Asian Journal of Mathematics and Computer Research

Volume 31, Issue 2, Page 96-108, 2024; Article no.AJOMCOR.12093
ISSN: 2395-4205 (P), ISSN: 2395-4213 (O)

Anomaly Detection for Industrial Time Series Data Based on Correlation Analysis and CNN-BiLSTM with Self-attention

Xinyi Yu [a], Bingbing Zeng [a] and Lidong Wang [a*]

[a] College of Engineering, Hangzhou Normal University, Hangzhou, 310018, China.

Authors' contributions

This work was carried out in collaboration among all authors. Author XY designed the study, performed the statistical analysis, and wrote the first draft of the manuscript. Authors BZ and LW managed the analyses of the study. All authors read and approved the final manuscript.

Article Information

DOI: https://doi.org/10.56557/ajomcor/2024/v31i28697

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://prh.ikprress.org/review-history/12093

Original Research Article

Received: 02/03/2024
Accepted: 06/05/2024
Published: 14/05/2024

Abstract

This paper aims to propose an anomaly detection model for industrial time series data based on correlation analysis and CNN-BiLSTM with self-attention to solve the problem of abnormal data detection in the field of industrial data analysis. Industrial data anomaly detection is an important task in the industrial field, which can help people to timely understand the production operation status and real-time record and perception of the operating environment. This paper introduces two key technologies: correlation analysis and CNN-BiLSTM with self-attention, and how to combine them to build an effective anomaly detection model for industrial time series data. Through experimental evaluation, this paper proves the effectiveness and superiority of the proposed model in industrial data anomaly detection tasks.

Keywords: Anomaly detection; CNN-BiLSTM; self-attention; industrial time series data.

**Corresponding author: Email: wld@hznu.edu.cn;*

Cite as: *Yu, X., Zeng, B., & Wang, L. (2024). Anomaly Detection for Industrial Time Series Data Based on Correlation Analysis and CNN-BiLSTM with Self-attention. Asian Journal of Mathematics and Computer Research, 31(2), 96–108. https://doi.org/10.56557/ajomcor/2024/v31i28697*

1 Introduction

Over the past decades, with the advancement of industrialization and modernization, China's manufacturing industry has sustained and rapid development. As an important application scenario of industrial big data,intelligent manufacturing is not only the carrier and source of data, but also the final application scenario and goal of data products formed by industrial big data [1-3]. Modern manufacturing production lines and intelligent products have achieved real-time recording and perception of production running status and operating environment through sensors, controllers, and intelligent instruments [4], and have accumulated and are producing a large number of industrial time series data. Through the analysis and mining of multi-dimensional time series data based on the collected time points, it is possible to control, analyze, make decisions and plan the running state of industrial systems [5], and diagnose, early warning, handle and repair the industrial failure problems that are monitored and the production hidden trouble problems that are predicted and calculated. The above process forms a positive cycle of effective industrial knowledge generation, extraction and application. However, industrial time series data often have the characteristics of high dimensionality, strong coupling and non-stationary, which brings great challenges to the analysis and prediction of industrial time series data. If the abnormal, fault, and crisis situation in the industry cannot be effectively identified in time, it is likely to bring associated losses to the entire manufacturing system.

At present, complex abnormal conditions in high-dimensional time series data have gradually attracted attention, and the occurrence of an abnormal condition often interacts with multi-dimensional series, so these abnormal patterns are more difficult to detect and identify.

The mainstream models commonly used anomaly detection [6] mostly directly model abnormal data, so as to realize the detection of abnormal data In industrial time series, but in the time series data in the actual industrial production process,abnormal data only accounts for a small proportion, so it is difficult to directly model abnormal data, and there is usually nonlinear correlation between the data. The above characteristics will further enhance the difficulty of data processing and anomaly detection.

This experiment carries out anomaly detection research from two perspectives of multi- dimensional time series data correlation and deep learning prediction model to detect abnormal data in big data. Combined with application scenarios, the Harbin Institute of Technology team designed and developed a comprehensive and intelligent industrial time series data cleaning system called Cleanits [7], which can effectively identify multi-dimensional time series. The method based on similarity measurement calculates the similarity between the standardized series to determine whether there is abnormal data, but this method has large time overhead. In the method of rule constraints, researchers have proposed sequential dependencies [8], speed constraints [9], which can effectively use the time series characteristics in time series to repair highly abnormal data. However, these methods are usually difficult to meet the requirements of sequential anomaly detection with variable patterns to effectively identify and repair sequence anomalies.

At present, domestic and foreign reviews [10-14] mostly discuss the methods of generalized anomaly detection, trying to cover various data forms such as images,videos, tables and sequences, etc. The literature [11,12] comprehensively summarizes and analyzes the anomaly detection methods based on deep learning, but lacks discussion on industrial scenarios. Although the literature [15,16] is reviewed with industrial production as the background,They mainly focus on traditional methods and system control, Literature [17] systematically summarize deep learning-based surface defect detection methods, but mainly review supervised methods, and recently, many new results have emerged based on unsupervised, semi-supervised and other Settings. However, there is no comprehensive and detailed review on the field of industrial timing anomaly detection, so this review aims to fill this gap and focuses on the introduction and summary of such new methods.

For industrial big data, which has the characteristics of large volume, strong multi-source heterogeneity, continuous sampling, low value density and strong dynamics [1,3,4], machine learning and deep learning were combined, and the data correlation analysis was used. Firstly, the multi-dimensional time series were segmented and standardized to obtain the correlation matrix and extract the quantitative correlation. Then, a time series correlogram model is established, and time series cliques are divided by the correlation strength on the time series correlogram. Embedding CNN-BiLSTM with self-attention model can first capture the spatial characteristics of the input fault sequence through the CNN layer, then capture the temporal correlation of the fault sequence through the BiLSTM layer, and finally better understand the internal correlation of the sequence through the self-attention layer. This comprehensive structure can deal with the fault sequence data more comprehensively. In this experiment, correlation analysis and CNN-BiLSTM with self-attention model are built to simulate the abnormal data detection model of industrial labeled feature data.

2 Related Works

2.1 Anomaly detection of industrial time series data

In the field of industrial data anomaly analysis, the research on time series anomaly detection can be divided into three research tasks: point anomaly, subsequence anomaly, and pattern anomaly. Industrial data anomaly analysis methods mainly include: (1) Statistical models [18] (such as ARIMA， GARCH, etc.); (2) Clustering [19,20] (such as k-means,EM,SVM models, etc.); (3) Similarity measure [21]; (4) Constraint rules [22].

2.1.1 Statistical model

The length detection based on statistical model predicts future data points by analyzing the statistical characteristics of industrial time series data. ARIMA (Autoregressive Integral Moving Average model) is suitable for analyzing time series data, while GARCH(Generalized Autoregressive Conditional Heteroscedastic model) is specialized for analyzing volatile time series data.

The advantages of this approach include high accuracy and interpretability. High accuracy is the ability to accurately predict the future direction of the data if the data follows the assumptions of the model. Interpretability means that the model parameters have clear economic or practical significance,which is easy to understand and interpret.

2.1.2 Clustering model

Clustering methods identify outliers by grouping data into groups with high intra-group data similarity and low inter-group data similarity. K-means and EM (Expectation Maximization) are methods for partitioning the data set, while SVM (Support Vector Machine) can be used for anomaly detection by building a model that separates outliers from normal points. It works on all types of data and can handle nonlinear relationships. It does not need to assume the data distribution in advance, so it has wide applicability. The disadvantage is that it may be too sensitive to outliers, resulting in unstable clustering results. It is difficult to choose the right number of clusters. Comparative clustering methods are superior to statistical models in dealing with large and complex datasets, but may require more computational resources.

2.1.3 Similarity measure

This method identifies anomalies by calculating the similarity between time series. Commonly used similarity measures include Euclidean distance, Dynamic Time Warping (DTW), etc. It has strong adaptability, and can adapt to changing data patterns and structures by comparing the similarity of time series. The similarity measure provides an intuitive way to evaluate the difference between time series. Meanwhile, a variety of distance measures (such as Euclidean distance,Manhattan distance, DTW,etc.) can be selected to adapt to different data characteristics.

2.1.4 Constraint-based analysis

By defining rules that data should satisfy under normal operating conditions, data points that violate these rules are identified as anomalies. Rules can be formulated based on expertise or the results of data exploration. The advantage is immediate feedback, since the data that violates the rules can be identified immediately, which is suitable for real-time monitoring systems. The rules can be customized to specific industrial processes, which enhances the adaptability and accuracy of the model. Compared with other methods, rule- based methods are more suitable for real-time monitoring and detection of specific conditions, but may be inferior to data-driven methods in terms of adaptability and generalization.

2.2 Correlation analysis

Correlation analysis refers to the analysis of two or more variable elements with correlation, so as to measure the degree of correlation between two variable factors. Correlation analysis can only be carried out if there is a certain connection or probability between the elements of correlation.

In industrial data forecasting, correlation analysis plays a crucial role. It involves assessing and understanding the interdependence and correlation between different variables in an industrial data set. Through in-depth analysis of these relationships, the performance of prediction models can be optimized to improve production efficiency and decision quality.

The methods and techniques are as follows:

2.2.1 Statistical methods

Pearson's correlation coefficient: It quantifies the strength and direction of the linear correlation between two continuous variables.

Spearman's rank correlation coefficient: It is used to assess the correlation between the rank order of two variables and is suitable for nonlinear relationships.

2.2.2 Deep learning methods

Neural networks: Deep neural networks can identify complex nonlinear relationships between variables, especially for high-dimensional data.

Convolutional Neural Networks (CNNs): It is suitable for correlation analysis of spatial data such as images.

Recurrent Neural Network (RNN) : It is good at processing sequential data (such as time series), analyzing the dependence of variables over time.

2.2.3 Machine learning algorithms

Random forest: It can be used to evaluate feature importance, which indirectly reflects the correlation between variables.

Principal Component Analysis (PCA) : It reveals the main variable correlations in the data through dimensionality reduction.

By identifying and exploiting associations between variables, we can build more accurate predictive models. Correlation analysis helps to identify redundant features, simplify the model, and reduce the computational burden. Meanwhile, in-depth understanding of the relationship between variables provides more abundant information for decision makers to support more reasonable decisions.

Correlation analysis is a powerful tool in industrial data forecasting. It not only helps to identify intrinsic patterns and relationships in the data, but also uses these insights to guide practices, optimizes production processes, and improves efficiency and quality.

2.3 CNN-BiLSTM with self-attention

The CNN-BiLSTM with self-attention fault recognition model combines three parts: Convolutional neural network (CNN)， bidirectional long Short-Term memory network (BiLSTM) and Multi-head self-attention. Each part plays different roles in the feature extraction process of fault recognition.

Convolutional neural network (CNN) is a kind of feedforward neural network with deep structure including convolution calculation, which is one of the representative algorithms of deep learning. CNN is mainly composed of five structures: input layer, conv layer, pool layer, full connected layer and softmax layer. The training process of CNN is composed of forward propagation and back propagation. The forward propagation mainly extracts different features of the input data, and the back propagation mainly optimizes the parameters of the forward propagation [23].

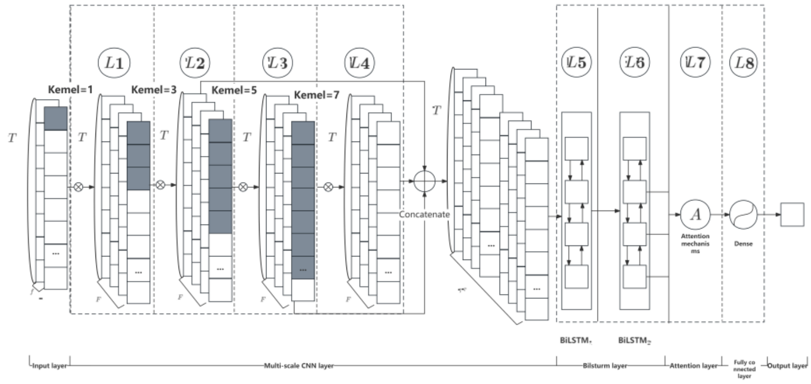

Fig. 1. The structure of multi-scale CNN-BiLSTM with self-attention

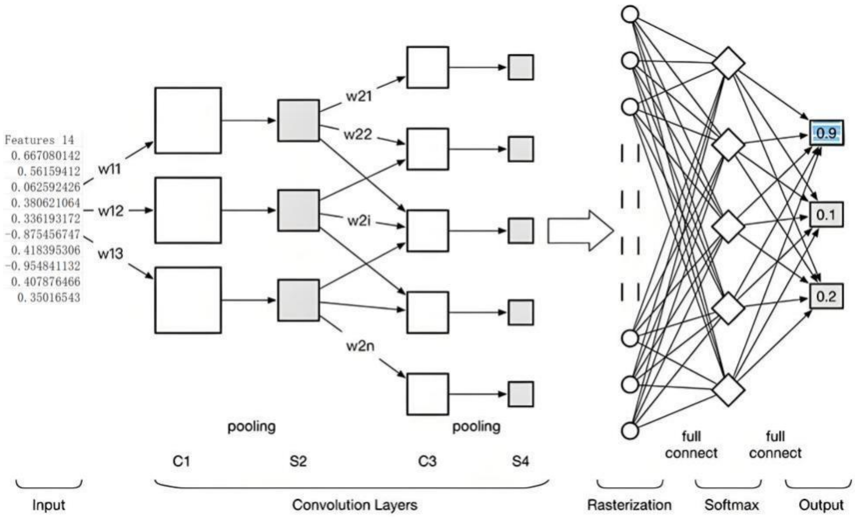

Fig. 2. Convolution neural network (CNN)

Long short-term memory (LSTM) is a variant of recurrent neural network architecture, which can solve the gradient explosion and gradient disappearance problems encountered in recurrent neural networks [24]. The standard LSTM network structure only has forward propagation operation and lacks the logic of before and after. On the basis of LSTM, Bi-directional Long Short-Term memory network (BiLSTM) uses known time series and reverses position sequence to deepen the feature extraction of the original sequence and improves the accuracy of the model output results through forward and back propagation bidirectional operation. The final output of the Bi-LSTM neural network is the sum of the output results of the forward and back propagation LSTM. The BiLSTM network structure is shown in Fig. 3.

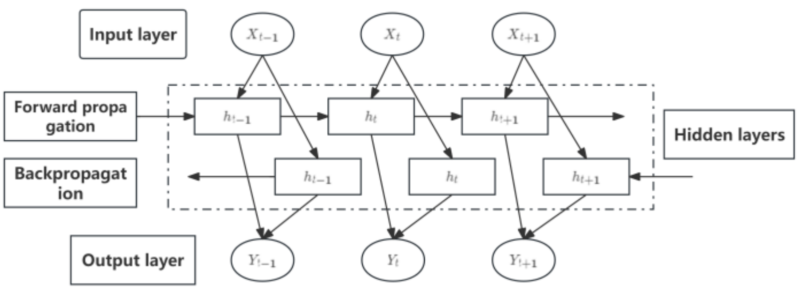

Fig. 3. Schematic diagram of BiLSTM network structure

Multi-head self-attention has become an epoch-making technique in deep learning, especially in natural language processing (NLP). It is an extension of the self-attention mechanism, which can process information at multiple positions in the sequence simultaneously, thus capturing richer context relationships. The self-attention mechanism is able to capture long-distance dependencies by calculating the attention degree of each element to other elements in the sequence, while the multi-head mechanism further extends this ability by executing multiple self-attention processes (i.e.,"heads") in parallel, each focusing on a different representation subspace of the input data, thus enriching the information processing ability of the model.

The core components of multi-head self-attention include the Query, Key, and Value, which represent different vector representations that are used to compute and update the output vector at each position. By performing a weighted sum of these vectors, the model is able to learn a different feature representation at each head and merge the outputs of these different heads at the end to obtain a more comprehensive representation. This mechanism not only enhances the ability of the model to capture the complex relationships within the sequence, but also significantly improves the efficiency of processing long sequence data.

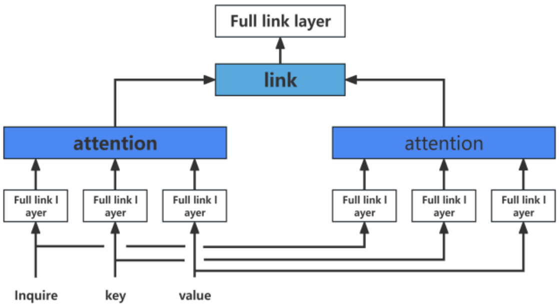

Fig. 4. Multi-head self-attention structure

Combined with the structure of these three layers, the model first captures the spatial features of the input fault sequence through the CNN layer, then captures the temporal correlation of the fault sequence through the BiLSTM layer, and finally better understands the internal correlation of the sequence through the self-attention layer. This comprehensive structure can deal with the fault sequence data more comprehensively.

3 Experimental Procedure

3.1 Data Preprocessing

3.1.1 Data collection

We collect appropriate abnormal industrial big data from github and major databases in China, and import it into xlsx files for packaging.

Features 15	Features 16	Features 17	Features 18	Features 19	Features 20	Fault class
0. 539568391	−1. 879025569	1. 215754157	2. 862283389	1. 833214762	−0. 398000389	5
0. 908866799	0. 792844643	0. 813485403	−1. 753415181	−1. 036712032	−0. 255248649	8
−0. 324674753	−0. 664449194	1. 757565223	1. 976397213	0. 035429072	−0. 101073442	5
−0. 18607562	0. 140151713	0. 676129096	1. 439987845	0. 445622068	0. 430213031	5
−0. 435222304	−1. 25915766	0. 081533997	0. 721443308	0. 745487604	0. 498692821	6
−0. 081069896	−1. 372722901	−0. 486970281	−0. 378703936	1. 121061365	0. 063101566	6
0. 902028527	1. 580546776	0. 8465687	2. 590489348	−0. 402585727	0. 491222233	4
0. 564022245	−2. 024028128	0. 81202579	1. 224190084	−0. 019567477	−0. 728463113	3
0. 392412159	1. 055199931	−0. 191573117	0. 750630626	0. 959652883	1. 472195731	7
−1. 132245775	−0. 067338434	0. 527744945	1. 562698589	0. 072356732	−0. 055352492	5
−0. 813970932	1. 064884753	0. 803454075	0. 935878775	1. 344827974	1. 546453009	7
0. 381918413	−1. 409653423	0. 270419287	0. 841780707	0. 366885368	0. 004410218	3
−1. 119621255	1. 302697425	0. 564635029	−0. 001328664	0. 506079702	1. 324758248	7
0. 058015649	0. 068948127	1. 201425958	2. 185398642	0. 636537862	−0. 781338692	5
0. 140412749	−0. 003659908	0. 729657925	1. 398876181	0. 57054248	−0. 108984814	5
0. 122201064	−0. 181211334	0. 246707028	1. 299475465	0. 881529439	−0. 529105323	3
0. 442347274	−0. 390343294	1. 305566847	1. 335795221	0. 495016694	0. 25494866	5
0. 237142377	−0. 389579484	0. 967027388	2. 160824324	−0. 337188611	0. 28759971	5

Fig. 5. Data samples in our experiment

Data cleaning: We need to clean the data before importing data. Verify the data dimension and data type, remove duplicate rows, and make sure that the float value is within the normal threshold, and this step can screen out some abnormal data. In addition, we have to check whether there are missing values. If there are missing values, we use the mean of the column to fill the missing values, to ensure that the data fit the experimental requirements.

Data dividing: We randomly divide training set and test set. In order to ensure the robustness of the later training model, we randomly divide the obtained data into training set and test set.

3.2 Data correlation analysis

3.2.1 Correlation calculation

In this step, the historical data accumulated in industry are used to model and calculate the correlation between the series. The sensors of industrial equipment are usually continuously sampled, so the research object of this experiment (industrial time series) can be regarded as the stream data that is coming continuously. It is necessary to divide the accumulated large amount of historical series data into appropriate length time series segments for subsequent calculation processing.

For the reconstructed time subsequence group $S = \{S_1^m, S_2^m \ldots\ldots S_k^m\}$, the correlation between the time subsequences should be measured and expressed as the sequence correlation of the time subsequence group. In this step, the covariance matrix (Pearson coefficient matrix) is used to initially calculate the correlation between the series in the time subsequence group. Specifically, in the l segment (default length is n) time subsequency group of sensor group S, the k-th time subsequency is denoted as $S_k^l = \{s(1)_k^l, \ \ldots\ldots, s(n)_k^l\}$. In this sequence time period, we define the correlation coefficient matrix $R^l_{\square}$, used to measure the parameter l on sensor group S Correlation of K sequences over time. Each element in the correlation coefficient matrix R_{ij}^l in the correlation coefficient matrix is as follows:

$$R_{ij}^l = \frac{\sum_{m=1}^{n}\left(s(m)_i^l - \overline{s_i^l}\right)\left(s(m)_j^l - \overline{s_j^l}\right)}{n-1}, \text{ where, } \overline{s_i^l} = \frac{\sum_{m=1}^{n} s(m)_i^l}{n}, \overline{s_j^l} = \frac{\sum_{m=1}^{n} s(m)_j^l}{n}$$

where $s(m)_i^l$ represents the data value of time series S_i^l at the m-th time point in this time period, and $\overline{s_i^l}$ and $\overline{s_j^l}$ are the mean values of all series data points of S_i^l and S_j^l in this time period, respectively.

3.2.2 Construction of temporal correlation graph model

After obtaining the correlation coefficient matrix of K time series, in order to effectively represent the correlation between the series, we intend to use the series correlation graph model to further calculate the correlation of the series according to the value of the elements in the matrix. For the k-dimensional time series data on a given sensor group S_m, an undirected time series correlation graph $G_r(S) = (V, E)$ is established. The vertex set records all the series, and the edge set records whether there is correlation information between the series greater than a threshold. Then, the undirected temporal correlation graph $G_r = (V, E)$ is initialized, and an undirected edge is connected between two vertices whose correlation parameter is greater than or equal to the θ_c. The specific constructed model is shown in the following figure:

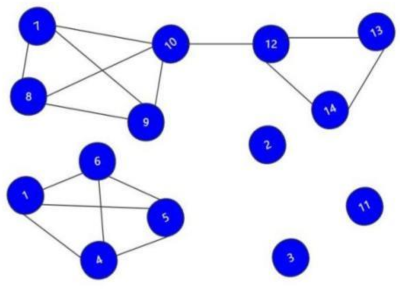

Fig. 6. Time series correlation graph

3.2.3 Partition of time series cliques based on graphical model

According to the basic concept of graph model, an undirected graph is called sum-connected if there is a path from vertex to vertex. A graph is connected if any two vertices are connected. Otherwise, the graph is disconnected. The connected component of G_r is defined as a maximal connected subgraph, and a maximal subgraph is one that contains the largest number of vertices. In order to further analyze the groups of time series with different correlation strengths on the graph, we plan to divide the time series correlation cliques, and mine and identify abnormal patterns in multi-dimensional series by calculating the connected components of the graph and performing necessary pruning processing, so as to improve the efficiency of anomaly detection methods.

First, we maintain a label for each vertex on the graph. Then, we iterate over each vertex and use breadth-first search to find all connected vertices with edge weights greater than a given threshold, and add them to the same temporal clique. After adding all possible vertices to a clique, the clique is determined according to the satisfied conditions. If the degree of every vertex in the clique is greater than half of the number of vertices in the clique, and the weight of every edge in the clique is greater than a given threshold, then the clique is formed. If there are points that do not meet the above conditions, the current clique is pruned and deleted until all the temporally related cliques satisfying the conditions $C = \{C_1, C_2, \cdots, C_m\}$ are found.

3.3 Design and training of CNN-BiLSTM-Attention model

The CNN-BiLSTM with self-attention model constructs a powerful and flexible network structure by skillfully combining CNN, BiLSTM and self-attention mechanism, which is designed to deal with sequence data with complex dependencies. With an appropriate regularization strategy, the model can not only effectively learn the features of the data, but also has a strong generalization ability. This design is an effective extension of the existing sequence processing models.

Table 1. Model parameters

Parameters	Settings
imageInputLayer	[20 1 1]
convolution2dLayer	[2 1], 16
batchNormalizationLayer	No specific parameter
Relulayer	No specific parameter
maxPooling2dLayer	[2 1]
Convolution2dLayer	[2 1] 32
batchNormalizationLayer	No specific parameter
reluLayer	No specific parameter
maxPooling2dLayer	[2 1]
flattenLayer	No specific parameter
bilstmLayer	128
dropoutLayer	0.5
selfAttentionLayer	1,64
fullyConnectedLayer	8
softmaxLayer	No specific parameter
classificationLayer	No specific parameter

The CNN-BilSTM with self-attention model designed in this experiment combines the local feature extraction ability of CNN, the sequence data processing ability of BiLSTM, and the information screening advantages of self-attention mechanism. Now, let us explore the structure in Table 1 and its parameter Settings.

3.3.1 Input layer

[20 1 1] means the input is 20 dimensional sequence data. This setup is particularly effective for dealing with small sequence data with a fixed number of feature dimensions, such as text classification, time series analysis, and other scenarios.

3.3.2 CNN layers

Convolutional layers: The initial convolutional layers use the settings of [2 1], 16 and the second layer uses [2 1], 32 to capture more complex features by increasing the number of filters. This configuration helps the model to be able to learn rich local features from sequential data while remaining computationally efficient.

Batch Normalization: Using batch normalization after each convolutional layer helps to speed up the training process and improve the stability of the model over different data distributions.

ReLU activation function: Helps the network to capture complex features by introducing nonlinearities.

Max pooling: Pooling windows is set to [2 1] to help reduce computation and reduce the risk of overfitting while retaining important information.

3.3.3 BiLSTM layer

Parameter setting: 128 indicates that 128 LSTM units are used, this number balances the complexity of the model and the ability to handle long sequences. The BiLSTM layer is able to learn information from both directions of the sequence, which enhances the model's ability to capture long-term dependencies in time series data.

3.3.4 Dropout layer

Parameter setting: A ratio of 0.5 is a commonly used setting to effectively prevent overfitting and improve the generalization ability of the model.

3.3.5 Self-attention layer

Parameter setting: 1, 64 means using a single attention head, and the dimension of the head is 64. This setting enhances the model's focus on important information while maintaining a balance between the number of parameters and computational efficiency of the model.

3.3.6 Fully connected layer and output layer

Fully connected layer: The output size is 8, suitable for dealing with classification problems with 8 classes. This design shows a clear orientation to the target task.
Softmax & Classification layer: Converts the output of the fully connected layer into a probability distribution and provides the final classification result.

4 Results and Discussion

4.1 Experimental results

Accuracy, precision, recall, and F1 score were used to evaluate the model performance. Fig. 7 shows the running results of our program.

4.1.1 Accuracy

Accuracy is the fraction of the total number of samples that the model predicts correctly. In the case of anomaly detection for industrial time-series data, this means the rate at which the model correctly labels normal and abnormal states. Accuracy gives a visual indication of the overall performance of the model, but this metric can be biased if the data is imbalanced.

4.1.2 Precision

Precision is the proportion of samples predicted by the model to be abnormal that are actually abnormal. It measures how accurate the model is in flagging anomalies.

4.1.3 Recall

Recall is the proportion of all true anomaly samples that the model correctly identifies. This reflects the model's ability to catch anomalies.

4.1.4 F1 score

The F1 score is the harmonic mean of precision and recall, and it tries to strike a balance between the precision and recall. The F1 score is a composite metric that is especially suitable for datasets with imbalanced classes. In industrial applications, it helps ensure that the model neither misses too many true anomalies nor misclassifies too many normal cases as anomalies.

These metrics in experimental results allow a comprehensive assessment of how the model performs in real-world settings. The ideal model will have high accuracy, precision, recall, and F1 score, ensuring that it can reliably detect anomalies while maintaining the correct judgment of normal conditions. However, in practice, it is often necessary to make a trade-off between these metrics and adjust the model performance according to the specific business needs.

The ideal model will have high accuracy, high precision, high recall, and high F1 score to ensure that it can reliably detect anomalies. The larger the precision, recall, and F1 score, the higher the accuracy of anomaly detection in industrial data, and vice versa. As shown in Fig. 8 and Fig. 9, the predicted samples on the test set match the true samples perfectly. All samples are classified correctly, and achieve 100% accuracy.

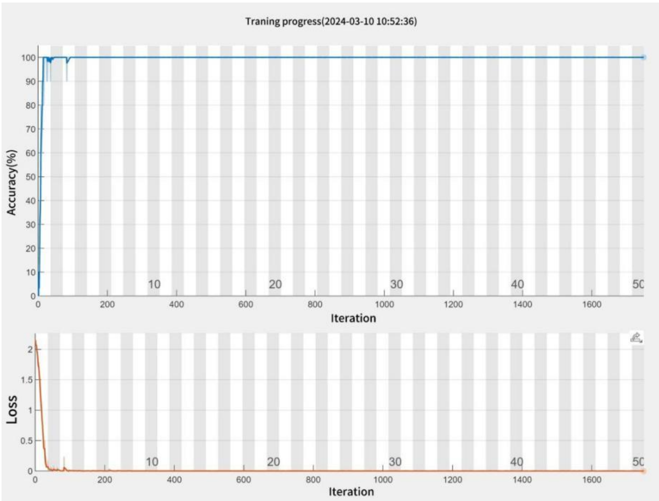

Fig. 7. Learning iteration process

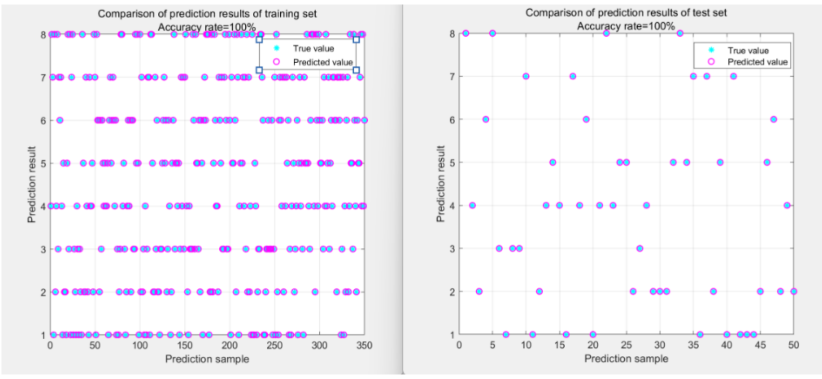

Fig. 8. The prediction results (accuracy) of the training set

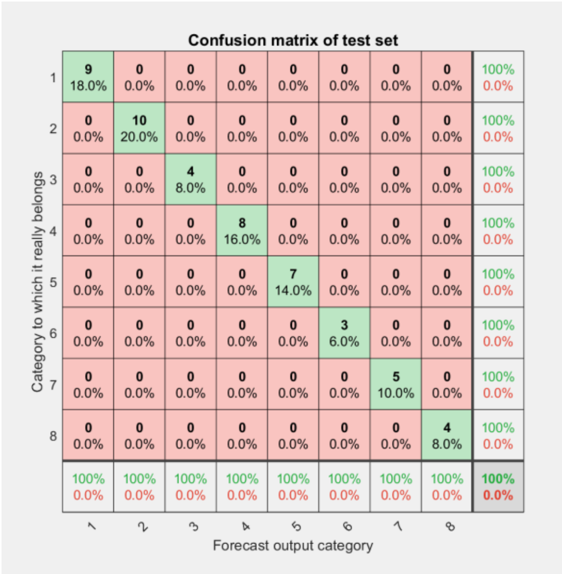

Fig. 9. Confusion matrix of test set

4 Conclusion

This paper makes use of the collected industrial failure data, the correlation analysis in deep learning and the learning model constructed by CNN-BiLSTM with self-attention to predict and judge the abnormal data in the industrial time series, and classifies the comments on microblog topics. By combining the correlation analysis of data and the CNN-BiLSTM with self-attention model, the anomaly detection model of industrial relaxing data can accurately determine the abnormal existence of data in big data more accurately. Correlation analysis can use time series cluster division based on correlation calculation of time series data to effectively mine the correlation relationship of each dimension sequence. CNN-BiLSTM with self-attention combines the structure of these three layers. The model first captures the spatial features of the input fault sequence through the CNN layer, then captures the temporal correlation of the fault sequence through the BiLSTM layer, and finally better understands the internal correlation of the sequence through the self-attention layer. This comprehensive structure can deal with the fault sequence data more comprehensively, so as to capture the abnormal value in the large industrial data efficiently. This integrated fusion model can better analyze and judge the correlation between industrial data and the accuracy of anomaly detection and prediction, improve the accuracy and robustness of anomaly detection of industrial time series data, and provide a new perspective and method for in-depth analysis and judgment of industrial time series anomaly data.

Acknowledgements

This study is supported by Key Lab of Film and TV Media Technology of Zhejiang Province (No.2020E10015), Zhejiang Provincial High-Education Teaching Reform Project under Grant No.jg20220770, Medical and Health Technology Plan of Zhejiang Province (No. 2022507615).

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Zhang Jie, Qin Wei, Bao Jinsong,et al. Manufacturing big data. Shanghai: Shanghai Science and Technology Press; 2016.

[2] Industrial Big Data AD Hoc group of industrial internet industry alliance. Industrial big data technology and application practice.Beijing: Publishing House of Electronics Industry; 2017.

[3] National Manufacturing Power Building Strategy Advisory Committee. Made in China2025 Key Areas Technology Roadmap; 2015.

[4] Wang Jianmin. White paper on China's industrial big data technology and application. Beijing: Industrial Internet Industry Alliance; 2017.

[5] Wang Jianmin. Overview of industrial big data technology. Big Data. 2022,(6):3-14.

[6] Ewusie JE，Soobiah C，Blondal E，et al. Methods,applica- tions and challenges in the analysis of interrupted time series data: A scoping review [J] .Journal of Multidisciplinary Healthcar. 2020；13:411.

[7] Ding XO, Wang HZ, Su JX, et al. Cleanits: A data cleaning system for industrial time series. Proc, of the VLDB Endowment. 2019;12(12):1786-1789.

[8] Golab L, Karloff H, Korn F, et al. Sequential dependencies. Proc, of the VLDB Endowment. 2019;2(1):574-585.

[9] Song S, Zhang A, Wang J, et al. SCREEN: Stream data cleaning under speed constraints. In: Proc of the ACM SIGMOD International Conference on Management of Data. ACM. 2015;827-841.

[10] Chandola V, Banerjee A, Kumar V. Anomaly detection: a survey. ACM Comput Surv. 2009;41:1-58.

[11] Pang G, Shen C, Cao L, et al. Deep learning for anomaly detection: A review. 2020. ArXiv; 2007;02500.

[12] Chalapathy R, Chawla S. Deep learning for anomaly detection: A survey. 2019.ArXiv:1901.03407.

[13] Bulusu S, Kailkhura B, Li B, et al. Anomalous instance detection in deep learning: A survey. In: Proceedings of the 42nd lEEE Symposium on Security and Privacy Princeton; 2020.

[14] Ruff L, Kauffmann JR, Vandermeulen R A, et al. A unifying review of deep and shallow anomaly detection.Proc IEEE. 2021;109:756-795.

[15] Czimmermann T, Ciuti G, Milazzo M, et al. Visual-based defect detection and classification approaches for industrial applications--A survey. Sensors. 2020;20:1459.

[16] Zhou DH, Wei MH, Si XS. A survey on anomaly detection, life prediction and maintenance decision for industrial processes. Acta Automatica Sin. 2013;39:711-722.

[17] Tao X, Hou W, Xu D. A survey of surface defect detection methods based on deep learning. Acta Autom Sin. 2021;47:1017–1034.

[18] Enders W, et al. Du Jiang, Yuan Jingan: Trans. Applied econometrics: Time series analysis. 2nd Ed. Beijing: Higher Education Press; 2006.

[19] Eamoon JK, Kaushik C, Sharad M, Michael JP, Locally adaptive dimensionality reduction for indexing large time series databases. ACM Trans, on Database Systems. 2012;27(2):188-228.

[20] Ma J, Ma J, Perkins S, et al. Time-series novelty detection using one-class support vector machines. In: Proc of the International Joint Conference on Neural Networks. IEEE; 2013.

[21] Chandola V, Mithal V, Kumar V. Comparative evaluation of anomaly detection techniques fbr sequence data. In: Proc of the 8th IEEE International Conf on Data Mining. IEEE Computer Society; 2008.

[22] Li X, Han J, Kim S, et al. ROAM: Rule- and motiAbased anomaly detection in massive moving object data sets. In: Proc, of the 7th SIAM Int'l Conf, on Data Mining. 2007;273-284.

[23] Alex K, Ilya S, Geoffrey EH. Image Net classification with deep convolutional neural networks ［J］. Neural Information Processing Systems. 2017；60(6):84-90.

[24] Hochreiter S, Schmidhuber J. Long shortterm menory ［J］. Neural Computation. 1997；9(9):1735-1780.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
https://prh.ikprress.org/review-history/12093