM network, sparse regularized multi-head self-attention mechanism, and unidirectional LSTM network. The structure of the proposed model is shown in Figure.

Temporal attention layer: In order to enhance the nonlinear expression ability of the model, a temporal attention layer is added after stacking each layer of the LSTM. This layer receives the hidden state and unit state of the previous LSTM layer as input, and realizes the sparsity of the weight by calculating the attention weight of each node and introducing L1 regularization. A penalty term based on the sum of the absolute values of the attention weights is added to encourage the weight to shrink toward zero, so as to realize the sparsity of the weights, so that the model can focus on the time series features that really contribute to the prediction. No temporal attention layer is added after the last LSTM layer, which can reduce the excessive increase in the complexity of the model. The final weighted results can better represent the structural relationship between different LSTM layers and the global feature extraction ability, so as to improve the nonlinear expression ability of the model.

In the multi-head attention mechanism, the attention weight is calculated as:

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

The standard attention score is calculated as:

$$S = \frac{QK^T}{\sqrt{d_k}}$$

The standard attention weights are calculated as:

$$\text{attention\_weight} = \text{softmax}(S)$$

Where Q is the query, K is the key, and V is the value, which are the three key components in the attention mechanism. d_k is the dimension of the key vector, softmax similarity is normalized, the weight of each key vector is calculated, then the weight is multiplied by the value vector, and finally the weighted sum is performed to obtain the attention output.

When the sparse regularized multi-head attention mechanism is introduced into the SARIMA-LTSM hybrid model, in order to optimize the allocation of attention weights and make it more streamlined and efficient, the L1 regularization strategy is incorporated. By imposing an absolute penalty on each nonzero element of the attention weight matrix, L1 regularization encourages the model to automatically ignore weights that contribute little to the prediction during the learning process, thereby achieving the sparsity of the attention weights. The L1 regularization loss is calculated as:

$$\text{l1\_loss} = \sum_i |\text{attention\_weight}_i|$$

The total loss is calculated as follows:

$$\text{loss} = \text{l1\_loss} \times \text{l1\_reg}$$

The stacked LSTM model based on sparse regularized multi-head attention mechanism combines two complementary time series prediction models, SARIMA and LSTM, which not only overcomes the limitations of a single model when dealing with complex time series data, but also fully maintains their respective advantages. The model deeply mines the nonlinear features in the time series, combines the sparse attention mechanism to improve the capture ability of key information, and realizes a more accurate and stable prediction.

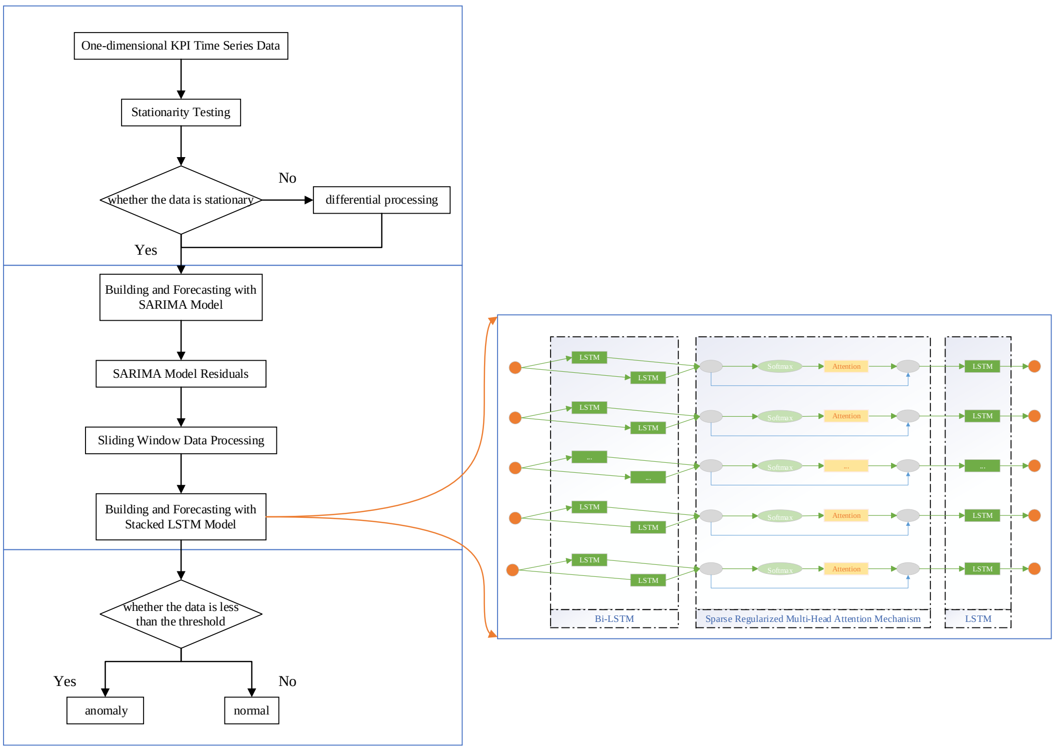

Figure 3. Diagram of the stacked LSTM architecture

4. Experiment

4.1. Data Analysis

In the experimental part, the univariate Key performance indicator (KPI) data were used to evaluate the efficacy of the proposed time series anomaly detection model. The KPI data is derived from the AIOPS data competition and consists of multiple KPI curves, and its abnormal labels are collected from multiple Internet companies such as Sogou, Tencent, and eBay. Most KPI curves have an interval of 1 minute between two adjacent data points, while some have an interval of 5 minutes. These datasets cover time series at different time intervals and cover a wide range of time series patterns, which are commonly used to evaluate the performance of time series anomaly detection. The dataset contains a total of 214,885 data points, and for training and testing purposes, the dataset is divided in a ratio of 8:2, with the first 80% of the data used as the training set and the last 20% as the test set. The plot of the sequence residuals for the KPI dataset is shown in FIG. 4. It can be seen from Figure 4 that this group of data has obvious non-stationarity characteristics. In order to meet the needs of modeling, the sequence is smoothed by differential processing.

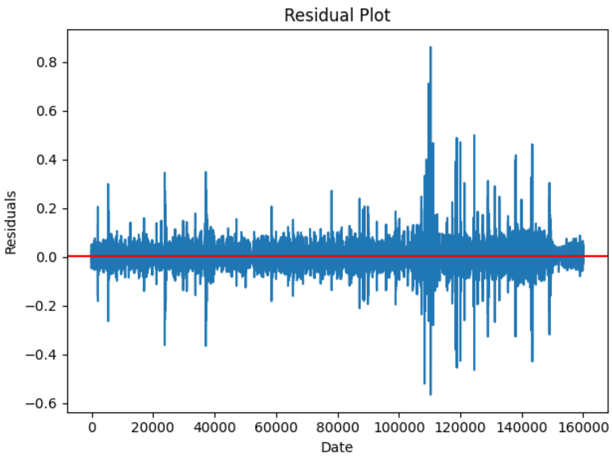

Figure 4. Residual plot

The difference series is tested by ADF, and the p-value is less than 0.05, which rejects the null hypothesis with more than 95% probability, indicating that the difference series is stationary. According to the Autocorrelation Function diagram and Partial Autocorrelation Function diagram of the series, it can be seen from Figure 5 that most of the data lie within the confidence interval, indicating that the series is stable and can be used for modeling.

The Augmented Dickey-Fuller (ADF) test was performed on the differentially processed sequences. The test results show that the p-value is below the 0.05 significance level to reject the null hypothesis with more than 95% confidence. From this, it is certain that the differenced sequence has been transformed into a stationary sequence. In order to visually confirm the stationarity of the series, the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots of the series are drawn, as shown in Figure 5. In these two figures, most of the autocorrelation coefficients and partial autocorrelation coefficients corresponding to the lag orders are within their confidence intervals, indicating that the series after difference has effectively removed the trend and seasonal components, showing the typical characteristics of a stationary series.

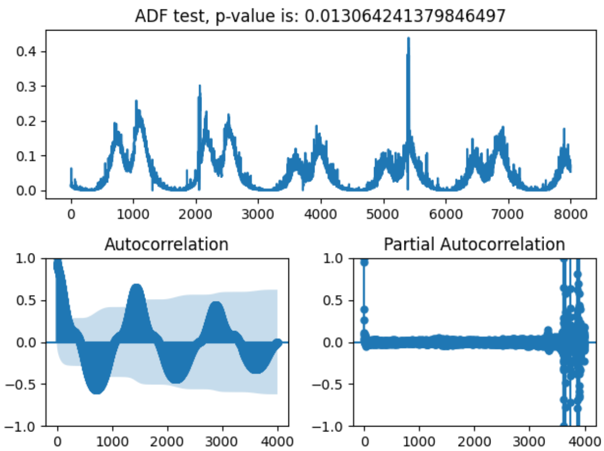

Figure 5. Sequence stationary diagram

4.2. Model Prediction

According to Akaike Information Criterion (AIC) and Bayesian Information criterion (BIC), the quality of SARIMA model parameters can be judged. Through rigorous testing, SARIMA (2,1,2) x (1,1,1,12) is finally determined as the optimal model configuration. This parameter setting not only effectively captures the seasonal and non-seasonal characteristics of time series, but also ensures a good balance between model simplicity and fitting degree.

After determining the parameters of the SARIMA model, the corresponding model was constructed and fitted to the data. In order to improve the prediction performance, the residuals after model fitting are retained in the experiment, which may contain complex dynamic information that is not fully captured by the SARIMA model. Subsequently, an improved stacked LSTM model was constructed in the experiment, and the advantage of LSTM in capturing long-term dependencies was used to learn the residual. Through multiple rounds of experimental comparison, the prediction results are as follows:

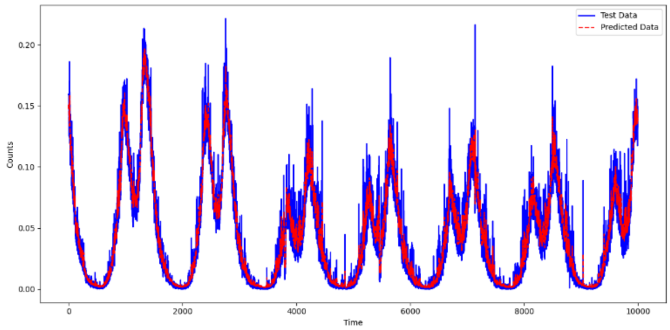

Figure 6. SARIMA-LSTM model prediction results

As you can see from the above figure, the combined SARIMA-LSTM model prediction results are shown. It can be seen from the figure that the proposed model can effectively capture the periodic patterns and nonlinear trends of time series data, while modeling the short-term dependence through the LSTM network, thus improving the accuracy of the prediction. However, there is a certain difference between the predicted data and the actual data in some areas, which is because the original data contains outliers. Outliers can be found by prediction and setting thresholds, so as to achieve the purpose of anomaly detection

4.3. Model Evaluation

Model evaluation is an important step to ensure model performance and quality. Commonly used evaluation metrics are Precision, Recall, F1 score, and Acc(Accuracy).

$$\mathrm{Acc} = \frac{\mathrm{T_p} + \mathrm{T_n}}{T_p + F_p + T_n + F_n}$$

$$\mathrm{Precision} = \frac{\mathrm{T_p}}{T_p + F_p}$$

$$\mathrm{Recall} = \frac{\mathrm{T_p}}{T_p + F_n}$$

$$\mathrm{F_1} = 2 \times \frac{\mathrm{PR}}{P + R}$$

Where T_p represents the number detected as normal and actually normal, F_p represents the number detected as normal but actually abnormal, T_n represents the number detected as abnormal and actually also abnormal, and F_n represents the number detected as abnormal but actually normal.

According to the comparison results in Table 1, when SARIMA model is used alone, although it can deal with part of the time series characteristics, its prediction accuracy is relatively limited. The stacked LSTM model improved the prediction accuracy by 0.155 due to its advantages in dealing with nonlinear time series. The increase in accuracy is 0.025. The combined SARIMA-LSTM model shows good prediction accuracy by fusing the advantages of SARIMA in capturing seasonality and trend and the advantages of LSTM in deep feature extraction. Compared with the LSTM model, the accuracy is improved by 0.082. The increase in accuracy is 0.058.

Table 1. Results of model anomaly detection metrics

Model	Precision	Recall	F1-score	Accuracy
LSTM	0.859	0.83	0.839	0.854
SARIMA	0.704	0.815	0.64	0.829
SARIMA-LSTM	0.887	0.868	0.904	0.912

5. Conclusion

Time series prediction is an important topic in the field of data analysis, especially when dealing with data containing nonlinear factors, LSTM shows significant advantages. However, when a single LSTM model is faced with a complex time series containing multiple factors, it is often difficult to fully capture all the feature information in the data. In addition, although the traditional SARIMA model can effectively deal with seasonal and non-seasonal patterns in time series data, it has limitations in dealing with nonlinear relationships and outliers, and is complicated in model selection and parameter estimation.

This paper proposes a hybrid anomaly detection model of SARIMA and LSTM time series data. Firstly, the nonlinear fitting capability of stacked LSTM is used to solve the learning problem of nonlinear and non-stationary temporal features. Secondly, the introduction of sparse regularization multi-head attention mechanism can better learn important global temporal features. Finally, in order to verify the validity of the model, the constructed SARMIA-LTSM hybrid model was validated by using KPI data. In the experimental results, the anomaly detection accuracy of the SARMIa-LSTM hybrid model is 91.2%, and the F1-Score is 0.904, indicating that the SARMIA-LSTM hybrid model has a better performance in the prediction accuracy than the SARIMA model and the LSTM model.

Data Availability

The link of dataset is https://smileyan.lanzoul.com/ixpcU03lp97g. The dataset is publicly available and accessible.

Conflicts of Interets

The author declares no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation (62273290) and Programs for Science and Technology Development of Henan province No. 222102210078.

References

[1] Li Bin, Wang Yijie, and Cheng Li. Adaptive Enhancement for Active Anomaly Detection in Dynamic Network Traffic[J].Frontiers of Information Technology & Electronic Engineering,2024.

[2] Chen Bofeng, Li Jingdong, Lu Xingjian, et al. Survey of deep learning based graph anomaly detection methods[J]. Journal of Computer Research and Development, 2021.

[3] Jiang Jiangguo, Chen Jiuming, Gu Tianbo, et al. Anomaly detection with graph convolutional networks for insider threat and fraud detection[C]//Proc of IEEE Military Communications Conf. Piscataway, NJ: IEEE, 2019.

[4] Yamauchi M, Ohsita Y, Murata M, Ueda K, and Kato Y. Anomaly detection in smart home operation from user behaviors and home conditions, IEEE Trans. Consum. Electron., 2020.

[5] Espeholt, L., and Coauthors, Deep learning for twelve hourprecipitation forecasts. Nature Communications, 2022.

[6] Wang, Y. B., H. X. Wu, J. J. Zhang, Z.F. Gao, J. M. Wang, P. S.Yu, and M. S. Long,PredRNN: A recurrent neural network for spatiotemporal predictive learning. lEEE Transac.tions on Pattern Analysis and Machine Intelligence, 2023.

[7] Han Z, Zhao J, Leung H, Ma K. F., and Wang W. A review of deep learning models for time series prediction[J], IEEE Sensors J, 2021.

[8] Alamri R, Murugesan R. K., Man M., Abdulateef A. F. , M. l-Sharafi A. A, and Alkahtani A. A.. A review of machine learning and deep learning techniques for anomaly detection in IoT data, Appl. Sci., 2021.

[9] Bai DX, Lu GY, Zhu ZQ, et al. Using time series analysis and dual-stage attention-based recurrent neural network to predict landslide displacement[J]. Environ Earth Sci, 2022.

[10] Ren, Y. B., and X. F. Li. Predicting the daily sea ice concentration on a subseasonal scale of the pan-arctic during the melting season by a deep learning model. IEEE Trans. Geosci. Remote Sens,2023,

[11] SHEERAZ A, ALI K Z, MUHAMMAD M S, et al. Effective and efficient DDoS attack detection using deep learning algorithm, multi-layer perceptron[J]. Future Internet, 2023.

[12] Nahid F A, Ongsakul W, Nimal M M. Very Short Term Wind Speed Forecasting Using Convolutional Long Short Term Memory Recurrent Neural Network[C]// 2020 International Conference and Utility Exhibition on Energy, Environment and Climate Change (ICUE), 2021.

[13] Dou S, Yang K, and Poor H. V. PC2A: Predicting collective contextual anomalies via LSTM with deep generative model," IEEE Internet Things J., 2019.

[14] Karevan Z, Suykens J. Transductive LSTM for time-series prediction: An application to weather forecasting[J]. Neural Networks, 2020.

[15] Zhang Mingmin, et al. Hourly prediction of PM 2.5 concentration in Beijing based on Bi-LSTM neural network[J]. Multimedia Tools and Applications, 2021.

[16] Huang, J.,Y.Luo, J. Shi, X.Ma, Q.Q.Li, and Y.Y. Li, Rapid modeling of the sound speed field in the South ChinaSea based on a comprehensive optimal LM-BP artificial neural network. Journal of Marine Science and Engineering, 2021.

[17] Li B. Y., and J. S. Zhai. A novel sound speed profile prediction method based on the convolutional long-short term memory network. Journal of Marine Science and Engineering. 2022.

[18] Xiao, C. J., N. C. Chen, C. L. Hu, K. Wang, J. Y. Gong, and ZO. Chen. Short and mid-term sea surface temperature

prediction using time-series satellite data and LSTMAdaBoost combination approach. Remote Sensing ofEnviron, 2019.

[19] Gharehbaghi Amin, Ghasemlounia Redvan, Ahmadi Farshad, et al. Groundwater level prediction with meteorologically sensitive Gated Recurrent Unit (GRU) neural networks[J]. Journal of Hydrology, 2022.

[20] Jasmir J, Nurmaini S, Malik R F, etal. Improving Eligibility Classification on Clinical Trials Document using Bidirectional Long Short Term Memory Recurrent Neural Network[J]. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 2021.

[21] Liu D, Cao Y, Yang J, et al. SM-CycleGAN: crop image data enhancement method based on self-attention mechanism CycleGAN[J]. Scientific Reports, 2024.