Computers and Electrical Engineering 79 (2019) 106458

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Real-time anomaly detection based on long short-Term memory and Gaussian Mixture Model ☆

Nan Ding*, HaoXuan Ma, Huanbo Gao, YanHua Ma*, GuoZhen Tan

Computer Science and Technology, Dalian University of Technology, No.2 Linggong Road, Ganjingzi District, Dalian, Liaoning, China

ARTICLE INFO

Article history:
Received 29 December 2018
Revised 4 September 2019
Accepted 4 September 2019

Keywords:
Anomaly detection
Long short term memory
Gaussian mixture model
Multivariate sensing time series

ABSTRACT

Anomaly detection is a long-standing problem in system designation. High-quality anomaly detection can benefit plenty of applications (e.g. system monitoring, disaster precaution and intrusion detection). Most of the existing anomalies detection algorithms are less competent for both effectiveness and real-time capability requirements simultaneously. Therefore, in this paper, the LGMAD, a real-time anomaly detection algorithm based on Long-Short Term Memory (LSTM) and Gaussian Mixture Model (GMM)is proposed. Specifically, we evaluate the real-time anomalies of each univariate sensing time-series via LSTM model, and then a Gaussian Mixture Model is adopted to give a multidimensional joint detection of possible anomalies. Both NAB dataset and self-made dataset are employed to verify our approach. Extensive experiments are conducted to demonstrate the superiority of LGMAD compared to existing anomaly detection algorithms.

1. Introduction

Anomaly detection for time series data has continuously become a hot issue in academia and industry. Detecting and locating anomaly points or anomaly areas can provide important information at critical moments so as to prevent or take action on impending faults.

The anomaly detection for time series data has a crucial importance in the field of industrial, financial, military, medical, insurance, critical system security, robotics, multi-agent, network security, and Internet of Things [1,2]. Melvin Gauci et al. from Stanford University have composed 100 agents into a system. And corresponding simulation experiments proved that unrestricted single short-term anomalies will spread rapidly within the group and eventually lead to a system collapse [3], which indicated that the effect of anomaly detection is the core of secure interaction in many scenarios. The importance of anomaly detection in the practical applications accelerate its development by featuring not only accuracy but also immediacy [4,5].

In this paper, aiming at the application scenario where low-dimension parameters are used to characterize anomaly states of the system, we propose the LGMAD algorithm. Based on the univariate anomaly detection using our designed LSTM-BP algorithm, the Gaussian model is used to perform the anomaly detection for multivariate time series data. The proposed algorithm can be applied to anomaly detection without any prior anomaly knowledge, and improves the anomaly detection

☆ This paper is for CAEE special section SI-aisec.Reviews processed and recommended for publication to the Editor-in-Chief by Guest Editor Dr. James J. Park.

* Corresponding authors.
E-mail addresses: dingnan@dlut.edu.cn (N. Ding), mayanhua@dlut.edu.cn (Y. Ma).

https://doi.org/10.1016/j.compeleceng.2019.106458
0045-7906/

rate effectively with outstanding temporal performance. Besides, it can adapt to the anomaly detection requirements for current multivariate time series data. The experimental results demonstrated that the LGMAD algorithm performed better than the HTM algorithm, and is also superior to the LSTM-based Predictive Data Model algorithm. This paper focuses on the low-dimensional application scenarios for characterizing system anomaly detection instead of involving high-dimensional or super-high-dimensional anomaly detection scenarios. For the case of high-dimensional and super-high-dimensional anomaly detection, by dimension reduction, such as PCA, the proposed method can also be used to detect anomalies [5]. However, the dimension reduction and its methods are beyond the scope of this paper.

The contribution of this paper is mainly reflected in the following three aspects:

1. Aiming at the application scenario where low-dimension parameters are used to characterize anomaly states of the system, we propose the LGMAD algorithm, a real-time anomaly detection algorithm based on Gaussian Mixture Model, to ensure both accuracy and real-time requirements at the same time.
2. Traditional LSTM cannot achieve the same good result consistently on all data sets. Therefore, we propose the LSTM-BP based on the LTSM, and improve the internal structure of LSTM to make it suitable for processing data time series anomaly detection.
3. In order to improve the efficiency of the algorithm, we introduce the health factor α, which evaluate the performance of the system.

This paper is organized as follows. The second chapter (related work), a brief introduction of the anomaly detection method is presented. In the third chapter, the LSTM-BP algorithm which is based on LSTM improved anomaly detection algorithm for univariate time series, and the multivariate Gaussian model are addressed. The fourth chapter introduces the overall architecture and the detailed flow of anomaly detection algorithm LGMAD proposed in this paper for multivariate time series. The fifth chapter (experimental part) compares the LGMAD algorithm with LSTM-based Predictive Data Model algorithm proposed by Pavel Filonov et al. [6], HTM algorithm proposed by Numenta, and LSTM-BP algorithm proposed in this paper. And the comparative verification between the public data set and the real data set proves that the LGMAD algorithm has better performance in most cases. The sixth chapter summarizes the work of this paper and prospects for the future work.

2. Related work

The traditional anomaly detection methods for time series data mainly focus on a one-dimensional scenario, and judge anomalies according to the correlation between the data samples at different time points. The work in this field has been relatively advanced after years of development, among which relatively simple methods include adaptive thresholding, clustering and exponential smoothing. Smith et al. use the three-exponential smoothing method to detect anomalies by different features in historical data to speculate current data values and this method is very effective in the commercial field [7]. Stanway et al. proposed the Skyline project for anomaly detection for streaming data which is made up of a simple detector and a voting scheme so as to output a final anomaly score. This project performs effectively in monitoring real-time anomalies on high-traffic websites [8]. ARIMA algorithm presented by Bianco et al. is a general technique for seasonal time-series data modeling, which works well for detecting regular data, but can not dynamically determiner anomalies in seasonal data [9]. In addition, there are many model-based methods applied to specific fields, examples include detection for cloud data center temperatures [10], anomaly detection in aircraft engine measurements [11] and ATM fraud detection [12] and so on.

Recently, Long Short-Term Memory(LSTM) has received extensive attention because of its advantage in processing time series data. The LSTM improves the shortcomings of RNN's(Recurrent Neural Network) inability to handle long-range correlations. Numenta proposed a RNN-based HTM algorithm and a public dataset NAB(Numenta Anomaly Benchmark) to verify the performance of HTM algorithm [13]. The comparison between our algorithm and HTM algorithm has been included in the experiment. The anomaly detection algorithm implemented by Pankaj Malhotra et al. using LSTM has been proved to acquire excellent results in a dataset of four different fields [14]. Suche Chauhan et al. defined five different types of anomalies and modified a LSTM variant to distinguish anomalies [15]. By adding GRU to transform LSTM, Anvardh Nanduri implemented a similar work to perform anomaly detection for aircraft flights [16]. Jihyun Kim et al. presented an unsupervised anomaly detection method, which is illustrated on real industrial datasets [17]. Overall, due to the effectiveness of the LSTM algorithm in the anomaly detection for time series data, it is chosen to be the basis for anomaly detection for univariate time series.

In recent years, with the rapid growth of data sample magnitude and data dimension, the demand for anomaly detection for multivariate time series has been increased. Many scholars have made great progress in the research of anomaly detection for multivariate time series. Pavel Filonov et al. implemented the method of synthesizing multivariate vectors into a univariate vector to manipulate multivariate data, and then used conventional anomaly detection method for univariate time series to detect anomalies. This method of converting multivariate data into univariate data and then performing anomaly detection is feasible in the case of low dimensions, and usually requires a certain correlation between data of different dimensions. Similar work includes that Han Bao et al. transformed the sequence into characteristic vectors in a lossless way through the multi-dimensional characteristic sequence transformation algorithm and incremental characteristic selection algorithm, and then detected anomalies based on the C-SVM anomaly detection algorithm Chauhan and Vig [18].

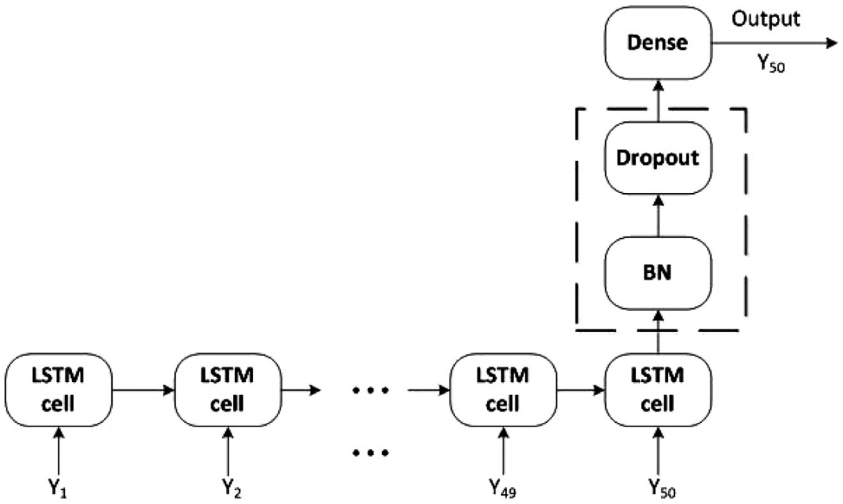

Fig. 1. Overall architecture of LSTM-BP algorithm.

Recently, some academic experts have also proposed some innovative methods. Jones, M et al. extended 8-dimensional data to 32-dimensional, and combined with one-dimensional anomaly detection method to perform multi-dimensional anomaly detection considering the correlative changes between different dimensions [19]. This method can be applied to both univariate time series and multivariate time series, which is a creative notion.

To summarize, we propose an LGMAD anomaly detection method for multivariate time series based on LSTM and multivariate Gaussian model. The LGMAD can analyze the correlation between multivariate data without loss of temporal performance, and demonstrates outstanding detection results.

3. Performance problems and model

3.1. LSTM-BP model based on traditional LSTM

LSTM is a temporal recursive neural network. It has been proved to be extremely suitable for processing time series data. The basic idea of anomaly detection in this paper is to predict the value of the next time of the time series data with the LSTM algorithm, after comparing the predicted value with the actual value to judge the deviation degree of value, the anomalies are determined. Actually, the LSTM algorithm has evolved many variants in recent years. The most thorough analysis of its various variants is the study by Klaus Greff et al. [20], which conducted a large-scale analysis of 5400 experiments on eight different types of LSTM variants. They came to the conclusion that no LSTM variant can significantly improve performance, but some of these variants can save time costs without sacrificing performance. Rafal Jozefowicz et al. from Google conducted a comprehensive architectural search to evaluate more than 10,000 different RNN/LSTM architectures [21]. The final experimental results also support the study by Greff's work, suggesting that an architecture that consistently outperforms GRU cannot be found, but a variant that is superior to other LSTM architectures in some tasks can be identified. Combining the analysis of these two excellent articles with our experimental results, this paper improves the structure of LSTM so as to make it more suitable for dealing with the problem of anomaly detection for time series data.

The traditional LSTM can not achieve the same excellent effect on all data sets. Thus, this paper proposes a LSTM-BP algorithm based on LSTM algorithm, which improves the internal structure of LSTM to make it suitable for processing data in the field of anomaly detection for time series data. The LSTM model we designed considers the influence of time series data on the internal cell state of LSTM, and adds several auxiliary layers to accelerate the convergence speed of the model training while maximally preventing over-fitting. The comparative experiment proves that our LSTM model has more accurate predictions and higher performance on most datasets. The experimental results are included in the experimental section. The flow of LSTM-BP model constructed in this paper is shown in Fig. 1. In the training procedure, single-layer LSTM often fails to meet the demand. This problem can be solved by building a deep LSTM layer with multiple recurrent LSTM layers. The deep LSTM allows time series data to be processed on different time scales with better training results. Based on this, an LSTM unit is set for each input. A total of 50 LSTM units are used to model the network. The output from LSTM is the input to BatchNormalization(BN) layer for normalization. Then through the dropout layer 20% of the neurons are discarded randomly to prevent over-fitting. Finally, the dense layer outputs the predicted results produced by our LSTM-BP model.

The LSTM-BP algorithm proposed in this paper is improved by changing its activation function and adding the peephole connection, so that it can learn time series data better and significantly improve the convergence. The internal structure

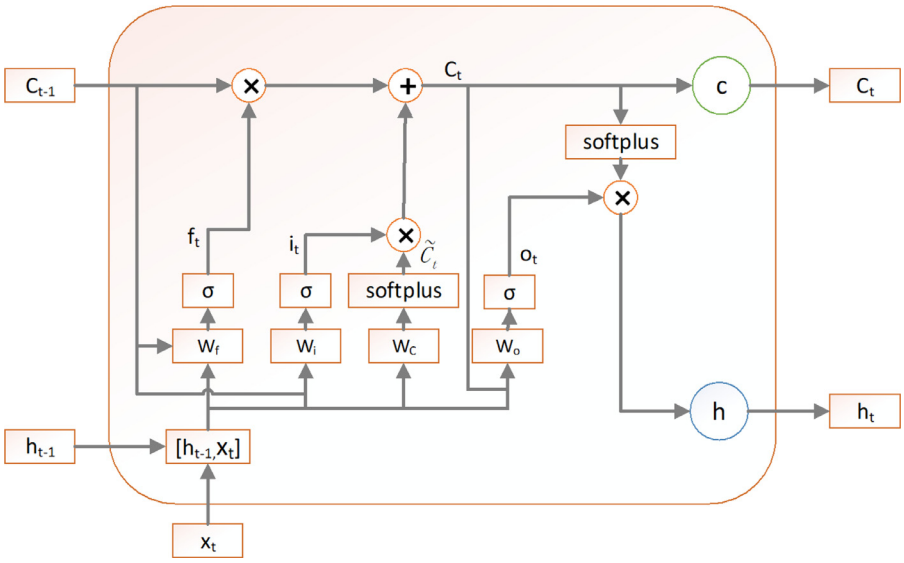

Fig. 2. Overall architecture of LSTM-BP algorithm.

of LSTM layer constructed in this paper is shown in Fig. 2. The implementation of the LSTM calculation unit in this paper is represented by (1)-(6), where f_t represents the forget gate, i_t represents the input gate, w_t represents the previous cell state, C_t represents the current cell state, o_t represents the output gate, and h_t represents the output of the current unit. h_{t-1} represents the output of the previous unit. σ is a logical sigmoid function. W_f, W_i, W_C, W_o are the corresponding weight matrices, b_f, b_i, b_c, b_o are the corresponding deviation vectors. In the LSTM networks, information flow control is implemented by means of the forget gate, the input gate and the output gate (f, i, o). The forget door determines whether to forget the previous memory h_{t-1}, as shown in (1). The input gate calculates the input ratio i_t, as shown in (2). The output of the forget gate f_t and the input gate i_t, decides the output of the current state of the cell C_t in (4). The output gate controls whether to output the calculation result of the current cell as shown in (5) and (6).

$$f_t = \sigma\left(W_f \cdot [C_{t-1}, h_{t-1}, x_t] + b_f\right) \tag{1}$$

$$i_t = \sigma\left(W_i \cdot [C_{t-1}, h_{t-1}, x_t] + b_i\right) \tag{2}$$

$$\widetilde{C}_t = softplus(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{3}$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{4}$$

$$o_t = \sigma\left(W_o \cdot [C_t, h_{t-1}, x_t] + b_o\right) \tag{5}$$

$$h_t = o_t * softplus(C_t) \tag{6}$$

This paper chooses the soft-plus function as the activation function of the input and output of the LSTM calculation unit, which is shown in (7):

$$y = log(1 + e^x) \tag{7}$$

The soft-plus function is similar to the activation frequency function of the brain neurons which is more suitable for controlling the conduction process between neurons in neural networks. It improves the convergence speed and alleviates the over-fitting and gradient disappearance problem compared with the original tanh function. Experimental part can also prove that it has a better convergence effect.

As can be seen from (1)–(6), this paper implements a bi-directional association between cell state and gates by inputting the state of each cell into the forget gate, input gate and output gate. Moreover, with the aid of the peep-hold connection, the LSTM network can record more correlations on time series data, which helps extract relevant information in time series data with certain fluctuation laws. Therefore the LSTM can learn and memorize time series data more accurately.

In order to improve the convergence speed and learning efficiency, as well as solve the over-fitting problem during the training process, we add the BatchNormalization (BN) layer to LSTM model. The BN layer is currently a prevalent technique

for accelerating convergence, which is proposed by Sergey Ioffe et al. in 2015 [22]. The advantage of BN layer is that it has the characteristic of accelerating convergence, which allows it to converge quickly even with a large learning rate, so there is no need to worry about the problem of choosing the appropriate learning rate. It can also improve network generalization ability, which allows it to control the over-fitting condition, helping reduce the reliance on the Dropout layer and reduce the neural network's insensitivity to initialization weights. For the requirements of LSTM network and anomaly detection for time series data, the introduction of BN layer can greatly improve the learning efficiency of LSTM on training set, reduce the use of Dropout layer, and greatly accelerate the convergence speed when training LSTM model.

We added Dropout layer to LSTM model to alleviate the overfitting condition. Dropout is the commonly used method to solve the overfitting problem [14]. Its processing is to discard neurons from the network with a certain probability when updating the parameters during the training process. In this way, the joint adaptability between the neuron nodes is weakened, the neural network is prevented from co-adaptation, and its generalization ability is enhanced, thereby preventing the over-fitting phenomenon to a certain extent. Benefitting from the BN layer, the over-fitting problem is effectively alleviated, so that LSTM model achieves the same good training effect while having better convergence speed.

This paper selects the Dense layer to be the predicted output of LSTM model. Dense layer is also called the fully connected layer. Each node of the Dense layer is connected to all the nodes in the upper layer. It is used to combine the characteristics extracted from the upper layer, which can greatly reduce the impact of characteristic location on classification [20]. The Dense layer extracts the output values and combines them into a single value to output the predicted results of LSTM model.

3.2. Multivariate gaussian distribution and analysis

The multivariate Gaussian distribution is a high-dimensional generalization of normal distribution. The traditional Gaussian distribution model can not directly utilize the correlation between characteristics, and needs to reconstruct the characteristics into new characteristic.

Different from the traditional Gaussian model, the multivariate Gaussian model can automatically capture the correlation between characteristic variables, so it is more suitable for processing continuous data. However, its computational cost is high, it is not suitable for large-scale characteristic classification, and only effective in the condition that the sample size is much larger than the dimension of characteristics. Compared with the traditional Gaussian model, the multi-Gaussian model is more demanding, but it can achieve better effect for processing time series data with low dimensions and correlations between characteristics.

3.2.1. Gaussian distribution

If a random variable x obeys a Gaussian distribution with a mathematical expectation of μ and a variance of σ_2, then x is assumed to follow Gaussian distribution:

$$x \sim N(\mu, \sigma^2) \tag{8}$$

Its probability density function is shown in (9):

$$P(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{9}$$

For a given dataset $(x_1, x_2, \ldots, x_m)$, where $x \in R$, the parameters are calculated by (10) and (11):

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x_i \tag{10}$$

$$\sigma^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu)^2 \tag{11}$$

Gaussian distribution can also be applied to analyze multivariate characteristic variables. Given an n-dimensional dataset $(x_1, x_2, \ldots, x_m)$, for each sample $x \in R$, it is assumed that each of the characteristic variables x follows a Gaussian distribution, as shown in (12):

$$x_1\ N(\mu_1, \sigma_1^2), x_2\ N(\mu_2, \sigma_2^2), \ldots, x_n\ N(\mu_n, \sigma_n^2) \tag{12}$$

Its probability density function is shown in (13):

$$P(x) = \prod_{j=1}^{n} P(x_j; \mu_j, \sigma_j^2) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_j} exp\left(-\frac{(x_j - \mu_j)^2}{2\sigma_j^2}\right. \tag{13}$$

After obtaining the probability density, by setting a threshold for comparison, the anomaly classification can be realized. However, the probability density obtained by the above method only considers the change of only one kind of characteristic, while the correlation information between the characteristic variables, which should not be neglected, will not be reflected. The multivariate Gaussian distribution model can automatically capture the relevance between different characteristics without establishing new characteristics.

3.2.2. Multivariate gaussian distribution

For a given n-dimensional dataset $(x_1, x_2, \ldots, x_m)$, where $x \in R$, the mathematical expectation μ of all characteristics were calculated by the multivariate Gaussian distribution and constructing a covariance matrix Σ for all characteristics, as shown in (14) and (15):

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x_i \tag{14}$$

$$\Sigma = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu)(x_i - \mu)^T \tag{15}$$

Its probability density function is shown in (16):

$$P(x, \mu, \sigma^2) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) \tag{16}$$

The probability density function calculated by (17) is used to analyze the new data, and the $P_{(x)}$ is compared with the adaptive threshold to locate anomalies.

4. LGMAD

For convenience, the relevant formulas in this chapter are shown as follows:

$$E_t = |y_t - x_t| \tag{17}$$

where x_t is the value of time t in the time series dataset $X = (x_1, x_2, \ldots, x_m)$, and y_t is the value of time t in the time series dataset $Y = (y_1, y_2, \ldots, y_m)$, and E_t is the difference between the predicted data and the actual data.

$$S_t = \frac{E_t}{max(E_1, E_2, \ldots, E_m)} \tag{18}$$

S_t is the anomaly score calculated through E_t.

$$H = (S_1, S_2, S_3, \ldots, S_n) \tag{19}$$

$$\alpha = ||H|| = \sqrt{(S_1 - \overline{S_1})^2 + (S_2 - \overline{S_2})^2 + \ldots + (S_n - \overline{S_n})^2} \tag{20}$$

α is the health factor, which represents the level of system operation.

The LSTM network introduced in Chapter 3 is very suitable for processing time series data. After training the model with normal samples, the possible values of the next time point of the time series data can be predicted. By using the predicted value of a certain time in the time series data to compare with the real value, according to the difference combined with the historical behavior, the anomaly point or the anomaly area can be effectively located. Based on this notion, this paper uses the improved LSTM algorithm to perform anomaly detection for univariate time series data. Based on this univariate anomaly detection algorithm, we add the multivariate Gaussian model to analyze the correlation in multivariate data and implement LGMAD algorithm, its state transition diagram is shown in Fig. 3.

The main process of LGMAD algorithm is as follows:

S0: Initialization state.
S1: Acquire a multivariate vector V of the anomalies to be detected.
S2: Enter the LSTM-BP model.

The LSTM-BP model designed in this paper is used to detect the acquired vector V. This paper uses a large amount of normal data to train the improved LSTM-BP model built in Chapter 3. Unlike most anomaly detection algorithms that require tagged anomaly datasets for training to make sure the accuracy of anomaly detection, our algorithm does not need any prior anomaly and utilizes fully normal datasets for training.

S3: Calculate the anomaly score S_t.

The time series dataset $X = (x_1, x_2, \ldots, x_m)$ is predicted using the LSTM-BP model, and the result obtained by the Dense layer is $Y = (y_1, y_2, \ldots, y_m)$, and the difference E_t between the predicted result and the actual sequence is calculated in (18). Then, we use E_t to calculate the anomaly score S_t at time t in the time series data, as shown in (19). After the model training and prediction, and the processing of (18) and (19), each point in the original time series data corresponds to an anomaly score.

S4: Calculate health factor α.

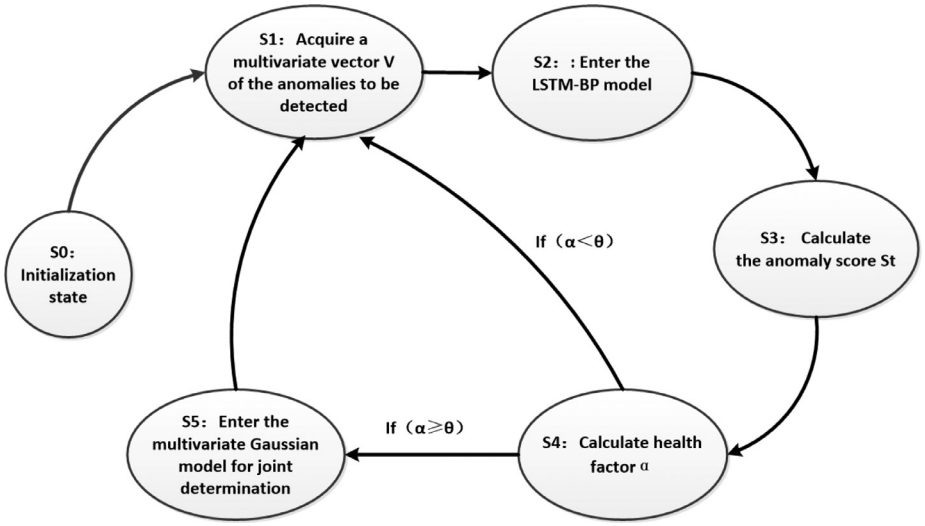

Fig. 3. Overall architecture of LSTM-BP algorithm.

In order to improve the efficiency of the algorithm, this paper introduces the concept of health factor α, which is defined as (20) and (21). Health factor denotes the system performance of the algorithm. The reason for introducing the health factor is that the level of the anomalies of different kind of systems is distinct. For a system with extremely low probability of anomalies in daily operation, if we import all data into the multivariate Gaussian model for joint detection, it is obviously a waste of system resources, even affects the temporal performance of the algorithm. In this paper, it is emphasized that only when α is higher than the threshold θ, the data is imported into the multivariate Gaussian model for joint detection, which can effectively improve the temporal performance of the algorithm. The threshold θ can be dynamically selected between 0 and 1 depending on the system state. When θ is equal to 0 ($\theta = 0$), it means that all the data is imported into the multivariate Gaussian model for joint determination, which corresponds to the condition where the system state is extremely unhealthy. When θ is equal to 1 ($\theta = 1$), it represents the system is completely normal, and no matter what value of α is, it is not necessary to enter the multivariate Gaussian model for joint determination.

In this algorithm, the n-dimensional vector $V = (v_1, \ldots, v_n)$ is used as the system input, and after preprocessing, it is converted into an n-dimensional vector $V = (v_1, \ldots, v_n)$ which can be directly processed by the LSTM-BE model. The trained LSTM-BE model is used to process the n-dimensional vector, output the anomaly score, calculate the value of α, and compare α with the threshold θ. If $\alpha < \theta$, the system is considered normal and no further step is needed. Otherwise, it is considered that anomalies may occur, then we should use multivariate Gaussian model for joint determination and get the final list of anomalies.

S5: Enter the multivariate Gaussian model for joint determination, thereby complete the detection of the anomaly points and the division of the anomaly region.

5. Experiment and analysis

5.1. Experimental scenario and datasets

The LGMAD algorithm proposed in this paper is designed for low-dimensional system characterization and state analysis. For the anomaly detection problem in the typical anomaly monitoring scenario, this algorithm is utilized to detect anomalies under the real scenario using computer system. In the experimental analysis, this paper uses two datasets to test and compare the performance of this algorithm and related algorithms.

(1) Dataset 1: NAB dataset [23]. This dataset was proposed by Numenta in 2015 and contains many different types of labeled anomaly data from real world. Its anomaly type consists of various fields and applications, including as many stream data anomaly types as possible with a certain amount of noise. For the anomaly detection for univariate time series data, this paper uses the proposed LSTM-BP algorithm to separately detect the data of CPU, memory and network rate in the NAB dataset, and compares with the HTM algorithm, the conventional LSTM algorithm, and the algorithm only adding BN layer based on conventional LSTM (shortened by LSTM-BN algorithm).
(2) Dataset 2: Self-made dataset based on the system state in real scenario. This dataset is used to perform anomaly characterization and detection of the system on multi-dimensional parameters. For a better reflection of the real-time state of the system, we use Neusoft RealInsight APM software to collect real-time parameters of system state (CPU utilization, system memory occupancy ratio, network speed). An example diagram is shown in Fig. 4

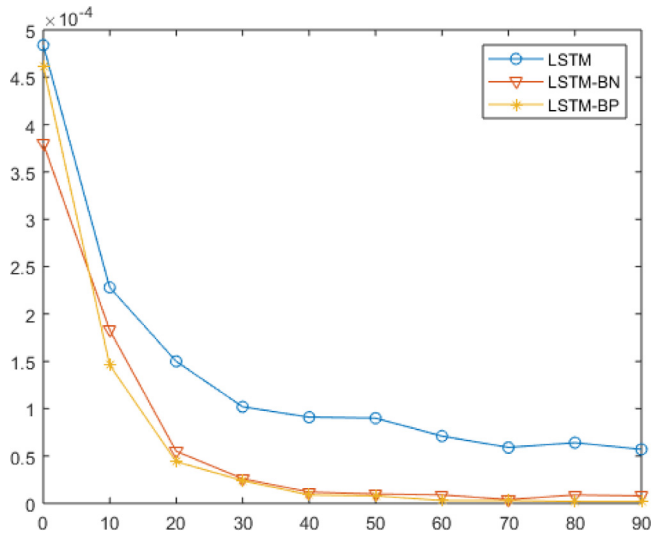

Fig. 4. Overall architecture of LSTM-BP algorithm.

5.2. Performance comparison and analysis for univariate anomaly detection

5.2.1. Efficiency analysis

In this experiment, the three system state parameters of CPU, MEM and NET in dataset 1 are separately detected by HTM algorithm, conventional LSTM algorithm, LSTM-BN algorithm, and LSTM-BP algorithm proposed in this paper and the performance of these algorithms are compared and analyzed. Before presenting the comparative results, we first briefly introduce the evaluation criteria of this experiment.

We use the accuracy rate (*Precision*), recall rate (*Recall*) and F_1 score as the evaluation criteria for the anomaly detection algorithm [24,25]. Their calculation methods are as shown in Formulas 21–23.

$$Precision = \frac{T_p}{T_p + F_p} \tag{21}$$

$$Recall = \frac{T_p}{T_p + F_n} \tag{22}$$

$$F_1 = \frac{2 * Preicision * Recall}{Precision + Recall} \tag{23}$$

Among them, T_p represents the number of correctly detected anomalies, F_p represents the number of false positives, and F_n represents the number of false negatives. As shown in Formulas 22–24, the number of T_p, F_p and F_n will determine the precision and recall results. Precision is used to judge the sensitivity of the anomaly detection algorithm to the anomalies. Recall reflects the ability of the algorithm to detect anomalies. Precision and Recall affect the final F_1 score. The F_1 score represents the overall performance of the anomaly detection algorithm.

The experimental results of anomaly detection for univariate time series are shown in Table 1. We can see that the conventional LSTM algorithm, LSTM-BN algorithm, and LSTM-BP algorithm perform better than HTM algorithm in three data types. The advantage of the Precision values is the most obvious, which is mainly due to the fact that LSTM algorithm can obtain fewer false positives. The advantages of Recall values are not as obvious as Precision, but is still superior to the HTM algorithm. The significantly higher Precision values in turn increase the values of F_1 score. There is the litter difference between the detection effects of the LSTM algorithm, the LSTM-BN algorithm, and the LSTM-BP algorithm, where the detection effect of LSTM-BP algorithm is slightly better than the former two algorithms.

Although the detection results of LSTM-BP algorithm proposed in this paper is slightly better than LSTM algorithm and LSTM-BN algorithm, it does not show obvious improvements in terms of detection effect. The main advantage of LSTM-BE algorithm is that it has better convergence when training model, as shown in Fig. 5. We set epoch to 100, batchsize to 100, and learning rate to 0.005 to train the model. The results show that the convergence of LSTM-BP is significantly better than the conventional LSTM algorithm and also little better than LSTM-BN algorithm.

5.2.2. Performance comparison and analysis for multivariate anomaly detection

This experiment uses dataset 2 to compare and analyze the proposed LGMAD algorithm and the LSTM-based Predictive Data Model(LSTM-PDM) proposed by Pavel Filonov [6]. The latter algorithm decomposes multidimensional data into one-dimensional data and uses univariate anomaly detection method for detection. It has achieved good results on industrial

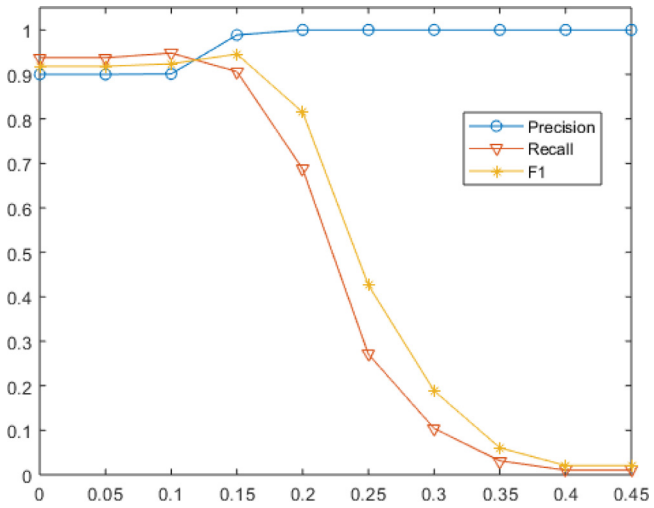

Fig. 5. Overall architecture of LSTM-BP algorithm.

datasets. The process of this algorithm is shown as Formula 24:

$$MSE(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2, \tag{24}$$

where X_i is the value of n-dimensional time series data $X_n = (x_1, x_2, \ldots, x_m)$ in the ith dimension, and Y_i is the value of its prediction result $Y_n = (y_1, y_2, \ldots, y_m)$ in the i_{th} dimension. n is the total data dimension. In this paper, n is set to 3. The MSE value corresponds to the anomaly score of HTM algorithm, and an anomaly is determined when it is higher than the adaptive threshold.

In order to avoid the influence of different health factor α on the performance of the algorithm, α is set to 0, that is, the overall data in the dataset needs to enter the Gaussian model for analysis. For more intuitive demonstration of the improvement of the LGMAD algorithm, this experiment, HTM algorithm and the anomaly detection algorithm are also adopted for comparison. The experimental results are shown in Table 2.

From Table 2 we can see that the LGMAD algorithm is superior to the LSTM-based Predictive Data Model(LSTM-PDM) in all evaluation criteria, and is also superior to the anomaly detection algorithm for univariate time series. Since anomaly types in some data sample are difficult to detect only through single dimension, it is necessary to combine other dimensions to locate anomalies. This is the main problem of anomaly detection for univariate time series, however, by using Gaussian model for joint detection of multiple dimensions we can solve this problem effectively and derived better detection results. It can also be seen from the comparison experiment that the recall value of the LGMAD algorithm is extremely outstanding, which is mainly due to using the joint detection of multiple dimensions to detect anomalies that cannot be detected only from single dimension, which significantly reduces the false negative rate, thus improves recall value and then increases the value of F_1 score.

The above experiments are all conducted with the condition $\alpha = 0$. For different values of α, the performance analysis of the LGMAD algorithm is shown in Fig. 6. α gradually increases from 0. When α is still at 0, there is no significant change

Table 1
Safety distance.

Group	Algorithm	Precision	Recall	F_1
CPU	HTM	0.835	0.808	0.821
	LSTM	0.938	0.878	0.907
	LSTM-BN	0.885	0.945	0.914
	LSTM-BP	0.933	0.913	0.923
DISK	HTM	0.850	0.885	0.867
	LSTM	0.907	0.793	0.846
	LSTM-BN	0.887	0.856	0.871
	LSTM-BP	0.845	0.869	0.857
NET	HTM	0.857	0.867	0.862
	LSTM	0.895	0.912	0.903
	LSTM-BN	0.907	0.954	0.925
	LSTM-BP	0.955	0.925	0.939

Table 2
Safety distance.

Algorithm	Precision	Recall	F_1
HTM-CPU	0.656	0.613	0.630
HTM-MEM	0.395	0.487	0.436
HTM-NET	0.605	0.525	0.562
LSTMBP-CPU	0.826	0.797	0.811
LSTMBP-MEM	0.445	0.558	0.495
LSTMBP-NET	0.664	0.707	0.685
LSTM-PDM	0.888	0.811	0.848
LGMAD	0.901	0.938	0.919

in the performance of the algorithm. When α is over 0.15, a small number of false positives are filtered out, resulting in a slight increase in the precision value, thereby the F_1 score increases. As α continues to fall, the correct detected points are filtered out, and recall begins to drop significantly until 0, and the F_1 score also drops to 0 following recall.

In this chapter, we utilize the LSTM-BP algorithm and the LGMAD algorithm proposed in this paper to perform anomaly detection separately on the NAB public dataset and the generated dataset. According to the experimental results, we can draw the following conclusions. The detection effect of LSTM-BP algorithm is optimal in anomaly detection for univariate time series with the best convergence. However, in some cases of multivariate anomaly detection, its performance degrades significantly. The values of precision, recall and F_1 scores of the LGMAD algorithm are optimal, and its detection effect is better than the LSTM-based Predictive Data Model, even it exhibits significant advantages compared to the univariate anomaly detection algorithms.

6. Conclusion

In this paper, a real-time anomaly detection algorithm, which combines the Long Short Term Memory algorithm and Gaussian Mixture Model effectively, is proposed for the complex systems. First, we use the modified Long Short Term Memory algorithm to detect anomalies of univariate time series data. Then, a Gaussian Mixture Model is adopted to give a multidimensional joint detection of possible anomalies in the first step. Besides, in order to improve the efficiency of the algorithm, the concept of the health factor α is proposed to describe the health level of the system. This method can greatly improve the performance of the algorithm in the health system. Finally, the experiment on two types of dataset validate that the Real-time Anomaly Detection algorithm performs. Furthermore, it will remarkably improve the performance of real-time anomaly detection in many domains. However, this paper mainly focuses on low-dimension anomaly detection issues, without considering the high-dimensional or super-high-dimensional anomaly detection scenarios. Our future work will include optimizing our algorithm further and improving the detection accuracy, we also attempt to enhance our algorithm to handle a higher dimensional scene.

Declaration of Competing Interest

None.

References

[1] Ahmad Subutai, Purdy Scott. Real-Time Anomaly Detection for Streaming Analytics[J]. Computer Science. AI 2016;1607:1–9.
[2] Sadeghi Reza, Hamidzadeh Javad. Automatic support vector data description. Soft Computing 2018;22(1):147–58.
[3] Gauci M, Chen J, Li W, et al. Self-organized aggregation without computation[J]. The International Journal of Robotics Research 2014;33(8):1145–61.
[4] Szmit M, Szmit A. Usage of Modified Holt-Winters Method in the Anomaly Detection of Network Traffic: Case Studies[J]. Journal of Computer Networks and Communications. 2012(8):1-5.
[5] Telangre KS. Anomaly Detection using multidimensional reduction Principal Component Analysis[J]. IOSR Journal of Computer Engineering 2014;16(1):86–90.
[6] Filonov P, Lavrentyev A, Vorontsov A. Multivariate Industrial Time Series with Cyber-Attack Simulation: Fault Detection Using an LSTM-based Predictive Data Model[J]. NIPS Time Series Workshop 2016.
[7] Makridakis S, Spiliotis E, Assimakopoulos V. Statistical and Machine Learning forecasting methods: Concerns and ways forward[J]. PloS one 2018;13(3):e0194889.
[8] Stanway, A. Etsy Skyline, 2013. https://github.com/etsy/skyline.
[9] Bianco A M, García Ben M, Martínez E J, Yohai V J. Outlier detection in regression models with ARIMA errors using robust estimates. Journal of Forecasting 2001;20(8):565–79.
[10] Chen Y. Design and Implementation of Network Resource Management and Configuration System based on Container Cloud Platform[C]. In: International Conference on Frontiers of Manufacturing Science and Measuring Technology; 2017. p. 331–5.
[11] Simon DL, Rinehart AW. A Model-Based Anomaly Detection Approach for Analyzing Streaming Aircraft Engine Measurement Data[J]. ASME Turbo Expo 2014: Turbine Technical Conference and Exposition 2014;6(6):32–43.
[12] Klerx T, Anderka M. Buning. Model-Based Anomaly Detection for Discrete Event Systems[C]. In: IEEE International Conference on TOOLS with Artificial Intelligence; 2014. p. 665–72.
[13] Hawkins J, Ahmad S, Dubinsky. HTM Cortical Learning Algorithms[R]. Redwood City: Numenta, Incorporation; 2011.
[14] Malhotra P, Vig L, Shroff G. Long short term memory networks for anomaly detection in time series[J]. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning 2015(8):89–94.

[15] Chauhan S, Vig L. Anomaly detection in ECG time signals via deep long short-term memory networks[C]. In: IEEE International Conference on Data Science and Advanced Analytics; 2015. p. 1–7.
[16] Nanduri A, Sherry L. Anomaly detection in aircraft data using Recurrent Neural Networks (RNN)[C]. In: IEEE Integrated Communications Navigation and Surveillance; 2016. 5C2-1-5C2-8.
[17] Kim J, Kim J, Thu HLT. Long Short Term Memory Recurrent Neural Network Classifier for Intrusion Detection[C]. In: IEEE International Conference on Platform Technology and Service; 2016. p. 1–5.
[18] Bao H, Wang Y. A C-SVM Based Anomaly Detection Method for Multi-Dimensional Sequence over Data Stream[C]. In: IEEE International Conference on Parallel and Distributed Systems; 2017. p. 948–55.
[19] Jones M, Nikovski D, Imamura M, et al. Exemplar learning for extremely efficient anomaly detection in real-valued time series[J]. Data Mining and Knowledge Discovery 2016;30(6):1427–54.
[20] Greff Klaus, Srivastava Rupesh K, Koutnik Jan. LSTM: A Search Space Odyssey[J]. IEEE Transactions on Neural Networks Learning Systems. 2017;28(10):2222–32.
[21] Jozefowicz R, Zaremba W, Sutskever I. An empirical exploration of recurrent network architectures[C]. In: International Conference on Machine Learning; 2015. p. 2342–50.
[22] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[J]. 2015:448-456.
[23] Lavin A, Ahmad S. Evaluating Real-Time Anomaly Detection Algorithms-The Numenta Anomaly Benchmark[C]. In: IEEE 14th International Conference on Machine Learning and Applications; 2015. p. 38–44.
[24] Zabihi Mahdieh, Vafaei Jahan M, Hamidzadeh Javad. A density based clustering approach to distinguish between web robot and human requests to a web server. The ISC International Journal of Information Security 2014;6(1):77–89.
[25] Zabihimayvan M, Doran D. Fuzzy Rough Set Feature Selection to Enhance Phishing Attack Detection[J]. arXiv preprint arXiv:1903.05675, 2019.

Nan Ding He received the B.S., M.S., and Ph.D. degrees in Computer Science and Engineering from Dalian University of Technology. He has been an Associate Professor in Computer Science and Engineering from Dalian University of Technology since 2014. His research interests include vehicular networks, wireless security, wireless sensor networks, cache management, and distributed fault tolerant computing.

Haoxuan Ma He was born in 1995. M.S. candidate. Computer Science and Engineering from Dalian University of Technology, Dalian, China. His research interests include network virtualization and vehicular networks.

Huanbo Gao He was born in 1993. M.S. candidate. Computer Science and Engineering from Dalian University of Technology, Dalian, China. His research interests include anomaly detection and wireless sensor networks.

Yanhua Ma She received the Ph.D. degree in control theory and control engineering from Beijing University of Aeronautics and Astronautics (BUAA) in 2011. She is currently an Associate Professor with DLUT. Her current research interests include system and control theory, computational intelligence, and fault diagnosis and fault-tolerant control.

Guozhen Tan He has been a Professor with the School of Computer Science and Technology, Dalian University of Technology and the Director of Engineering and Technology Research Center for the Internet of things and Collaborative Sensing, Liaoning province. His research interests include Internet of things, vehicular networking, intelligent transportation control, etc.